

JOEL CAMILO SOUZA CARNEIRO

**ANÁLISE ESTATÍSTICA MULTIVARIADA APLICADA À  
AVALIAÇÃO SENSORIAL DE ALIMENTOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós Graduação em Ciência e Tecnologia de Alimentos, para obtenção do título de “*Doctor Scientiae*”.

VIÇOSA  
MINAS GERAIS – BRASIL  
2005

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

T

C289a  
2005

Carneiro, Joel Camilo Souza, 1974-

Análise estatística multivariada aplicada à avaliação  
sensorial de alimentos / Joel Camilo Souza Carneiro.  
– Viçosa : UFV, 2005.  
xii, 89f. : il. ; 29cm.

Orientador: Valéria Paula Rodrigues Minim.  
Tese (doutorado) - Universidade Federal de  
Viçosa.

Inclui bibliografia.

1. Alimentos - Avaliação sensorial - Métodos  
estatísticos. 2. Análise multivariada. 3. Análise de  
componentes principais. 4. Análise fatorial.  
I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 664.07

JOEL CAMILO SOUZA CARNEIRO

**ANÁLISE ESTATÍSTICA MULTIVARIADA APLICADA À  
AVALIAÇÃO SENSORIAL DE ALIMENTOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós Graduação em Ciência e Tecnologia de Alimentos, para obtenção do título de "*Doctor Scientiae*".

Aprovada em: 24 de novembro de 2005.

---

Prof. Adair José Regazzi  
(Conselheiro)

---

Prof. Carlos Henrique Osório Silva  
(Conselheiro)

---

Prof. Luis Antonio Minim

---

Sérgio Henriques Saraiva

---

Profa. Valéria Paula Rodrigues Minim  
(Orientadora)

## **DEDICATÓRIA**

Aos meus pais José Maria (Ieié) e Maria de Lourdes (Filita).

À Vanderléia e Gabriela.

Aos meus irmãos.

Aos meus amigos.

## **AGRADECIMENTOS**

A Deus, pela saúde, paz e perseverança no meu dia a dia.

Aos meus familiares, pela boa convivência e receptividade, sempre.

Aos meus pais, meus irmãos, minha esposa e minha filha, pela amizade, auxílio e fonte de carinho inesgotável.

Aos meus amigos e colegas da época do Colégio Universitário, da graduação e do mestrado. Em especial ao meu irmão João de Deus e Manoel Messias, pela boa convivência e companheirismo.

Aos meus amigos de Calambau, Cruzes, Viçosa e Itapetinga.

Ao Sr. José Antônio Filho, Netinha, Valéria, Vágner, Ediana e Éder, pela amizade.

À família Bonomo (Paulo, Renata e Helena), ao João de Deus e Leo, pelo apoio, amizade e boa convivência, sempre.

À Universidade Federal de Viçosa, por intermédio do Departamento de Tecnologia de Alimentos, pela oportunidade de realizar este curso.

À Universidade Estadual do Sudoeste da Bahia, em especial às plenárias do DTRA e do Colegiado do Curso de Engenharia de Alimentos, que sempre me apoiaram permitindo que eu desse andamento ao meu curso de doutorado.

Ao CNPq e FAPESB, pelo apoio financeiro.

À professora Valéria Paula Rodrigues Minim, pela orientação, auxílio, convivência e amizade nos últimos anos.

Aos laboratoristas, secretárias e demais funcionários do Departamento de Tecnologia de Alimentos, pela ajuda e amizade.

À doutora Sílvia Yokota pela disponibilização dos dados experimentais.

Aos professores da UESB que me substituíram possibilitando que eu concluísse a tese.

Ao professor e amigo Adair José Regazzi pelos valiosos ensinamentos que muito tem contribuído para minha formação. Pela convivência e amizade.

Aos professores Carlos Henrique Hosório Silva e José Benício Paes Chaves pelas sugestões e atenção sempre que foi necessário.

Aos membros da banca de defesa de tese.

A todos os meus mestres, pela contribuição em minha formação, inclusive os do primário, onde tudo começou.

Aos professores Cosme Damião Cruz, Pedro Carneiro e Adair Regazzi, pela valiosa contribuição que têm dado para o uso cada vez crescente da análise estatística multivariada, seja pelas disciplinas lecionadas, materiais didáticos elaborados ou publicações científicas.

## **PREFÁCIO**

Esta tese é composta por três artigos redigidos de forma independente, embora as conclusões sejam interligadas. Além de introdução e conclusão geral.

Cada artigo ilustra uma das técnicas multivariadas: variáveis canônicas, análise de fatores e análise de componentes principais, e apresenta os programas para execução das análises no Statistical Analysis System (SAS) a partir de um único conjunto de dados.

## ÍNDICE

|                                                                                             | Página |
|---------------------------------------------------------------------------------------------|--------|
| RESUMO.....                                                                                 | viii   |
| ABSTRACT.....                                                                               | x      |
| INTRODUÇÃO GERAL.....                                                                       | 1      |
| ARTIGO 1 – VARIÁVEIS CANÔNICAS APLICADAS À AVALIAÇÃO<br>SENSORIAL DE ALIMENTOS.....         | 3      |
| RESUMO.....                                                                                 | 3      |
| SUMMARY.....                                                                                | 4      |
| 1. INTRODUÇÃO.....                                                                          | 5      |
| 2. MATERIAL E MÉTODOS.....                                                                  | 6      |
| 2.1. DADOS EXPERIMENTAIS.....                                                               | 6      |
| 2.2. METODOLOGIA.....                                                                       | 6      |
| 2.2.1. ANÁLISE DE VARIÂNCIA, MATRIZES E e H,<br>AUTOVALORES e AUTOVETORES (PROGRAMA 1)..... | 9      |
| 2.2.2. MÉDIAS CANÔNICAS E GRÁFICO DE DISPERSÃO<br>(PROGRAMA 2).....                         | 11     |
| MÉTODO DE AGRUPAMENTO DE TOCHER.....                                                        | 13     |



|                                                                                               |    |
|-----------------------------------------------------------------------------------------------|----|
| 2.2.3. IMPORTÂNCIA RELATIVA DOS ATRIBUTOS (PROGRAMA 3).....                                   | 14 |
| 2.2.4. ESTIMATIVAS DAS CORRELAÇÕES ENTRE OS ATRIBUTOS (PROGRAMA 4).....                       | 16 |
| 2.2.5. ANOVA E TESTE DE MÉDIAS PARA OS GRUPOS (PROGRAMA 5).....                               | 17 |
| 3. RESULTADOS E DISCUSSÃO.....                                                                | 20 |
| 4. CONCLUSÕES.....                                                                            | 33 |
| 5. REFERÊNCIAS BIBLIOGRÁFICAS.....                                                            | 35 |
| ARTIGO 2 – ANÁLISE DE FATORES APLICADA À AVALIAÇÃO SENSORIAL DE ALIMENTOS.....                | 36 |
| RESUMO.....                                                                                   | 36 |
| SUMMARY.....                                                                                  | 37 |
| 1. INTRODUÇÃO.....                                                                            | 38 |
| 2. MATERIAL E MÉTODOS.....                                                                    | 39 |
| 2.1. DADOS EXPERIMENTAIS.....                                                                 | 39 |
| 2.2. METODOLOGIA.....                                                                         | 39 |
| 2.2.1. ANÁLISES PRELIMINARES.....                                                             | 39 |
| a. ANÁLISE DE VARIÂNCIA, MÉDIAS DOS ATRIBUTOS E ANÁLISE DE CORRELAÇÃO (PROGRAMA 1).....       | 39 |
| b. PADRONIZAÇÃO DAS MÉDIAS (PROGRAMA 2).....                                                  | 42 |
| 2.2.2. ANÁLISE DE FATORES (PROGRAMA 3).....                                                   | 44 |
| MÉTODO DE AGRUPAMENTO DE TOCHER.....                                                          | 49 |
| 3. RESULTADOS E DISCUSSÃO.....                                                                | 49 |
| 4. CONCLUSÕES.....                                                                            | 59 |
| 5. REFERÊNCIAS BIBLIOGRÁFICAS.....                                                            | 60 |
| ARTIGO 3 - ANÁLISE DE COMPONENTES PRINCIPAIS APLICADA À AVALIAÇÃO SENSORIAL DE ALIMENTOS..... | 61 |
| RESUMO.....                                                                                   | 61 |
| SUMMARY.....                                                                                  | 62 |

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| 1. INTRODUÇÃO.....                                                         | 64 |
| 2. MATERIAL E MÉTODOS.....                                                 | 65 |
| 2.1. DADOS EXPERIMENTAIS.....                                              | 65 |
| 2.2. METODOLOGIA.....                                                      | 65 |
| 2.2.1. ANÁLISES PRELIMINARES.....                                          | 65 |
| a. ANÁLISE DE VARIÂNCIA E MÉDIAS DOS ATRIBUTOS<br>(PROGRAMA 1).....        | 65 |
| b. PADRONIZAÇÃO DAS MÉDIAS (PROGRAMA 2).....                               | 67 |
| 2.2.2. ANÁLISE DE COMPONENTES PRINCIPAIS (PROGRAMA<br>3).....              | 69 |
| MÉTODO DE AGRUPAMENTO DE TOCHER.....                                       | 73 |
| 2.2.3. ESTIMATIVAS DAS CORRELAÇÕES ENTRE OS<br>ATRIBUTOS (PROGRAMA 4)..... | 73 |
| 3. RESULTADOS E DISCUSSÃO.....                                             | 75 |
| 4. CONCLUSÕES.....                                                         | 84 |
| 5. REFERÊNCIAS BIBLIOGRÁFICAS.....                                         | 86 |
| CONCLUSÕES GERAIS.....                                                     | 88 |

## RESUMO

CARNEIRO, Joel Camilo Souza, D.S., Universidade Federal de Viçosa. Novembro de 2005. **Análise estatística multivariada aplicada à avaliação sensorial de alimentos**. Orientadora: Valéria Paula Rodrigues Minim. Conselheiros: Adair José Regazzi, Carlos Henrique Osório Silva e José Benício Paes Chaves.

Os dados experimentais obtidos por YOKOTA (2005), referentes a 14 cachaaas envelhecidas, avaliadas quanto a 13 atributos sensoriais pelo método Análise Descritiva Quantitativa (ADQ), foram submetidos à análise estatística pelas técnicas multivariadas: variáveis canônicas, análise de componentes principais e análise de fatores. Apresentou-se os programas necessários para execução dos procedimentos estatísticos no sistema Statistical Analysis System (SAS) e enfatizou-se, além dos referidos procedimentos, a interpretação e discussão dos resultados. O objetivo do presente trabalho foi detalhar a aplicação das referidas técnicas multivariadas na análise de dados provenientes da avaliação sensorial descritiva. O texto desta tese foi redigido em três artigos, cada um ilustra uma das técnicas e apresenta os programas para execução das análises no SAS, independentemente. A conclusão de cada artigo relaciona os resultados obtidos com os dos outros artigos. Observou-se, pelos resultados, que as três técnicas foram eficientes, pois permitiram a redução no conjunto de informação de 13 atributos para um número bem menor de dimensões e estas retiveram a maior proporção da variância total dos dados originais. Em variáveis canônicas foram necessárias apenas duas dimensões (duas variáveis canônicas) que retiveram

91,58% da variância. Em componentes principais e análise de fatores houve a redução para quatro dimensões (componentes ou fatores) que retiveram 88,69% da variância. Foi possível, também, avaliar a importância ou influência dos atributos sobre as primeiras dimensões consideradas na caracterização e comparação entre as cachaças. Nas três análises, a primeira e a segunda dimensão foram influenciadas por atributos relacionados à madeira e ao teor alcoólico, respectivamente. Ficou, então, evidenciado que os resultados das três técnicas foram equivalentes e que o teor alcoólico e as características atribuídas pela madeira são importantes na avaliação de cachaça envelhecida. A análise por variáveis canônicas apresentou algumas vantagens em relação aos componentes principais, ou seja, as primeiras variáveis canônicas retiveram maior proporção da variância do que os primeiros componentes e o agrupamento entre as cachaças similares foi mais consistente. Portanto, recomenda-se que para avaliar dados obtidos por Análise Descritiva Quantitativa, em que as informações são obtidas com repetições, seja utilizada preferencialmente a análise por variáveis canônicas. Já a análise por componentes principais é mais apropriada para conjuntos de dados sem repetições, ou seja, em que há apenas uma observação por variável em cada indivíduo avaliado. A rotação efetuada na análise de fatores facilitou a interpretação dos resultados em comparação à análise de componentes principais. Porém, não se pode dizer que a análise de fatores seja melhor, pois estas técnicas não são concorrentes. Enquanto na análise de componentes principais a ênfase é explicar a variância, na análise de fatores o objetivo é tentar explicar as covariâncias. Os resultados obtidos no presente estudo foram satisfatórios para as três técnicas empregadas, o que nos leva a concluir pela recomendação do emprego de tais técnicas a outros dados também obtidos por Análise Descritiva Quantitativa, com as seguintes ressalvas: (a) aplicar variáveis canônicas e análise de fatores preferencialmente à análise de componentes principais; (b) verificar se os resultados serão satisfatórios no sentido de permitir concluir conforme o presente estudo.

## ABSTRACT

CARNEIRO, Joel Camilo Souza, D.S., Universidade Federal de Viçosa. November, 2005. **Multivariate statistical analysis applied to sensory evaluation of food**. Adviser: Valéria Paula Rodrigues Minim. Committee Members: Adair José Regazzi, Carlos Henrique Osório Silva and José Benício Paes Chaves.

This work illustrates the application of three multivariate techniques, factor analysis, principal component analysis and analysis by canonical variables, applied to sensory evaluation of food. The data used was obtained by YOKOTA (2005), who evaluated 14 commercial cachaça brands in relation to 13 sensory attributes, using Quantitative Descriptive Analysis (QDA). The programs necessary to apply the statistical procedures using the Statistical Analysis System (SAS) were presented and emphasis was given to interpretation and discussion of the results. The goal of the present work was to furnish details on how to apply these multivariate techniques to analyses data obtained from descriptive sensory evaluation of food. This thesis comprised three articles, each one illustrating one of the techniques applied and presenting the programs to run SAS analysis. The conclusion of each article provides the results presented in the other articles. The results showed that the three techniques were efficient as they allowed the reduction of 13 attributes to a smaller dimension, still retaining a large proportion of the variance from the initial dataset. Canonical variables analysis showed that only two canonical variables retained 91.58% of the total variance, while principal

component analysis and factor analysis showed that only four dimensions (components or factors) retained 88.69% of the total variance. It was also possible to evaluate the importance or influence of the attributes on the first dimensions considered in the characterization and comparison of the cachaça brands. Attributes related to wood and alcoholic content, respectively, influenced the first and second dimensions in all three analyses. Results from the three techniques were equivalent in that they all showed that wood attributes and alcoholic content are important when evaluating aged cachaça. Canonical variables analysis had some advantages over the principal component analysis since the first canonical variables retained a greater proportion of the total variance than the first principal components, with the clustering of the brands into groups being more consistent. It was concluded that canonical variables analysis should be used for data with repetitions, i.e., when for each attribute, more than one observation is available per subject (cachaça brand in our study). On the other hand, principal component analysis should be used when there are no repetitions. Although principal components and factor analysis are not competing techniques, the rotation applied in factor analysis did help the interpretation of results. While in principal component analysis, emphasis is on explaining variances, in factor analysis the goal is to explain co-variances. The results obtained in the present study were satisfactory for the three techniques used. Thus, the use of these techniques to other data also obtained by Quantitative Descriptive Analysis is recommended, with the following observations: (a) canonical variables and factor analysis should be applied instead of principal component analysis; (b) verify if satisfactory results were obtained.

## INTRODUÇÃO GERAL

A análise multivariada, de um modo geral, refere-se a todos os métodos estatísticos que simultaneamente analisam múltiplas medidas sobre cada indivíduo ou objeto sob investigação. Porém, algumas técnicas multivariadas são unicamente projetadas para lidar com questões multivariadas, como análise fatorial e análise discriminante (HAIR et al., 2005). Variáveis canônicas e análise de fatores são, também, técnicas com estas características.

HAIR et al. (2005) citam que em muitos casos, técnicas multivariadas são um meio de executar em uma única análise aquilo que antes exigia múltiplas análises usando técnicas univariadas.

Técnicas analíticas multivariadas estão sendo amplamente aplicadas na indústria, no governo e em centros de pesquisa. E poucas áreas de estudo ou pesquisa têm falhado na integração de técnicas multivariadas em suas “ferramentas” analíticas. Para atender a esse interesse crescente, têm sido publicados muitos livros e artigos que tratam dos aspectos teóricos e matemáticos dessas ferramentas, e textos introdutórios têm surgido em quase todas as áreas. Poucos livros, porém, têm sido escritos para o pesquisador que não é especialista em matemática ou estatística. Menos livros ainda discutem as aplicações da estatística multivariada, oferecendo uma discussão conceitual dos métodos estatísticos (HAIR et al., 2005). Observa-se que em algumas áreas do conhecimento, tais como comportamento humano, administração e melhoramento genético, a análise estatística multivariada é mais difundida e muito utilizada. Em outras áreas, por exemplo, na análise sensorial que é alvo do presente estudo, há carência de trabalhos científicos que abordem a aplicação de técnicas estatísticas multivariadas, principalmente com relação

aos procedimentos implementados em pacotes estatísticos computacionais e na interpretação dos resultados.

Com a revolução na informática ocorrida nas últimas décadas, há vários pacotes estatísticos disponíveis, até mesmo, nos computadores pessoais. Estes oferecem aos pesquisadores os recursos necessários para abordar problemas multivariados. O que falta é o pesquisador atualizar-se e interar-se das particularidades inerentes à aplicação das técnicas multivariadas na sua área de atuação. Isto tem levado a uma maior demanda e valorização dos estudos que envolvem aplicações das técnicas de análise multivariada.

Este estudo teve como objetivo detalhar a aplicação das técnicas análise de componentes principais, análise de fatores e análise por variáveis canônicas, a dados provenientes da avaliação sensorial de alimentos, com ênfase nos procedimentos estatísticos utilizados, na interpretação e discussão dos resultados. A execução das análises é demonstrada com a utilização do sistema SAS (Statistical Analysis System).



## **ARTIGO 1**

### **VARIÁVEIS CANÔNICAS APLICADAS À AVALIAÇÃO SENSORIAL DE ALIMENTOS**

### **CANONICAL VARIABLES APPLIED TO SENSORY EVALUATION OF FOOD**

#### **RESUMO**

Este estudo teve como objetivo principal o detalhamento da aplicação da análise por variáveis canônicas, uma técnica estatística multivariada. Aplicou-se a técnica na análise de dados experimentais obtidos pela avaliação sensorial e apresentaram-se os programas necessários para a execução dos procedimentos estatísticos no sistema SAS (Statistical Analysis System), com ênfase na interpretação e discussão dos resultados. Foram utilizados dados experimentais, obtidos por YOKOTA (2005), de cachaça envelhecida de 14 marcas comerciais, avaliadas quanto a 13 atributos sensoriais, pelo método Análise Descritiva Quantitativa (ADQ). Obteve-se, como resultados, a alocação das cachaças em cinco grupos disjuntos e foi realizada a redução, no conjunto de informações, de 13 atributos para duas variáveis canônicas. Estas acumularam 91,58% da variabilidade presente nos atributos originais, indicando que tal análise foi satisfatória. Foram identificados para descarte os atributos sabor alcoólico residual, sabor alcoólico inicial, aroma de madeira, sabor de madeira inicial, gosto amargo, aroma de baunilha e sabor de madeira residual. Estes foram os de menor importância relativa, pois apresentaram pequena variabilidade ou estavam correlacionados com outros considerados no estudo. O aroma alcoólico, gosto doce, gosto ácido, adstringência, coloração

amarela e viscosidade foram suficientes para a caracterização sensorial das cachaças e/ou para a comparação entre elas. Dessa forma, conclui-se que o estudo por variáveis canônicas facilitou a visualização e interpretação dos resultados e evidenciou o potencial desta técnica para utilização na análise estatística de dados descritivos.

**PALAVRAS-CHAVE:** análise multivariada, análise sensorial, análise descritiva quantitativa, ADQ.

## **SUMMARY**

The main objective of this study was to provide a detailed description of the application of canonical variables analysis, a multivariate statistical technique. The technique was applied to the analysis of experimental data obtained by sensory evaluation with emphasis in interpretation and discussion of the results, and the necessary programs to run statistical analysis in SAS system were showed. The experimental data used were obtained by YOKOTA (2005), from 14 commercial cachaça brands in relation to 13 sensory attributes, using Quantitative Descriptive Analysis (QDA). To obtain an overall result from the analysis, the 14 commercial brands were allocated into five disjoint groups, and two canonical variables were found to retain 91.58% of the total variability present in the 13 attributes originally evaluated. Hence, the analysis was effective in reducing the dimensionality of the explanatory variables without loss of information. The attributes that can be disregarded were: alcoholic residual flavor, alcoholic initial flavor, wooden aroma, wooden initial flavor, bitter taste, vanilla aroma and residual wood flavor, which presented small relative importance, because of their small variability or correlation with other attributes included in the analysis. Alcoholic aroma, sweet taste, acid taste, astringency, yellowish color and viscosity were sufficient for sensory characterization and/or comparison of the brands. It was concluded that the study of canonical analysis helped visualize and interpret the results better and evidenced the potential use of his technique in statistical analysis of descriptive data.

**KEYWORDS:** multivariate analysis, sensory analysis, quantitative descriptive analysis, QDA.

## 1. INTRODUÇÃO

Variáveis canônicas e componentes principais são duas técnicas alternativas que permitem a simplificação no conjunto de dados pela redução da dimensão. Isto é, as informações, originalmente contidas em um grupo de  $p$  variáveis  $X_1, X_2, \dots, X_p$  explicativas são reduzidas em outras poucas variáveis denominadas variáveis canônicas. A utilização de variáveis canônicas em detrimento aos componentes principais apresenta a vantagem de considerar as variâncias e covariâncias residuais entre os dados. Assim, esta técnica deve ser utilizada, preferencialmente à de componentes principais, nas situações em que se dispõe de dados experimentais com repetições, em que se pode obter as variâncias e covariâncias residuais. A análise é realizada a partir das observações das  $p$  variáveis  $X_1, X_2, \dots, X_p$  em  $n$  tratamentos. São estimados os escores médios (médias canônicas) por tratamento, para as primeiras variáveis canônicas e estes valores dispostos em um gráfico de dispersão, possibilitam o agrupamento dos tratamentos similares e a interpretação dos resultados. Tal técnica permite, também, a identificação das variáveis de menor importância relativa no estudo, ou seja, daquelas passíveis de descarte (CRUZ e CARNEIRO, 2003; REGAZZI, 1997).

A análise por variáveis canônicas tem sido utilizada em várias áreas do conhecimento na solução de uma série de problemas práticos (ADHIKARI et al., 2003; MARTINS, 1999 e MARTINS FILHO, 1991).

Na ciência e tecnologia de alimentos, mais especificamente na análise sensorial, esta técnica é pouco explorada, e, geralmente, os estudos estão focados na discussão dos resultados. Assim, há carência de informações sobre os procedimentos estatísticos que os aplicativos computacionais utilizam na análise por variáveis canônicas.

Tendo-se em vista tais considerações e o potencial que a análise por variáveis canônicas tem demonstrado nas várias áreas do conhecimento, este estudo teve por objetivo ilustrar a aplicação desta técnica de análise para dados experimentais obtidos pela avaliação sensorial, com ênfase nos procedimentos estatísticos utilizados e interpretação dos resultados. A execução da análise é demonstrada com a utilização do sistema SAS (Statistical Analysis System).

## **2. MATERIAL E MÉTODOS**

### **2.1. DADOS EXPERIMENTAIS**

Foram utilizados dados experimentais obtidos por YOKOTA (2005), provenientes da avaliação sensorial, pelo método de Análise Descritiva Quantitativa (ADQ), de 14 cachaças envelhecidas em reservatórios de diferentes tipos de madeira, por 18 a 24 meses, que são os tempos com maiores índices de comercialização de cachaças envelhecidas.

Uma equipe de oito provadores treinados avaliou as 14 cachaças com relação aos seguintes atributos sensoriais: aroma alcoólico, aroma de madeira, aroma de baunilha, gosto doce, gosto ácido, gosto amargo, sabor alcoólico inicial, sabor alcoólico residual, sabor de madeira inicial, sabor de madeira residual, adstringência, coloração amarela e viscosidade. As definições e referências para os atributos estão apresentadas no Quadro 1.

As amostras foram apresentadas de forma monádica, com cinco repetições, totalizando 70 sessões por provador. No Quadro 2, a título de ilustração, é apresentada uma parte do arquivo que contém os dados analisados. O arquivo completo é formado por um total de 560 observações, com a ressalva de que as vírgulas devem ser trocadas por pontos nos números apresentados para se executar a análise com a utilização do SAS.

Maiores informações sobre a coleta dos dados experimentais estão disponíveis em YOKOTA (2005).

### **2.2. METODOLOGIA**

A análise por variáveis canônicas foi realizada com base nos trabalhos de CRUZ e CARNEIRO (2003), CRUZ (2001) e REGAZZI (1997).

