**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Uso do teste de Scott-Knott e da análise de agrupamentos, na
obtenção de grupos de locais para experimentos com
cana-de-açúcar**

Cristiane Mariana Rodrigues da Silva

**Dissertação apresentada, para obtenção do título de
Mestre em Agronomia, Área de concentração:
Estatística e Experimentação Agrônômica**

**Piracicaba
2007**

Cristiane Mariana Rodrigues da Silva

Uso do teste de Scott-Knott e da análise de agrupamentos, na obtenção de grupos de locais para experimentos com cana-de-açúcar

Orientador:

Prof. Dr. DÉCIO BARBIN

Dissertação apresentada, para obtenção do título de Mestre em Agronomia, Área de concentração: Estatística e Experimentação Agronômica

**Piracicaba
2007**

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Rodrigues da Silva, Cristiane Mariana

Uso do teste de Scott-Knott e da análise de agrupamentos, na obtenção de grupos de locais para experimentos com cana-de-açúcar / Cristiane Mariana Rodrigues da Silva. - - Piracicaba, 2007.

48 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2007.
Bibliografia.

1. Análise de conglomerados 2. Cana-de-açúcar I. Título

CDD 633.61

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

DEDICATÓRIA

Dedico este trabalho, primeiramente, à Ms. Maria Izalina F. Alves *“in memoriam”*, de quem partiu o incentivo para que eu pudesse começá-lo.

Dedico também ao meu marido Valter José de Moraes, à minha filha Daniela B. de Moraes e aos demais familiares, pela compreensão e paciência, principalmente por minha ausência.

Não poderia deixar de dedicar à Prof^a. Dra. Clarice Garcia Borges Demétrio, por todo apoio e incentivo, sem o qual talvez eu não concluísse mais esta etapa da minha vida, e à Solange Sabadin, também por todo o incentivo, dedicação, companheirismo e sincera amizade.

E aos amigos que me acompanharam e me incentivaram muito durante todo esse processo, mas sem citar nomes, pois poderia cometer injustiças, esquecendo-me de algum.

AGRADECIMENTOS

Agradeço, primeiramente, à meu orientador Prof. Dr. Décio Barbin, que além de desempenhar muito bem seu papel de orientador, é um amigo de valor inestimável.

Agradeço ao Prof. Dr. Carlos Tadeu dos Santos Dias, sempre disposto a me orientar nas análises multivariadas, com muita atenção e dedicação.

Agradeço ao Ms. Rubens L. do Canto Braga Jr, pela disposição em me orientar quando necessário e pela compreensão frente a todos os problemas encontrados e, ao Centro de Tecnologia Canavieira (CTC), pela oportunidade de trabalhar com esses dados, e acesso a esse grupo de grandes colaboradores.

Aos amigos, Ms. Vanderly Janeiro, Ms. Lúcio Borges de Araújo, e Ms. Ângela Coelho, pelo auxílio em momentos de dúvidas, e principalmente pela sincera amizade.

E principalmente à Deus, pelas oportunidades e pessoas maravilhosas apresentadas em minha vida.

*“O que antes parecia impossível torna-se possível quando alguém
se sente apto a transpor todos os obstáculos,
conhecendo a sua verdadeira grandeza interior”.*

*“Nada é mais importante que viver o momento presente intensamente,
porque essa é a verdadeira vida”.*

Nuno Cobra

SUMÁRIO

RESUMO	6
ABSTRACT	7
LISTA DE FIGURAS	8
LISTA DE TABELAS	9
LISTA DE QUADROS	10
LISTA DE ABREVIATURAS	11
1 INTRODUÇÃO	12
2 DESENVOLVIMENTO	14
2.1 Revisão Bibliográfica	14
2.1.1 Teste de Scott-Knott	14
2.1.2 Análise de Agrupamentos	17
2.1.3 Análise conjunta de Experimentos em Blocos Casualizados com alguns Tratamentos Comuns	23
2.1.4 Considerações sobre a cana-de-açúcar	26
2.1.5 Programas de Melhoramento de cana-de-açúcar	29
2.2 Material e Métodos	30
2.3 Resultados e Discussão	34
3 CONSIDERAÇÕES FINAIS	45
REFERÊNCIAS	46

RESUMO

Uso do teste de Scott-Knott e da análise de agrupamentos, na obtenção de grupos de locais para experimentos com cana-de-açúcar

O Centro de Tecnologia Canavieira (CTC), situado na cidade de Piracicaba, é uma associação civil de direito privado, criada em agosto de 2004, com o objetivo de realizar pesquisa e desenvolvimento em novas tecnologias para aplicação nas atividades agrícolas, logísticas e industriais dos setores canavieiro e sucroalcooleiro e desenvolver novas variedades de cana-de-açúcar. Há 30 anos, são feitos experimentos, principalmente no estado de São Paulo onde se concentra a maior parte dessas unidades produtoras associadas. No ano de 2004 foram instalados ensaios em 11 destas Unidades Experimentais dentro do estado de São Paulo, e há a necessidade de se saber se é possível a redução deste número, visando aos aspectos econômicos. Se se detectarem grupos de Unidades com dados muito similares, pode-se reduzir o número destas, reduzindo-se, conseqüentemente, o custo dessas pesquisas, e é através do teste estatístico de Scott-Knott e da Análise de Agrupamento, que essa similaridade será comprovada. Este trabalho tem por objetivo, aplicar as técnicas da Análise de Agrupamento ("Cluster Analysis") e o teste de Scott-Knott na identificação da existência de grupos de Unidades Industriais, visando à diminuição do número de experimentos do Centro de Tecnologia Canavieira (CTC) e, por conseguinte, visando ao menor custo operacional. Os métodos de comparação múltipla baseados em análise de agrupamento univariada, têm por objetivo separar as médias de tratamentos que, para esse estudo foram médias de locais, em grupos homogêneos, pela minimização da variação dentro, e maximização entre grupos e um desses procedimentos é o teste de Scott-Knott. A análise de agrupamento permite classificar indivíduos ou objetos em subgrupos excludentes, em que se pretende, de uma forma geral, maximizar a homogeneidade de objetos ou indivíduos dentro de grupos e maximizar a heterogeneidade entre os grupos, sendo que a representação desses grupos é feita num gráfico com uma estrutura de árvore denominado dendrograma. O teste de Scott-Knott, é um teste para Análise Univariada, portanto, mais indicado quando se tem apenas uma variável em estudo, sendo que a variável usada foi TPH5C, por se tratar de uma variável calculada a partir das variáveis POL, TCH e FIB. A Análise de Agrupamento, através do Método de Ligação das Médias, mostrou-se mais confiável, pois possuía-se, nesse estudo, três variáveis para análise, que foram: TCH (tonelada de cana por hectare), POL (porcentagem de açúcar), e FIB (porcentagem de fibra). Comparando-se o teste de Scott-Knott com a Análise de Agrupamentos, confirmam-se os agrupamentos entre os locais L020 e L076 e os locais L045 e L006. Conclui-se, portanto, que podem ser eliminadas dos experimentos duas unidades experimentais, optando por L020 (Ribeirão Preto) ou L076 (Assis), e L045 (Ribeirão Preto) ou L006 (Região de Jaú), ficando essa escolha, a critério do pesquisador, podendo assim, reduzir seu custo operacional.

Palavras-chave: Análise de Conglomerados; Cana-de-açúcar

ABSTRACT

Scott-Knott test and cluster analysis use in the obtainment of placement groups for sugar cane experiments

The Centre of Sugar Cane Technology (CTC), placed at the city of Piracicaba, is a private right civilian association, created in August of 2004, aiming to research and develop new technologies with application in agricultural and logistic activities, as well as industrial activities related to sugar and alcohol sectors, such as the development of new sugar cane varieties. Experiments have been made for 30 years, mainly at the state of São Paulo, where most of the associated unities of production are located. At the year of 2004, experiments were installed in 11 of those Experimental Unities within the state of São Paulo, and there is the need to know if it is possible the reduction of this number, aiming at the economical aspects. If it were detected groups of Unities with very similar data, it would be possible to eliminate some of these Unities, diminishing, consequently, the researches cost, and it is through the Scott-Knott statistical test and the Cluster Analysis that this similarity may be corroborated. This work aims to apply the Cluster Analysis techniques and the Scott-Knott test to the identification of the existence of groups of Industrial Unities, aiming at the reduction of the CTC's experiments number and, consequently, aiming at the smaller operational cost. The methods of multiple comparison based on univariate cluster analysis aim to split the treatments means in homogenous groups, for this work were used the placement groups means, through the minimization of the variation within, and the maximization amongst groups; one of these methods is the Scott-Knott test. The cluster analysis allows the classification of individual or objects in excludent groups; again, the idea is to maximize the homogeneity of objects or individual within groups and to maximize the heterogeneity amongst groups, being that these groups are represented by a tree structured graphic by the name of dendrogram. The Scott-Knott test is a Univariate Analysis test, therefore is appropriate for studies with only one variable of interest. The Cluster Analysis, through the Linkage of Means Method, proved to be more reliable, for, in this case, there were three variables of interest for analysis, and these were: TCH (weight, in tons, of sugar cane by hectare), POL (percentage of sugar) and FIB (percentage of fiber). By comparing the Scott-Knott test with the Cluster Analysis, two pairs of clustering are confirmed, these are: placements L020 and L076; and L045 and L006. Therefore it is concluded that two of the experimental unities may be removed, one can choose from L020 (Ribeirão Preto) or L076 (Assis), and L045 (Ribeirão Preto) or L006 (Região de Jaú), the choice lies with the researcher, and it can diminish the operational cost.