Para as análises estatísticas foram utilizados os procedimentos descritos em SAS (1999), do pacote computacional SAS, versão 8.1, licenciado para UFV em 2005. São apresentados os comandos e os respectivos comentários aparecem entre barras e asteriscos, (*/\*comentário\*/*), uma das alternativas comumente utilizadas para se inserir comentários na tela de procedimentos no SAS.

Quadro 1: Definições e referências para os atributos avaliados em cachaças envelhecidas de diferentes marcas comerciais.

| ATRIBUTOS           | DEFINIÇÕES                                                                                                                       | REFERÊNCIAS                                                                                                                                                                              |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Aparência</b>    |                                                                                                                                  |                                                                                                                                                                                          |
| COLORAÇÃO AMARELA   | Coloração amarela translúcida característica de cerveja tipo “pilsen” sem gás e sem espuma, em um cálice transparente.           | <b>Fraca:</b> solução hidroalcoólica a 30% com 0,0005% de corante caramelo. <b>Forte:</b> solução hidroalcoólica a 30% com 0,10112% de corante tartrazina e 0,0104% de corante caramelo. |
| VISCOSIDADE         | Resistência ao deslocamento de um líquido.                                                                                       | <b>Pouca:</b> etanol a 10°GL. <b>Muita:</b> etanol a 51°GL contendo 10% de glicerina.                                                                                                    |
| <b>Aroma</b>        |                                                                                                                                  |                                                                                                                                                                                          |
| ALCOÓLICO           | Aroma característico de etanol.                                                                                                  | <b>Fraco:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 51%.                                                                                                  |
| DE MADEIRA          | Aroma característico da madeira utilizada no tonel para envelhecimento da cachaça.                                               | <b>Nenhum:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 30%, contendo 17% de extrato de madeira (carvalho), marca Duas Rodas.                                |
| DE BAUNILHA         | Aroma exalado por uma solução alcoólica de baunilha.                                                                             | <b>Nenhum:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 30%, contendo 1% de vanília PA.                                                                      |
| <b>Gosto</b>        |                                                                                                                                  |                                                                                                                                                                                          |
| DOCE                | Gosto percebido pela mucosa oral, semelhante àquele provocado pela sacarose.                                                     | <b>Fraco:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 30%, contendo 1,5% de sacarose.                                                                       |
| ÁCIDO               | Gosto associado ao ácido acético (que lembra vinagre).                                                                           | <b>Fraco:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 30%, contendo 0,1% v/v de ácido acético p.a.                                                          |
| AMARGO              | Gosto de quinina, característico de água tônica.                                                                                 | <b>Nenhum:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 30%, contendo 0,005% de quinina.                                                                     |
| <b>Sabor</b>        |                                                                                                                                  |                                                                                                                                                                                          |
| ALCOÓLICO INICIAL   | Sabor alcoólico percebido no instante em que a amostra entra em contato com a mucosa oral.                                       | <b>Fraco:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 51%.                                                                                                  |
| ALCOÓLICO RESIDUAL  | Sabor alcoólico que permanece por um período de tempo após a ingestão da amostra.                                                | <b>Fraco:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 51%.                                                                                                  |
| DE MADEIRA INICIAL  | Sabor característico promovido pela madeira utilizada no tonel para envelhecimento da cachaça.                                   | <b>Nenhum:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 30%, contendo 17% de extrato de madeira (carvalho), marca Duas Rodas.                                |
| DE MADEIRA RESIDUAL | Sabor característico de madeira que permanece por um período de tempo após a ingestão da amostra.                                | <b>Nenhum:</b> solução hidroalcoólica a 30%. <b>Forte:</b> solução hidroalcoólica a 30%, contendo 17% de extrato de madeira (carvalho), marca Duas Rodas.                                |
| ADSTRINGÊNCIA       | Sensação de secura na mucosa oral, semelhante àquela causada de forma intensa por certas frutas verdes, como o caqui e a banana. | <b>Fraco:</b> solução hidroalcoólica a 30%, contendo 0,005% de ácido tânico. <b>Forte:</b> solução hidroalcoólica a 30%, contendo 0,150% de ácido tânico.                                |

Fonte: YOKOTA, 2005.

Quadro 2: Dados experimentais, em escala de 15 cm, de cachaças envelhecidas, avaliadas em blocos ao acaso (cada provador foi considerado um bloco), por oito provadores (PROV), com cinco repetições (REPE), relativos a 13 atributos.

| CACHAÇA | PROV | REPE | AALC | AMAD | ABAU | DOCE | ACID | AMAR | SALI | SALR | SMAI | SMAR | ADST | COLA | VISC |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1       | 1    | 1    | 12,8 | 11,8 | 3,4  | 3,2  | 11,2 | 9,8  | 11,8 | 11,4 | 11,8 | 10,9 | 9,6  | 10,0 | 11,5 |
| 1       | 1    | 2    | 11,0 | 12,0 | 3,5  | 3,2  | 11,3 | 10,9 | 10,6 | 10,6 | 12,4 | 11,7 | 9,4  | 8,6  | 12,0 |
| 1       | 1    | 3    | 11,5 | 12,5 | 3,5  | 8,8  | 10,8 | 9,4  | 10,6 | 10,1 | 12,5 | 12,4 | 10,3 | 8,3  | 11,5 |
| 1       | 1    | 4    | 12,5 | 13,1 | 3,3  | 3,0  | 10,8 | 10,6 | 11,9 | 11,9 | 12,2 | 11,6 | 9,4  | 8,6  | 12,2 |
| 1       | 1    | 5    | 12,3 | 11,9 | 3,6  | 3,3  | 10,9 | 10,6 | 11,5 | 11,4 | 12,2 | 11,9 | 9,4  | 7,8  | 11,7 |
| ...     | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| 1       | 8    | 1    | 8,8  | 11,9 | 4,7  | 3,4  | 11,5 | 4,1  | 8,2  | 7,9  | 11,3 | 10,5 | 7,4  | 9,7  | 14,0 |
| 1       | 8    | 2    | 6,2  | 8,1  | 0,0  | 2,8  | 11,6 | 5,1  | 7,3  | 7,1  | 7,5  | 7,1  | 7,9  | 10,3 | 14,0 |
| 1       | 8    | 3    | 4,2  | 12,6 | 0,0  | 3,6  | 12,2 | 5,5  | 4,6  | 4,5  | 8,7  | 9,7  | 7,0  | 10,3 | 14,0 |
| 1       | 8    | 4    | 4,1  | 11,8 | 0,0  | 5,9  | 11,5 | 5,2  | 4,1  | 4,1  | 11,9 | 12,0 | 6,7  | 9,8  | 14,0 |
| 1       | 8    | 5    | 4,1  | 11,4 | 0,0  | 4,8  | 10,8 | 4,1  | 4,4  | 4,5  | 11,0 | 10,8 | 7,4  | 9,8  | 14,0 |
| ...     | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| 14      | 1    | 1    | 12,3 | 12,0 | 9,4  | 9,2  | 10,5 | 4,1  | 11,9 | 11,7 | 12,6 | 12,1 | 9,8  | 9,1  | 11,3 |
| 14      | 1    | 2    | 12,7 | 10,3 | 8,3  | 8,9  | 9,4  | 8,8  | 11,8 | 11,6 | 11,1 | 10,3 | 9,1  | 3,1  | 10,6 |
| 14      | 1    | 3    | 12,4 | 11,3 | 9,7  | 3,7  | 11,0 | 10,6 | 12,4 | 12,3 | 10,8 | 10,7 | 10,4 | 6,9  | 11,9 |
| 14      | 1    | 4    | 12,5 | 12,4 | 9,0  | 8,8  | 9,7  | 3,0  | 12,1 | 12,0 | 12,5 | 12,5 | 9,3  | 9,3  | 12,2 |
| 14      | 1    | 5    | 11,5 | 10,1 | 3,5  | 3,7  | 10,0 | 9,1  | 10,9 | 11,0 | 10,7 | 10,5 | 9,3  | 0,5  | 11,7 |
| ...     | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| 14      | 8    | 1    | 4,2  | 11,6 | 5,4  | 6,9  | 12,1 | 4,1  | 4,5  | 4,5  | 11,5 | 11,5 | 9,5  | 8,0  | 14,0 |
| 14      | 8    | 2    | 4,9  | 6,7  | 0,1  | 4,4  | 11,1 | 4,3  | 4,2  | 4,2  | 11,6 | 11,6 | 7,9  | 6,8  | 13,5 |
| 14      | 8    | 3    | 5,3  | 4,8  | 3,2  | 2,4  | 9,2  | 4,9  | 4,1  | 3,8  | 12,2 | 12,1 | 5,6  | 2,9  | 13,1 |
| 14      | 8    | 4    | 8,1  | 7,1  | 0,0  | 3,9  | 11,4 | 3,9  | 6,8  | 6,7  | 8,1  | 8,6  | 7,3  | 6,2  | 14,0 |
| 14      | 8    | 5    | 8,2  | 5,6  | 0,0  | 2,4  | 12,6 | 3,9  | 8,6  | 6,9  | 7,2  | 7,2  | 6,8  | 5,2  | 13,7 |

AALC: aroma alcoólico, AMAD: aroma de madeira, ABAU: aroma de baunilha, DOCE: gosto doce, ACID: gosto ácido, AMAR: gosto amargo, SALI: sabor alcoólico inicial, SALR: sabor alcoólico residual, SMAI: sabor de madeira inicial, SMAR: sabor de madeira residual, ADST: adstringência, COLA: coloração amarela e VISC: viscosidade.

### 2.2.1. ANÁLISE DE VARIÂNCIA, MATRIZES E e H, AUTOVALORES E AUTOVETORES (PROGRAMA 1)

Foram realizadas, inicialmente, análises de variância uni e multivariada dos atributos avaliados nas cachaças. As fontes de variação consideradas foram as cachaças (CACHA), os provadores (PROV), a interação entre cachaça e provador (CACHA\*PROV) e o resíduo. Estas análises foram realizadas com o propósito de se verificar a variabilidade de cada atributo (análise univariada – ANOVA) e obtenção das matrizes de somas de quadrados e de produtos (análise multivariada – MANOVA) a serem utilizadas na análise por variáveis canônicas.

Foram estimadas, também, as matrizes E e H, e os autovalores de  $E^{-1}H$ , como solução do sistema  $\det(E^{-1}H - \lambda_i I) = 0$ , em que:

E: matriz de somas de quadrados e de produtos residuais (no modelo adotado  $E = CACHA * PROV$ );

H: matriz de somas de quadrados e de produtos de tratamentos (no modelo adotado  $H = CACHA$ );

$\lambda$ : autovalores da matriz  $E^{-1}H$ ;

I: matriz identidade.

A partir dos autovalores, foram obtidos os correspondentes autovetores, resolvendo-se o sistema  $(E^{-1}H - \lambda_i I)a_i^* = \phi$ , em que  $a_i^*$  é um autovetor associado ao autovalor  $\lambda_i$ . Cada autovetor  $a_i^*$  é normalizado, obtendo-se  $a_i$  tal que  $a_i' \frac{E}{n_e} a_i = 1$ , com  $n_e$  o número de graus de liberdade associado à matriz E. Os elementos de  $a_i$  são os coeficientes de ponderação associados aos atributos para a i-ésima variável canônica.

No programa 1, a seguir, é realizada a análise de variância uni e multivariada, obtendo-se as matrizes E e H, os autovalores e autovetores, com as seguintes particularidades inerentes ao conjunto de dados utilizados:

1. Na ANOVA, o valor de F, para testar o efeito de cachaça, foi obtido, dividindo-se o quadrado médio de CACHA pelo quadrado médio da interação entre cachaça e provador (CACHA\*PROV), e não pelo quadrado médio do resíduo. Este procedimento pode ser adotado, considerando que:
  - Foi detectado, para os atributos com interação CACHA\*PROV significativa, que se tratava de interação do tipo não grave (simples) (YOKOTA, 2005). Neste caso, a interação está, provavelmente, associada ao fato dos provadores, ao analisarem

as amostras, utilizarem partes diferentes da escala de avaliação. STONE e SIDEL (1993) recomendam, para esta situação, a utilização do quadrado médio da interação na obtenção do valor de F;

- Para os atributos com interação CACHA\*PROV não significativa, o quadrado médio do resíduo e o quadrado médio da interação apresentam valores muito semelhantes. Assim, o quadrado médio da interação pode ser usado na obtenção do valor de F.
2. Na MANOVA a matriz E é composta pelas somas de quadrados e somas de produtos da interação entre cachaça e provador (CACHA\*PROV), ao invés de somas de quadrados e produtos residuais, isto devido o modelo adotado.

**/\*Procedimento que importa os dados da planilha do Excel\*/**

/\*O arquivo de dados foi organizado em planilha do Excel como no Quadro 2, porém, os dados estão na íntegra e as vírgulas foram substituídas por pontos. As três primeiras colunas se referem aos tratamentos (cachaça, abreviada por CACHA), provador e repetição. As demais se referem aos atributos avaliados. Consta na primeira linha o nome de cada variável. O termo CACHA também foi utilizado nos procedimentos do SAS.\*/

```
PROC IMPORT out=dados_originais
             datafile="D:\canonica\programas\dado_ori.xls"/*1*/
             dbms=excel2000 replace;
             getnames=yes; RUN;
options nodate nonumber;
```

```
/*1*/ /*Informa que o arquivo de dados a ser lido, dado_ori.xls, está na
      pasta "D:\canonica\programas"*/
```

**/\* Imprimir, na tela, os dados lidos\*/**

```
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS'; RUN;
```

**/\*Análise de variância uni e multivariada, matrizes E e H, autovalores e autovetores normalizados\*/**

```
PROC GLM data=dados_originais;
title 'ANÁLISE DE VARIÂNCIA E ANÁLISE POR VARIÁVEIS CANONICAS';
class CACHA PROV REPE;
model AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA
VISC=CACHA PROV CACHA*PROV; /*2*/
test H=CACHA E=CACHA*PROV; /*3*/
manova H=CACHA E=CACHA*PROV/PRINTH PRINTE CANONICAL SHORT; /*4*/
RUN;
```

```
/*2*/ /*Considerando o delineamento em blocos casualizados, d.b.c, e cada
      provador (PROV) como um bloco.*/
```

```
/*3*/ /*Para obtenção do F (na ANOVA), para cachaça, utilizando como
      variância residual, o quadrado médio da interação CACHA*PROV*/
```



```
/*4*//*E=CACHA*PROV na MANOVA, informa que a matriz E, utilizada nos
testes multivariados para avaliar a igualdade dos vetores de médias
de cachaças e no estudo por variáveis canônicas, foi obtida
considerando as somas de quadrados e de produtos da interação
CACHA*PROV, e não as residuais*/
```

### 2.2.2. MÉDIAS CANÔNICAS E GRÁFICO DE DISPERSÃO (PROGRAMA 2)

As variáveis canônicas foram obtidas como combinações lineares dos 13 atributos (AALC, AMAD, ABAU, DOCE, ACID, AMAR, SALI, SALR, SMAI, SMAR, ADST, COLA e VISC) e escritas como:

$$\begin{aligned} \text{Can}_1 &= a_{1,1}\text{AALC} + a_{1,2}\text{AMAD} + \Lambda + a_{1,13}\text{VISC} \\ \text{Can}_2 &= a_{2,1}\text{AALC} + a_{2,2}\text{AMAD} + \Lambda + a_{2,13}\text{VISC} \\ &\dots \dots \dots \\ \text{Can}_{13} &= a_{13,1}\text{AALC} + a_{13,2}\text{AMAD} + \Lambda + a_{13,13}\text{VISC} \end{aligned}$$

Os coeficientes de ponderação destas variáveis são os elementos dos autovetores normalizados

$$a_1 = \begin{bmatrix} a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{1,13} \end{bmatrix}, a_2 = \begin{bmatrix} a_{2,1} \\ a_{2,2} \\ \vdots \\ a_{2,13} \end{bmatrix}, \dots, a_{13} = \begin{bmatrix} a_{13,1} \\ a_{13,2} \\ \vdots \\ a_{13,13} \end{bmatrix}.$$

Ao substituir  $a_1, a_2, \Lambda, a_{13}$  e AALC, AMAD,  $\Lambda$ , VISC pelos respectivos valores numéricos, obteve-se os escores para as variáveis canônicas. Foram utilizados os valores médios de AALC, AMAD,  $\Lambda$ , VISC, por cachaça. Então, os escores obtidos são valores médios e podem ser chamados de médias canônicas.

Foi obtido, também, o gráfico de dispersão, a partir dos escores para as primeiras variáveis canônicas, aquelas que retêm a maior parte da variação contida nas variáveis originais, o que permitiu comparar as cachaças, pelos seus posicionamentos no espaço.

No programa 2, a seguir, foram estimadas as médias canônicas e obtido o gráfico de dispersão.

```
/*Procedimento que importa os dados da planilha do Excel.*/
PROC IMPORT out=dados_originais
    datafile="D:\canonica\programas\dado_ori.xls"
    dbms=excel2000 replace;
    getnames=yes;
RUN; options nodate nonumber;
/*Procedimento que imprime, na tela, os dados lidos.*/
PROC PRINT data=dados_originais;
```

```

title 'LEITURA DOS DADOS';
RUN;

/*Procedimento que calcula as médias dos atributos por cachaça.*/
PROC MEANS;
output out=medias;
by CACHA; /*NOTA*/
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS DOS ATRIBUTOS';
RUN;
/*NOTA*/ /*O arquivo de dados já estava ordenado por cachaça. Se não
estivesse, seria necessário ordená-lo. Para tal, seria utilizado
o procedimento PROC SORT; BY CACHA;, antes do PROC MEANS.

/*Procedimento que organiza as médias para serem usadas em
procedimentos a seguir.*/
DATA medias; set medias(where=(__STAT_='MEAN'))
keep= CACHA __STAT_ AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR
ADST COLA VISC);
RUN;

/*Procedimento que imprime, na tela, as médias organizadas
previamente.*/
PROC PRINT data=medias;
title 'MÉDIAS DOS ATRIBUTOS POR CACHAÇA';
RUN;

/*Procedimento que estima os valores para as duas primeiras variáveis
canônicas, Can1 e Can2.*/

/*São determinados os valores de Can1 e Can2, para cada cachaça,
substituindo-se os valores médios dos atributos AALC, AMAD, ... VISC,
que constam no arquivo chamado medias.*/
DATA Can1eCan2; set medias;
Can1=0.08409885*AALC-0.0203075*AMAD+0.02398392*ABAU-
0.02497044*DOCE+0.04029627*ACID-0.02097523*AMAR+0.21296284*SALI-
0.25943151*SALR+0.04081351*SMAI+0.07597381*SMAR-
0.04251445*ADST+0.47476288*COLA-0.07953288*VISC; /*5*/

Can2=-0.0976703*AALC-0.02402243*AMAD+0.00068391*ABAU-
0.0700265*DOCE+0.22398466*ACID+0.25432446*AMAR-0.04607967*SALI-
0.03732882*SALR+0.04064412*SMAI+0.08612089*SMAR+0.2893113*ADST-
0.07027558*COLA-0.19949123*VISC; /*6*/
RUN;

/*5*/ /*Can1 é uma combinação linear dos atributos AALC, AMAD, ..., VISC.
Os coeficientes de ponderação associados a estes atributos são
elementos do primeiro autovetor normalizado. Estes valores foram
obtidos no programa 1.*/

```

```

/*6*//*Can2 é uma combinação linear dos atributos AALC, AMAD, ..., VISC.
Os coeficientes de ponderação associados a estes atributos são
elementos do segundo autovetor normalizado. Estes valores foram,
também, obtidos no programa 1.*/

/*Foram determinadas apenas as duas primeiras variáveis canônicas,
já que juntas, elas retiveram a maior parte da variação disponível
nos dados originais, 91,58%, Quadro 7.*/

/*Procedimento que imprime, na tela, os valores de Can1 e Can2, para
cada cachaça. Estes valores são chamados de médias canônicas.*/
PROC PRINT; var CACHA Can1 Can2;
title 'ESCORES PARA AS VARIÁVEIS CANÔNICAS Can1 E Can2 (MÉDIAS
CANÔNICAS), POR CACHAÇA';
RUN;

/*Procedimento que gera o gráfico de dispersão das cachaças (Can1
versus Can2).*//*7*/
PROC PLOT;
plot Can2*Can1 = CACHA / vpos=20; /*8*/
title 'GRÁFICO DE DISPERÇÃO PARA AS CACHAÇAS (Can1 X Can2)';
RUN;

/*7*//*O gráfico de dispersão foi feito, também, no EXCEL, a partir das
médias canônicas, pois o autor é mais familiarizado com a edição de
gráficos no aplicativo Excel.*/
/*8*//*O gráfico é gerado a partir das médias canônicas.*/

```

## MÉTODO DE AGRUPAMENTO DE TOCHER

A avaliação visual do posicionamento das cachaças, no gráfico de dispersão, permite agrupá-las, de forma que as cachaças que estão mais próximas entre si fazem parte de um mesmo grupo. O estabelecimento destes grupos pode variar de um analista para outro, sendo, portanto, uma avaliação subjetiva. Para evitar a subjetividade, o agrupamento das cachaças foi realizado pelo Método de Agrupamento de Otimização proposto por Tocher, citado por RAO (1952). Este método faz parte dos procedimentos do Programa Genes (Cruz, 2001), o qual foi utilizado no estabelecimento dos grupos.

O agrupamento, neste método, é realizado com base nas distâncias de Mahalanobis, que é uma medida de dissimilaridade entre as cachaças. As distâncias de Mahalanobis, por sua vez, são obtidas a partir das médias originais por cachaça e das estimativas das variâncias e covariâncias residuais (neste estudo, variâncias e covariâncias da interação entre cachaça e provador) para os atributos estudados.

### 2.2.3. IMPORTÂNCIA RELATIVA DOS ATRIBUTOS (PROGRAMA 3)

Considerou-se, como atributos de menor importância relativa, aqueles que apresentam coeficientes de ponderação de maior magnitude, em valor absoluto, em pelo menos uma das últimas 11 variáveis canônicas.

Os atributos de menor importância relativa apresentam baixa variabilidade ou estão correlacionados com outros que fazem parte do estudo. A baixa variabilidade indica que eles contribuem pouco para a discriminação entre as cachacas. Já os atributos correlacionados com outro(s) têm seu comportamento explicado por este(s). Portanto, os atributos de menor importância relativa são passíveis de descarte em experimentos futuros de mesma natureza, ou até podem ser desconsiderados neste estudo.

Recomenda-se, que, para estudo da importância relativa, sejam utilizados os autovetores obtidos da análise com os atributos padronizados, eliminando-se os efeitos do uso de diferentes escalas de avaliação. Há duas formas, equivalentes, de obter os autovetores cujos elementos estarão associados aos atributos padronizados:

- 1) Estimar os autovetores a partir dos atributos originais e multiplicar cada um de seus elementos pelo desvio padrão residual (no modelo adotado: desvio padrão da interação CACHA\*PROV) do atributo com o qual está associado; ou,
- 2) estimar os autovetores a partir dos atributos padronizados (dados padronizados). Os atributos são padronizados, dividindo-se os valores observados em cada atributo pelo seu respectivo desvio padrão residual (raiz quadrada do quadrado médio do resíduo. No modelo adotado: raiz quadrada do quadrado médio da interação CACHA\*PROV). Assim, tem-se

$$(a) Z_j = \frac{X_j}{\sqrt{\hat{\sigma}_j^2}} \text{ ou } (b) Z_j = \frac{X_j - \bar{X}_j}{\sqrt{\hat{\sigma}_j^2}}, \text{ em que:}$$

$Z_j$ : atributo padronizado;

$X_j$ : atributo original;

$\bar{X}_j$ : média do atributo original;

$\sqrt{\hat{\sigma}_j^2}$ : desvio padrão residual (no modelo adotado: desvio padrão da interação CACHA\*PROV) associado a  $X_j$ .

Foi considerada, neste estudo, a segunda forma (2a) de estimar os autovetores.

O programa 3, a seguir, foi usado, inicialmente, para padronizar os atributos, e em seguida, estimar os autovetores.

```

/*Procedimento que importa os dados da planilha do Excel.*/
PROC IMPORT out=dados_originais
            datafile="D:\canonica\programas\dado_ori.xls"
            dbms=excel2000 replace;
            getnames=yes;
RUN;
options nodate nonumber;

/*Procedimento que imprime, na tela, os dados lidos.*/
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS ORIGINAIS';
RUN;

/*Procedimento para padronizar os dados*/
DATA variaveis_padronizadas;
set dados_originais;
AALCp=AALC/2.07871571; /*9*/
AMADp=AMAD/2.92763112;
ABAUp=ABAU/3.15895853;
DOCEp=DOCE/2.35548657;
ACIDp=ACID/3.19256809;
AMARp=AMAR/2.43672499;
SALIp=SALI/2.08976027;
SALRp=SALR/2.21456881;
SMAIp=SMAI/2.71794132;
SMARp=SMAR/3.10452057;
ADSTp=ADST/2.21726746;
COLAp=COLA/2.11710982;
VISCp=VISC/2.03121442;
/*9*//*AALCp: atributo aroma alcoólico padronizado;*/
/*AALC: atributo aroma alcoólico original;*/
/* 2.07871571: desvio padrão da interação CACHA*PROV associado a
AALC. Este valor foi obtido extraíndo-se a raiz quadrada do
quadrado médio da interação CACHA*PROV, 4.3211, associado a AALC e
obtido pela análise de variância (programa 1). Procedimento similar
foi usado para padronizar os demais atributos.*/

/*Procedimento que imprime, na tela, os dados padronizados*/
PROC PRINT;
var CACHA PROV REPE AALCp AMADp ABUp DOCEp ACIDp AMARp SALIp SALRp
SMAIp SMARp ADSTp COLAp VISCp;
title 'VARIÁVEL PADRONIZADA = VARIÁVEL ORIGINAL/RAIS QUADRADA DO Q.M.
DE CACHA*PROV';
RUN;

/*Procedimento para estimar os autovetores associados aos atributos
padronizados*/
PROC GLM data=variaveis_padronizadas; /*10*/

```

```

title 'ESTIMATIVAS DOS AUTOVETORES ASSOCIADOS AOS ATRIBUTOS
PADRONIZADOS';
class CACHA PROV REPE;
model AALCp AMADp ABAUp DOCEp ACIDp AMARp SALIp SALRp SMAIp SMARp
ADSTp COLAp VISCp= CACHA PROV CACHA*PROV;
test H=CACHA E=CACHA*PROV;
manova H=CACHA E=CACHA*PROV/PRINTH PRINTE CANONICAL SHORT;
RUN;

/*10*//*Foram estimados vários outros parâmetros, mas o interesse foi
apenas nos autovetores associados aos atributos padronizados.*/

```

#### 2.2.4. ESTIMATIVAS DAS CORRELAÇÕES ENTRE OS ATRIBUTOS (PROGRAMA 4)

Uma vez identificados os atributos de menor importância relativa (aqueles que apresentaram baixa variabilidade ou que estão correlacionados com outros), é importante ter-se uma idéia da variabilidade dos mesmos e das possíveis correlações deles com os atributos de maior importância relativa no estudo.

A variabilidade dos atributos pode ser observada nos resultados das análises de variância e as correlações (coeficientes de correlação de Pearson) entre os atributos, inclusive entre os de menor e de maior importância relativa, a partir de seus valores médios, foram determinadas no programa 4, a seguir.

```

/*Procedimento que importa os dados da planilha do Excel.*/
PROC IMPORT out=dados_originais
            datafile="D:\canonica\programas\dado_ori.xls"
            dbms=excel2000 replace;
            getnames=yes;
RUN;
options nodate nonumber;

/*Procedimento que imprime, na tela, os dados lidos.*/
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS';
RUN;

/*Procedimento que calcula as médias dos atributos, por cachaça.*/
PROC MEANS;
output out=medias;
by CACHA;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇA';
RUN;

```

```

/*Procedimento que organiza as médias para serem usadas em
procedimentos a seguir. */
DATA medias; set medias(where=(_STAT_='MEAN'))
keep= CACHA _STAT_ AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR
ADST COLA VISC);
RUN;

/*Procedimento que imprime, na tela, as médias.*/
PROC PRINT data=medias;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇA';
RUN;

/*Procedimento que obtêm as correlações (Coeficiente de Correlação de
Pearson) entre os atributos.*/
PROC CORR data=medias; /*11*/
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'CORRELAÇÕES ENTRE OS ATRIBUTOS';
RUN;

/*11*/ /*As correlações entre os atributos foram determinadas a partir de
seus valores médios, por cachaça. */

```

## 2.2.5. ANOVA E TESTE DE MÉDIAS PARA OS GRUPOS (PROGRAMA 5)

Os grupos de cachaças estabelecidos pelo método de agrupamento de Tocher apresentam homogeneidade dentro dos grupos. Assim, pode-se passar a discutir os resultados, pela comparação dos grupos entre si, em relação às variáveis canônicas e, também, em relação aos atributos. Para tal, os dados dos grupos foram submetidos a análise de variância e teste de comparação entre médias. O programa 5, a seguir, ilustra tal procedimento.

As observações para um determinado grupo foram obtidas como médias das observações das cachaças que compõem o referido grupo. Assim, o arquivo com os dados das cachaças, “dado\_ori.XLS” apresentado no Quadro 2, foi reorganizado no Excel, a fim de obter-se o arquivo com dados dos grupos, “grupos.XLS”, apresentado no Quadro 3 e usado para as análises seguintes:

```

/*Procedimento que importa os dados da planilha do Excel.*/
PROC IMPORT out=dados_grupos
            datafile="D:\canonica\programas\grupos.xls"
            dbms=excel2000 replace;
            getnames=yes;
RUN;
options nodate nonumber;

/*Procedimento que imprime, na tela, os dados lidos.*/
PROC PRINT data=dados_grupos;

```

```

title 'LEITURA DOS DADOS';
RUN;

/*Procedimento para ANOVA e teste de comparação entre médias (Teste de
Duncan). */
PROC ANOVA data=dados_grupos;
title 'ANOVA E TESTE DE MÉDIAS PARA OS GRUPOS COMPOSTOS PELAS
CACHAÇAS';
class GRUPO PROV REPE;
model AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA
VISC= GRUPO PROV GRUPO*PROV;
test H=GRUPO E=GRUPO*PROV;
means GRUPO/duncan alpha=0.05 E=GRUPO*PROV;
RUN;

```



Quadro 3: Dados experimentais, em escala de 15 cm, de 5 grupos (GRUPO) de cachaças avaliadas em blocos ao acaso, por oito provadores (PROV), com cinco repetições (REPE), em relação a 13 atributos.

| GRUPO | PROV | REPE | AALC | AMAD | ABAU | DOCE | ACID | AMAR | SALI | SALR | SMAI | SMAR | ADST | COLA | VISC |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1     | 1    | 1    | 11,8 | 11,5 | 4,5  | 5,9  | 10,2 | 6,8  | 11,2 | 11,1 | 11,6 | 11,2 | 9,6  | 7,1  | 11,6 |
| 1     | 1    | 2    | 11,5 | 10,9 | 5,3  | 6,2  | 10,1 | 9,5  | 11,7 | 11,5 | 11,5 | 11,1 | 9,7  | 7,7  | 11,2 |
| 1     | 1    | 3    | 12,1 | 10,9 | 7,5  | 5,4  | 10,9 | 8,2  | 11,5 | 11,4 | 11,1 | 10,9 | 9,6  | 7,3  | 11,6 |
| 1     | 1    | 4    | 11,8 | 11,0 | 5,3  | 7,3  | 10,3 | 5,5  | 11,8 | 11,5 | 11,4 | 11,2 | 9,5  | 7,4  | 11,7 |
| 1     | 1    | 5    | 12,1 | 10,5 | 7,4  | 4,7  | 9,1  | 8,5  | 11,0 | 11,0 | 10,9 | 10,6 | 8,5  | 6,0  | 11,6 |
| ...   | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| 1     | 8    | 1    | 4,2  | 10,3 | 1,7  | 4,4  | 10,6 | 5,1  | 4,9  | 4,8  | 11,2 | 11,1 | 7,8  | 8,1  | 13,4 |
| 1     | 8    | 2    | 4,7  | 8,6  | 0,7  | 3,6  | 11,8 | 4,9  | 5,2  | 5,3  | 10,3 | 10,2 | 7,3  | 6,9  | 13,7 |
| 1     | 8    | 3    | 5,1  | 8,0  | 2,0  | 3,2  | 11,2 | 4,3  | 4,4  | 4,4  | 11,4 | 11,1 | 6,7  | 6,6  | 13,6 |
| 1     | 8    | 4    | 4,5  | 10,0 | 0,7  | 3,3  | 9,7  | 5,3  | 4,8  | 4,9  | 11,0 | 11,1 | 7,1  | 8,1  | 13,9 |
| 1     | 8    | 5    | 5,7  | 7,3  | 1,1  | 3,3  | 12,0 | 6,2  | 5,3  | 5,0  | 10,1 | 10,0 | 7,4  | 6,1  | 13,8 |
| ...   | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| 5     | 1    | 1    | 12,2 | 11,0 | 9,2  | 3,8  | 10,9 | 9,2  | 12,4 | 12,3 | 12,2 | 11,9 | 10,1 | 3,3  | 11,2 |
| 5     | 1    | 2    | 11,6 | 11,9 | 3,4  | 8,7  | 11,4 | 8,3  | 12,1 | 12,1 | 11,8 | 11,6 | 8,7  | 4,0  | 10,6 |
| 5     | 1    | 3    | 12,2 | 11,1 | 3,2  | 2,8  | 11,1 | 10,4 | 12,5 | 12,5 | 11,7 | 11,4 | 10,2 | 3,5  | 11,2 |
| 5     | 1    | 4    | 12,8 | 11,3 | 4,2  | 8,2  | 10,0 | 10,0 | 12,1 | 11,9 | 11,7 | 11,5 | 11,4 | 5,0  | 11,1 |
| 5     | 1    | 5    | 12,8 | 11,0 | 4,3  | 3,2  | 10,1 | 9,6  | 12,0 | 11,9 | 11,5 | 11,0 | 10,0 | 6,2  | 10,7 |
| ...   | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |
| 5     | 8    | 1    | 5,8  | 8,1  | 0,0  | 3,5  | 11,8 | 5,8  | 4,7  | 4,9  | 10,5 | 10,6 | 7,4  | 2,3  | 13,0 |
| 5     | 8    | 2    | 6,2  | 7,2  | 0,0  | 3,4  | 11,1 | 5,9  | 6,1  | 6,1  | 7,1  | 7,2  | 7,4  | 2,9  | 13,1 |
| 5     | 8    | 3    | 4,5  | 6,9  | 0,0  | 4,1  | 12,3 | 5,2  | 4,4  | 4,3  | 8,0  | 8,1  | 8,2  | 0,8  | 12,5 |
| 5     | 8    | 4    | 7,8  | 3,7  | 0,0  | 3,2  | 12,6 | 5,4  | 5,0  | 5,1  | 7,9  | 7,9  | 7,0  | 3,0  | 14,0 |
| 5     | 8    | 5    | 7,7  | 4,1  | 0,0  | 1,4  | 12,1 | 6,2  | 5,2  | 5,0  | 6,2  | 6,0  | 7,0  | 1,7  | 14,0 |

AALC: aroma alcoólico, AMAD: aroma de madeira, ABAU: aroma de baunilha, DOCE: gosto doce, ACID: gosto ácido, AMAR: gosto amargo, SALI: sabor alcoólico inicial, SALR: sabor alcoólico residual, SMAI: sabor de madeira inicial, SMAR: sabor de madeira residual, ADST: adstringência, COLA: coloração amarela e VISC: viscosidade.

### 3. RESULTADOS E DISCUSSÃO

No Quadro 4 estão apresentados os resultados das análises de variância dos 13 atributos avaliados nas 14 cachaças envelhecidas. Observa-se diferença significativa ( $P = 0,05$ ) entre as cachaças em relação a todos os atributos, exceto o gosto ácido. É importante ressaltar que entre os atributos avaliados, alguns podem apresentar maior importância relativa na caracterização sensorial das cachaças e/ou na comparação entre elas. O estudo por variáveis canônicas possibilita identificar tais atributos.

Este estudo teve início com a obtenção das matrizes  $E$  e  $H$ , apresentadas nos Quadros 5 e 6, respectivamente. Os elementos da diagonal principal da matriz  $E$  são exatamente as somas de quadrados para a interação entre cachaça e provador ( $SQC \times P$ ), apresentadas no Quadro 4. Fora da diagonal, têm-se as somas de produtos entre cada par de atributos. A diagonal da matriz  $H$  é formada pelas somas de quadrados para cachaças ( $SQC$ ) apresentadas no Quadro 4. Fora da diagonal estão as somas de produtos para cada par de atributos.

Estão apresentadas, no Quadro 7, as estimativas dos 13 autovalores não nulos da matriz  $E^{-1}H$  e as proporções de variância retidas por cada variável canônica. Observa-se que a proporção da variância decresce das primeiras para as últimas variáveis canônicas, chegando a valores próximos a zero. A primeira variável canônica ( $Can_1$ ) retém 88,69% da variância contida nas 13 variáveis canônicas, e conseqüentemente, 88,69% da variância apresentada pelos atributos originais AALC, AMAD, ABAU, DOCE, ACID, AMAR, SALI, SALR, SMAI, SMAR, ADST, COLA e VISC. As duas primeiras variáveis canônicas acumulam a maior parte da variância (91,58%), o que possibilita comparar as cachaças entre si, considerando apenas estas duas variáveis. Dessa forma, é feita uma simplificação no conjunto de dados, com a menor perda de informação possível, sendo satisfatório o estudo por variáveis canônicas.

No Quadro 7 também estão apresentados, dispostos por linhas, os autovetores normalizados, correspondentes aos autovalores  $\lambda_i$ . Cada linha é composta pelos elementos (coeficientes de ponderação para os atributos AALC, AMAD, ..., VISC) de um autovetor. Dessa forma, as variáveis  $Can_1$  e  $Can_2$ , combinações lineares dos atributos, podem ser escritas como

$$Can_1 = 0,0841 * AALC - 0,0203 * AMAD + \Delta - 0,0795 * VISC$$

$$Can_2 = -0,0977 * AALC - 0,0240 * AMAD + \Delta - 0,1995 * VISC .$$

Quadro 4: Resumo das análises de variância univariadas dos dados correspondentes aos 13 atributos avaliados nas 14 cachaças envelhecidas.

| Atributos                 | SQC     | SQC*P  | SQR     | QMC    | QMC*P              | QMR  | F <sub>c</sub> | Prob.> F <sub>c</sub> | Média |
|---------------------------|---------|--------|---------|--------|--------------------|------|----------------|-----------------------|-------|
| Aroma alcoólico           | 110,68  | 393,22 | 1092,39 | 8,51   | 4,32*              | 2,44 | 1,97           | 0,03                  | 8,59  |
| Aroma de madeira          | 1969,03 | 779,96 | 1114,24 | 151,46 | 8,57*              | 2,49 | 17,67          | <0,00                 | 8,32  |
| Aroma de baunilha         | 464,45  | 908,09 | 1326,77 | 35,73  | 9,98*              | 2,96 | 3,58           | 0,00                  | 3,73  |
| Gosto doce                | 130,24  | 504,90 | 2046,19 | 10,02  | 5,55 <sup>ns</sup> | 4,57 | 1,81           | 0,05                  | 6,13  |
| Gosto ácido               | 190,70  | 927,52 | 2847,42 | 14,67  | 10,19*             | 6,36 | 1,44           | 0,16                  | 6,38  |
| Gosto amargo              | 194,10  | 540,32 | 2337,98 | 14,93  | 5,94 <sup>ns</sup> | 5,22 | 2,51           | 0,01                  | 6,66  |
| Sabor alcoólico inicial   | 103,13  | 397,41 | 918,82  | 7,93   | 4,37*              | 2,05 | 1,82           | 0,05                  | 7,84  |
| Sabor alcoólico residual  | 114,99  | 446,29 | 1018,44 | 8,85   | 4,90*              | 2,27 | 1,80           | 0,05                  | 8,79  |
| Sabor de madeira inicial  | 1685,40 | 672,24 | 1090,21 | 129,65 | 7,39*              | 2,43 | 17,55          | <0,00                 | 7,94  |
| Sabor de madeira residual | 1983,27 | 877,06 | 1238,46 | 152,56 | 9,64*              | 2,76 | 15,83          | <0,00                 | 8,25  |
| Adstringência             | 149,41  | 447,38 | 1616,42 | 11,49  | 4,92*              | 3,61 | 2,34           | 0,01                  | 6,60  |
| Coloração amarela         | 7181,99 | 407,88 | 1225,10 | 552,46 | 4,48*              | 2,73 | 123,26         | <0,00                 | 6,24  |
| Viscosidade               | 126,63  | 375,45 | 888,86  | 9,74   | 4,13*              | 1,98 | 2,36           | 0,01                  | 10,15 |
| G.L.                      | 13      | 91     | 448     |        |                    |      |                |                       |       |

SQC: Somas de quadrados para cachaça;

SQC\*P: Somas de quadrados para a interação entre cachaça e provador;

SQR: Somas de quadrados do resíduo;

QMC: Quadrados médios para cachaça;

QMC\*P: Quadrados médios para a interação entre cachaça e provador;

QMR: Quadrados médios do resíduo;

F<sub>c</sub>: valor de F para cachaça ( $F_c = QMC/QMC*P$ );

Prob.: Valores de probabilidade (valor-P);

\*Interação entre cachaça e provador significativa ( $P \leq 0,05$ );

<sup>ns</sup> Interação entre cachaça e provador não significativa ( $P > 0,05$ ).

#### Esquema da ANOVA

| F.V.           | G.L. |
|----------------|------|
| Cachaça (C)    | 13   |
| Provador (P)   | 7    |
| Interação(C*P) | 91   |
| Resíduo        | 448  |

F.V.: fonte de variação

G.L.: graus de liberdade

Quadro 5: Matriz E, de somas de quadrados e somas de produtos da interação entre cachaça e provador (CACHAÇA\*PROV), para os 13 atributos avaliados nas 14 cachaças envelhecidas.

| Atributos                 | Atributos       |                  |                   |     |             |
|---------------------------|-----------------|------------------|-------------------|-----|-------------|
|                           | Aroma alcoólico | Aroma de madeira | Aroma de baunilha | ... | Viscosidade |
| Aroma alcoólico           | 393,22          | -204,57          | -38,17            | ... | 6,50        |
| Aroma de madeira          | -204,57         | 779,96           | 303,02            | ... | 127,73      |
| Aroma de baunilha         | -38,17          | 303,02           | 908,09            | ... | 22,76       |
| Gosto doce                | -75,97          | 21,45            | 37,88             | ... | 10,40       |
| Gosto ácido               | 94,63           | -23,34           | -114,76           | ... | 122,74      |
| Gosto amargo              | 125,46          | -21,09           | -94,98            | ... | 81,76       |
| Sabor alcoólico inicial   | 255,63          | -54,64           | 11,50             | ... | -5,49       |
| Sabor alcoólico residual  | 237,01          | -19,63           | 59,40             | ... | 1,96        |
| Sabor de madeira inicial  | -164,04         | 548,06           | 354,36            | ... | 162,81      |
| Sabor de madeira residual | -151,69         | 604,84           | 448,79            | ... | 179,50      |
| Adstringência             | 26,43           | 53,29            | 49,83             | ... | 70,09       |
| Coloração amarela         | -55,99          | 142,75           | -16,94            | ... | 65,76       |
| Viscosidade               | 6,50            | 127,73           | 22,76             | ... | 375,45      |

Quadro 6: Matriz H, de somas de quadrados e somas de produtos de cachaça, para os 13 atributos avaliados nas 14 cachaças envelhecidas.

| Atributos                 | Atributos       |                  |                   |     |             |
|---------------------------|-----------------|------------------|-------------------|-----|-------------|
|                           | Aroma alcoólico | Aroma de madeira | Aroma de baunilha | ... | Viscosidade |
| Aroma alcoólico           | 110,68          | -305,43          | -111,36           | ... | -55,93      |
| Aroma de madeira          | -305,43         | 1969,03          | 738,79            | ... | 388,57      |
| Aroma de baunilha         | -111,36         | 738,79           | 464,45            | ... | 113,94      |
| Gosto doce                | 39,00           | -149,03          | 14,06             | ... | -35,24      |
| Gosto ácido               | 28,78           | -156,17          | -89,75            | ... | -2,29       |
| Gosto amargo              | -27,88          | 285,82           | 62,92             | ... | 83,92       |
| Sabor alcoólico inicial   | 71,75           | -41,08           | -30,41            | ... | 14,96       |
| Sabor alcoólico residual  | 81,06           | -109,72          | -52,80            | ... | -5,87       |
| Sabor de madeira inicial  | -282,29         | 1800,19          | 662,83            | ... | 370,61      |
| Sabor de madeira residual | -291,26         | 1941,91          | 721,25            | ... | 391,67      |
| Adstringência             | -17,76          | 305,69           | 129,40            | ... | 51,43       |
| Coloração amarela         | -573,45         | 3654,43          | 1301,88           | ... | 751,25      |
| Viscosidade               | -55,93          | 388,57           | 113,94            | ... | 126,63      |

Quadro 7: Estimativas dos autovalores e coeficientes de ponderação (autovetores normalizados) dos 13 atributos avaliados nas 14 cachaças envelhecidas.

| i  | Variáveis canônicas<br>(Can <sub>i</sub> ) | Autovalores $\lambda_i$<br>de $E^{-1}H$ | Proporção da<br>variância (%) * | Proporção<br>acumulada (%) | Coeficientes de ponderação (autovetores normalizados) |                  |     |             |
|----|--------------------------------------------|-----------------------------------------|---------------------------------|----------------------------|-------------------------------------------------------|------------------|-----|-------------|
|    |                                            |                                         |                                 |                            | Aroma alcoólico                                       | Aroma de madeira | ... | Viscosidade |
| 1  | Can <sub>1</sub>                           | 20,7639                                 | 88,69                           | 88,69                      | 0,0841                                                | -0,0203          | ... | -0,0795     |
| 2  | Can <sub>2</sub>                           | 0,6759                                  | 2,89                            | 91,58                      | -0,0977                                               | -0,0240          | ... | -0,1995     |
| 3  | Can <sub>3</sub>                           | 0,5919                                  | 2,53                            | 94,11                      | 0,3082                                                | -0,1179          | ... | -0,1619     |
| 4  | Can <sub>4</sub>                           | 0,4149                                  | 1,77                            | 95,88                      | -0,0550                                               | -0,2323          | ... | 0,0443      |
| 5  | Can <sub>5</sub>                           | 0,2947                                  | 1,26                            | 97,14                      | 0,2045                                                | 0,2079           | ... | 0,0011      |
| 6  | Can <sub>6</sub>                           | 0,2209                                  | 0,94                            | 98,08                      | -0,1101                                               | -0,1660          | ... | 0,2053      |
| 7  | Can <sub>7</sub>                           | 0,1971                                  | 0,84                            | 98,92                      | 0,1391                                                | 0,0708           | ... | 0,1974      |
| 8  | Can <sub>8</sub>                           | 0,1153                                  | 0,49                            | 99,42                      | -0,2971                                               | -0,1626          | ... | 0,1432      |
| 9  | Can <sub>9</sub>                           | 0,0754                                  | 0,32                            | 99,74                      | 0,1333                                                | -0,0581          | ... | -0,3378     |
| 10 | Can <sub>10</sub>                          | 0,0290                                  | 0,12                            | 99,86                      | -0,1199                                               | -0,3356          | ... | -0,0660     |
| 11 | Can <sub>11</sub>                          | 0,0173                                  | 0,07                            | 99,94                      | -0,2147                                               | 0,1327           | ... | 0,0769      |
| 12 | Can <sub>12</sub>                          | 0,0110                                  | 0,05                            | 99,98                      | 0,4327                                                | -0,0042          | ... | 0,1642      |
| 13 | Can <sub>13</sub>                          | 0,0038                                  | 0,02                            | 100,00                     | -0,0452                                               | -0,1137          | ... | 0,1023      |

$$\text{Can}_1 = +0,0841 * \text{AALC} - 0,0203 * \text{AMAD} + \Lambda - 0,0795 * \text{VISC}$$

$$\text{Can}_2 = -0,0977 * \text{AALC} - 0,0240 * \text{AMAD} + \Lambda - 0,1995 * \text{VISC}$$

...

$$\text{Can}_{13} = -0,0452 * \text{AALC} - 0,1137 * \text{AMAD} + \Lambda + 0,1023 * \text{VISC}$$

\* Proporção da variância (%) para  $\text{Can}_i = \frac{\lambda_i}{\sum_{i=1}^{13} \lambda_i} \cdot 100\%$

Substituindo-se, em  $Can_1$  e  $Can_2$ , os valores de AALC,  $AMAD_K$ , VISC, pelos do Quadro 8, obtêm-se as coordenadas para  $Can_1$  e  $Can_2$ . Estas coordenadas são chamadas de escores das variáveis canônicas ou médias canônicas, já que os valores do Quadro 8 são médias por cachaça. As médias canônicas estão apresentadas no Quadro 9 e no gráfico de dispersão (Figura 1).