Keywords: Cluster Analysis; Sugar Cane

LISTA DE FIGURAS

Figura 1 – Dendrograma da análise de agrupamentos realizada a partir dos dados apresentados na Tabela 2	22
Figura 2 – Sistema radicular da cana planta estabelecido.....	27
Figura 3 – Porção basal de uma touceira	27
Figura 4 – Cana-planta no estágio inicial de brotação com os dois tipos de raízes	28
Figura 5 – Cana-planta em brotação mostrando dois tipos de raízes.....	28
Figura 6 – Rizomas de cana-de-açúcar formadores de touceiras em cana soca e ressoca	28
Figura 7 – Delineamento experimental em uma Unidade Experimental utilizado pelo Centro de Tecnologia Canavieira (CTC).....	31
Figura 8 – Dendrograma do processo hierárquico de agrupamentos de Locais, através das variáveis TCH, POL e FIB – Método de Ligação das Médias	40
Figura 9 – Dendrograma do processo hierárquico de agrupamentos de Locais, através da variável TPH5C – Método de Ligação das Médias	42

LISTA DE TABELAS

Tabela 1 – Distância Euclidiana entre 10 pares de empresas públicas americanas	21
Tabela 2 – Dados coletados de 10 empresas públicas americanas, no ano de 1975	21
Tabela 3 – Análise Conjunta de Variância da Variável TCH	35
Tabela 4 – Análise Conjunta de Variância da Variável POL	35
Tabela 5 – Análise Conjunta de Variância da Variável FIB	36
Tabela 6 – Análise Conjunta de Variância da Variável TPH5C	36
Tabela 7 – Matriz de Similaridades entre Locais, para as variáveis TCH, POL e FIB, obtida através da Distância Euclidiana	39
Tabela 8 – Matriz Cofenética baseada no dendrograma da Figura 8	41
Tabela 9 – Matriz de Similaridades entre Locais, para a variável TPH5C, obtida através da Distância Euclidiana	42
Tabela 10 – Matriz Cofenética baseada no dendrograma da Figura 9	43

LISTA DE QUADROS

Quadro 1 – Matrizes de dados brutos e padronizados ou relativizados	19
Quadro 2 – Fórmulas utilizadas na análise dos ensaios pelo CTC	33
Quadro 3 – Análise de Agrupamentos para Locais, através das variáveis TCH, POL e FIB, realizada através do programa SAS	40
Quadro 4 – Análise de Agrupamentos para Locais, através da variável TPH5C, realizada através do programa SAS.....	43

LISTA DE ABREVIATURAS

CTC – Centro de Tecnologia Canavieira

MMD – Método das Médias das Distâncias

MLS – Método da Ligação Simples

MLC – Método da Ligação Completa

TCH – Tonelada de Cana por Hectare

POL – Porcentagem de açúcar da cana

FIB – Porcentagem de fibra da cana

TPH5C – Tonelada de Pol por Hectare estimada para 5 cortes

1 INTRODUÇÃO

O Centro de Tecnologia Canavieira (CTC), situado na cidade de Piracicaba, é uma associação civil de direito privado, criada em agosto de 2004, com o objetivo de realizar pesquisa e desenvolvimento em novas tecnologias para aplicação nas atividades agrícolas, logísticas e industriais dos setores canavieiro e sucroalcooleiro e desenvolver novas variedades de cana-de-açúcar.

O Melhoramento Genético de Cana-de-açúcar integra um dos 11 programas de pesquisa, que estão em desenvolvimento no CTC, com o objetivo de obter variedades mais produtivas, resistentes às principais doenças e pragas e com maior teor de sacarose.

Esse programa de Melhoramento Genético em Cana-de-açúcar, segundo Tecnologias CTC (2006), consiste em ensaios de competição em cerca de 20 usinas, perfazendo 13 anos a cada ciclo ou série, desde o cruzamento até a liberação comercial, incluindo os testes de reação às principais pragas e doenças e caracterização quanto à adaptação aos diversos ambientes de produção e época de colheita, com fases de seleção nas estações experimentais de Piracicaba e Jaú. Cada variedade liberada, destaca-se por produtividade, precocidade, teor de sacarose, época de colheita da safra, resistência às principais pragas e doenças, ambientes de produção, florescimento e/ou isoporização.

Existem hoje, mais de 150 associados (unidades industriais produtoras de cana, açúcar e álcool e plantadores de cana-de-açúcar e suas respectivas associações) ao CTC inseridos nessas pesquisas, nos estados da região Centro-Sul (Paraná, São Paulo, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Goiás e Espírito Santo).

Há 30 anos, são feitos experimentos, principalmente no estado de São Paulo onde se concentra a maior parte dessas unidades produtoras associadas.

No ano de 2004 foram instalados ensaios em 11 dessas Unidades Experimentais dentro do estado de São Paulo, e há a necessidade de se saber se é possível a redução desse número, visando aos aspectos econômicos.

Se se detectarem grupos de Unidades com dados muito similares, pode-

se reduzir o número destas, reduzindo-se, conseqüentemente, o custo dessas pesquisas, e é através dos testes estatísticos de Scott-Knott e Análise de Agrupamento, que essa similaridade será verificada.

Sendo assim, este trabalho tem por objetivo, aplicar as técnicas da Análise de Agrupamento ("*Cluster Analysis*") e o teste de Scott-Knott na identificação da existência de grupos de Unidades Industriais, visando à diminuição do número de experimentos do Centro de Tecnologia Canavieira (CTC) e, por conseguinte, visando ao menor custo operacional.

2 DESENVOLVIMENTO

2.1 Revisão Bibliográfica

Os pesquisadores têm facilidade em utilizar os testes de comparações múltiplas, pois a bibliografia é extensa e a disponibilidade nos programas computacionais é grande. Porém, nos procedimentos de comparações de médias mais comuns, como o Teste de Tukey, Duncan, LSD, entre outros, apresentam resultados de difícil interpretação, devido a ambigüidade apresentada. A ambigüidade ocorre quando dois tratamentos, tidos como diferentes entre si, não diferem de um terceiro. Os métodos de comparação múltipla baseados em análise de agrupamento univariada, eliminam este tipo de problema, pois têm por objetivo separar as médias de tratamentos em grupos homogêneos, pela minimização da variação dentro, e maximização entre grupos e um desses procedimentos é o teste de Scott-Knott (SANTOS, 2000).

A quantidade de métodos de comparações múltiplas é grande, sendo que sua aplicação basicamente consiste em examinar um conjunto de médias e ver quais delas possuem diferenças significativas entre si, mas dois ou três métodos são os mais comuns, tais como o Método de Ligação Simples, Método de Ligação Completa e Método de Ligação Média, que são métodos hierárquicos de agrupamento.

2.1.1 Teste de Scott-Knott

Segundo Ramalho, et al (2000), o procedimento de Scott-Knott utiliza a razão de verossimilhança para testar a significância de que n tratamentos podem ser divididos em dois grupos que maximizem a soma de quadrados entre grupos. Sejam, por exemplo, 3 tratamentos, A, B e C. O processo consiste em determinar uma partição, em dois grupos, que maximize a soma de quadrados, sendo que o número de partições possível é dado por $(2^{n-1} - 1)$. Nesse exemplo, são possíveis $(2^{3-1} - 1) = (2^{2-1} - 1) = (4 - 1) = 3$ grupos, isto é, A vs B e C, B vs A e C e C vs A e B. Com

um número pequeno de tratamentos como o do exemplo, é fácil obter todos os grupos. Contudo, quando o número de tratamentos é grande, o número de grupos cresce exponencialmente, dificultando a aplicação do teste. Para atenuar esse problema, basta ordenar as médias dos tratamentos.