Quadro 8: Estimativas das médias, por cachaça, para os 13 atributos avaliados.

| Cachaças | Médias          |                  |     |             |
|----------|-----------------|------------------|-----|-------------|
|          | Aroma alcoólico | Aroma de madeira | ... | Viscosidade |
| 1        | 8,09            | 11,55            | ... | 10,58       |
| 2        | 7,74            | 10,83            | ... | 10,88       |
| 3        | 9,02            | 5,81             | ... | 10,11       |
| 4        | 8,50            | 8,62             | ... | 10,34       |
| 5        | 8,46            | 10,51            | ... | 11,21       |
| 6        | 8,59            | 8,64             | ... | 9,75        |
| 7        | 8,87            | 6,70             | ... | 9,66        |
| 8        | 8,14            | 9,27             | ... | 9,87        |
| 9        | 8,79            | 5,49             | ... | 9,60        |
| 10       | 8,40            | 9,07             | ... | 10,20       |
| 11       | 9,66            | 7,49             | ... | 9,96        |
| 12       | 8,89            | 8,57             | ... | 10,08       |
| 13       | 8,63            | 5,43             | ... | 9,51        |
| 14       | 8,53            | 8,49             | ... | 10,39       |

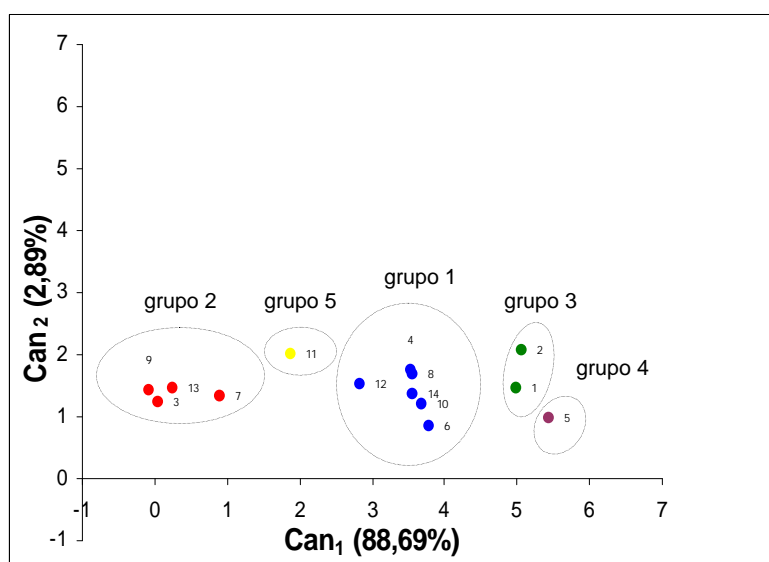
Observa-se, na Figura 1, que as 14 cachaças estão distribuídas em cinco grupos distintos, em que há homogeneidade intragrupo e heterogeneidade intergrupos, em relação às variáveis canônicas  $Can_1$  e  $Can_2$ . Como estas variáveis canônicas são combinações lineares dos atributos em estudo, pode-se dizer, indiretamente, que há homogeneidade intragrupo e heterogeneidade intergrupo em relação a tais atributos. Vale ressaltar que o grau desta homogeneidade e/ou heterogeneidade varia de uma variável canônica para a outra e, também, de um atributo para outro.

Cachaças mais próximas entre si, em um mesmo grupo são as mais homogêneas do grupo. Pode-se destacar as marcas 3, 9 e 13 no grupo dois e as 4, 8, 10 e 14 no grupo um, como as mais homogêneas em cada grupo (Figura 1).

Observa-se também que os grupos estão bem dispersos em relação ao eixo x ( $Can_1$ ) e bem próximos em relação ao y ( $Can_2$ ), indicando que a variabilidade entre os grupos, em relação à  $Can_1$  é muito maior que a da  $Can_2$ . Isto se deve ao fato da  $Can_1$  reter 88,69% da variação contra apenas 2,89% da  $Can_2$ .

Quadro 9: Estimativas das médias canônicas (valores por cachaça, para as variáveis canônicas  $Can_1$  e  $Can_2$ ).

| Cachaças | Médias canônicas |         |
|----------|------------------|---------|
|          | $Can_1$          | $Can_2$ |
| 1        | 5,00             | 1,47    |
| 2        | 5,06             | 2,06    |
| 3        | 0,05             | 1,24    |
| 4        | 3,54             | 1,74    |
| 5        | 5,45             | 0,96    |
| 6        | 3,80             | 0,84    |
| 7        | 0,92             | 1,33    |
| 8        | 3,56             | 1,68    |
| 9        | -0,08            | 1,43    |
| 10       | 3,68             | 1,20    |
| 11       | 1,88             | 1,98    |
| 12       | 2,83             | 1,51    |
| 13       | 0,25             | 1,45    |
| 14       | 3,58             | 1,36    |



| Grupos* | Cachaças             |
|---------|----------------------|
| 1       | 4, 6, 8, 10, 12 e 14 |
| 2       | 3, 7, 9 e 13         |
| 3       | 1 e 2                |
| 4       | 5                    |
| 5       | 11                   |

\*Observação: Os cinco grupos foram estabelecidos pelo método de Tocher.

Figura 1: Dispersão das 14 cachaças em relação às duas primeiras variáveis canônicas.

O sabor alcoólico residual, sabor alcoólico inicial, aroma de madeira, sabor de madeira inicial, gosto amargo, aroma de baunilha e sabor de madeira residual foram considerados de menor importância relativa na caracterização sensorial das cachaças e/ou na comparação entre elas, pois apresentaram coeficientes de ponderação de maior magnitude, em valor absoluto, em pelo menos uma das últimas 11 variáveis canônicas (coeficientes destacados no Quadro 10), que retiveram, juntas, menos de 10% da variância total.



Quadro 10: Estimativas dos coeficientes de ponderação (elementos dos autovetores) associados aos atributos padronizados.

| Atributo      | Can <sub>1</sub> * | Can <sub>2</sub> | Can <sub>3</sub> | Can <sub>4</sub> | Can <sub>5</sub> | Can <sub>6</sub> | Can <sub>7</sub> | Can <sub>8</sub> | Can <sub>9</sub> | Can <sub>10</sub> | Can <sub>11</sub> | Can <sub>12</sub> | Can <sub>13</sub> |
|---------------|--------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-------------------|-------------------|-------------------|-------------------|
| AALCp         | 0,17               | -0,20            | 0,64             | -0,11            | 0,43             | -0,23            | 0,29             | -0,62            | 0,28             | -0,25             | -0,45             | 0,90              | -0,09             |
| <b>AMADp</b>  | -0,06              | -0,07            | -0,35            | -0,68            | 0,61             | -0,49            | 0,21             | -0,48            | -0,17            | <b>-0,98</b>      | 0,39              | -0,01             | -0,33             |
| <b>ABAU</b> p | 0,08               | 0,00             | 0,01             | <b>-0,88</b>     | 0,09             | 0,51             | -0,43            | 0,44             | -0,04            | 0,10              | -0,21             | 0,17              | -0,06             |
| DOCEp         | -0,06              | -0,16            | 0,73             | -0,23            | 0,12             | 0,23             | 0,51             | -0,11            | 0,22             | -0,02             | 0,27              | -0,34             | 0,15              |
| ACIDp         | 0,13               | 0,72             | 0,22             | -0,02            | -0,43            | 0,36             | -0,12            | 0,00             | 0,45             | -0,31             | 0,19              | 0,31              | -0,22             |
| <b>AMAR</b> p | -0,05              | 0,62             | -0,23            | -0,53            | -0,03            | -0,01            | <b>0,58</b>      | 0,13             | 0,29             | 0,22              | -0,25             | -0,16             | 0,05              |
| <b>SAL</b> Ip | 0,45               | -0,10            | 0,37             | 0,50             | 0,31             | 1,11             | 0,10             | 0,28             | -0,80            | -0,40             | -0,59             | <b>-1,58</b>      | -1,17             |
| <b>SAL</b> Rp | -0,57              | -0,08            | -1,09            | 0,10             | 0,29             | -0,71            | -0,40            | 0,07             | 0,84             | 0,67              | <b>0,96</b>       | 0,95              | <b>1,18</b>       |
| <b>SMA</b> Ip | 0,11               | 0,11             | -0,71            | 0,02             | <b>-0,65</b>     | <b>1,41</b>      | -0,47            | <b>-1,99</b>     | <b>0,84</b>      | 0,71              | -0,61             | 0,07              | 0,75              |
| <b>SMAR</b> p | 0,24               | 0,27             | <b>1,33</b>      | 0,82             | 0,36             | -1,29            | 0,48             | 1,45             | -0,80            | 0,47              | 0,67              | 0,04              | -0,95             |
| ADSTp         | -0,09              | 0,64             | 0,14             | 0,20             | 0,25             | 0,01             | -0,34            | -0,07            | -0,22            | -0,11             | -0,05             | -0,19             | 0,61              |
| COLAp         | 1,01               | -0,15            | 0,01             | 0,06             | -0,10            | -0,01            | -0,04            | 0,27             | 0,17             | -0,08             | -0,13             | 0,01              | 0,13              |
| VISCp         | -0,16              | -0,41            | -0,33            | 0,09             | 0,00             | 0,42             | 0,40             | 0,29             | -0,69            | -0,13             | 0,16              | 0,33              | 0,21              |

AALCp: aroma alcoólico, AMADp: aroma de madeira, ABAUp: aroma de baunilha, DOCEp: gosto doce, ACIDp: gosto ácido, AMARp: gosto amargo, SALIp: sabor alcoólico inicial, SALRp: sabor alcoólico residual, SMAIp: sabor de madeira inicial, SMARp: sabor de madeira residual, ADSTp: adstringência, COLAp: coloração amarela e VISCp: viscosidade.

\*Can<sub>1</sub> = 0,17AALCp - 0,06 AMADp + 0,08 ABAUp - 0,06 DOCEp + 0,13 ACIDp - 0,05 AMARp + 0,45 SALIp - 0,57 SALRp + 0,11 SMAIp + 0,24 SMARp - 0,09 ADSTp + 1,01 COLAp - 0,16 VISCp

...

...

...

Can<sub>13</sub> = -0,09AALCp - 0,33 AMADp - 0,06 ABAUp + 0,15 DOCEp - 0,22 ACIDp + 0,05 AMARp - 1,17 SALIp + 1,18 SALRp + 0,75 SMAIp - 0,95 SMARp + 0,61 ADSTp + 0,13 COLAp + 0,21 VISCp

Observação: Cada coluna representa os coeficientes de ponderação, associados aos atributos padronizados, em uma variável canônica. Tais coeficientes expressam a importância relativa dos atributos em estudo.

A avaliação iniciou-se na última variável canônica, identificando o sabor alcoólico residual com o maior dos coeficientes (1,18), em valor absoluto, e, portanto, considerado como um dos atributos de menor importância relativa. Em seguida, observaram-se os coeficientes da penúltima variável canônica ( $Can_{12}$ ), identificando, também, o sabor alcoólico inicial como atributo de menor importância relativa. E assim, sucessivamente, foram verificados os coeficientes de cada variável canônica, finalizando na variável  $Can_3$  e concluindo a identificação dos atributos de menor importância relativa.

Vale ressaltar que quando em uma variável canônica de menor variância, o maior coeficiente de ponderação está associado a um caráter já previamente descartado, tem-se optado por não fazer nenhum outro descarte com base nos coeficientes daquela variável, mas prosseguir a identificação da importância relativa dos caracteres na outra variável canônica de variância imediatamente superior (CRUZ e REGAZZI, 1997). Isto aconteceu para o atributo sabor de madeira inicial e sabor alcoólico residual (Quadro 10).

Observa-se, nos Quadros 4 e 11, que os atributos indicados para descarte (os de menor importância relativa) apresentam, relativamente, baixa variabilidade e/ou estão correlacionados com outros que fazem parte do estudo.

Quadro 11: Correlações (coeficientes de correlação de Pearson) entre os 13 atributos avaliados nas 14 cachaças envelhecidas.

| Atributos de menor importância relativa | Atributos de maior importância relativa |               |                        |                     |            |             |
|-----------------------------------------|-----------------------------------------|---------------|------------------------|---------------------|------------|-------------|
|                                         | Aroma alcoólico                         | Adstringência | Coloração amarela      | Viscosidade         | Gosto doce | Gosto ácido |
| Sabor alcoólico residual                | 0,72<br>0,00                            | —             | —                      | —                   | —          | —           |
| Sabor alcoólico inicial                 | 0,67<br>0,01                            | —             | —                      | —                   | —          | —           |
| Aroma de madeira                        | -0,65<br>0,01                           | 0,56<br>0,04  | <b>0,97</b><br><0,0001 | <b>0,78</b><br>0,00 | —          | —           |
| Sabor de madeira inicial                | -0,65<br>0,01                           | 0,58<br>0,03  | <b>0,98</b><br><0,0001 | <b>0,80</b><br>0,00 | —          | —           |
| Gosto amargo                            | —                                       | —             | —                      | 0,54<br>0,05        | —          | —           |
| Aroma de baunilha                       | —                                       | —             | 0,71<br>0,00           | —                   | —          | —           |
| Sabor de madeira residual               | -0,62<br>0,02                           | 0,63<br>0,01  | <b>0,98</b><br><0,0001 | <b>0,78</b><br>0,00 | —          | —           |

Observação: Estão apresentadas apenas as correlações significativas ( $P = 0,05$ ).

O sabor alcoólico inicial e o residual estão correlacionados com o aroma alcoólico (Quadro 11) e estão entre os atributos de menor variabilidade, ou seja, estão entre aqueles que apresentam os menores valores da razão de variância (Quadro 4).

O aroma de madeira, o sabor de madeira inicial e o residual estão correlacionados com os atributos aroma alcoólico, adstringência, coloração amarela e viscosidade, destacando-se uma maior relação com estes dois últimos (Quadro 11).

O gosto amargo está correlacionado com a viscosidade e o aroma de baunilha com a coloração amarela (Quadro 11).

Os demais atributos avaliados: aroma alcoólico, gosto doce, gosto ácido adstringência, coloração amarela e viscosidade, considerados de maior importância relativa, apresentaram variabilidade entre as cachaças ou não estiveram correlacionados com outros no estudo. Houve, para aroma alcoólico, adstringência, coloração amarela e viscosidade, variabilidade entre as cachaças. Isto pode ser observado pelo efeito significativo de cachaça, para estes atributos (Quadro 4). O gosto doce e o gosto ácido não estão correlacionados ( $P > 0,05$ ) com nenhum outro atributo. Esta pode ser a causa deles não serem indicados para descarte, apesar de estarem entre os atributos com menor variabilidade (menores valores de F, Quadro 4).

Os coeficientes de ponderação, em valor absoluto, apresentados no Quadro 12, que inclui os atributos não indicados para descarte, permitem identificar os atributos mais relacionados com as duas primeiras variáveis canônicas, e, conseqüentemente, aqueles que mais contribuíram para a variabilidade entre os grupos.

A coloração amarela apresentou elevado coeficiente em  $Can_1$  (1,01) e apenas 0,15, em valor absoluto, em  $Can_2$ , indicando que a coloração está relacionada com  $Can_1$  e, portanto, é um dos atributos responsáveis pela variabilidade entre os grupos, observada ao longo do eixo x, Figura 1. Como o coeficiente 1,01 é positivo, indica que a coloração da cachaça aumenta no mesmo sentido do eixo x, ou seja, quanto mais à direita estiver posicionado o grupo, Figura 1, maior será a coloração da cachaça. A coloração amarela é, também, o atributo com o maior dos coeficientes em  $Can_1$ , indicando que este é o principal atributo relacionado à primeira variável canônica e, portanto o de maior variância. No Quadro 13 observa-se que as médias, por grupo, para a coloração amarela aparecem na mesma seqüência em que os grupos estão dispostos ao longo do eixo x, Figura 1, e há diferença significativa ( $P = 0,05$ ) entre todos os grupos.

Os atributos aroma de madeira, sabor de madeira inicial, aroma de baunilha e sabor de madeira residual, estão correlacionadas com a coloração amarela ( $P = 0,01$ ), Quadro 11, e, portanto, apresentam comportamento semelhante ao deste atributo. Ou seja, há variabilidade entre os grupos para tais atributos, sendo que os grupos dispostos mais a direita do eixo x, Figura 1, apresentam, além de uma coloração amarela mais intensa, maior intensidade nas características aroma de madeira, sabor de madeira inicial, aroma de baunilha e sabor de madeira residual.

Quadro 12: Coeficientes de ponderação (elementos de autovetores) associados aos atributos padronizados de maior importância relativa (não indicados para descarte).

| Atributos                  | Can <sub>1</sub> | Can <sub>2</sub> |
|----------------------------|------------------|------------------|
| Aroma alcoólico            | <b>0,17</b>      | <b>-0,20</b>     |
| Gosto doce                 | -0,06            | <b>-0,16</b>     |
| Gosto ácido                | 0,13             | <b>0,72</b>      |
| Adstringência              | -0,09            | <b>0,64</b>      |
| Coloração amarela          | <b>1,01</b>      | -0,15            |
| Viscosidade                | -0,16            | <b>-0,41</b>     |
| Proporção da variância (%) | 88,69            | 2,89             |

Quadro 13: Médias, por grupo, para os atributos de maior importância relativa, aroma alcoólico (AALC), gosto doce (DOCE), gosto ácido (ACID), adstringência (ADST), coloração amarela (COLA) e viscosidade (VISC), avaliados nas 14 cachaças envelhecidas.

| Aroma alcoólico |         | Gosto doce |       | Gosto ácido |       |
|-----------------|---------|------------|-------|-------------|-------|
| Grupo           |         | Grupo      |       | Grupo       |       |
| 5               | 9,7 a   | 1          | 6,3 a | 5           | 7,3 a |
| 2               | 8,8 b   | 4          | 6,2 a | 2           | 6,4 a |
| 1               | 8,5 b c | 5          | 6,2 a | 3           | 6,4 a |
| 4               | 8,5 b c | 2          | 6,2 a | 1           | 6,3 a |
| 3               | 7,9 c   | 3          | 5,5 a | 4           | 5,9 a |

| Adstringência |         | Coloração amarela |        | Viscosidade |          |
|---------------|---------|-------------------|--------|-------------|----------|
| Grupo         |         | Grupo             |        | Grupo       |          |
| 5             | 7,3 a   | 4                 | 11,4 a | 4           | 11,2 a   |
| 3             | 7,0 a b | 3                 | 10,6 b | 3           | 10,7 a b |
| 1             | 6,8 a b | 1                 | 7,5 c  | 1           | 10,1 b c |
| 4             | 6,5 b c | 5                 | 4,3 d  | 5           | 10,0 b c |
| 2             | 6,0 c   | 2                 | 1,3 e  | 2           | 9,7 c    |

Médias seguidas de pelo menos uma mesma letra, na coluna, dentro de cada atributo, não diferem entre si ( $P > 0,05$ ), pelo teste de Duncan.

A adstringência e a viscosidade, atributos com maiores coeficientes em Can<sub>2</sub> do que em Can<sub>1</sub> (Quadro 12), são atributos relacionados à segunda variável canônica, contribuindo para a variabilidade entre os grupos, observada

ao longo do eixo y, Figura 1. Como estes dois atributos estão relacionadas a  $Can_2$ , variável canônica com proporção de variância igual a apenas 2,89%, a variabilidade entre os grupos, em relação aos mesmos é bem inferior, quando comparada à da coloração amarela. O coeficiente positivo, 0,64 para adstringência, indica que os grupos posicionados mais acima, Figura 1, apresentam maior intensidade neste atributo. Para a viscosidade, ocorre o contrário. Seu coeficiente negativo, -0,41, indica que os grupos posicionados mais abaixo na Figura 1 são os de maior viscosidade. Nota-se na Figura 1, ao longo do eixo y, uma maior variabilidade entre os grupos quatro e cinco, em relação à adstringência e viscosidade. O comportamento destes dois atributos pode ser observado, também, no Quadro 13.

O aroma de madeira, gosto amargo, sabor de madeira inicial e sabor de madeira residual, correlacionados, também, com a viscosidade, Quadro 11, apresentam comportamento semelhante ao deste atributo. Assim, o grupo quatro destaca-se dos demais, principalmente do cinco, por apresentar, além de uma maior viscosidade, maior intensidade no aroma de madeira, gosto amargo, sabor de madeira inicial e sabor de madeira residual.

Os coeficientes 0,17 em  $Can_1$  e -0,20 em  $Can_2$  (Quadro 12), associados ao aroma alcoólico, são valores muito próximos (em valor absoluto), indicando que a relação entre aroma alcoólico e as duas variáveis canônicas é equivalente. Observa-se, também, que estes coeficientes são muito pequenos quando comparados aos valores destacados no Quadro 12, e também, em relação a outros coeficientes associados ao aroma alcoólico, em outras variáveis canônicas, por exemplo, o valor 0,64 em  $Can_3$ , 0,43 em  $Can_5$  e -0,62 em  $Can_8$  (Quadro 10). Assim, o aroma alcoólico apresenta maior relação com estas variáveis canônicas do que com  $Can_1$  e  $Can_2$ , e, portanto, o comportamento dos grupos em relação ao aroma alcoólico não pode ser observado apenas nas duas primeiras variáveis canônicas (Figura 1). Tal comportamento pode ser observado no Quadro 13. O grupo cinco e o três destacam-se dos demais por apresentarem o maior e o menor aroma alcoólico, respectivamente. A variabilidade desta característica é menor para os demais grupos.

O sabor alcoólico residual e o sabor alcoólico inicial, correlacionados com o aroma alcoólico, Quadro 11, apresentam comportamento semelhante ao deste atributo, em relação à variabilidade dos grupos.

O gosto doce e o gosto ácido apresentaram maiores coeficientes em  $Can_2$  do que em  $Can_1$  (Quadro 12), indicando que eles estão associados à segunda variável canônica. Porém, não significa que estes atributos contribuam

para a variabilidade apresentada por ela. É preciso lembrar que o fato do gosto doce e o ácido estarem entre os atributos de maior importância relativa não implica, necessariamente, que eles apresentem alta variabilidade. Provavelmente eles estão entre os de maior importância relativa por não estarem correlacionados ( $P > 0,05$ ) com nenhum dos outros atributos avaliados.

O gosto doce e o gosto ácido estão entre os atributos de menores variabilidades, como pode ser observado no Quadro 4 (menores valores de F) e no Quadro 13, onde são mostradas as médias por grupo de cachaça.

#### 4. CONCLUSÕES

Constatou-se que para a caracterização sensorial e/ou comparações entre as 14 cachaças envelhecidas, distribuídas em cinco grupos distintos, foram necessários seis dos 13 atributos avaliados. São eles: aroma alcoólico, gosto doce, gosto ácido, adstringência, coloração amarela e viscosidade. A coloração amarela foi o que apresentou maior variância, sendo seguido pelo aroma alcoólico, adstringência e viscosidade. Os gostos doce e ácido foram os de menor variabilidade e não apresentaram redundância, ou seja, não se correlacionaram ( $P > 0,05$ ) com nenhum dos atributos avaliados.

Os sabores alcoólico inicial e residual tiveram seu comportamento explicado pelo aroma alcoólico, com o qual estão correlacionados. O aroma de madeira e os sabores de madeira inicial e residual tiveram seu comportamento explicado pela coloração amarela e viscosidade, com os quais estão correlacionados. Portanto, na avaliação das cachaças envelhecidas, destacaram-se os atributos relacionados ao teor alcoólico e à madeira.

Num próximo estudo de mesma natureza, por exemplo, se o produtor de cachaça quiser acompanhar, periodicamente, as características sensoriais de seu produto, basta avaliar apenas os seis atributos de maior importância relativa: aroma alcoólico, gosto doce, gosto ácido, adstringência, coloração amarela e viscosidade. Reduzindo, assim, tempo, custos e facilitando a avaliação sensorial.

O uso, em conjunto, do método de agrupamento de Tocher e das variáveis canônicas deu um bom resultado, evidenciando, que sempre que for possível, é interessante usar mais de uma técnica de análise multivariada, pois elas se complementam.

A análise de variância univariada também foi útil, sendo aplicada após o estabelecimento dos grupos e identificação dos atributos de maior importância relativa.

Os objetivos da análise estatística por variáveis canônicas foram alcançados neste estudo, ou seja, foi possível fazer a redução de 13 atributos para duas variáveis canônicas, com a menor perda de informação possível. Pôde-se, também, identificar e desconsiderar aqueles atributos de menor importância relativa. Assim, esta técnica de análise foi viável para analisar os dados provenientes da Análise Descritiva Quantitativa.

A análise estatística por variáveis canônicas pode ser estendida para outros conjuntos de dados obtidos pela Análise Descritiva Quantitativa. Porém, há de considerar, que as primeiras variáveis canônicas precisam acumular a

maior proporção da variância apresentada pelos atributos (pelo menos 70 ou 80% da variância).



## 5. REFERÊNCIAS BIBLIOGRÁFICAS

ADHIKARI, K.; HEYMANN, H.; HUFF, H.E. Textural characteristics of lowfat, fullfat and smoked cheeses: sensory and instrumental approaches. **Food Quality and Preference**. 14, p. 211 – 218. 2003.

CRUZ, C.D. **Programa Genes: versão Windows; aplicativo computacional em genética e estatística**. 1. ed. Viçosa: Editora UFV, 2001. 648 p.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. v. 2, 1. ed. Viçosa: Editora UFV, 2003. 585 p.

CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. v. 1, 1. ed. Viçosa: Editora UFV, 1997. 309p.

MARTINS FILHO, S. **Análise dialélica da resistência da soja (*Glicine max* (L.) Merrill) à *Cercospora sojina* Hara**. 1991. 70f. Dissertação (Mestrado em Genética e Melhoramento) - Universidade Federal de Viçosa. Viçosa.

MARTINS, I.S. **Comparação entre métodos uni e multivariados aplicados na seleção em *Eucalyptus grandis***. 1999. 94f. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa. Viçosa.

RAO, R.C. **Advanced statistical methods in biometric research**. New York: John Wiley and Sons, 1952. 390 p.

REGAZZI, A. J. INF 766 - Análise Multivariada (notas manuscritas). DPI - UFV. Viçosa – MG, 1997.

SAS/STAT users' guide. Version 8. Cary, NC: **SAS Institute Inc.**, 1999.

STONE, H.; SIDEL, J. L. **Sensory Evaluation Practices**. London: Academic Press: New York, 1993.

YOKOTA, S.R.C. **Avaliação sensorial descritiva de cachaça envelhecida entre 18 e 24 meses por 4 painéis de julgadores**. 2005. 102f. Tese (Doutorado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa. Viçosa.