Fisher (1958) demonstrou, que é necessário considerar apenas $(p-1)$ partições obtidas pela ordenação das médias e sua divisão entre duas partições sucessivas. Esta simplificação torna possível a aplicação do teste, por meio de uma calculadora, até com 11 ou 12 tratamentos. Após encontrar a melhor separação entre dois grupos, repete-se o processo em cada subgrupo. A partir daí, prossegue-se com a subdivisão até que os grupos resultantes sejam considerados estatisticamente iguais pelo teste de χ^2 .

Segundo Scott-Knott (1974), uma vez ordenadas as médias, procede-se do seguinte modo, fazendo inicialmente o número de tratamentos $g = n$:

- i. Determinar a partição entre dois grupos que maximize a soma de quadrados entre grupos. Essa soma de quadrados será definida por B_0 , e será estimada da seguinte forma: sejam T_1 e T_2 os totais dos dois grupos com k_1 e k_2 tratamentos em cada um.

$$B_0 = \frac{T_1^2}{k_1} + \frac{T_2^2}{k_2} - \frac{(T_1 + T_2)^2}{k_1 + k_2}$$

$$T_1 = \sum_{i=1}^{k_1} \bar{Y}_{(i)} \quad \text{e} \quad T_2 = \sum_{i=k_1+1}^g \bar{Y}_{(i)}$$

em que $\bar{Y}_{(i)}$ é a média do tratamento da posição ordenada i . Os dois grupos deverão ser identificados por meio da inspeção das somas de quadrados das $g-1$ partições possíveis, sendo g o número de tratamentos envolvidos no grupo de médias considerado.

- ii. Determinar o valor da estatística λ da seguinte forma:

$$\lambda = \frac{\pi}{2(\pi-2)} \times \frac{B_0}{\hat{\sigma}_0^2}$$

em que $\hat{\sigma}_0^2$ é o estimador de máxima verosimilhança de σ_Y^2 .

Seja $s_Y^2 = \frac{QME}{r}$ o estimador não viesado de σ_Y^2 e v os graus de liberdade associados a este estimador.

$$\hat{\sigma}_0^2 = \frac{1}{g+v} \left[\sum_{i=1}^g (\bar{Y}_{(i)} - \bar{Y})^2 + v s_Y^2 \right]$$

- iii. Se $\lambda \geq \chi_{(\alpha; g/(\pi-2))}^2$, rejeita-se a hipótese de que os dois grupos são idênticos em favor da hipótese alternativa de que os dois grupos diferem.
- iv. No caso de rejeitar essa hipótese, os dois subgrupos formados serão, independentemente, submetidos aos passos (i) a (iii), fazendo, respectivamente, $g = k_1$ e $g = k_2$. O processo em cada subgrupo se encerra ao aceitar H_0 no passo (iii) ou se cada subgrupo contiver apenas uma média.

Segundo Silva (1998), a grande vantagem deste teste é a ausência de ambigüidade presente nos procedimentos de comparações múltiplas. Para $p = 3$ (3 tratamentos considerados), é necessário repartir as médias ordenadas, em dois grupos, envolvendo as duas médias de maior amplitude para determinar qual a partição que maximiza B_0 .

Ainda segundo esse mesmo autor, para um experimento com p ou $n = 5$ tratamentos, com as médias ordenadas, A, B, C, D e E, existem $(2^{n-1} - 1) = (2^{5-1} - 1) = 15$ partições possíveis destas 5 médias em dois grupos distintos. Pela simplificação de Fisher (1958) é necessário considerar as $(5 - 1) = 4$ partições das médias ordenadas, em dois grupos distintos:

1ª partição:	A	versus	B	C	D	E
2ª partição:	A	B	versus	C	D	E
3ª partição:	A	B	C	versus	D	E
4ª partição:	A	B	C	D	versus	E

Na aplicação do teste, quando muitas médias são consideradas, dificilmente a operação termina com apenas uma partição. Após encontrar a melhor separação entre dois grupos, repete-se o processo em cada subgrupo. A partir daí, prossegue-se com a subdivisão até que os grupos resultantes sejam considerados estatisticamente iguais pelo teste de χ^2 .

Suponha, como ilustração, que a 2ª partição, que maximiza B_0 , foi considerada estatisticamente diferente pelo teste de χ^2 e as demais foram consideradas estatisticamente iguais. Tomando-se então o primeiro subgrupo A B, formam-se novos subgrupos se a partição A versus B for significativa, caso contrário A e B pertencerão ao mesmo grupo.

Em seguida, procede-se a partição do segundo subgrupo C D E da seguinte forma:

1. C versus D E
2. C D versus E

Busca-se a partição que maximize B_0 e calcula-se a estatística λ , verificando se esta foi significativa ou não, em um nível preestabelecido. Se foi significativa, repete-se o processo nos subgrupos formados. Os grupos que forem classificados como estatisticamente iguais receberão as mesmas letras.

2.1.2 Análise de Agrupamentos

Para Bussab, et all (1990), a Análise de Agrupamentos (A.A.) engloba uma variedade de técnicas e algoritmos cujo objetivo é encontrar e separar objetos em grupos similares e segundo Santos (2000), os procedimentos baseados em análise de agrupamento eliminam as ambigüidades dos resultados de comparações de tratamentos, uma vez que nesses métodos trabalha-se com partição de grupos de médias, e não mais com comparações de médias duas a duas, como nos testes mais conhecidos.

Usualmente, pretende-se agrupar objetos semelhantes, segundo suas características (variáveis), mas nada impede que o interesse seja o de agrupar variáveis segundo os valores obtidos pelos objetos, sendo muito importante a definição do objeto, e a variável correspondente (BUSSAB, et all, 1990).

As técnicas de análise multivariada avaliam, segundo Demétrio (1985), simultaneamente, um conjunto de características, levando-se em consideração as correlações existentes entre elas, o que permite que inferências sobre o conjunto de

características sejam feitas em um nível de significância conhecido.

A análise de agrupamento permite classificar indivíduos ou objetos em subgrupos excludentes, em que se pretende, de uma forma geral, maximizar a homogeneidade de objetos ou indivíduos dentro de grupos e maximizar a heterogeneidade entre os grupos, sendo que a representação desses grupos é feita num gráfico com uma estrutura de árvore denominado dendrograma (FERRAUDO, 2006).

A estrutura básica da aplicação de técnicas de A.A. (Análise de Agrupamento), segundo Bussab, et al (1990), pode ser decomposta nas seguintes etapas:

- (i.) Definição de objetivos, critérios, escolha de variáveis e objetos;
- (ii.) Obtenção dos dados;
- (iii.) Tratamento dos dados;
- (iv.) Escolha de critérios de similaridade ou dissimilaridade (parecença);
- (v.) Adoção e execução de um algoritmo de A.A.;
- (vi.) Apresentação dos resultados;
- (vii.) Avaliação e interpretação dos resultados.

Como essas etapas não são independentes, às vezes, torna-se necessário voltar a etapas anteriores para corrigir e aprimorar etapas posteriores.

i. Definição de objetivos, critérios, escolha de variáveis e objetos:

Após investigar o assunto, o pesquisador deve concluir quais as variáveis que serão indicadores próximos da sua característica de interesse. Assim, o objetivo operacional passa a ser o de agrupar os indivíduos da população alvo segundo às variáveis mais facilmente mensuráveis.

Esta é a fase mais importante da A.A., a de fixação dos critérios de homogeneidade, que depende dos objetivos a serem alcançados.

ii. Obtenção dos dados:

A matriz de dados é o material básico para a aplicação das técnicas de Análise de Agrupamentos, que indica os valores das características por objetos de

interesse, convencionadas, geralmente, por indicar os objetos nas linhas e as variáveis nas colunas. Observa-se no Quadro 1, que se pode trabalhar com os dados brutos ou padronizados.

Quadro 1 – Matrizes de dados brutos e padronizados

(a) Brutos	(b) Padronizados
$X = \begin{matrix} & X_1 & X_2 & \cdots & X_p \\ \begin{matrix} o_1 \\ o_2 \\ \vdots \\ o_n \end{matrix} & \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \end{matrix}$	$Z = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{pmatrix}$

Geralmente, pretende-se agrupar objetos semelhantes segundo suas variáveis (características), mas pode-se agrupar variáveis segundo os valores obtidos pelos objetos, se este for o interesse.

iii. Tratamento dos dados:

As técnicas de A.A. podem ser aplicadas a qualquer conjunto de dados, brutos ou padronizados, sendo que a escolha depende dos objetivos, mas é evidente que os resultados são distintos, e dependendo das variáveis envolvidas, o uso dos dados originais pode tornar muito difícil a interpretação do conceito de homogeneidade. Então, o objetivo desta etapa é a padronização da matriz de dados. Segundo Ferraudo (2006), as unidades associadas aos atributos podem, arbitrariamente, afetar o grau de similaridade entre os objetos e com a padronização o efeito dessa arbitrariedade é eliminado, pois esta faz com que os atributos contribuam com o mesmo peso no cálculo do coeficiente de similaridade entre objetos.

iv. Escolha de critérios de similaridade ou dissimilaridade (parecença):

Segundo Johnson & Wichern (2002), quando objetos são agrupados, a proximidade está indicada, geralmente, por alguma medida de similaridade e segundo Ferraudo (2006), a escolha de um coeficiente que quantifique o quanto dois objetos são similares é um conceito fundamental. Têm-se então, duas categorias para esse

conceito: medidas de similaridade e medidas de dissimilaridade.

Segundo Bussab, et all (1990), uma medida de similaridade muito utilizada é o coeficiente de correlação, pois quanto maior seu valor maior a associação, e uma medida de dissimilaridade, também muito utilizada, é a distância euclidiana, pois quanto menor seu valor, mais próximo está um do outro. Para verificar se um objeto A é mais parecido com B ou com C , e se o número de atributos envolvidos for pequeno, a inspeção visual, ou seja, uma representação gráfica da matriz Z , pode responder. Seguindo-se o conceito natural de distância, usar-se-á a distância euclidiana entre os objetos A e B para um espaço bidimensional, definida por:

$$d = \left[(z_1(A) - z_1(B))^2 + (z_2(A) - z_2(B))^2 \right]^{1/2} \quad (1)$$

Generalizando essa distância euclidiana para um espaço de dimensão p , tem-se:

$$d(A, B) = \left[\sum_{i=1}^p (z_i(A) - z_i(B))^2 / p \right]^{1/2} \quad (2)$$

Segundo Ferraudo (2006), a Distância Euclidiana é a mais utilizada, mas tem-se ainda a Distância Euclidiana Média, a Distância de Manhattan, a Distância de Chebychev, a Distância de Potência etc.

A Tabela 1, ilustra como esses agrupamentos podem ser realizados, simplesmente reorganizando uma matriz de distâncias, baseada nos dados alistados na Tabela 2, depois que foram padronizados.

Tabela 1 – Distância Euclidiana entre 10 pares de empresas públicas americanas

Empresa	Ariz	Bost	Loui	Com m	Cons	Flor	Hawa	Idah	Kent	Madi
Ariz	0,00									
Bost	3,10	0,00								
Loui	3,68	4,92	0,00							
Comm	2,46	2,16	4,11	0,00						
Cons	4,12	3,85	4,47	4,13	0,00					
Flor	3,61	4,22	2,99	3,20	4,60	0,00				
Hawa	3,90	3,45	4,22	3,97	4,60	3,35	0,00			
Idah	2,74	3,89	4,99	3,69	5,16	4,91	4,36	0,00		
Kent	3,25	3,96	2,75	3,75	4,49	3,73	2,80	3,59	0,00	
Madi	3,10	2,71	3,93	1,49	4,05	3,83	4,51	3,67	3,57	0,00

Fonte: Adaptado de Johnson & Wichern, 2002, p.672.

Tabela 2 – Dados coletados de 10 empresas públicas americanas, no ano de 1975

Empresa		Variáveis							
		X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Ariz	Arozina Public Service	1,06	9,2	151	54,4	1,6	9077	0,0	0,628
Bost	Boston Edison Co.	0,89	10,3	202	57,9	2,2	5088	25,3	1,555
Loui	Central Louisiana Eletric Co.	1,43	15,4	113	53,0	3,4	9212	0,0	1,058
Comm	Commonwealth Edison Co.	1,02	11,2	168	56,0	0,3	6423	34,3	0,700
Cons	Consolidated Edison Co. (N.Y.)	1,49	8,8	192	51,2	1,0	3300	15,6	2,044
Flor	Florida Power & Light Co.	1,32	13,5	111	60,0	-2,2	11127	22,5	1,241
Hawa	Hawaiian Eletric Co.	1,22	12,2	175	67,6	2,2	7642	0,0	1,652
Idah	Idaho Power Co.	1,10	9,2	245	57,0	3,3	13082	0,0	0,309
Kent	Kentucky Utilities Co.	1,34	13,0	168	60,4	7,2	8406	0,0	0,862
Madi	Madison Gas & Eletric Co.	1,12	12,4	197	53,0	2,7	6455	39,2	0,623

Fonte: Adaptado de Johnson & Wichern, 2002, p.687.

Em que X₁: Proporção analisada (Rendimentos/Débitos);

X₂: Taxa de retorno de capital;

X₃: Custo por capacidade de quilowatt no país;

X₄: Fator anual de carga;

X₅: Crescimento demanda em kWh de 1974 para 1975;

X₆: Vendas (uso do kWh/ano);

X₇: Percentual nuclear; e X₈: Custos total de combustível (centavos/kWh)

v. Adoção e execução de um algoritmo de Análise de Agrupamento:

A escolha de um particular algoritmo de agrupamento exige o conhecimento de suas propriedades aliado aos objetivos da pesquisa. Citam-se dentre muitos, o Método das Médias das Distâncias (MMD), Método da Ligação Simples ou do Vizinho Mais Próximo (MLS) e o Método da Ligação Completa, ou do Vizinho Mais Longe (MLC), que são Métodos Hierárquicos de Agrupamento. Têm-se ainda, os Métodos Não Hierárquicos de Agrupamento que são mais sensíveis a partição inicial, em que iniciando-se o processo com partições diferentes, obtêm-se soluções diferentes.

vi. Apresentação dos resultados:

Difícilmente dois objetos serão considerados exatamente iguais, mas sendo condescendentes no critério de igualdade, pode-se aceitar que sejam “parecidos”.

Geralmente, apresentam-se os dados em uma tabela resumo, que possui uma representação gráfica muito usada em Análise de Agrupamento, conhecida por dendrograma (gráfico com uma estrutura de árvore).

Figura 1 – Dendrograma da análise de agrupamentos, de empresas públicas americanas, realizada a partir dos dados apresentados na Tabela 2

vii. Avaliação e interpretação dos resultados:

O dendrograma pode ser considerado a representação simplificada da matriz de similaridade, mas questiona-se se esta é ou não uma “boa” simplificação. Para responder a esta questão, verifica-se a capacidade do dendrograma em reproduzir a matriz de similaridade, construindo uma matriz cofenética, que é a matriz de distâncias entre os objetos obtidos a partir do dendrograma.

Segundo Bussab, et al (1990), o Coeficiente de Correlação Cofenética é o coeficiente de correlação entre os valores da matriz de similaridade e os correspondentes da matriz de cofenética produzida por uma classificação hierárquica, e quanto mais próximo da unidade melhor será a representação, e quanto mais próximo de zero será pior.

2.1.3 Análise Conjunta de Experimentos em Blocos Casualizados com alguns Tratamentos Comuns

Segundo Pimentel Gomes (1990), o fato se terem apenas alguns tratamentos comuns a todos os experimentos analisados, não impede que a análise conjunta seja realizada, mas exige a utilização de métodos adequados.

O autor apresenta um exemplo para a montagem do Esquema da Análise de Variância deste tipo de experimento, considerando 40 novos híbridos de milho a ensaiar, que se deseja comparar com três variedades ou híbridos já conhecidos, tidos como padrões, A, B e C. Repartindo-se os novos híbridos em 4 grupos de 10, agregam-se as variedades A, B e C, obtendo-se então, os seguintes grupos:

1º grupo: 1, 2, ..., 10, A, B, C;

2º grupo: 11, 12, ..., 20, A, B, C;

3º grupo: 21, 22, ..., 30, A, B, C;

4º grupo: 31, 32, ..., 40, A, B, C.

Obter-se-á de cada grupo, um ensaio em blocos casualizados, com 3 repetições, por exemplo. Para cada um destes grupos, teremos, pois, o seguinte esquema da análise da variância.

Causa de variação	G.L.
Blocos	2
Tratamentos	12
Resíduo	24
Total	38

Após realizada a análise de cada grupo individualmente, verificam-se os valores do maior (QM_1) e do menor (QM_2) quadrado médio do resíduo. Para realizar a análise conjunta destes experimentos, o quociente entre estes valores, segundo Pimentel Gomes (1990), deve ser menor que 7, ou seja,

$$\frac{QM_1}{QM_2} < 7 \quad (3)$$

Supondo-se que no exemplo citado acima, o quociente entre o maior e o menor quadrado médio do resíduo seja menor que 7, obteremos 4 ensaios, e o conjunto destes corresponderá a um experimento em blocos incompletos com $v = 43$ tratamentos (40 novos híbridos e 3 variedades padrões), em blocos de $k = 13$ parcelas, com 3 repetições para os híbridos padrões. A sua análise conjunta obedeceria ao seguinte esquema.