## **ARTIGO 2**

### **ANÁLISE DE FATORES APLICADA À AVALIAÇÃO SENSORIAL DE ALIMENTOS**

### **FACTOR ANALYSIS APPLIED TO SENSORY EVALUATION OF FOOD**

#### **RESUMO**

Os dados experimentais referentes a cachaças envelhecidas de 14 marcas comerciais, avaliadas em 13 atributos sensoriais pelo método Análise Descritiva Quantitativa – ADQ (YOKOTA, 2005), foram submetidos à análise de fatores com o objetivo de detalhar a aplicação de tal técnica na análise de dados provenientes da avaliação sensorial descritiva. Apresentou-se os programas necessários para a execução dos procedimentos estatísticos no sistema SAS (Statistical Analysis System) e deu-se ênfase à interpretação e discussão dos resultados. As cachaças foram caracterizadas e comparadas em função de apenas quatro fatores comuns. O primeiro recebeu o nome de efeito da madeira por correlacionar-se com os atributos aroma de madeira e aroma de baunilha, sabor de madeira inicial e sabor de madeira residual, coloração amarela e viscosidade. O segundo fator foi nomeado de teor alcoólico por correlacionar-se com os atributos aroma alcoólico e com o sabor alcoólico inicial e sabor alcoólico residual. Foi nomeado, também, o fator 3 como fator gosto doce por correlacionar-se com o atributo gosto doce e o fator 4 como fator gosto ácido por correlacionar-se com o gosto ácido. Apenas o gosto amargo e a adstringência não se correlacionaram com nenhum dos fatores. Vale ressaltar que as comunalidades foram altas para todos os atributos e que os quatro fatores comuns acumularam, juntos, 88,69% da variância dos dados

originais. Assim, obteve-se um pequeno número de fatores comuns e estes retiveram grande proporção da variância existente nos dados originais, as communalidades foram altas para os atributos e a maioria deles se correlacionou com um dos fatores comuns. Portanto, os resultados da análise de fatores foram satisfatórios e tal técnica pode ser utilizada para outros conjuntos de dados obtidos por análise descritiva.

**PALAVRAS-CHAVE:** análise multivariada, análise sensorial, análise descritiva quantitativa, ADQ.

## **SUMMARY**

The experimental data relative to aged cachaça from 14 commercial brands, evaluated in relation to 13 sensory attributes by Quantitative Descriptive Analysis - QDA (YOKOTA, 2005) were submitted to factor analysis to provide a detailed description of application of this technique to analyze data obtained from descriptive sensory evaluation. The programs needed to apply the statistical procedures in the SAS (Statistical Analysis System) package were shown, emphasizing result interpretation and discussion. The cachaça brands were characterized and compared using only four common factors. The first factor, wood effect, was correlated to the attributes wood aroma and vanilla aroma, initial and residual wood flavor, yellowish color and viscosity. The second factor, alcohol content, was correlated with the attributes alcoholic aroma as well as initial and residual alcoholic flavor. The third factor, sweet taste, was correlated with the attribute sweet taste and the fourth factor, acid taste equally correlated with acid taste. The attributes bitter taste and astringency were the only ones, which were not significantly correlated to any of the four factors. In addition, all communalities showed high values for all attributes, with the four factors together retaining 88.69% of the data variance. Thus, a small number of common factors was obtained retaining high proportion of the original data variance, communalities were high for the attributes and most of them were significantly correlated with one of the common factors. It was concluded that factor analysis produced good results and should be applied to data obtained from descriptive analysis.

**KEYWORDS:** multivariate analysis, sensory analysis, quantitative descriptive analysis, QDA.

## 1. INTRODUÇÃO

A análise fatorial ou análise de fatores é um nome genérico dado a uma classe de métodos estatísticos multivariados cujo propósito principal é definir a estrutura subjacente em uma matriz de dados. Em termos gerais, a análise fatorial aborda o problema de analisar a estrutura das inter-relações (correlações) entre um grande número de variáveis, como por exemplo, escores de testes, itens de testes, respostas de questionários. O objetivo é definir um conjunto de dimensões latentes comuns, chamadas de fatores. Com a análise fatorial, o pesquisador pode primeiro identificar as dimensões e então determinar o grau em que cada variável é explicada por cada dimensão. Uma vez que essas dimensões e a explicação de cada variável estejam determinadas, os dois principais usos da análise fatorial – resumo e redução dos dados – podem ser conseguidos. Ao resumir os dados, a análise fatorial define dimensões latentes que, quando interpretadas e compreendidas, descrevem os dados em um número muito menor de conceitos do que as variáveis individuais originais. A redução de dados pode ser conseguida calculando-se escores para cada dimensão latente e substituindo-se as variáveis originais pelos mesmos (HAIR et al., 2005).

A análise de fatores tem sido utilizada em várias áreas do conhecimento (CRUZ e CARNEIRO, 2003; MARQUES, 2000; MARTINS, 1999 e POLIGNANO et al., 1999). Entretanto, a utilização desta técnica na área de avaliação sensorial ainda é pequena. Assim, este estudo teve por objetivo ilustrar a aplicação da análise de fatores a dados provenientes da avaliação sensorial, com ênfase nos procedimentos estatísticos utilizados no sistema SAS (Statistical Analysis System) e na interpretação e discussão dos resultados.

## **2. MATERIAL E MÉTODOS**

### **2.1. DADOS EXPERIMENTAIS**

Foram utilizados os mesmos dados experimentais da Análise Descritiva Quantitativa de cachaças envelhecidas (YOKOTA, 2005), como descritos no primeiro artigo.

### **2.2. METODOLOGIA**

A análise de fatores foi realizada com base nos trabalhos de REGAZZI (2004), CRUZ e CARNEIRO (2003), CRUZ (2001), MARQUES (2000), MARTINS (1999), e JOHNSON e WICHERN (1992).

As análises estatísticas foram executadas no programa SAS (1999), licenciado para UFV em 2005. São apresentados os procedimentos estatísticos e comentários a respeito de suas funções. Os comentários aparecem entre barras e asteriscos, /\*comentário\*/, uma alternativa para se inserir comentários na janela editora do programa.

#### **2.2.1. ANÁLISES PRELIMINARES**

##### **a. ANÁLISE DE VARIÂNCIA, MÉDIAS DOS ATRIBUTOS E ANÁLISE DE CORRELAÇÃO (PROGRAMA 1)**

Os dados experimentais foram inicialmente submetidos a análises de variância univariadas (ANOVA), tendo-se cachaças (CACHA), provadores (PROV), a interação (CACHA\*PROV) e o resíduo como fontes de variação, apenas para ter-se uma idéia da variabilidade entre as 14 cachaças em relação aos atributos avaliados. As médias dos atributos por cachaça apresentadas no Quadro 1 foram utilizadas nas análises de correlação e na análise de fatores. A análise de fatores pode ser realizada para um conjunto de dados constituídos por  $n$  indivíduos avaliados em  $p$  variáveis, ou para médias oriundas de  $n = 14$  tratamentos avaliados em  $p = 13$  variáveis, como é o caso deste estudo (Quadro 1). É comum dispor os  $n$  indivíduos (ou médias) em linhas e as  $p$  variáveis em colunas.

A partir das médias dos atributos (Quadro 1) por tipo de cachaça foi estimada a média geral para cada atributo e os respectivos desvios-padrão, utilizados na padronização dos atributos (Quadro 2), para numa outra etapa, estimar os escores fatoriais.

Quadro 1: Médias por cachaça nos 13 atributos sensoriais.

| CACHAÇA | N  | AALC | AMAD | ABAU | DOCE | ACID | AMAR | SALI | ... | VISC |
|---------|----|------|------|------|------|------|------|------|-----|------|
| 1       | 40 | 8,1  | 11,5 | 5,8  | 6,1  | 5,6  | 7,4  | 7,8  | ... | 10,6 |
| 2       | 40 | 7,7  | 10,8 | 3,9  | 4,9  | 7,1  | 8,1  | 7,6  | ... | 10,9 |
| 3       | 40 | 9,0  | 5,8  | 2,5  | 6,2  | 6,3  | 6,7  | 8,1  | ... | 10,1 |
| 4       | 40 | 8,5  | 8,6  | 3,4  | 6,0  | 6,8  | 6,6  | 7,9  | ... | 10,3 |
| 5       | 40 | 8,5  | 10,5 | 3,8  | 6,2  | 5,9  | 6,7  | 8,0  | ... | 11,2 |
| 6       | 40 | 8,6  | 8,6  | 3,6  | 5,9  | 5,5  | 6,0  | 8,0  | ... | 9,7  |
| 7       | 40 | 8,9  | 6,7  | 3,0  | 5,7  | 5,6  | 6,6  | 7,8  | ... | 9,7  |
| 8       | 40 | 8,1  | 9,3  | 4,1  | 6,2  | 6,4  | 6,2  | 7,1  | ... | 9,9  |
| 9       | 40 | 8,8  | 5,5  | 2,6  | 6,0  | 7,2  | 5,9  | 8,0  | ... | 9,6  |
| 10      | 40 | 8,4  | 9,1  | 4,0  | 6,9  | 6,1  | 6,4  | 7,4  | ... | 10,2 |
| 11      | 40 | 9,7  | 7,5  | 3,0  | 6,2  | 7,3  | 7,5  | 8,9  | ... | 10,0 |
| 12      | 40 | 8,9  | 8,6  | 4,9  | 6,3  | 6,3  | 6,3  | 7,9  | ... | 10,1 |
| 13      | 40 | 8,6  | 5,4  | 2,9  | 6,8  | 6,5  | 6,5  | 7,1  | ... | 9,5  |
| 14      | 40 | 8,5  | 8,5  | 4,8  | 6,5  | 6,8  | 6,4  | 8,1  | ... | 10,4 |

N: número de observações;

AALC: aroma alcoólico, AMAD: aroma de madeira, ABAU: aroma de baunilha, DOCE: gosto doce, ACID: gosto ácido, AMAR: gosto amargo, SALI: sabor alcoólico inicial,...,VISC: viscosidade.

A análise de correlação foi realizada com o intuito de se verificar a existência de relações lineares entre os atributos, tendo-se em vista que muitas correlações significativas entre os atributos é um indício de que a análise de fatores será bem sucedida, pois esta análise busca a separação dos atributos de forma que cada grupo de atributos correlacionados entre si compõe um dos fatores.

A seguir estão apresentadas as análises e estimativas mencionadas.

**/\*Procedimento que importa os dados da planilha do Excel\*/**

```
PROC IMPORT out=dados_originais
            datafile="D:\fatorial\programas\dados_ori.xls" /*1*/
            dbms=excel2000 replace;
            getnames=yes;
RUN;
options nodate nonumber;
/*1*/ /*Informa que o arquivo de dados a ser lido, dados_ori.xls, está na
      pasta "D:\fatorial\programas". Este arquivo é o mesmo utilizado no
      primeiro artigo.*/
```

**/\*Procedimento que imprime, na tela, os dados lidos\*/**

```
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS ORIGINAIS';
RUN;
```

Quadro 2: Médias originais e padronizadas para os atributos em estudo.

| CACHAÇAS                            | Médias originais ( $X_{ij}$ ) |      |      |      |     |      |
|-------------------------------------|-------------------------------|------|------|------|-----|------|
|                                     | AALC                          | AMAD | ABAU | DOCE | ... | VISC |
| 1                                   | 8,1                           | 11,5 | 5,8  | 6,1  | ... | 10,6 |
| 2                                   | 7,7                           | 10,8 | 3,9  | 4,9  | ... | 10,9 |
| 3                                   | 9,0                           | 5,8  | 2,5  | 6,2  | ... | 10,1 |
| 4                                   | 8,5                           | 8,6  | 3,4  | 6,0  | ... | 10,3 |
| 5                                   | 8,5                           | 10,5 | 3,8  | 6,2  | ... | 11,2 |
| 6                                   | 8,6                           | 8,6  | 3,6  | 5,9  | ... | 9,7  |
| 7                                   | 8,9                           | 6,7  | 3,0  | 5,7  | ... | 9,7  |
| 8                                   | 8,1                           | 9,3  | 4,1  | 6,2  | ... | 9,9  |
| 9                                   | 8,8                           | 5,5  | 2,6  | 6,0  | ... | 9,6  |
| 10                                  | 8,4                           | 9,1  | 4,0  | 6,9  | ... | 10,2 |
| 11                                  | 9,7                           | 7,5  | 3,0  | 6,2  | ... | 10,0 |
| 12                                  | 8,9                           | 8,6  | 4,9  | 6,3  | ... | 10,1 |
| 13                                  | 8,6                           | 5,4  | 2,9  | 6,8  | ... | 9,5  |
| 14                                  | 8,5                           | 8,5  | 4,8  | 6,5  | ... | 10,4 |
| Médias por atributo ( $\bar{X}_j$ ) | 8,6                           | 8,3  | 3,7  | 6,1  | ... | 10,2 |
| Desvio padrão $s(X_j)$              | 0,5                           | 1,9  | 0,9  | 0,5  | ... | 0,5  |

| CACHAÇAS | Médias padronizadas ( $Z_{ij}$ ) |       |                   |                   |     |                   |
|----------|----------------------------------|-------|-------------------|-------------------|-----|-------------------|
|          | AALCp                            | AMADp | ABAU <sub>p</sub> | DOCE <sub>p</sub> | ... | VISC <sub>p</sub> |
| 1        | -1,1                             | 1,7   | 2,2               | -0,1              | ... | 0,9               |
| 2        | -1,9                             | 1,3   | 0,2               | -2,6              | ... | 1,5               |
| 3        | 0,9                              | -1,3  | -1,3              | 0,1               | ... | -0,1              |
| 4        | -0,2                             | 0,2   | -0,3              | -0,3              | ... | 0,4               |
| 5        | -0,3                             | 1,1   | 0,0               | 0,2               | ... | 2,1               |
| 6        | 0,0                              | 0,2   | -0,2              | -0,4              | ... | -0,8              |
| 7        | 0,6                              | -0,8  | -0,8              | -0,8              | ... | -1,0              |
| 8        | -1,0                             | 0,5   | 0,4               | 0,1               | ... | -0,6              |
| 9        | 0,4                              | -1,5  | -1,2              | -0,3              | ... | -1,1              |
| 10       | -0,4                             | 0,4   | 0,3               | 1,6               | ... | 0,1               |
| 11       | 2,3                              | -0,4  | -0,7              | 0,2               | ... | -0,4              |
| 12       | 0,6                              | 0,1   | 1,2               | 0,2               | ... | -0,1              |
| 13       | 0,1                              | -1,5  | -0,9              | 1,4               | ... | -1,3              |
| 14       | -0,1                             | 0,1   | 1,2               | 0,8               | ... | 0,5               |

$$* Z_{ij} = \frac{X_{ij} - \bar{X}_j}{s(X_j)};$$

AALC: aroma alcoólico, AMAD: aroma de madeira, ABAU: aroma de baunilha, DOCE: gosto doce,...,VISC: viscosidade.

```

/*Procedimento que realiza análise de variância*/
PROC ANOVA data=dados_originais;
class CACHA PROV REPE;
model AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC
= CACHA PROV CACHA*PROV; /*2*/
test H=CACHA E=CACHA*PROV; /*3*/
title 'ANÁLISE DE VARIÂNCIA DOS ATRIBUTOS';
RUN;

```

```

/*2*/ /* O modelo adotado na ANOVA é para o delineamento em blocos
casualizados (d.b.c) e cada provador (PROV) foi considerado como
um bloco.*/

/*3*/ /* No modelo adotado, o F para cachaça foi obtido considerando o
quadrado médio da interação CACHA*PROV como variância residual.*/

/*Procedimento que calcula as médias dos atributos, por cachaça, para
serem utilizadas na análise de correlação e de fatores*/
PROC MEANS data=dados_originais;
output out=medias_por_cacha;
by CACHA;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS DOS ATRIBUTOS, POR CAHAÇA';
RUN;

/*Procedimento que organiza as médias para serem usadas na análise de
correlação e de fatores*/ /*4*/
DATA medias_por_cacha; set medias_por_cacha(where=(_STAT_='MEAN'))
keep= CACHA _STAT_ AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR
ADST COLA VISC);
RUN;

/*4*/ /*As médias dos atributos, por cachaça, estão apresentadas no
Quadro 1.*/

/*Procedimento que imprime, na tela, as médias*/
PROC PRINT data=medias_por_cacha;
title 'MÉDIAS DOS ATRIBUTOS POR CAHAÇA';
RUN;

/*Procedimento que estima a média geral para cada atributo e
respectivo desvio padrão*/ /*5*/
PROC MEANS data=medias_por_cacha;
output out=medias_por_atributo;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS POR ATRIBUTO';
RUN;

/*5*/ /*Valores apresentados no Quadro 2*/

/*Procedimento que imprime, na tela, as médias por atributo*/
PROC PRINT data=medias_por_atributo;
title 'MÉDIAS POR ATRIBUTO';
RUN;

/*Procedimento que obtêm as correlações (Coeficiente de Correlação de
Pearson) entre as médias dos atributos*/
PROC CORR data=medias_por_cacha; /*6*/
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'CORRELAÇÕES ENTRE OS ATRIBUTOS MÉDIOS';
RUN;

/*6*/ /*As correlações foram obtidas a partir das médias dos atributos
por cachaça (Quadro 1).*/

```

## **b. PADRONIZAÇÃO DAS MÉDIAS (PROGRAMA 2)**

As médias dos atributos, por cachaça (Quadro 1), foram padronizadas para média zero e variância um:



$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{s(X_j)}, \text{ sendo:}$$

$i = 1, 2, \dots, n = 14$  cachaaas;

$j = 1, 2, \dots, p = 13$  atributos;

$Z_{ij}$ : médias padronizadas dos atributos;

$X_{ij}$ : médias originais dos atributos;

$\bar{X}_j$ : média geral do  $j$ -ésimo atributo;

$s(X_j)$ : desvio-padrão associado ao  $j$ -ésimo atributo.

Os procedimentos para a padronização estão apresentados a seguir e os resultados dispostos no Quadro 2.

```

/*Procedimento que importa os dados da planilha do Excel*/
PROC IMPORT out=dados_originais
    datafile="D:\fatorial\programas\dado_ori.xls"
    dbms=excel2000 replace;
    getnames=yes;
RUN;
options nodate nonumber;

/*Procedimento que imprime, na tela, os dados lidos*/
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS ORIGINAIS';
RUN;

/*Procedimento que calcula as médias dos atributos, por cachaa*/
PROC MEANS data=dados_originais;
output out=medias_por_cacha;
by CACHA;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS DOS ATRIBUTOS POR CACHAAS';
RUN;

/*Procedimento que organiza os atributos médios para serem utilizados
no próximo procedimento*/
DATA medias_por_cacha; set medias_por_cacha(where=( _STAT_='MEAN' )
keep= CACHA _STAT_ AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR
ADST COLA VISC);
RUN;

/*Procedimento que imprime, na tela, as médias*/
PROC PRINT data=medias_por_cacha;
title 'MÉDIAS DOS ATRIBUTOS POR CACHA';
RUN;

/*Procedimento para padronizar as médias dos atributos*/***/
PROC STANDARD data=medias_por_cacha mean=0 std=1
out=medias_padronizadas;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
RUN;

***/** Padronização para média igual a zero e variância igual a um.*/

```

```

/*Procedimento que imprime, na tela, as médias padronizadas*/
PROC PRINT data=medias_padronizadas;
var CACHA AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA
VISC;
title 'MÉDIAS PADRONIZADAS';
RUN;

```

### 2.2.2. ANÁLISE DE FATORES (PROGRAMA 3)

A análise de fatores pode ser resumida nas seguintes etapas:

1. determinação da matriz de correlação entre as médias dos atributos em estudo;
2. escolha do modelo;
3. determinação do número de fatores comuns necessários para representar os dados;
4. cálculo das cargas fatoriais iniciais;
5. rotação dos fatores;
6. cálculo dos escores fatoriais;
7. caracterização sensorial e comparação entre as cachaças baseada nos escores fatoriais.

O procedimento PROC FACTOR foi executado a partir do arquivo “medias\_por\_cacha” contendo as médias originais dos atributos por cachaça. Tais médias foram estimadas a partir do arquivo “dado\_ori.xls” contendo os dados experimentais. O primeiro passo do PROC FACTOR foi determinar a matriz de correlação, R, entre as médias originais dos atributos, a partir da qual a análise de fatores foi prosseguida. Vale ressaltar que se optou por trabalhar com dados padronizados. Neste caso a matriz de covariâncias (S) entre as médias padronizadas é igual à matriz de correlação (R) entre as médias originais. Outra opção seria usar a matriz S estimada a partir das médias dos dados originais.

A exemplo dos estudos de REGAZZI (2004), CRUZ (2001), MARQUES (2000), MARTINS (1999), e JOHNSON e WICHERN (1992), foi utilizado o seguinte modelo fatorial ortogonal, ou seja, de fatores não correlacionados:

$$X_j = I_{j1}F_1 + I_{j2}F_2 + \dots + I_{jm}F_m + e_j, \text{ em que:}$$

$X_j$  = j-ésimo atributo, com  $j = 1, 2, \dots, p = 13$ ;

$I_{j1}, I_{j2}, \dots, I_{jm}$  : cargas dos m fatores comuns, para o j-ésimo atributo;

$F_1, F_2, \dots, F_m$  : m fatores comuns;

$e_j$  = fator específico do j-ésimo atributo.

Define-se  $h_j^2$  como sendo a comunalidade de uma variável  $X_j$ , dada por:  $h_j^2 = l_{j1}^2 + l_{j2}^2 + \dots + l_{jm}^2$ . A comunalidade expressa a proporção da variância de  $X_j$  que é explicada pelos fatores comuns. Adicionalmente, define-se especificidade de  $X_j$ , a proporção da variância de  $X_j$  dada pela ação dos fatores específicos.

A análise de fatores procura explicar o máximo da variação em  $X_j$  com o menor número possível de fatores, tornando  $e_j$  mínimo. Optou-se, neste estudo, por determinar o número de fatores comuns ( $m$ ) igual ao número de autovalores de  $R$  maiores do que um (JOHNSON e WICHERN, 1992), uma maneira usual de se determinar o número de fatores. Inicialmente, a análise foi realizada com seis fatores e foi observado que existiam quatro autovalores maiores que a unidade. Procedeu-se, então, à análise definitiva, na qual se especificou o número de fatores igual a quatro.

O método dos Componentes Principais foi utilizado na obtenção dos autovalores, autovetores e das cargas fatoriais iniciais. Tais autovalores ( $\lambda_i$ ), ou raízes características da matriz de correlação ( $R$ ), foram estimados resolvendo-se a expressão  $\det(R - \lambda_i I) = 0$ , sendo  $I$  a matriz identidade. Cada autovetor ( $a_i^*$ ), associado a um autovalor ( $\lambda_i$ ), foi obtido a partir de  $(R - \lambda_i I)a_i^* = \phi$ , em que  $\phi$  é um vetor nulo. Os autovetores ( $a_i^*$ ) foram normalizados, obtendo-se ( $a_i$ ) tal que  $a_i' \cdot a_i = 1$ . As cargas fatoriais iniciais são as correlações (coeficientes de correlação de Pearson) dos quatro fatores comuns com os 13 atributos em estudo. Tais cargas foram obtidas como:

$$l_{ij} = \sqrt{\lambda_i} \cdot a_{ij}, \text{ em que:}$$

$l_{ij}$ : carga fatorial inicial do  $i$ -ésimo fator correspondente ao  $j$ -ésimo atributo;

$\lambda_i$ :  $i$ -ésimo autovalor maior que a unidade, obtido da matriz de correlação ( $R$ );

$a_{ij}$ :  $j$ -ésimo valor do  $i$ -ésimo autovetor.

Cada fator possui cargas individuais para todas as variáveis, o que esclarece a razão da denominação “fator comum”, empregada na teoria de análise de fatores (MARTINS, 1999).

A rotação dos fatores é uma transformação ortogonal correspondente a uma rotação dos eixos das coordenadas e, portanto, uma transformação ortogonal das cargas fatoriais e, conseqüentemente dos fatores (REGAZZI, 2004). Após a rotação, algumas cargas fatoriais (correlações) se aproximam dos limites extremos 1 ou -1, e outras de zero. E alguns valores intermediários tendem a desaparecer. Isto facilita a interpretação, pois cada fator apresentará

correlação relativamente forte com um ou mais atributos e correlação relativamente fraca com os demais atributos. Foi escolhido, para a rotação, o método ortogonal Varimax. Como o nome já sugere, trata-se da rotação que torna máxima uma variância. Desse modo consegue-se que cada fator inclua a maior porcentagem possível da variância existente nos atributos originais por ele representados. Foram estimadas, com a rotação, as cargas fatoriais finais (correlações entre atributos e fatores) e estas usadas para identificar os atributos que compunham cada fator. Considera-se neste procedimento, que os atributos correlacionados com um mesmo fator são os que influenciam e compõem tal fator.

Vale ressaltar que as comunalidades ( $h_j^2$ ) não se alteram com a rotação dos fatores e podem, agora, serem escritas como  $h_j^2 = l_{j1}^{*2} + l_{j2}^{*2} + \dots + l_{jm}^{*2}$ , em que:

$l_{j1}^*, l_{j2}^*, \dots, l_{jm}^*$ : Cargas fatoriais finais, após a rotação, associadas ao j-ésimo atributo, nos m fatores comuns.