Causa de variação	G.L.
Experimentos (E)	3
Blocos dentro de Experimentos	8
Tratamentos (ajustados)	42
Interação Trat. Comuns X Experimentos	6
Resíduo	96
Total	155

Generalizando, então, para **g** grupos, cada um com **z** tratamentos regulares e **c** tratamentos comuns, num total de **v = gz + c** tratamentos, cada grupo seria ensaiado com **r** blocos de **k = z + c** parcelas e seu esquema da análise seria o seguinte.

Causa de variação	G.L.
Blocos	$r - 1$
Tratamentos	$z + c - 1$
Resíduo	$(r - 1) (z + c - 1)$
Total	$(z + c) r - 1$

A análise conjunta, por sua vez, teria o seguinte esquema.

Causa de variação	G.L.
Experimentos (E)	$g - 1$
Blocos dentro de Experimentos	$g (r - 1)$
Tratamentos (ajustados)	$gz + c - 1$
Interação Trat. Comuns X Experimentos	$(c - 1) (g - 1)$
Redíduo	$g (r - 1) (z + c - 1)$
Total	$(z + c) gr - 1$

As comparações entre tratamentos, segundo esse mesmo autor, seguem as regras seguintes:

a.) contraste entre dois tratamentos regulares do mesmo grupo:

$$\hat{V} = (\hat{m}_i - \hat{m}_j) = \frac{2}{r} s^2$$

b.) contraste entre dois tratamentos regulares de grupos diferentes:

$$\hat{V} = (\hat{m}_i - \hat{m}_j) = \frac{2}{r} \left(1 + \frac{1}{c} \right) s^2$$

c.) contraste entre dois tratamentos comuns:

$$\hat{V} = (\hat{m}_i - \hat{m}_j) = \frac{2}{gr} s^2$$

d.) contraste entre um tratamento comum e um regular:

$$\hat{V} = (\hat{m}_i - \hat{m}_j) = \frac{1}{r} \left(1 + \frac{1}{c} + \frac{1}{g} - \frac{1}{cg} \right) s^2$$

em que s^2 é o QM_{Res} (Quadrado Médio do Resíduo).

As estimativas das médias de tratamentos se obtêm da seguinte maneira:

a.) Para os tratamentos comuns: média aritmética dos dados respectivos, sem nenhum ajuste.

b.) Para os tratamentos regulares: média aritmética dos dados respectivos, da qual se subtrai uma correção **K** que se calcula assim:

K = Média dos tratamentos comuns no experimento respectivo –
Média geral dos tratamentos comuns

2.1.4 Considerações sobre a cana-de-açúcar

Segundo Cesnik & Miocque (2004), a cana-de-açúcar se desenvolve caracteristicamente em forma de touceira. Possui, como a maioria das espécies, uma parte aérea, formada por colmos, folhas e inflorescências, e uma outra subterrânea, constituída de raízes e rizomas.

- **Raízes:** São fasciculadas e muito bem desenvolvidas, podendo atingir até 4 m de profundidade (Figura 2). Segundo Ripoli, et all (2006), o sistema radicular inicia seu desenvolvimento logo após o plantio do rebolo da muda (Figura 3), com as raízes originárias dos primórdios radiculares situados na zona radicular do internódio plantado.

O sistema radicular da cana é totalmente renovado a cada corte, pois

logo após este, o sistema radicular antigo ainda se mantém ativo por um tempo, suprindo as necessidades da touceira, até que os novos perfilhos possam emitir suas próprias raízes (RIPOLI, et all, 2006).

Figura 2 – Sistema radicular da cana planta estabelecido. **Fonte:** Ripoli, et all (2006, p.83)

Figura 3 – Porção basal de uma touceira. **Fonte:** Ripoli, et all (2006, p.80)

- **Rizomas:** Estes possuem nódios, internódios e gemas. Estas são as responsáveis pelo aparecimento dos perfilhos (Figuras 4 e 5), formados na touceira. São os rizomas (Figura 6) que irão rebrotar, após a colheita, dando origem às novas touceiras tanto da soca como da ressoca.

A cana-de-açúcar, é segundo Ripoli, et all (2006), uma planta que perfilha, e esta característica influencia as áreas de plantio pois, define o número de gemas por metro e a capacidade que cada variedade tem de rebrota, por meio das gemas não brotadas, no que restou dos colmos cortados. Cada perfilho comporta-se como uma planta independente, possuindo um sistema radicular próprio.

2.1.5 Programas de Melhoramento da cana-de-açúcar

Segundo Cesnik & Miocque (2004), os programas de melhoramento da cana-de-açúcar são desenvolvidos, em diferentes países, por instituições governamentais e particulares ou em sistema cooperativo formado pelos próprios produtores. Mesmo sendo empreendimentos onerosos, cujos resultados são obtidos a longo prazo, chegando até mais de 15 anos, são muito necessários, uma vez que, as variedades, depois de algum tempo de reprodução vegetativa, perdem o vigor e a produtividade, adquirindo um aumento de suscetibilidade às principais pragas e doenças.

Segundo o artigo de Silva publicado no II Seminário de Biotecnologia Agrícola (1984), muitos fatores contribuem na obtenção de alta produtividade da cana-de-açúcar, mas a utilização de variedades geneticamente melhoradas é o sistema mais eficiente para se obterem melhores resultados, mas como cada variedade exige um manejo diferenciado, torna-se necessária uma tecnologia específica.

O sistema reprodutivo da variedade vegetal determina as linhas gerais da metodologia a ser empregada no seu processo de melhoramento e considerando-se esse sistema reprodutivo, as espécies vegetais podem ser classificadas em autógamas e alógamas. Estas se referem às plantas que se reproduzem por polinização cruzada, e aquelas, às que o fazem por autofecundação. Os procedimentos de seleção utilizados para o melhoramento das variedades alógamas são os mesmos daqueles utilizados para os das autógamas, porém, os resultados obtidos no final do melhoramento são diferentes, uma vez que a estrutura populacional entre os dois tipos de reprodução também diferem entre si (CESNIK & MIOCQUE, 2004).

A cana-de-açúcar apresenta dois tipos de reprodução: sexuada e assexuada. A reprodução assexuada é a maneira comercial de multiplicá-la, e a sexuada é utilizada em programas de melhoramento, mas que não prescindem, de maneira alguma, da assexuada (CESNIK & MIOCQUE, 2004).

Alguns programas brasileiros de melhoramento de cana-de-açúcar adotaram o método do “melting” havaiano para os seus cruzamentos, cujo processo é mais econômico, quando comparado ao tradicional policruzamento, e de eficiência redobrada. Estes programas, não descartaram, porém, a utilização de cruzamentos biparentais. No processo de “melting pot”, necessita-se de um número muito menor de operários do que os utilizados nos policruzamentos, e a germinação das sementes é melhor naqueles do que nestes. No “melting pot”, os colmos são colocados em baldes ou em calhas com solução nutritiva, trocada periodicamente (CESNIK & MIOCQUE, 2004).

2.2 Material e Métodos

Segundo Raizer, et all (2006), as amostragens para análise tecnológica nos ensaios devem ser feitas preferencialmente em cana crua, retirando-se em todas as repetições, 10 colmos por parcela; deve-se adentrar pelo menos 1 metro na parcela para retirar os 10 colmos seguidos de cada repetição. Se a presença de brotões for uma característica geral da parcela, estes devem fazer parte do feixe de amostragem. Normalmente, essa amostragem é feita antes da queima da cana, porém em condições de muito tombamento, principalmente na cana-planta instalada em solos de alta fertilidade. Pode-se fazer a coleta das amostras após a queima, mas nesse caso, as amostras devem ser processadas e analisadas o mais rápido possível. Os colmos devem ser despalhados, identificando-se cada feixe antes de mandar para análise, que deverá ser feita no máximo 48 horas após o corte.

Ainda segundo esses autores, as amostras devem ser recebidas pelo laboratório em feixes de 10 colmos, que são desintegrados no desintegrador tipo forrageira, devidamente regulado de forma a obter, o que se denomina de, um bom

preparo. Em cada amostra desintegrada é feita a homogeneização das diferentes partes dos colmos (pé, meio e ponta), o que é de fundamental importância para a obtenção de resultados analíticos confiáveis.

Os experimentos, referentes a este estudo, foram conduzidos pelo Centro de Tecnologia Canavieira, considerando-se um Delineamento em Blocos Casualizados com quatro Tratamentos Comuns, como apresentado na Figura 7. Os tipos T0051, T0052 e T0053 apresentados nesse delineamento, são um controle local, ou seja, uma divisão horizontal da área experimental, e os quadrados coloridos na figura, representam os tratamentos comuns. Tem-se, portanto, 31 tratamentos diferentes em cada tipo, totalizando 93 variedades regulares, com duas repetições (blocos) e 4 tratamentos comuns, repetidos nos blocos e nos tipos.

Figura 7 – Delineamento experimental em uma Unidade Experimental utilizado pelo Centro de Tecnologia Canavieira (CTC)

Para a realização das análises é feita a extração do caldo na cana desintegrada através da prensa hidráulica devidamente regulada (tempo e pressão). O bagaço resultante é denominado de **bolo úmido**, o qual é pesado e depois seco, pelo menos, por 6 horas a 105-110°C, resultando no **bolo seco**, através do qual por

diferença determina-se a **fibra%cana**.

No caldo extraído são realizadas as seguintes análises tecnológicas: **Brix%caldo**, que é medido em refratômetro e **Pol%caldo**, medido em sacarímetro.

O Centro de Tecnologia Canavieira (CTC) planta, conduz, colhe e analisa, nas usinas associadas, em torno de 400 ensaios de competição de variedades/clones por ano, seguindo todos os princípios básicos da experimentação.

Os dados cedidos pelo Centro de Tecnologia Canavieira, compreendem 11 Locais, 3 Tipos, 2 Blocos, 97 Variedades, sendo 4 variedades padrões ou comuns e 93 variedades regulares, e as variáveis TCH, POL, FIB e TPH5C, para a colheita do ano de 2004, totalizando 2310 observações, em que:

- TCH – Tonelada de cana-de-açúcar por hectare;
- POL – Porcentagem de açúcar, obtida através da fórmula apresentada no quadro 2, envolvendo a variável FIB;
- FIB – Porcentagem de fibra;
- TPH5C – Tonelada de POL por hectare estimada para 5 cortes, obtida em função das variáveis POL e TCH, através da fórmula apresentada no Quadro 2.

Os 11 Locais, descritos como Unidades Experimentais pertencentes ao estado de São Paulo, são:

- L006 – Região de Jaú
- L013 – Catanduva
- L020 – Ribeirão Preto
- L029 – Piracicaba
- L045 – Ribeirão Preto
- L055 – Assis
- L057 – Ribeirão Preto
- L058 – Piracicaba
- L075 – Piracicaba
- L076 – Assis
- L078 – Ribeirão Preto

Todos os resultados das análises são então encaminhados para o banco de dados, após o que são elaboradas as análises estatísticas. As variáveis utilizadas para este estudo são Local, Tipo, Bloco, Variedade, TCH, POL, FIB e TPH5C.

Quadro 2 – Fórmulas utilizadas na análise dos ensaios pelo CTC

<p>PBU = peso do bolo úmido (kg)</p> <p>PBS = peso do bolo seco (kg)</p> <p>Lpol = leitura de pol do sacarímetro (quantidade de açúcar)</p> <p>BCE = brix % caldo do refratômetro</p> <p>TCH = tonelada de cana por hectare</p>	<p>POL = pol % cana</p> <p>PUR = pureza do caldo (%)</p> <p>FIB = fibra % cana</p>
<p> $\text{POL (pol\%cana)} = \text{SCE} \cdot (1 - 0,01 \cdot \text{FIB}) \cdot (1,0313 - 0,00575 \cdot \text{FIB}) \cdot 4$ em que: $\text{SCE (pol\%caldo)} = \text{Lpol} \cdot (0,26047 - 0,0009882 \cdot \text{BCE})$ $\text{FIB} = \text{fibra\%cana}$ <p>a) se $\text{PBS} = 0 \Rightarrow \text{FIB} = 0,09079 \cdot \text{PBU} - 0,03503 \cdot \text{BCE} - 0,09745$</p> <p>b) se $\text{PBS} \neq 0 \Rightarrow \text{FIB} = 100 \cdot \text{PBS} - \text{PBU} \cdot \text{BCE} / 5 \cdot (100 - \text{BCE})$</p> $\text{PUR (pureza do caldo)} = 100 \cdot \text{SCE} / \text{BCE}$ </p>	
<p>TPH = toneladas de pol por hectare</p> $\text{TPH} = \text{TCH} \times \text{Pol\%cana}$	
<p>TCH5c = toneladas de cana estimada para 5 cortes</p> $\text{TCH5c (2 cortes)} = (\text{TCH c1} + 3,5 \cdot \text{TCH c2}) / 5$ $\text{TCH5c (3 cortes)} = (\text{TCH c1} + \text{TCH c2} + 2,8 \cdot \text{TCH c3}) / 5$	
<p>TPH5c = toneladas de pol por hectare estimada para 5 cortes</p> $\text{TPH5c} = \text{TCH5c} \cdot \text{POL(médio)} / 100$	

Fonte: RAIZER, et all (2006, p.24).

Os métodos de comparações múltiplas baseados em análise de agrupamento univariada, têm por objetivo separar as médias de tratamentos em grupos homogêneos, pela minimização da variação dentro, e maximização entre grupos e um desses procedimentos é o teste de Scott-Knott (SANTOS, 2000).

Este teste será aplicado, através do programa SISVAR, sempre que o teste F, para locais, for significativo, para as variáveis TCH, POL, FIB e TPH5C. Esta última envolve em sua fórmula as anteriores.

Antes do teste ser aplicado, é necessário, fazer as análises de variância das variáveis TCH, POL, FIB e TPH5C, para cada tipo, em cada local, a fim de observar a relação entre os QM_{Res} (Quadrado Médio do Resíduo), tendo em vista a realização da análise de variância conjunta para os três tipos (T0051, T0052 e T0053), com tratamentos comuns.

Da mesma maneira, deve-se obter o quociente entre o maior e o menor quadrado médio do resíduo dessas análises conjuntas para tipos, em cada local, visando a possibilidade de se reunirem os 11 locais em uma análise conjunta.

O método para Análise Multivariada de Agrupamento utilizado foi o Método de Ligação Média, que é um método hierárquico de agrupamento. A vantagem desse método, está em que a análise foi feita para as variáveis TCH, POL e FIB ao mesmo tempo, através do programa SAS, e também para a variável TPH5C como um comparativo ao teste de Scott-Knott.

2.3 Resultados e Discussão

Verificou-se que estão satisfeitas as condições para as análises conjuntas, reunindo-se os 3 tipos em cada local, pois, como sugere Pimentel Gomes (1990), o quociente entre o maior e o menor QM_{Res} deve ser inferior a sete. Sendo assim, deu-se prosseguimento à análise de variância para cada local. Do mesmo modo, o quociente entre o maior e o menor QM_{Res} das análises conjuntas para tipos foi menor que sete, para todas as variáveis em estudo, o que permitiu que se realizassem as análises conjuntas para locais.

As Análises Conjuntas de Variância, das variáveis TCH, POL, FIB e TPH5C para locais, são apresentadas, respectivamente, nas Tabela 3, 4, 5 e 6. Embora o objetivo principal dessas análises seja a obtenção de uma estimativa de σ^2 , pelo QM_{Res} , para que se possa aplicar o teste de Scott-Knott, pode-se verificar que o

QM_{Locais} é significativamente superior ao $QM_{Int L \times V}$. Esse fato permite que se estudem as médias de locais, independentemente de variedades.