A partir dos 4 fatores comuns foram estimados os escores para serem utilizados na caracterização e comparação entre as cachaças, como se fossem os valores de novas variáveis, relacionadas com esses fatores. Tais escores fatoriais foram estimados com a equação:

$$F^{*'} = (L^{*} L^{*'})^{-1} L^{*} Z' = \beta Z' \quad (\text{CRUZ e CARNEIRO, 2003}), \text{ sendo:}$$

$F^{*}$ : matriz de dimensão mxn (4x14) dos escores fatoriais;

$L^{*}$ : matriz de dimensão mxp (4x13) das cargas fatoriais finais (após a rotação dos fatores);

$Z$ : matriz de dimensão nxp (14x13) dos atributos médios padronizados;

$\beta$ : matriz de dimensão mxp (4x13) das cargas canônicas (coeficientes de ponderação dos atributos médios padronizados, para obtenção dos escores fatoriais). E  $\beta = (L^{*} L^{*'})^{-1} L^{*}$ .

Neste caso, obtém-se:

$$F_k^{*} = b_{1k}Z_1 + b_{2k}Z_2 + \dots + b_{pk}Z_p, \text{ sendo:}$$

$F_k^{*}$ : escores fatoriais;

$b_{jk}$ : elemento da matriz  $\beta$  (cargas canônicas);

$k = 1, \dots, m = 4$  fatores comuns;

$Z_1, Z_2, \dots, Z_p$ : atributos médios padronizados.

Assim, as novas variáveis associadas aos 4 fatores comuns foram escritas como:

$$F_1^{*} = b_{1,1}Z_1 + b_{2,1}Z_2 + \dots + b_{13,1}Z_{13}$$

$$F_2^{*} = b_{1,2}Z_1 + b_{2,2}Z_2 + \dots + b_{13,2}Z_{13}$$

$$F_3^{*} = b_{1,3}Z_1 + b_{2,3}Z_2 + \dots + b_{13,3}Z_{13}$$

$$F_4^* = b_{1,4}Z_1 + b_{2,4}Z_2 + \dots + b_{13,4}Z_{13}$$

Com a substituição de  $b_{jk}$  e  $Z_1, Z_2, \dots, Z_p$  pelos respectivos valores numéricos obtiveram-se os escores fatoriais que foram representados em gráfico de dispersão das cachaças. A caracterização sensorial e comparação entre as cachaças foram feitas em função destes escores.

A seguir estão os procedimentos utilizados para obter as estimativas mencionadas nas etapas da análise de fatores.

```

/*Procedimento que importa os dados da planilha do Excel*/
PROC IMPORT out=dados_originais
            datafile="D:\fatorial\programas\dados_ori.xls" /*8*/
            dbms=excel2000 replace;
            getnames=yes;
RUN;
options nodate nonumber;

/*8*/ /*Os dados do arquivo "dados_ori.xls" são os mesmos já utilizados
      nos programas anteriores.*/

/*Procedimento que imprime, na tela, os dados lidos*/
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS ORIGINAIS';
RUN;

/*Procedimento que calcula as médias dos atributos, por cachaça, para
serem utilizadas na análise de fatores*/ /*9*/
PROC MEANS data=dados_originais;
output out=medias_por_cacha;
by CACHA;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇA';
RUN;

/*9*/ /*Tais médias estão apresentadas no Quadro 1.*/

/*Procedimento que organiza as médias para serem usadas na análise de
fatores*/
DATA medias_por_cacha; set medias_por_cacha(where=( _STAT_='MEAN' )
keep= CACHA _STAT_ AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR
ADST COLA VISC);
RUN;

/*Procedimento que imprime, na tela, as médias*/
PROC PRINT data=medias_por_cacha;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇA';
RUN;

/*Procedimento que realiza a análise de fatores*/
PROC FACTOR data=medias_por_cacha NFACTOR=4 METHOD=P SCORE SCREE R=V
OUTSTAT=SAIDAL OUT=RESULT; /*10*/
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'ANÁLISE DE FATORES';
RUN;

```

```

/*10*/  /*A análise de fatores foi feita a partir do arquivo
"medias_por_cacha" que contém as médias dos atributos por
cachaça. Tais médias estão apresentadas, também, no Quadro 1.*/

/*NFACTOR=4 informa que serão 4 os fatores comuns;*/

/*METHOD=P informa que para estimar os autovalores, autovetores
e as cargas fatoriais iniciais será utilizado o método dos
Componentes Principais;*/

/*Foi omitida a letra S após o procedimento PROC FACTOR,
indicando que a análise de fatores foi realizada a partir da
matriz de correlação entre os atributos e não da matriz de
covariância S;*/

/*R=V informa que a rotação dos fatores será feita pelo método
Varimax.*/

/*Procedimento que imprime, na tela, os escores fatoriais médios*//*11*/
PROC PRINT data=RESULT;
var FACTOR1-FACTOR4;
title 'ESCORES FATORIAIS MÉDIOS';
RUN;

/*11*/ /*Estes escores fatoriais médios foram obtidos direto pelo PROC
FACTOR, a partir dos atributos médios padronizados e das cargas
canônicas.*/

/*Procedimento que gera o gráfico de dispersão das cachaças em função
dos fatores Factor1 e Factor2, a partir dos escores fatoriais
médios)*//*12*/
PROC PLOT data=RESULT;
plot FACTOR2*FACTOR1 = CACHA / vpos=20;
title 'GRAFICO DE DISPERSÃO DAS CACHAÇAS (Factor1 x Factor2)';
RUN;

/*12*//*O gráfico de dispersão foi feito, também, no EXCEL, a partir dos
escores fatoriais médios, pois o autor é mais familiarizado com a
edição de gráficos no aplicativo Excel.*/

/*Procedimento que imprime, na tela, todos os resultados de forma
resumida*/
PROC PRINT data=SAIDA1;
title 'RESULTADOS RESUMIDOS';
RUN;

```

## **MÉTODO DE AGRUPAMENTO DE TOCHER**

A avaliação visual do posicionamento das cachaaas, no gráfico de dispersão, permite agrupá-las, de forma que as cachaaas que estão mais próximas entre si fazem parte de um mesmo grupo. O estabelecimento destes grupos pode variar de um analista para outro, sendo portanto uma avaliação subjetiva. Para evitar a subjetividade, o agrupamento das cachaaas foi realizado pelo Método de Agrupamento de Otimização proposto por Tocher, citado por RAO (1952). Este método faz parte dos procedimentos do Programa Genes (Cruz, 2001), o qual foi utilizado no estabelecimento dos grupos.

O agrupamento, neste método, foi realizado com base na distância Euclidiana Média, que é uma medida de dissimilaridade entre as cachaaas. A distância Euclidiana Média, por sua vez, foi obtida a partir dos atributos médios padronizados (CRUZ, 2001). Vale ressaltar que foi utilizada a padronização para média igual a zero e variância igual a um.

## **3. RESULTADOS E DISCUSSÃO**

Observa-se, pelos resultados da análise de variância (Quadro 3), que existe diferença significativa ( $P = 0,05$ ) entre as médias das cachaaas, para 12 dos 13 atributos avaliados. Portanto, há diferenças entre as cachaaas e faz-se necessário investigá-las. Uma das alternativas é usar a análise de fatores. O fato de existirem correlações significativas ( $P = 0,05$ ) entre vários dos atributos (Quadro 4) sinaliza que esta análise pode ser bem sucedida.

Estão apresentadas, no Quadro 5, as estimativas dos autovalores da matriz de correlação (R), a proporção da variância retida por cada fator e a proporção acumulada. Quando os autovalores são obtidos a partir da matriz R, basta considerar na análise os fatores associados aos autovalores maiores do que um. Assim, fizeram parte da análise apenas os primeiros quatro fatores, o que garante que são estes fatores que retêm a maior proporção da variância dos dados originais (88,69 %).

Quadro 3: Resumo das análises de variância univariadas dos dados correspondentes aos 13 atributos avaliados nas 14 cachaças envelhecidas.

| Atributos                 | QMC    | QMC*P              | QMR  | F <sub>C</sub> | Prob.> F <sub>C</sub> | Média |
|---------------------------|--------|--------------------|------|----------------|-----------------------|-------|
| Aroma alcoólico           | 8,51   | 4,32*              | 2,44 | 1,97           | 0,03                  | 8,59  |
| Aroma de madeira          | 151,46 | 8,57*              | 2,49 | 17,67          | <0,00                 | 8,32  |
| Aroma de baunilha         | 35,73  | 9,98*              | 2,96 | 3,58           | 0,00                  | 3,73  |
| Gosto doce                | 10,02  | 5,55 <sup>ns</sup> | 4,57 | 1,81           | 0,05                  | 6,13  |
| Gosto ácido               | 14,67  | 10,19*             | 6,36 | 1,44           | 0,16                  | 6,38  |
| Gosto amargo              | 14,93  | 5,94 <sup>ns</sup> | 5,22 | 2,51           | 0,01                  | 6,66  |
| Sabor alcoólico inicial   | 7,93   | 4,37*              | 2,05 | 1,82           | 0,05                  | 7,84  |
| Sabor alcoólico residual  | 8,85   | 4,90*              | 2,27 | 1,80           | 0,05                  | 8,79  |
| Sabor de madeira inicial  | 129,65 | 7,39*              | 2,43 | 17,55          | <0,00                 | 7,94  |
| Sabor de madeira residual | 152,56 | 9,64*              | 2,76 | 15,83          | <0,00                 | 8,25  |
| Adstringência             | 11,49  | 4,92*              | 3,61 | 2,34           | 0,01                  | 6,60  |
| Coloração amarela         | 552,46 | 4,48*              | 2,73 | 123,26         | <0,00                 | 6,24  |
| Viscosidade               | 9,74   | 4,13*              | 1,98 | 2,36           | 0,01                  | 10,15 |

QMC: Quadrados médios para cachaça;

QMC\*P: Quadrados médios para a interação entre cachaça e provador;

QMR: Quadrados médios do resíduo;

F<sub>C</sub>: valor de F para cachaça ( $F_C = QMC/QMC*P$ );

Prob.: Valores de probabilidade (valor-P);

\*Interação entre cachaça e provador significativa ( $P \leq 0,05$ );

<sup>ns</sup>Interação entre cachaça e provador não significativa ( $P > 0,05$ ).

#### Esquema da ANOVA

| F.V.           | G.L. |
|----------------|------|
| Cachaça (C)    | 13   |
| Provador (P)   | 7    |
| Interação(C*P) | 91   |
| Resíduo        | 448  |

F.V.: fontes de variação

G.L.: graus de liberdade



Quadro 4: Matriz de correlação (coeficientes de correlação de Pearson) entre os atributos médios em estudo.

| ATRIBUTOS | AALC                 | AMAD                   | ABAU                | DOCE          | ACID          | AMAR                | SALI                   | SALR          | SMAI                   | SMAR                   | ADST         | COLA                | VISC |
|-----------|----------------------|------------------------|---------------------|---------------|---------------|---------------------|------------------------|---------------|------------------------|------------------------|--------------|---------------------|------|
| AALC      | 1,00                 |                        |                     |               |               |                     |                        |               |                        |                        |              |                     |      |
| AMAD      | <b>-0,65</b><br>0,01 | 1,00                   |                     |               |               |                     |                        |               |                        |                        |              |                     |      |
| ABAU      | -0,49<br>0,07        | <b>0,77</b><br>0,00    | 1,00                |               |               |                     |                        |               |                        |                        |              |                     |      |
| DOCE      | 0,32<br>0,26         | -0,29<br>0,31          | 0,06<br>0,85        | 1,00          |               |                     |                        |               |                        |                        |              |                     |      |
| ACID      | 0,20<br>0,50         | -0,25<br>0,38          | -0,30<br>0,29       | -0,15<br>0,60 | 1,00          |                     |                        |               |                        |                        |              |                     |      |
| AMAR      | -0,19<br>0,51        | 0,46<br>0,10           | 0,21<br>0,47        | -0,49<br>0,08 | 0,26<br>0,37  | 1,00                |                        |               |                        |                        |              |                     |      |
| SALI      | <b>0,67</b><br>0,01  | -0,09<br>0,76          | -0,14<br>0,64       | -0,17<br>0,57 | 0,25<br>0,38  | 0,22<br>0,45        | 1,00                   |               |                        |                        |              |                     |      |
| SALR      | <b>0,72</b><br>0,00  | -0,23<br>0,43          | -0,23<br>0,43       | -0,17<br>0,57 | 0,19<br>0,51  | 0,17<br>0,56        | <b>0,96</b><br><0,0001 | 1,00          |                        |                        |              |                     |      |
| SMAI      | <b>-0,65</b><br>0,01 | <b>0,99</b><br><0,0001 | <b>0,75</b><br>0,00 | -0,26<br>0,37 | -0,20<br>0,50 | 0,42<br>0,14        | -0,10<br>0,74          | -0,26<br>0,37 | 1,00                   |                        |              |                     |      |
| SMAR      | <b>-0,62</b><br>0,02 | <b>0,98</b><br><0,0001 | <b>0,75</b><br>0,00 | -0,24<br>0,40 | -0,18<br>0,54 | 0,43<br>0,12        | -0,09<br>0,76          | -0,25<br>0,38 | <b>0,99</b><br><0,0001 | 1,00                   |              |                     |      |
| ADST      | -0,14<br>0,64        | <b>0,56</b><br>0,04    | 0,49<br>0,07        | -0,25<br>0,39 | 0,34<br>0,24  | 0,46<br>0,09        | 0,14<br>0,64           | -0,02<br>0,93 | <b>0,58</b><br>0,03    | <b>0,63</b><br>0,01    | 1,00         |                     |      |
| COLA      | <b>-0,64</b><br>0,01 | <b>0,97</b><br><0,0001 | <b>0,71</b><br>0,00 | -0,24<br>0,41 | -0,25<br>0,39 | 0,36<br>0,20        | -0,09<br>0,77          | -0,24<br>0,40 | <b>0,98</b><br><0,0001 | <b>0,98</b><br><0,0001 | 0,50<br>0,07 | 1,00                |      |
| VISC      | -0,47<br>0,09        | <b>0,78</b><br>0,00    | 0,47<br>0,09        | -0,27<br>0,34 | -0,01<br>0,96 | <b>0,54</b><br>0,05 | 0,13<br>0,66           | -0,05<br>0,87 | <b>0,80</b><br>0,00    | <b>0,78</b><br>0,00    | 0,37<br>0,19 | <b>0,79</b><br>0,00 | 1,00 |

AALC: aroma alcoólico, AMAD: aroma de madeira, ABAU: aroma de baunilha, DOCE: gosto doce, ACID: gosto ácido, AMAR: gosto amargo, SALI: sabor alcoólico inicial, SALR: sabor alcoólico residual, SMAI: sabor de madeira inicial, SMAR: sabor de madeira residual, ADST: adstringência, COLA: coloração amarela e VISC: viscosidade.

Quadro 5: Estimativas dos autovalores da matriz R e variância retida pelos fatores.

| i  | Autovalores $\lambda_i$<br>de R | Fatores       | Proporção da<br>variância (%)* | Proporção<br>Acumulada (%) |
|----|---------------------------------|---------------|--------------------------------|----------------------------|
| 1  | <b>6,4332</b>                   | <b>Fator1</b> | <b>49,49</b>                   | <b>49,49</b>               |
| 2  | <b>2,7143</b>                   | <b>Fator2</b> | <b>20,88</b>                   | <b>70,37</b>               |
| 3  | <b>1,3724</b>                   | <b>Fator3</b> | <b>10,56</b>                   | <b>80,92</b>               |
| 4  | <b>1,0102</b>                   | <b>Fator4</b> | <b>7,77</b>                    | <b>88,69</b>               |
| 5  | 0,5602                          | Fator5        | 4,31                           | 93,00                      |
| 6  | 0,4530                          | Fator6        | 3,48                           | 96,49                      |
| 7  | 0,2791                          | Fator7        | 2,15                           | 98,63                      |
| 8  | 0,1270                          | Fator8        | 0,98                           | 99,61                      |
| 9  | 0,0241                          | Fator9        | 0,19                           | 99,80                      |
| 10 | 0,0133                          | Fator10       | 0,10                           | 99,90                      |
| 11 | 0,0065                          | Fator11       | 0,05                           | 99,95                      |
| 12 | 0,0060                          | Fator12       | 0,05                           | 100,00                     |
| 13 | 0,0006                          | Fator13       | 0,00                           | 100,00                     |

\* Proporção da variância(%) para o Fator<sub>i</sub> =  $\frac{\lambda_i}{\sum_{i=1}^{13} \lambda_i} \cdot 100\% = \frac{\lambda_i}{13} \cdot 100\%$

R: Matriz de correlação entre os atributos médios em estudo (Quadro 4).

As cargas fatoriais iniciais, as comunalidades e as cargas fatoriais finais, após a rotação pelo método Varimax, estão apresentadas no Quadro 6. Nota-se que as comunalidades são altas para todos os atributos, portanto, a maior parte da variância para cada um deles é explicada pelos 4 fatores comuns.

As cargas fatoriais representam as correlações (coeficientes de correlação de Pearson) entre cada fator comum (escores fatoriais) e cada um dos 13 atributos. Assim, cada uma das cargas fatoriais, destacadas no Quadro 6, indica que há correlação entre um atributo e um fator. Observa-se que os atributos aroma alcoólico, aroma de madeira, aroma de baunilha, sabor de madeira inicial e de madeira residual, coloração amarela e viscosidade estão correlacionados com o fator 1. O sabor alcoólico inicial e o residual com o fator 2. Porém, a adstringência e os gostos doce, ácido e amargo correlacionaram-se fracamente (apresentaram coeficientes de correlação com valores intermediários) com mais de um fator. Isto sugere que a rotação dos fatores pode simplificar os resultados.

Verifica-se, após a rotação dos fatores, que as comunalidades não se alteram e que apenas o gosto amargo e a adstringência continuam correlacionando-se fracamente com mais de um fator (Quadro 6). Portanto, a solução após a rotação dos fatores é melhor do que a primeira e será a considerada.

Quadro 6: Transposta da matriz de cargas fatoriais iniciais ( $L'$ ), communalidades e transposta da matriz de cargas fatoriais finais ( $L^{*}$ ), após rotação, estimadas nos atributos em estudo, avaliados nas cachaças envelhecidas.

| Atributos | Cargas fatoriais originais ( $L'$ ) |             |        |        | Comunalidades<br>Originais | Cargas fatoriais após rotação ( $L^{*}$ ) |                  |                  |                  | Comunalidades<br>após rotação |
|-----------|-------------------------------------|-------------|--------|--------|----------------------------|-------------------------------------------|------------------|------------------|------------------|-------------------------------|
|           | Fator1                              | Fator2      | Fator3 | Fator4 |                            | Fator1<br>FEMAD                           | Fator2<br>FTALCO | Fator3<br>FGDOCE | Fator4<br>FGACID |                               |
| AALC      | <b>-0,72</b>                        | 0,50        | 0,35   | 0,20   | 0,94                       | -0,50                                     | <b>0,75</b>      | 0,34             | 0,12             | 0,94                          |
| AMAD      | <b>0,98</b>                         | 0,04        | 0,12   | -0,05  | 0,99                       | <b>0,96</b>                               | -0,13            | -0,21            | -0,02            | 0,99                          |
| ABAU      | <b>0,77</b>                         | -0,11       | 0,35   | 0,25   | 0,79                       | <b>0,85</b>                               | -0,12            | 0,23             | 0,00             | 0,79                          |
| DOCE      | -0,31                               | -0,40       | 0,54   | 0,54   | 0,84                       | -0,11                                     | -0,03            | <b>0,90</b>      | -0,07            | 0,84                          |
| ACID      | -0,18                               | 0,48        | -0,61  | 0,48   | 0,86                       | -0,27                                     | 0,11             | -0,19            | <b>0,86</b>      | 0,86                          |
| AMAR      | 0,49                                | 0,54        | -0,37  | -0,10  | 0,68                       | 0,38                                      | 0,16             | -0,61            | 0,37             | 0,68                          |
| SALI      | -0,18                               | <b>0,90</b> | 0,35   | -0,07  | 0,98                       | 0,02                                      | <b>0,98</b>      | -0,12            | 0,11             | 0,98                          |
| SALR      | -0,33                               | <b>0,85</b> | 0,34   | -0,18  | 0,97                       | -0,14                                     | <b>0,97</b>      | -0,14            | 0,01             | 0,97                          |
| SMAI      | <b>0,98</b>                         | 0,04        | 0,11   | 0,01   | 0,98                       | <b>0,96</b>                               | -0,15            | -0,18            | 0,02             | 0,98                          |
| SMAR      | <b>0,98</b>                         | 0,06        | 0,10   | 0,06   | 0,98                       | <b>0,97</b>                               | -0,14            | -0,15            | 0,08             | 0,98                          |
| ADST      | 0,59                                | 0,39        | -0,17  | 0,56   | 0,85                       | 0,60                                      | 0,06             | -0,02            | 0,70             | 0,85                          |
| COLA      | <b>0,96</b>                         | 0,01        | 0,16   | -0,05  | 0,95                       | <b>0,95</b>                               | -0,14            | -0,17            | -0,06            | 0,95                          |
| VISC      | <b>0,80</b>                         | 0,26        | 0,03   | -0,15  | 0,73                       | <b>0,77</b>                               | 0,05             | -0,37            | 0,03             | 0,73                          |

AALC: aroma alcoólico, AMAD: aroma de madeira, ABAU: aroma de baunilha, DOCE: gosto doce, ACID: gosto ácido, AMAR: gosto amargo, SALI: sabor alcoólico inicial, SALR: sabor alcoólico residual, SMAI: sabor de madeira inicial, SMAR: sabor de madeira residual, ADST: adstringência, COLA: coloração amarela e VISC: viscosidade.

Estimativa da comunalidade, por exemplo, para a primeira variável avaliada (atributo AALC):

$$h_1^2 = l_{11}^2 + l_{12}^2 + l_{13}^2 + l_{14}^2 = (-0,72)^2 + (0,50)^2 + (0,35)^2 + (0,20)^2 = 0,94$$

$$h_1^2 = l_{11}^{*2} + l_{12}^{*2} + l_{13}^{*2} + l_{14}^{*2} = (-0,50)^2 + (0,75)^2 + (0,34)^2 + (0,12)^2 = 0,94, \text{ em que:}$$

$h_1^2$ : Comunalidade para a primeira variável avaliada (atributo AALC);

$l_{11}, l_{12}, l_{13}, l_{14}$ : Cargas fatoriais iniciais associadas ao primeiro atributo, AALC;

$l_{11}^*, l_{12}^*, l_{13}^*, l_{14}^*$ : Cargas fatoriais finais, após a rotação, associadas ao primeiro atributo, AALC.

As cargas fatoriais após a rotação, destacadas no Quadro 6, indicam os atributos que estão correlacionados com um determinado fator e baseando-se nestas cargas pode-se tentar compreender o significado dos fatores e até mesmo nomeá-los.

O fator rotacionado 1 tem altas cargas positivas para o aroma de madeira e de baunilha, sabor de madeira inicial e de madeira residual, coloração amarela e viscosidade (Quadro 6). Como a maioria destes atributos está associada à madeira, indica que o fator 1 reflete o efeito da madeira e recebeu o nome de **“fator efeito da madeira” (FEMAD)**.

O fator 2, após rotação, correlacionou-se positivamente com os atributos aroma alcoólico e com os sabores alcoólico inicial e residual (Quadro 6). Atributos, estes, relacionados ao conteúdo de álcool das cachaças. Portanto, o fator 2 foi nomeado de **“fator teor alcoólico” (FTALCO)**.

Foi nomeado, também, o fator 3 como **“fator gosto doce” (FGDOCE)** por correlacionar-se com o atributo gosto doce e o fator 4 como **“fator gosto ácido” (FGACID)** por correlacionar-se com o gosto ácido.

Não foi possível atribuir a influência do gosto amargo e da adstringência sobre nenhum dos 4 fatores, pois estes fatores não se correlacionaram com nenhum dos dois atributos (Quadro 6).

Neste estudo foi possível nomear os fatores. Embora, nem sempre é possível dar nomes aos fatores. Isto requer imaginação e um pouco de prática. Os fatores precisam ter um certo sentido prático em relação ao estudo que está sendo realizado.

A análise de fatores reduziu o número de atributos (13) para um menor número de fatores (4), que retêm as informações mais importantes dos dados originais. Assim, foi possível uma redução no conjunto de informações, de modo a possibilitar se estudar as cachaças envelhecidas em relação aos fatores efeito da madeira, teor alcoólico, gosto doce e gosto ácido.

As cargas canônicas, utilizadas como coeficientes de ponderação dos atributos médios padronizados, para obtenção dos escores dos fatores, estão apresentadas no Quadro 7. Estes coeficientes, aplicados aos dados padronizados (Quadro 2), geraram os escores fatoriais que estão apresentados no Quadro 8 e nas Figuras 1, 2 e 3. A análise destes escores permitiu a caracterização sensorial e comparações entre as cachaças envelhecidas.