Tabela 3 – Análise Conjunta de Variância da variável TCH

Causa de Variação	gL	SQ	QM
Locais (L)	10	739745,516883	73974,5517
Tipos d. Locais	22	11438,600000	519,9364
Variedades (V)	96	163192,287446	1699,9197
Int. VarComuns X Tipos	6	345,310606	57,5518
Int. L X V	960	140536,664935	146,3924
Blocos d. Tipos d. Locais	33	7843,742858	237,6892
Int. VarComuns X Tipos X Locais	60	3007,022727	50,1170
Resíduo Médio	1122	40276,657142	35,8972
Total	2309	1106385,802597	-----

CV (%) = 6,76

Média geral: 96,09

Tabela 4 – Análise Conjunta de Variância da variável POL

Causa de Variação	gL	SQ	QM
Locais (L)	10	482,363761	48,2364
Tipos d. Locais	22	46,220331	2,1009
Variedades (V)	96	1061,179337	11,0540
Int. VarComuns X Tipos	6	0,374681	0,0624
Int. L X V	960	422,612754	0,4402
Blocos d. Tipos d. Locais	33	8,465495	0,2565
Int. VarComuns X Tipos X Locais	60	11,452694	0,1909
Resíduo Médio	1122	156,750865	0,1397
Total	2309	2189,419918	-----

CV (%) = 2,53

Média geral: 15,06

Tabela 5 – Análise Conjunta de Variância da variável FIB

Causa de Variação	gL	SQ	QM
Locais (L)	10	408,840605	40,8841
Tipos d. Locais	22	22,950270	1,0432
Variedades (V)	96	2233,528325	23,2659
Int. VarComuns X Tipos	6	1,783216	0,2972
Int. L X V	960	372,906710	0,3884
Blocos d. Tipos d. Locais	33	17,614957	0,5338
Int. VarComuns X Tipos X Locais	60	8,485459	0,1414
Resíduo Médio	1122	137,225615	0,1223
Total	2309	3203,335157	-----

CV (%) = 3,48

Média geral: 10,57

Tabela 6 – Análise Conjunta de Variância da variável TPH5C

Causa de Variação	gL	SQ	QM
Locais (L)	10	14513,881378	1451,3881
Tipos d. Locais	22	264,663171	12,0301
Variedades (V)	96	2619,382019	27,2852
Int. VarComuns X Tipos	6	13,227575	2,2046
Int. L X V	960	3782,763425	3,9404
Blocos d. Tipos d. Locais	33	171,784587	5,2056
Int. VarComuns X Tipos X Locais	60	67,843408	1,1307
Resíduo Médio	1122	997,940843	0,8894
Total	2309	22431,486406	-----

CV (%) = 8,32

Média geral: 12,16

Pelo teste de Scott-Knott, para a variável TCH, foram formados 9 agrupamentos para os 11 Locais estudados, da seguinte maneira:

- Grupo 1 L029
- Grupo 2 L058
- Grupo 3 L055
- Grupo 4 L075
- Grupo 5 L045 e L020
- Grupo 6 L076 e L006
- Grupo 7 L057
- Grupo 8 L078
- Grupo 9 L013

Para a variável POL, foram formados 6 agrupamentos para os 11 Locais estudados, por esse mesmo teste, da seguinte maneira:

- Grupo 1 L013
- Grupo 2 L076, L058, L055 e L078
- Grupo 3 L020, L029 e L075
- Grupo 4 L045
- Grupo 5 L057
- Grupo 6 L006

E para a variável FIB, foram formados 8 agrupamentos para os 11 Locais estudados, da seguinte maneira:

- Grupo 1 L029
- Grupo 2 L075 e L076
- Grupo 3 L078
- Grupo 4 L013 e L058
- Grupo 5 L045
- Grupo 6 L006 e L020
- Grupo 7 L057
- Grupo 8 L055

Como o teste de Scott-Knott é realizado individualmente para cada variável, obtiveram-se 3 agrupamentos completamente diferentes, dificultando a decisão para qual melhor agrupamento.

Sendo assim, optou-se em aplicar este mesmo teste considerando-se apenas a variável TPH5C, por se tratar de uma variável linearmente dependente de TCH, POL e FIB. Observa-se que foram formados 9 agrupamentos para os 11 Locais estudados, da seguinte maneira:

Grupo 1	L029
Grupo 2	L058
Grupo 3	L055
Grupo 4	L075
Grupo 5	L076 e L020
Grupo 6	L045 e L006
Grupo 7	L057
Grupo 8	L078
Grupo 9	L013

Para a Análise de Agrupamentos, obteve-se a Matriz de Similaridades, através da Distância Euclidiana, a partir das três variáveis em estudo: TCH, POL e FIB, descrita na Tabela 7.

Seguindo-se os critérios de uma Análise de Agrupamento, obteve-se um dendrograma, pelo Método de Ligação das Médias, descrito na Figura 8, gerado pelo programa SAS.

Tabela 7 – Matriz de Similaridades entre Locais, para as variáveis TCH, POL e FIB, obtida através da Distância Euclidiana

Locais	L006	L013	L020	L029	L045	L055	L057	L058	L075	L076	L078
L006	0,00000										
L013	1,24616	0,00000									
L020	1,17102	0,07515	0,00000								
L029	3,00319	1,75702	1,83217	0,00000							
L045	1,15402	0,09214	0,01700	1,84917	0,00000						
L055	0,82870	0,41746	0,34232	2,17449	0,32532	0,00000					
L057	0,28382	1,52999	1,45484	3,28701	1,43784	1,11252	0,00000				
L058	2,50106	1,25489	1,33004	0,50213	1,34704	1,67236	2,78488	0,00000			
L075	1,91802	0,67185	0,74700	1,08517	0,76400	1,08932	2,20184	0,58304	0,00000		
L076	2,07987	0,83370	0,90885	0,92332	0,92585	1,25117	2,36369	0,42119	0,16185	0,00000	
L078	1,37469	0,12852	0,20367	1,62850	0,22067	0,54599	1,65851	1,12637	0,54333	0,70518	0,00000

Os agrupamentos com menor distância entre os Locais, são observados através dessa matriz de similaridades, e isto ocorre entre os locais 020 e 045, formando-se um primeiro grupo. Isso pode ser observado, mais facilmente, através do dendrograma (Figura 8) e da análise de agrupamento gerado pelo programa SAS (Quadro 3).

O dendrograma, apresentado na Figura 8, gerado a partir da Matriz de Similaridades, considerando-se uma distância máxima de 0,50, apresenta os seguintes agrupamentos:

- Grupo 1 L006 e L057
- Grupo 2 L020, L045 e L075
- Grupo 3 L076 e L078
- Grupo 4 L029
- Grupo 5 L058
- Grupo 6 L055
- Grupo 7 L013

Figura 8 – Dendrograma do processo hierárquico de agrupamentos de Locais, para as variáveis TCH, POL e FIB – Método de Ligação das Médias

Quadro 3 – Análise de Agrupamento para Locais, para as variáveis, TCH, POL e FIB, realizada através do programa SAS

Cluster History				
NCL	Clusters Joined		Freq.	Dist.
10	L020	L045	2	0,1448
9	L006	L057	2	0,3458
8	L076	L078	2	0,3927
7 CL10		L075	3	0,4409
6 CL7	CL8		5	0,5621
5	L029	L058	2	0,6680
4 CL6	CL5		7	0,7847
3 CL9	CL4		9	1,0142
2 CL3		L055	10	1,2410
1 CL2		L013	11	1,3250

Baseado nesta análise (Quadro 3) e no dendrograma (Figura 8), gerou-se uma matriz cofenética, conforme descrito na Tabela 8, para confirmar se o agrupamento gerado apresenta um bom ajuste.

Segundo Bussab, et all (1990), deve-se verificar a proximidade das duas matrizes, através do coeficiente de correlação entre os valores da matriz de similaridades (Tabela 7) e os correspondentes da matriz cofenética (Tabela 8). Quanto mais próximo da unidade será melhor a representação, e quanto mais próximo de zero será pior.

Tabela 8 – Matriz Cofenética baseada no dendrograma da Figura 8

Locais	006	013	020	029	045	055	057	058	075	076	078
006	0,0000										
013	1,3250	0,0000									
020	1,0142	1,3250	0,0000								
029	1,0142	1,3250	0,7847	0,0000							
045	1,0142	1,3250	0,1448	0,7847	0,0000						
055	1,2410	1,3250	1,2410	1,2410	1,2410	0,0000					
057	0,3458	1,3250	1,0142	1,0142	1,0142	1,2410	0,0000				
058	1,0142	1,3250	0,7847	0,6680	0,7847	1,2410	1,0142	0,0000			
075	1,0142	1,3250	0,4409	0,7847	0,4409	1,2410	1,0142	0,7847	0,0000		
076	1,0142	1,3250	0,5621	0,7847	0,5621	1,2410	1,0142	0,7847	0,5621	0,0000	
078	1,0142	1,3250	0,5621	0,7847	0,5621	1,2410	1,0142	0,7847	0,5621	0,3927	0,0000

O coeficiente de correlação obtido para estas matrizes, $\text{corr}(\text{Simil}, \text{Cofen}) = 0,216364$, não pode ser considerado um bom ajuste, pois está muito distante da unidade. Logo, o agrupamento entre os locais, apresentado no dendrograma da Figura 8, não pode ser considerado como um bom agrupamento.

Sendo assim, optou-se em refazer a análise de agrupamentos, para a variável TPH5C, e analisar o ajuste obtido, obtendo-se então a matriz de similaridades para esta variável, observada na Tabela 9.

Seguindo-se os critérios de uma Análise de Agrupamento, obteve-se um dendrograma, descrito na Figura 9, e uma análise de agrupamento pelo método de

Ligação das Médias, descrita no Quadro 4, gerados através do programa SAS.

Tabela 9 – Matriz de Similaridades entre Locais, para a variável TPH5C, obtida através da Distância Euclidiana

Locais	006	013	020	029	045	055	057	058	075	076	078
006	0,00000										
013	2,03599	0,00000									
020	0,15983	2,19582	0,00000								
029	1,49179	3,52778	1,33195	0,00000							
045	0,05328	2,08927	0,10656	1,43851	0,00000						
055	0,75351	2,78949	0,59367	0,73828	0,70023	0,00000					
057	0,95520	1,08079	1,11504	2,44699	1,00848	1,70871	0,00000				
058	0,84484	2,88083	0,68501	0,64695	0,79156	0,09133	1,80004	0,00000			
075	0,35392	2,38991	0,19408	1,13787	0,30064	0,39959	1,30912	0,49092	0,00000		
076	0,19028	2,22627	0,03044	1,30151	0,13700	0,56323	1,14548	0,65456	0,16364	0,00000	
078	1,11884	0,91715	1,27868	2,61063	1,17212	1,87235	0,16364	1,96368	1,47276	1,30912	0,00000

Figura 9 – Dendrograma do processo hierárquico de agrupamento de Locais, para a variável TPH5C – Método de Ligação das Médias

Baseado na análise apresentada no Quadro 4, e no dendrograma apresentado na Figura 9, gerou-se uma matriz cofenética, conforme descrito na Tabela 10, para

confirmar se o agrupamento gerado apresenta um bom ajuste.

Quadro 4 – Análise de Agrupamento para Locais, para a variável TPH5C, realizada através do programa SAS

Cluster History					
NCL	Clusters Joined		Freq.	Dist.	
10	L020	L076	2	0,0215	
9	L006	L045	2	0,0377	
8	L055	L058	2	0,0646	
7 CL9	CL10		4	0,1072	
6	L057	L078	2	0,1157	
5 CL7		L075	5	0,1871	
4 CL5	CL8		7	0,4672	
3	L013	CL6	3	0,7087	
2 CL4		L029	8	0,8460	
1 CL2	CL3		11	1,3934	

Tabela 10 – Matriz Cofenética baseada no dendrograma da Figura 9

Locais	006	013	020	029	045	055	057	058	075	076	078
006	0,00000										
013	1,39340	0,00000									
020	0,10720	1,39340	0,00000								
029	0,84600	1,39340	0,84600	0,00000							
045	0,03770	1,39340	0,10720	0,84600	0,00000						
055	0,46720	1,39340	0,46720	0,84600	0,46720	0,00000					
057	1,39340	0,70870	1,39340	1,39340	1,39340	1,39340	0,00000				
058	0,46720	1,39340	0,46720	0,84600	0,46720	0,06460	1,39340	0,00000			
075	0,18710	1,39340	0,18710	0,84600	0,18710	0,46720	1,39340	0,46720	0,00000		
076	0,10720	1,39340	0,02150	0,84600	0,10720	0,46720	1,39340	0,46720	0,18710	0,00000	
078	1,39340	0,70870	1,39340	1,39340	1,39340	1,39340	0,11570	1,39340	1,39340	1,39340	0,00000

O coeficiente de correlação obtido para essas matrizes, $\text{corr}(\text{Simil}, \text{Cofen}) = 0,856423$, permite que se considere como um bom ajuste. A partir do dendrograma apresentado na Figura 9, e considerando-se uma distância máxima de 0,50, apresentam-se, então, o agrupamento de locais, da seguinte maneira:

Grupo 1 L006, L045, L020, L076, L055 e L058

Grupo 2 L029

Grupo 3 L013

Grupo 4 L057 e L078

Tomando-se, as menores distâncias entre locais, obtêm-se os agrupamentos entre os locais L020 e L076, entre os locais L006 e L045 e entre os locais L055 e L058.

3 CONSIDERAÇÕES FINAIS

Comparando-se o teste de Scott-Knott com a Análise de Agrupamentos, confirmam-se os agrupamentos entre os locais L020 e L076 e os locais L045 e L006.

Por se tratar de um teste subjetivo, é o pesquisador que determina a distância máxima para os agrupamentos, dependendo de seu interesse.

O teste de Scott-Knott, é um teste para Análise Univariada, portanto, indicado quando se tem apenas uma variável em estudo. E ao se trabalhar apenas com a variável TPH5C, este teste mostrou-se muito rígido, pois apresentou um número menor de agrupamentos que a Análise de Agrupamentos.

Face aos resultados, conclui-se que podem ser eliminadas dos experimentos duas unidades experimentais, optando por L020 (Ribeirão Preto) ou L076 (Assis), e L045 (Ribeirão Preto) ou L006 (Região de Jaú), ficando essa escolha, a critério do pesquisador, podendo assim, reduzir seu custo operacional.

REFERÊNCIAS

BUSSAB, W. O.; MIAZAKI, E. S.; ANDRADE, D. F. Introdução à Análise de Agrupamentos. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9., 1990, São Paulo. **Anais...** São Paulo: ABE, 1990, 105p.

CAMPOS, H. **Estatística Aplicada à Experimentação com Cana-de-açúcar**. São Paulo: FEALQ, p.89-98, 1984.

CENTRO DE TECNOLOGIA CANAVIEIRA. Disponível em: <<http://www.ctcanavieira.com.br/php/pagina.php?doc=home>>. Acesso em: 26 jun. 2006.

CESNIK, R.; MIOCQUE, J. **Melhoramento da Cana-de-açúcar**. Brasília: Embrapa Informação Tecnológica, 2004, 307p.

COBRA, N. **A Semente da Vitória**. 58ª ed. São Paulo: Editora Senac São Paulo, 2004. 223p.

DEMÉTRIO, C. G. B. **Análise multidimensional para dados de cana-de-açúcar**. 2002. 126p. Tese (Doutorado em Estatística e Experimentação Agrônômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2002.

FERRAUDO, A. S. **Técnicas de Análise Multivariada**. 2ª ed. São Caetano do Sul: StatSoft South America, p.15-27, 2006.

FERREIRA, D. F. **Manual do Sistema Sisvar para Análises Estatísticas**. Lavras: Universidade Federal de Lavras, 2000, 66p.

FISHER, W. K. On grouping for maximum homogeneity. **Journal American Statistical Association**, Washington, v.53, p.789-798, Dec., 1958.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 5th ed. New Jersey: Prentice Hall, p.668-694, 2002.

PIMENTEL GOMES, F. **Curso de Estatística Experimental**. 13^a ed. Piracicaba: Nobel, p.316-333, 1990.

RAIZER, A. J.; SORDI, R. A.; BRAGA JR, R. L. C. **Técnicas de Plantio, condução e análise de ensaios de cana-de-açúcar**. Piracicaba: Centro de Tecnologia Canavieira, 2006, 26p.

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em genética e melhoramento de plantas**. Lavras: UFLA, p.87-134, 2000.

RIPOLI, T. C. C.; RIPOLI, M. L. C.; CASAGRANDE, D. V.; IDE, B. Y. **Plantio de Cana-de-Açúcar: estado da arte**. Piracicaba: T.C.C. Ripoli, p.80-84, 2006.

SANTOS, C. **Novas Alternativas de Testes de Agrupamentos Avaliadas por meio de Simulação Monte Carlo**. 2000. 85p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2000.

SAS SYSTEM. **The SAS System for Windows Release 8.02**. Licensed to Massey University, Cary, 1999-2001.

SCOTT, A. J.; KNOTT, M. A. A cluster analysis method for grouping means in the analysis of variance. **Biometrics**. Raleigh, v.30, n.3, p.507-512, Sept. 1974.

SILVA, E. C. **Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo**. 1998. 54p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 1998.

SILVA, W. M. Perspectiva da cultura da cana-de-açúcar no Brasil. In: SEMINÁRIO DE BIOTECNOLOGIA AGRÍCOLA, 1984, Piracicaba. **Anais...** Piracicaba: ESALQ, 1984. p.47-56.

SISVAR. **Sisvar versão 4.6 (Build 63)**. Copyright Daniel Furtado Ferreira, DEX/UFLA, 1999-2003.

TECNOLOGIAS CTC. **Índice de Tecnologias Disponíveis na Área Agrícola**. Disponível em: <<http://www.sistemas.ctc.com.br/Portal/PortPublic?acao=ListaTecno>>. Acesso em: 26 jun. 2006.