Quadro 7: Transposta da matriz de cargas canônicas (transposta da matriz de coeficientes de ponderação ( $\beta'$ ) dos atributos médios padronizados, para obtenção dos escores fatoriais).

| Atributos | Fatores           |                |            |             |
|-----------|-------------------|----------------|------------|-------------|
|           | Efeito da madeira | Teor alcoólico | Gosto doce | Gosto ácido |
| AALC      | 0,01              | 0,29           | 0,25       | 0,08        |
| AMAD      | 0,17              | 0,02           | -0,02      | -0,06       |
| ABAU      | 0,21              | 0,04           | 0,30       | 0,04        |
| DOCE      | 0,12              | 0,04           | 0,66       | 0,12        |
| ACID      | -0,10             | -0,10          | -0,01      | 0,66        |
| AMAR      | -0,01             | 0,02           | -0,32      | 0,16        |
| SALI      | 0,08              | 0,42           | -0,01      | -0,06       |
| SALR      | 0,04              | 0,41           | -0,07      | -0,14       |
| SMAI      | 0,17              | 0,01           | 0,01       | -0,02       |
| SMAR      | 0,18              | 0,01           | 0,04       | 0,02        |
| ADST      | 0,12              | -0,01          | 0,21       | 0,54        |
| COLA      | 0,17              | 0,03           | 0,00       | -0,09       |
| VISC      | 0,12              | 0,07           | -0,15      | -0,07       |

AALC: aroma alcoólico, AMAD: aroma de madeira, ABAU: aroma de baunilha, DOCE: gosto doce, ACID: gosto ácido, AMAR: gosto amargo, SALI: sabor alcoólico inicial, SALR: sabor alcoólico residual, SMAI: sabor de madeira inicial, SMAR: sabor de madeira residual, ADST: adstringência, COLA: coloração amarela e VISC: viscosidade.

Quadro 8: Escores fatoriais médios, relativos aos fatores efeito da madeira (FEMAD), teor alcoólico (FTALCO), gosto doce (FGDOCE) e gosto ácido (FGACID) para as cachaças envelhecidas.

| Cachaças | Fatores           |                |            |             |
|----------|-------------------|----------------|------------|-------------|
|          | Efeito da madeira | Teor alcoólico | Gosto doce | Gosto ácido |
| 1        | 1,68              | 0,10           | -0,08      | -0,70       |
| 2        | 0,62              | -0,93          | -2,81      | 0,97        |
| 3        | -1,27             | 0,78           | -0,48      | -0,73       |
| 4        | 0,28              | -0,24          | -0,07      | 0,96        |
| 5        | 1,19              | 0,37           | -0,26      | -0,97       |
| 6        | 0,03              | 0,42           | -0,11      | -1,91       |
| 7        | -0,94             | 0,21           | -0,59      | -1,05       |
| 8        | 0,25              | -1,65          | 0,61       | 0,80        |
| 9        | -1,64             | 0,13           | -0,29      | 0,16        |
| 10       | 0,45              | -0,74          | 1,17       | -0,37       |
| 11       | -0,21             | 2,30           | 0,14       | 1,78        |
| 12       | 0,50              | 0,35           | 1,11       | 0,46        |
| 13       | -1,50             | -1,49          | 0,76       | 0,17        |
| 14       | 0,56              | 0,39           | 0,90       | 0,44        |

O posicionamento relativo das cachaças possibilita a comparação entre elas, ou seja, quanto mais próximas mais semelhantes são as cachaças e quanto mais distantes, mais distintas são (Figuras 1, 2 e 3). Observa-se, portanto, que as 14 cachaças estão distribuídas em quatro grupos distintos, em que há homogeneidade intragrupo e heterogeneidade intergrupos, em relação

aos atributos em estudo e, também, em relação aos fatores efeito da madeira, teor alcoólico, gosto doce e gosto ácido.

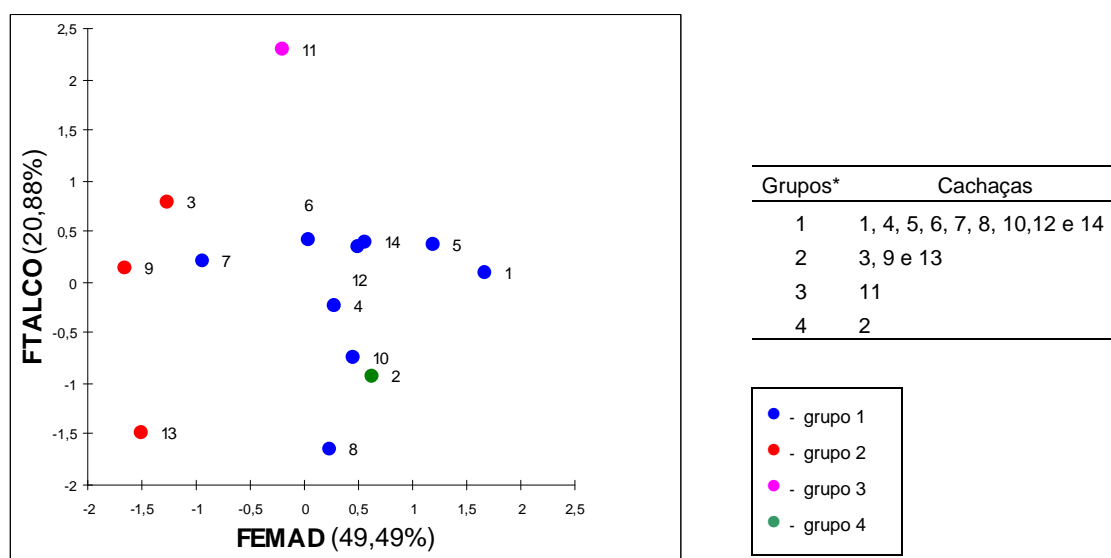
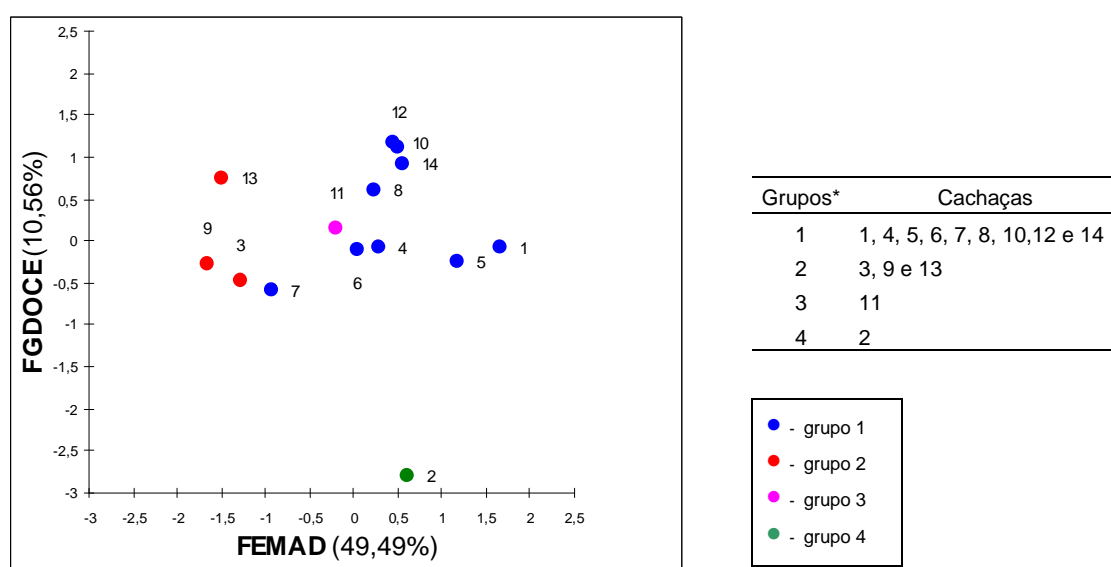


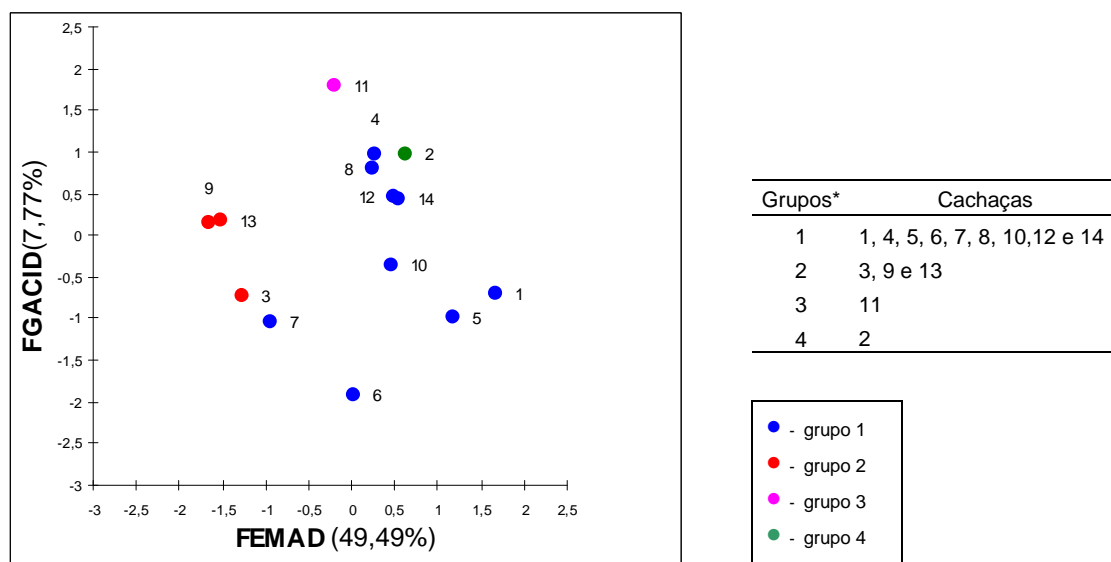
Figura 1: Dispersão das 14 cachaças em relação aos fatores comuns efeito da madeira (FEMAD) e teor alcoólico (FTALCO).



\*Observação: Os grupos foram estabelecidos pelo método de agrupamento de Tocher.

Figura 2: Dispersão das 14 cachaças em relação aos fatores comuns efeito da madeira (FEMAD) e gosto doce (FGDOCE).

Para evitar a subjetividade do agrupamento visual das cachaças, o estabelecimento dos grupos foi realizado pelo Método de Agrupamento de Tocher, a partir dos atributos médios padronizados (padronização para média igual a zero e variância igual a um).



\*Observação: Os grupos foram estabelecidos pelo método de agrupamento de Tocher.

Figura 3: Dispersão das 14 cachaaas em relação aos fatores comuns efeito da madeira (FEMAD) e gosto ácido (FGACID).

As cachaaas que compõem o grupo um foram caracterizadas por maior intensidade no fator efeito da madeira, quando comparadas às 3, 9 e 13 (grupo 2) (eixo x das Figuras 1, 2 e 3). Nota-se, também, que o efeito da madeira foi o principal fator responsável pela variabilidade entre os grupos um e dois.

Os coeficientes de correlação entre o fator efeito da madeira e os atributos aromas de madeira e de baunilha, sabor de madeira inicial e de madeira residual, coloração amarela e viscosidade são todos positivos (Quadro 6), indicando que a relação entre tal fator e os referidos atributos é direta. Assim, cachaaas ou grupos posicionados mais à direita no eixo x das Figuras 1, 2 e 3 possuem maiores escores para o fator efeito da madeira e são caracterizadas por maior intensidade nos aromas de madeira e de baunilha, nos sabores de madeira inicial e residual, na coloração amarela e na viscosidade.

A cachaa 11 (grupo 3) apresentou maiores escores em relação aos fatores teor alcoólico e gosto ácido. Estes foram os principais fatores responsáveis pela discriminação entre a cachaa 11 e os demais grupos (eixo y das Figuras 1 e 3). O fator teor alcoólico correlacionou-se positivamente com o aroma alcoólico, sabor alcoólico inicial e alcoólico residual; e o fator gosto ácido com o atributo gosto ácido (Quadro 6). Portanto, a cachaa 11 ficou caracterizada por maior intensidade nos atributos aroma alcoólico, sabor alcoólico inicial, sabor alcoólico residual e gosto ácido.

O fator gosto doce foi responsável pela discriminação entre a cachaça 2 (grupo 4) e os demais grupos, caracterizando-a como a cachaça de menor escore para este fator (eixo y da Figura 2) e também menor intensidade no atributo gosto doce com a qual se correlacionou positivamente (Quadro 6).



#### 4. CONCLUSÕES

A análise de fatores permitiu a caracterização e comparação entre as cachaças em relação aos fatores comuns e, indiretamente, em relação a 11 dos 13 atributos originais avaliados, sendo os resultados apresentados em gráficos de dispersão bidimensionais de fácil visualização.

O agrupamento das cachaças pelo Método de Tocher evitou a subjetividade do estabelecimento dos grupos apenas por exame visual do posicionamento das cachaças no gráfico de dispersão. Assim, tal técnica é um bom complemento à análise de fatores e reforça a tese de muitos autores de que muitas das técnicas de análise estatística multivariada são complementares e sempre que possível é interessante usá-las juntas.

A análise de fatores foi satisfatória para avaliar os dados da Análise Descritiva Quantitativa das 14 cachaças envelhecidas em reservatórios de diferentes tipos de madeira, por 18 a 24 meses, tendo-se em vista que:

- foi possível explicar o comportamento de 11 dos 13 atributos em apenas 4 fatores comuns (efeito da madeira, teor alcoólico, gosto doce e gosto ácido);
- não foi possível atribuir a influência de apenas dois dos atributos (gosto amargo e adstringência) sobre nenhum dos quatro fatores comuns;
- as comunalidades (proporção da variância de cada atributo explicada pelos fatores comuns) foram altas para todos os atributos;
- e os quatro fatores comuns retiveram juntos 88,69% da variância dos dados originais, o que indica que a redução no conjunto de dados se deu com pouca perda de informação.

É possível aplicar a análise de fatores para outros conjuntos de dados obtidos, também, pela Análise Descritiva Quantitativa, desde que se verifique sua viabilidade, ou seja, que se encontre um pequeno número de fatores comuns e que estes retenham grande proporção da variância existente nos dados originais; que as comunalidades sejam altas para os atributos e que todos os atributos, ou a maioria deles, se correlacionem com um dos fatores comuns.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

CRUZ, C.D. **Programa Genes: versão Windows; aplicativo computacional em genética e estatística**. 1. ed. Viçosa: Editora UFV, 2001. 648 p.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. v. 2, 1. ed. Viçosa: Editora UFV, 2003. 585 p.

HAIR, J.F, et al. **Análise multivariada de dados**. v. 1, Tradução de Adonai S. Sant'Anna, Anselmo C. Neto. Porto Alegre: Editora Bookman, 2005. 583 p.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4. ed. Upper Saddle River, New Jersey-USA: Prentice-Hall, 1998. 816 p.

MARQUES, M.J.B.S.G.S.M. **Número mínimo de famílias de meios-irmãos de milho de pipoca, critérios de seleção e predição de ganhos por seleção**. 2000. 236 f. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa. Viçosa.

MARTINS, I.S. **Comparação entre métodos uni e multivariados aplicados na seleção em *Eucalyptus grandis***. 1999. 94f. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa. Viçosa.

POLIGNANO, L. A. C.; DRUMOND, F. B.; CHENG, L. C. Utilização dos mapas de percepção e preferência como técnicas auxiliares do QFD durante o desenvolvimento de produtos alimentícios. In: CONGRESSO BRASILEIRO DE GESTÃO DO DESENVOLVIMENTO DE PRODUTO. 1., 1999, Belo Horizonte. **Anais...** 1999. p. 274-284.

RAO, R.C. **Advanced statistical methods in biometric research**. New York: John Wiley and Sons, 1952. 390 p.

REGAZZI, A. J. INF 766 - Análise Multivariada (notas de aula). DPI - UFV. Viçosa – MG, 2004.

SAS/STAT users' guide. Version 8. Cary, NC: **SAS Institute Inc.**, 1999.

YOKOTA, S.R.C. **Avaliação sensorial descritiva de cachaça envelhecida entre 18 e 24 meses por 4 painéis de julgadores**. 2005. 102f. Tese (Doutorado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa. Viçosa.

### **ARTIGO 3**

#### **ANÁLISE DE COMPONENTES PRINCIPAIS APLICADA À AVALIAÇÃO SENSORIAL DE ALIMENTOS**

#### **PRINCIPAL COMPONENT ANALYSIS APPLIED TO SENSORY EVALUATION OF FOOD**

##### **RESUMO**

Os dados experimentais referentes a 14 cachaças envelhecidas, avaliadas quanto a 13 atributos sensoriais pelo método Análise Descritiva Quantitativa – ADQ (YOKOTA, 2005), foram submetidos à análise de componentes principais (ACP) com o objetivo de detalhar a aplicação desta técnica na análise de dados provenientes da avaliação sensorial. As análises estatísticas foram realizadas com o pacote computacional SAS (Statistical Analysis System) e deu-se ênfase aos procedimentos estatísticos utilizados, à interpretação e à discussão dos resultados. A ACP foi satisfatória, pois permitiu a redução no conjunto de dados de 13 atributos para quatro componentes principais, que retiveram 88,69% da variância total disponível. Os atributos sabor de madeira inicial, sabor alcoólico residual, adstringência, gosto doce e gosto ácido foram identificados como os de maior importância relativa e, portanto, suficientes para a caracterização sensorial e comparação entre as cachaças. Os atributos sabor de madeira inicial e sabor alcoólico residual foram os que apresentaram maior variância, seguidos pelo atributo adstringência. Os gostos doce e ácido foram os de menor variabilidade e não apresentaram redundância, ou seja, não se correlacionaram ( $P > 0,05$ ) com nenhum dos demais atributos avaliados. A maioria dos atributos de menor importância relativa, ou seja, aqueles passíveis

de descarte em estudos futuros de mesma natureza, se correlacionaram, principalmente, com os atributos sabor de madeira inicial e sabor alcoólico residual, tendo seu comportamento explicado por estes. A análise por variáveis canônicas, apresentada no primeiro artigo, apresentou algumas vantagens em relação à ACP: as duas primeiras variáveis canônicas retiveram, juntas, 91,58% da variabilidade total disponível, contra 70,37% retida pelos dois primeiros componentes principais. Assim, houve menor perda de informação na redução do conjunto de dados; e também os grupos estabelecidos foram mais consistentes do que em componentes principais, no sentido de que as cachaças alocadas em um mesmo grupo distaram pouco entre si, o que não deixou dúvida quanto à sua alocação no grupo. Portanto, recomenda-se que, para avaliar dados obtidos por ADQ, em que as informações são obtidas com repetições, seja utilizada preferencialmente a análise por variáveis canônicas. Já a análise por componentes principais é mais apropriada para conjuntos de dados sem repetições, ou seja, em que há apenas uma observação por variável em cada indivíduo avaliado. Foi efetuada, também, uma comparação entre a ACP e a análise de fatores realizada no segundo artigo. Ambas foram eficientes, embora não se pode afirmar que uma ou outra tenha sido a melhor, pois elas não são concorrentes. Enquanto na ACP o interesse foi identificar os atributos que mais contribuíram para a variabilidade observada entre as cachaças, na análise de fatores buscou-se separar os atributos em grupos (fatores), de forma que cada grupo reuniu os atributos correlacionados entre si. Nesta análise, o interesse foi nas inter-relações entre os atributos avaliados.

**PALAVRAS-CHAVE:** ACP, análise multivariada, análise sensorial, análise descritiva quantitativa, ADQ.

## **SUMMARY**

The experimental data relative to 14 commercial brands of aged cachaça, evaluated with respect to 13 sensory attributes by Quantitative Descriptive Analysis - QDA method (YOKOTA, 2005), were submitted to Principal Component Analysis (PCA) to make a detailed description of the application of this technique to analyze data obtained from sensory evaluation of food. Statistical analyses were performed with SAS (Statistical Analysis System) computational package and the main goal of this paper was study the statistical procedures used, their interpretation and discussion of results. PCA allowed reduce data dimension from the 13 attributes to four principal components that retained 88.69% of the total available variance. The attributes initial wood flavor,

residual alcoholic flavor, astringency, sweet taste and acid taste showed the greatest relative importance and can be used for sensorial characterization and comparison among the cachaça brands. The attributes initial wood flavor, residual alcoholic flavor, followed by astringency, presented the largest variance. Sweet and acid tastes presented the smallest variance and did not show significant correlation ( $P > 0.05$ ) with any other attribute evaluated in this study. Most of the attributes of less relative importance, i.e., the ones that can be disregarded in future studies of the same nature, were correlated mainly with the attributes initial wood flavor and residual alcoholic flavor, having their behavior explained by them. The analysis using canonical variables, presented in the first article, showed some advantages over PCA: the first two canonical variables retained 91.58% of the total available variance against 70.37% for the first two principal components, which means that data reduction occurred with greater loss of information with PCA; in addition, the cachaça brand groups formed using canonical variables showed to be more consistent, i.e., with smaller distance values for brands belonging to the same group, raising no doubt about their position in the group. Therefore, it is recommended that, in order to evaluate data obtained by QDA, in which information is obtained with repetitions, analysis with canonical variables should be used. PCA is appropriate to data without repetitions, what means that there is only one observation per variable in each evaluated subject. It was also present a comparative study between PCA and factor analysis, the technique illustrated in the second article. Both techniques showed to be equally satisfactory in terms of results, although they are not competing techniques. While PCA is aimed at identifying those attributes that contribute the most to the variability among brands, factor analysis is aimed at forming groups or factors with the attributes that present higher correlation with each other. The main goal in PCA is to understand the inter relationships among attributes.

**KEYWORDS:** PCA, multivariate analysis, sensory analysis, quantitative descriptive analysis, QDA.

## 1. INTRODUÇÃO

A análise de componentes principais (ACP) é associada à idéia de redução da massa de dados. Procura-se redistribuir a variação observada nas variáveis (eixos originais) de forma a obter um conjunto de eixos não correlacionados. Esta análise tem como principal objetivo reduzir a dimensionalidade do conjunto original de variáveis, com a menor perda de informação possível, além de permitir o agrupamento de indivíduos (tratamentos, genótipos, etc.) similares, mediante exames visuais em dispersões gráficas no espaço bi ou tridimensional de fácil interpretação geométrica (REGAZZI, 1997; JOHNSON e WICHERN, 1992). A ACP permite, também, verificar a importância relativa ou influência de cada variável em estudo sobre os componentes. Esta importância, ou influência, é dada pelas correlações entre as variáveis e os componentes principais.

Esta técnica multivariada tem sido bastante utilizada na análise de dados provenientes da avaliação sensorial de alimentos (PONTES, 2004, CARNEIRO et al., 2003, FREITAS, 2002, CARNEIRO, 2001 e POLIGNANO et al., 1999), embora ainda haja carência por parte de muitos profissionais da área sobre conhecimentos relacionados aos procedimentos estatísticos computacionais e princípios matemáticos utilizados. Portanto, este estudo teve como objetivo ilustrar a aplicação da análise de componentes principais a dados provenientes da avaliação sensorial, com ênfase nos procedimentos estatísticos utilizados no sistema SAS (Statistical Analysis System), na interpretação e discussão dos resultados.

## **2. MATERIAL E MÉTODOS**

### **2.1. DADOS EXPERIMENTAIS**

Foram utilizados os mesmos dados experimentais da Análise Descritiva Quantitativa de cachaças envelhecidas (YOKOTA, 2005), como descritos no primeiro artigo.

### **2.2. METODOLOGIA**

A análise de componentes principais foi realizada com base nos trabalhos de CRUZ e CARNEIRO (2003), CRUZ (2001), REGAZZI (1997), CRUZ e REGAZZI (1997) e JOHNSON e WICHERN (1992).

As análises estatísticas foram executadas no sistema SAS (1999), licenciado para UFV em 2005. São apresentados os comandos da janela do editor de textos necessários para a execução dos procedimentos estatísticos com respectivos comentários a respeito de suas funções. Os comentários aparecem entre barras e asteriscos, /\*comentário\*/, uma alternativa para inserir comentários na janela do editor do programa SAS.

#### **2.2.1. ANÁLISES PRELIMINARES**

##### **a. ANÁLISE DE VARIÂNCIA E MÉDIAS DOS ATRIBUTOS (PROGRAMA 1)**

Os dados experimentais foram submetidos a análise de variância univariada (ANOVA) para ter-se uma idéia da variabilidade entre as 14 cachaças em relação aos atributos avaliados.

A ACP pode ser realizada para um conjunto de dados constituídos por  $n$  indivíduos avaliados em  $p$  variáveis. Ou para médias oriundas de  $n$  tratamentos avaliados em  $p$  variáveis, como é o caso deste estudo, em que se tem  $n = 14$  cachaças avaliadas em  $p = 13$  atributos. Assim, o passo inicial foi obter as médias para os atributos. A partir destas médias calculou-se a média geral para cada atributo, e o respectivo desvio padrão (Quadro 1).

Os dados do Quadro 1 foram usados, também, para obter as médias padronizadas dos atributos, já que é mais comum fazer a ACP com dados padronizados. A referida padronização está apresentada no item b (programa 2).

Quadro 1: Médias por cachaças, média e respectivo desvio padrão por atributo, em relação aos atributos em estudo.

| CACHAÇAS                            | Médias originais ( $\bar{X}_{ij}$ ) |                  |                   |            |     |             |
|-------------------------------------|-------------------------------------|------------------|-------------------|------------|-----|-------------|
|                                     | Aroma alcoólico                     | Aroma de madeira | Aroma de baunilha | Gosto doce | ... | Viscosidade |
| 1                                   | 8,1                                 | 11,5             | 5,8               | 6,1        | ... | 10,6        |
| 2                                   | 7,7                                 | 10,8             | 3,9               | 4,9        | ... | 10,9        |
| 3                                   | 9,0                                 | 5,8              | 2,5               | 6,2        | ... | 10,1        |
| 4                                   | 8,5                                 | 8,6              | 3,4               | 6,0        | ... | 10,3        |
| 5                                   | 8,5                                 | 10,5             | 3,8               | 6,2        | ... | 11,2        |
| 6                                   | 8,6                                 | 8,6              | 3,6               | 5,9        | ... | 9,7         |
| 7                                   | 8,9                                 | 6,7              | 3,0               | 5,7        | ... | 9,7         |
| 8                                   | 8,1                                 | 9,3              | 4,1               | 6,2        | ... | 9,9         |
| 9                                   | 8,8                                 | 5,5              | 2,6               | 6,0        | ... | 9,6         |
| 10                                  | 8,4                                 | 9,1              | 4,0               | 6,9        | ... | 10,2        |
| 11                                  | 9,7                                 | 7,5              | 3,0               | 6,2        | ... | 10,0        |
| 12                                  | 8,9                                 | 8,6              | 4,9               | 6,3        | ... | 10,1        |
| 13                                  | 8,6                                 | 5,4              | 2,9               | 6,8        | ... | 9,5         |
| 14                                  | 8,5                                 | 8,5              | 4,8               | 6,5        | ... | 10,4        |
| Médias por atributo ( $\bar{X}_j$ ) | 8,6                                 | 8,3              | 3,7               | 6,1        | ... | 10,2        |
| Desvio padrão $s(X_j)$              | 0,5                                 | 1,9              | 0,9               | 0,5        | ... | 0,5         |

Estão a seguir, os procedimentos para ANOVA e para estimar as médias dos atributos.

**/\*Procedimento que importa os dados da planilha do EXCEL\*/**

```
PROC IMPORT out=dados_originais
            datafile="D:\ACP\programas\dados_ori.xls" /*1*/
            dbms=excel2000 replace;
            getnames=yes;
RUN;
options nodate nonumber;
/*1*/ /*Informa que o arquivo de dados a ser lido, dados_ori.xls, está na
      pasta "D:\ACP\programas". Este arquivo é o mesmo usado no
      primeiro artigo.*/
```

**/\*Procedimento que imprime os dados na tela OUTPUT\*/**

```
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS ORIGINAIS';
RUN;
```

**/\*Procedimento que realiza análise de variância\*/**

```
PROC ANOVA data=dados_originais;
class CACHA PROV REPE;
model AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC
= CACHA PROV CACHA*PROV; /*2*/
test H=CACHA E=CACHA*PROV; /*3*/
title 'ANÁLISE DE VARIÂNCIA DOS ATRIBUTOS';
RUN;
```



```

/*2*/ /*O modelo adotado na ANOVA é para o delineamento em blocos
casualizados (d.b.c) e cada provador (PROV) foi considerado como
um bloco. O termo cachaça foi abreviado por cacha.*/

/*3*/ /*No modelo adotado, o F para cachaça foi obtido considerando-se o
quadrado médio da interação CACHA*PROV como variância residual.*/

/*Procedimento que calcula as médias dos atributos por cachaça */
PROC MEANS data=dados_originais;
output out=medias_por_cacha;
by CACHA;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇA';
RUN;

/*Procedimento que organiza as médias *//*4*/
DATA medias_por_cacha; set medias_por_cacha(where=(_STAT_='MEAN'))
keep= CACHA _STAT_ AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR
ADST COLA VISC);
RUN;

/*4*/ /*As médias dos atributos por cachaça estão apresentadas no Quadro
1.*/

/*Procedimento que imprime as médias na janela OUTPUT*/
PROC PRINT data=medias_por_cacha;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇA';
RUN;

/*Procedimento que estima a média geral para cada atributo e
respectivo desvio padrão*/
PROC MEANS data=medias_por_cacha;
output out=medias_por_atributo;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIA POR ATRIBUTO E RESPECTIVO DESVIO PADRÃO';
RUN;

/*Procedimento que imprime, na tela, as médias por atributo*/
PROC PRINT data=medias_por_atributo;
title 'MÉDIA POR ATRIBUTO';
RUN;

```

## b. PADRONIZAÇÃO DAS MÉDIAS (PROGRAMA 2)

As médias dos atributos, por cachaças, (Quadro 1) foram padronizadas para se obter média igual a zero e variância igual a um:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{s(X_j)}, \text{ sendo:}$$

$i = 1, 2, \dots, n = 14$  cachaças;

$j = 1, 2, \dots, p = 13$  atributos;

$Z_{ij}$ : médias padronizadas dos atributos por cachaças;

$X_{ij}$ : médias originais dos atributos por cachaças;

$\bar{X}_j$ : média geral do j-ésimo atributo;

$s(X_j)$ : desvio padrão associado ao j-ésimo atributo.

Os procedimentos para a padronização estão apresentados a seguir e os resultados dispostos no Quadro 2.

Quadro 2: Médias padronizadas para os atributos em estudo.

| CACHAÇAS | Médias padronizadas ( $Z_{ij}$ ) |                  |                   |            |     |             |
|----------|----------------------------------|------------------|-------------------|------------|-----|-------------|
|          | Aroma alcoólico                  | Aroma de madeira | Aroma de baunilha | Gosto doce | ... | Viscosidade |
| 1        | -1,1                             | 1,7              | 2,2               | -0,1       | ... | 0,9         |
| 2        | -1,9                             | 1,3              | 0,2               | -2,6       | ... | 1,5         |
| 3        | 0,9                              | -1,3             | -1,3              | 0,1        | ... | -0,1        |
| 4        | -0,2                             | 0,2              | -0,3              | -0,3       | ... | 0,4         |
| 5        | -0,3                             | 1,1              | 0,0               | 0,2        | ... | 2,1         |
| 6        | 0,0                              | 0,2              | -0,2              | -0,4       | ... | -0,8        |
| 7        | 0,6                              | -0,8             | -0,8              | -0,8       | ... | -1,0        |
| 8        | -1,0                             | 0,5              | 0,4               | 0,1        | ... | -0,6        |
| 9        | 0,4                              | -1,5             | -1,2              | -0,3       | ... | -1,1        |
| 10       | -0,4                             | 0,4              | 0,3               | 1,6        | ... | 0,1         |
| 11       | 2,3                              | -0,4             | -0,7              | 0,2        | ... | -0,4        |
| 12       | 0,6                              | 0,1              | 1,2               | 0,2        | ... | -0,1        |
| 13       | 0,1                              | -1,5             | -0,9              | 1,4        | ... | -1,3        |
| 14       | -0,1                             | 0,1              | 1,2               | 0,8        | ... | 0,5         |

```

/*Procedimento que importa os dados da planilha do EXCEL*/
PROC IMPORT out=dados_originais
    datafile="D:\ACP\programas\dado_ori.xls"
    dbms=excel2000 replace;
    getnames=yes;
RUN;
options nodate nonumber;

/*Procedimento que imprime, na tela, os dados lidos*/
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS ORIGINAIS';
RUN;

/*Procedimento que calcula as médias dos atributos por cachaça (médias originais)*/
PROC MEANS data=dados_originais;
output out=medias_por_cacha;
by CACHA;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇAS';
RUN;

/*Procedimento que organiza as médias */
DATA medias_por_cacha; set medias_por_cacha(where=(_STAT_='MEAN'))
keep= CACHA _STAT_ AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR
ADST COLA VISC);
RUN;

/*Procedimento que imprime, na tela, as médias*/
PROC PRINT data=medias_por_cacha;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇAS';
RUN;

```

```

/*Procedimento para padronizar as médias dos atributos*/**5*/
PROC STANDARD data=medias_por_cacha mean=0 std=1
out=medias_padronizadas;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
RUN;

/*5*/**Padronização para média igual a zero e variância igual a um.*/

/*6*/**As médias que constam no arquivo medias_por_cacha são as mesmas
apresentadas no Quadro 1, porém, as vírgulas foram substituídas por
pontos.*/

/*Procedimento que imprime, na tela, as médias padronizadas dos
atributos*/**7*/
PROC PRINT data=medias_padronizadas;
var CACHA AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA
VISC;
title 'MÉDIAS PADRONIZADAS';
RUN;

/*7*/**Tais médias estão apresentadas no Quadro 2.*/

```

### 2.2.2. ANÁLISE DE COMPONENTES PRINCIPAIS (PROGRAMA 3)

A análise de componentes principais consistiu em transformar o conjunto das médias padronizadas  $Z_1, Z_2, \dots, Z_p$  (Quadro 2) em um novo conjunto de variáveis  $CP_1, CP_2, \dots, CP_p$ , que sejam não correlacionadas e suas variâncias ordenadas para que seja possível comparar as cachacas com base apenas nas novas variáveis que apresentem maior variância. As novas variáveis  $CP_1, CP_2, \dots, CP_p$  são os componentes principais e foram obtidas como combinações lineares de  $Z_1, Z_2, \dots, Z_p$ , tal que:

a)  $CP_i = a_{i,1}Z_1 + a_{i,2}Z_2 + \dots + a_{i,13}Z_{13}$ , com  $i = 1, 2, \dots, 13$ , em que :

$CP_i$ : i-ésimo componente principal;

$a_{i,1}, a_{i,2}, \dots, a_{i,13}$ : elementos do i-ésimo autovetor ( $a_i$ ) normalizado, associado ao i-ésimo autovalor ( $\lambda_i$ ) do i-ésimo componente principal ( $CP_i$ ).

b)  $Var(CP_i) = \lambda_i$ , portanto,  $Var(CP_1) \geq Var(CP_2) \geq \dots \geq Var(CP_{13})$ ;

c)  $\sum_{i=1}^{13} Var(CP_i) = \sum_{j=1}^{13} Var(Z_j) = \sum_{i=1}^{13} \lambda_i$ ;

d)  $Cov(CP_i, CP_{i'}) = 0$  para  $i \neq i'$ .

Há duas possibilidades ao trabalhar com os dados padronizados: - estimar os autovalores da matriz de covariância (S) entre as médias padronizadas dos atributos (Quadro 2) ou estimá-los da matriz de correlação

(R) entre as médias originais dos atributos (Quadro 1), pois neste caso, as matrizes S e R são idênticas. Optou-se por estimar os autovalores a partir de R.

Os autovalores ( $\lambda_i$ ), ou raízes características da matriz de correlação (R), foram, então, estimados resolvendo-se a expressão  $\det(R - \lambda_i I) = 0$ , sendo I a matriz identidade. Cada autovetor ( $a_i^*$ ), associado a um autovalor ( $\lambda_i$ ), foi obtido a partir de  $(R - \lambda_i I)a_i^* = \phi$ , em que  $\phi$  é um vetor nulo. Os autovetores ( $a_i^*$ ) foram normalizados, obtendo-se ( $a_i$ ) tal que  $a_i' a_i = 1$ .

A análise de componentes principais é associada à idéia de redução no conjunto de informações, ou seja, redução de 13 atributos para k componentes principais, com  $k < 13$ . Assim, fez-se necessário definir o número de componentes principais (valor K) a serem considerados para a caracterização e comparação entre as cachaças. Geralmente são considerados os primeiros k componentes principais que retêm, juntos, a maior proporção de variância dos atributos originais (acima de 70 ou 80% da variância). Segundo CRUZ e REGAZZI (1997), em estudos da diversidade genética têm-se considerado os dois primeiros componentes quando eles envolvem pelo menos 80% da variação total. Nos casos em que este limite não é atingido nos dois primeiros, a análise é completada com a dispersão gráfica em relação ao terceiro e quarto componentes. Optou-se, neste estudo, por considerar os primeiros k componentes com variância maior ou igual à variância média dos atributos originais. Como, neste estudo, os componentes principais foram extraídos da matriz de correlação, equivale a considerar os primeiros k componentes associados a autovalores maiores ou iguais a um, que é a variância média dos atributos padronizados, dada por:

$$\text{Variância média dos } Z_j = \frac{\sum_{j=1}^{13} \text{Var}(Z_j)}{13} = \frac{\sum_{i=1}^{13} \text{Var}(CP_i)}{13} = \frac{\sum_{i=1}^{13} \lambda_i}{13} = \frac{\text{traço}(R)}{13} = \frac{13}{13} = 1.$$

A importância relativa de um componente principal é avaliada pela porcentagem da variância total que ele explica e a soma dos primeiros k autovalores dividida pela soma de todos os p autovalores ( $(\lambda_1 + \lambda_2 + \dots + \lambda_k) / (\lambda_1 + \lambda_2 + \dots + \lambda_p)$ ) representa a proporção da variância total explicada pelos primeiros k componentes principais, ou seja, a proporção da informação retida na redução da dimensão de p para k (REGAZZI, 1997).

Os primeiros k componentes principais foram, então, escritos como:

$$CP_1 = a_{1,1}Z_1 + a_{1,2}Z_2 + \dots + a_{1,13}Z_{13};$$

$$CP_2 = a_{2,1}Z_1 + a_{2,2}Z_2 + \dots + a_{2,13}Z_{13};$$

$$\dots$$

$$CP_k = a_{k,1}Z_1 + a_{k,2}Z_2 + \dots + a_{k,13}Z_{13}.$$

Tomando-se os elementos  $a_{ij}$  dos autovalores já determinados e substituindo em  $CP_1, CP_2, \dots, CP_k$ , os  $Z_1, Z_2, \dots, Z_{13}$  pelos valores numéricos relativos à cachaça número um (Quadro 2), obteve-se os escores dos  $k$  componentes principais para a primeira cachaça. O mesmo foi feito para as outras cachaças, até obter-se os escores para os componentes principais, com 14 linhas e  $k$  colunas. A caracterização sensorial e comparação entre as cachaças foram feitas a partir de tais escores representados em gráfico de dispersão.

Para que a análise dos dados por meio de componentes principais seja completa é necessária a interpretação de cada componente, que é feita verificando-se a importância ou influência que cada variável tem sobre o componente. Esta importância é dada pela correlação (coeficiente de correlação de Pearson) entre cada variável e o componente que está sendo interpretado. Baseado no princípio de que a importância relativa dos componentes principais decresce do primeiro para o último, tem-se que os últimos componentes são responsáveis pela explicação de uma pequena porção da variância disponível. Assim, a variável que apresentar maior correlação, em valor absoluto, com o componente de menor autovalor, será considerada de menor importância para explicar a variabilidade do material estudado, sendo, portanto, passível de descarte. A seguir, o próximo componente de menor autovalor é considerado, podendo-se descartar a variável com a maior correlação (em valor absoluto) no componente e a qual não tenha sido previamente descartada (CRUZ e REGAZZI, 1997). Optou-se, neste estudo, por descartar os atributos de maior correlação em cada um dos últimos  $p$  menos  $k$  ( $p - k$ ) componentes principais. Tal análise é feita dos últimos para os primeiros componentes.

Vale ressaltar que os atributos de menor importância relativa apresentam baixa variabilidade e/ou estão correlacionados com outros que fazem parte do estudo.

A seguir estão os procedimentos usados para obter as estimativas mencionadas.

```
/*Procedimento que importa os dados da planilha do EXCEL*/
PROC IMPORT out=dados_originais
    datafile="D:\ACP\programas\dado_ori.xls"
    dbms=excel2000 replace;
    getnames=yes;
RUN;
options nodate nonumber;
```

```

/*Procedimento que imprime, na tela, os dados lidos*/
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS ORIGINAIS';
RUN;

/*Procedimento que calcula as médias dos atributos, por cachaça, para
serem utilizadas na análise de componentes principais*/
PROC MEANS data=dados_originais;
output out=medias_por_cacha;
by CACHA;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇA';
RUN;

/*Procedimento que organiza as médias para serem usadas na análise de
componentes principais*/
DATA medias_por_cacha; set medias_por_cacha(where=(_STAT_='MEAN'))
keep= CACHA _STAT_ AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR
ADST COLA VISC);
RUN;

/*Procedimento que imprime, na tela, as médias*/
PROC PRINT data=medias_por_cacha;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇAS';
RUN;

/*ANÁLISE DE COMPONENTES PRINCIPAIS*/

/*Procedimento que estima a matriz de correlação (R), autovalores,
autovetores e os escores dos componentes*/
PROC PRINCOMP data=medias_por_cacha OUT=SAIDA; /*8*/
VAR AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
TITLE 'ANÁLISE DE COMPONENTES PRINCIPAIS A PARTIR DA MATRIZ DE
CORRELAÇÃO';
RUN;

/*8*/ /*A opção COV não foi especificada, portanto, os componentes
principais foram obtidos a partir da matriz de correlação (R) entre
as médias originais dos atributos (Quadro 1).*/

/*Procedimento para análise de correlação entre os atributos e os
componentes principais*/
PROC CORR DATA=SAIDA;
VAR AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
WITH PRIN1 PRIN2 PRIN3 - PRIN13;
TITLE 'CORRELAÇÕES DOS ATRIBUTOS COM OS COMPONENTES';

/*Procedimento que imprime os escores dos quatro primeiros componentes
principais*/
PROC PRINT DATA=SAIDA;
VAR CACHA PRIN1 PRIN2 PRIN3 PRIN4;
TITLE 'ESCORES DOS COMPONENTES PRINCIPAIS';
RUN;

/*Procedimento que gera o gráfico de dispersão das cachaças em relação
a dois dos componentes principais*/ /*9*/
PROC PLOT DATA=SAIDA;
PLOT PRIN2*PRIN1=CACHA / VPOS=20;
TITLE ' GRÁFICO DOS DOIS PRIMEIROS COMPONENTES PRINCIPAIS';
RUN;

```

```

/*9*/ /*Foi especificado para gerar o gráfico em relação aos dois
primeiros componentes principais. Tal procedimento pode ser usado
para gerar o gráfico em relação a qualquer par de componentes de
interesse. Os gráficos de dispersão foram feitos, também, no
Excel, aplicativo em que o autor está familiarizado a editar
gráficos.*/

```

## **MÉTODO DE AGRUPAMENTO DE TOCHER**

A avaliação visual do posicionamento das cachaaas, no gráfico de dispersão, permite agrupá-las, de forma que as cachaaas que estão mais próximas entre si fazem parte de um mesmo grupo. O estabelecimento destes grupos pode variar de um analista para outro, sendo portanto uma avaliação subjetiva. Para evitar a subjetividade, o agrupamento das cachaaas foi realizado pelo Método de Agrupamento de Otimização proposto por Tocher, citado por RAO (1952). Este método faz parte dos procedimentos do Programa Genes (Cruz, 2001), o qual foi utilizado no estabelecimento dos grupos.

O agrupamento, neste método, foi realizado com base na distância Euclidiana Média, que é uma medida de dissimilaridade entre as cachaaas. A distância Euclidiana Média, por sua vez, foi obtida a partir dos atributos médios padronizados (CRUZ, 2001). Vale ressaltar que foi utilizada a padronização para média igual a zero e variância igual a um.

### **2.2.3. ESTIMATIVAS DAS CORRELAÇÕES ENTRE OS ATRIBUTOS (PROGRAMA 4)**

Muitos dos atributos de menor importância na caracterização e comparação entre as cachaaas podem estar correlacionados com outro(s), e assim, ter seu comportamento explicado por ele(s). Tais correlações (coeficientes de correlação de Pearson) e os respectivos níveis de significância foram estimados nos procedimentos a seguir.

```

/*Procedimento que importa os dados da planilha do EXCEL.*/
PROC IMPORT out=dados_originais
    datafile="D:\ACP\programas\dado_ori.xls"
    dbms=excel2000 replace;
    getnames=yes;
RUN;
options nodate nonumber;

/*Procedimento que imprime, na tela, os dados lidos.*/
PROC PRINT data=dados_originais;
title 'LEITURA DOS DADOS';RUN;

```

```

/*Procedimento que calcula as médias dos atributos.*/
PROC MEANS;
output out=medias_por_cacha;
by CACHA;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇA';
RUN;

/*Procedimento que organiza as médias.*/
DATA medias_por_cacha; set medias_por_cacha(where=(_STAT_='MEAN'))
keep= CACHA _STAT_ AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR
ADST COLA VISC);
RUN;

/*Procedimento que imprime, na tela, as médias.*/
PROC PRINT data=medias_por_cacha;
title 'MÉDIAS DOS ATRIBUTOS, POR CACHAÇA';
RUN;

/*Procedimento que obtêm as correlações (coeficiente de correlação de
Pearson) entre as médias dos atributos.*/
PROC CORR data=medias_por_cacha;
var AALC AMAD ABAU DOCE ACID AMAR SALI SALR SMAI SMAR ADST COLA VISC;
title 'CORRELAÇÕES ENTRE AS MÉDIAS DOS ATRIBUTOS';
RUN;

```



### 3. RESULTADOS E DISCUSSÃO

Observa-se, pelos resultados da análise de variância (Quadro 3), que existem diferenças significativas ( $P = 0,05$ ) entre as médias das cachaças, para 12 dos 13 atributos avaliados. Portanto, há diferenças entre as cachaças e faz-se necessário investigá-las. Uma das alternativas é usar a análise de componentes principais.

Estão apresentados, no Quadro 4, as estimativas dos autovalores da matriz de correlação (R), a proporção da variância retida por cada componente principal, a proporção acumulada e os coeficientes de ponderação usados na obtenção dos escores dos componentes principais. Foram considerados, para a caracterização e comparação entre as cachaças, os primeiros quatro componentes principais, pois cada um deles apresentou variância (autovalor) superior a uma unidade, que é a variância média dos atributos padronizados. Estes quatro primeiros componentes principais retiveram, juntos, 88,69% da variância dos dados.

Os escores para os quatro primeiros componentes principais estão apresentados no Quadro 5 e nas Figuras 1, 2 e 3. O posicionamento relativo das cachaças nestas figuras permitiu a comparação entre elas, ou seja, quanto mais próximas mais semelhantes são as cachaças e quanto mais distantes, mais distintas são. Observa-se, portanto, que as 14 cachaças foram distribuídas em quatro grupos distintos, em que há homogeneidade intragrupo e heterogeneidade intergrupos, em relação aos atributos em estudo e, também, em relação aos quatro primeiros componentes principais. Para evitar a subjetividade do agrupamento visual das cachaças, o estabelecimento dos grupos foi realizado pelo Método de Agrupamento de Tocher, a partir das médias padronizadas dos atributos (padronização para média igual a zero e variância igual a um).

O primeiro componente principal ( $CP_1$ ), eixo horizontal das Figuras 1, 2 e 3, discriminou os quatro grupos entre si e os grupos posicionados mais à direita apresentaram maiores escores em relação a tal componente.

A cachaça 11 (grupo 3) apresentou maior escore em relação ao segundo componente principal. Componente este, que foi o principal responsável pela discriminação entre a referida cachaça e os demais grupos (eixo vertical da Figura 1).

O terceiro componente principal foi responsável pela discriminação entre a cachaça 2 (grupo 4) e os demais grupos, caracterizando-a como a cachaça de menor escore para este componente (eixo vertical da Figura 2).

Quadro 3: Resumo das análises de variância univariadas dos dados correspondentes aos 13 atributos avaliados nas 14 cachaças envelhecidas.

| Atributos                 | QMC    | QMC*P              | QMR  | F <sub>C</sub> | Prob.> F <sub>C</sub> | Média |
|---------------------------|--------|--------------------|------|----------------|-----------------------|-------|
| Aroma alcoólico           | 8,51   | 4,32*              | 2,44 | 1,97           | 0,03                  | 8,59  |
| Aroma de madeira          | 151,46 | 8,57*              | 2,49 | 17,67          | <0,00                 | 8,32  |
| Aroma de baunilha         | 35,73  | 9,98*              | 2,96 | 3,58           | 0,00                  | 3,73  |
| Gosto doce                | 10,02  | 5,55 <sup>ns</sup> | 4,57 | 1,81           | 0,05                  | 6,13  |
| Gosto ácido               | 14,67  | 10,19*             | 6,36 | 1,44           | 0,16                  | 6,38  |
| Gosto amargo              | 14,93  | 5,94 <sup>ns</sup> | 5,22 | 2,51           | 0,01                  | 6,66  |
| Sabor alcoólico inicial   | 7,93   | 4,37*              | 2,05 | 1,82           | 0,05                  | 7,84  |
| Sabor alcoólico residual  | 8,85   | 4,90*              | 2,27 | 1,80           | 0,05                  | 8,79  |
| Sabor de madeira inicial  | 129,65 | 7,39*              | 2,43 | 17,55          | <0,00                 | 7,94  |
| Sabor de madeira residual | 152,56 | 9,64*              | 2,76 | 15,83          | <0,00                 | 8,25  |
| Adstringência             | 11,49  | 4,92*              | 3,61 | 2,34           | 0,01                  | 6,60  |
| Coloração amarela         | 552,46 | 4,48*              | 2,73 | 123,26         | <0,00                 | 6,24  |
| Viscosidade               | 9,74   | 4,13*              | 1,98 | 2,36           | 0,01                  | 10,15 |

QMC: Quadrados médios para cachaça;

QMC\*P: Quadrados médios para a interação entre cachaça e provador;

QMR: Quadrados médios do resíduo;

F<sub>C</sub>: valor de F para cachaça (F<sub>C</sub> = QMC/QMC\*P);

Prob.: Valores de probabilidade ou nível descritivo do teste (valor-P);

\*Interação entre cachaça e provador significativa (P≤0,05);

<sup>ns</sup>Interação entre cachaça e provador não significativa (P>0,05).

#### Esquema da ANOVA

| F.V.           | G.L. |
|----------------|------|
| Cachaça (C)    | 13   |
| Provador (P)   | 7    |
| Interação(C*P) | 91   |
| Resíduo        | 448  |

F.V.: fontes de variação

G.L.: graus de liberdade

Quadro 4: Estimativas dos autovalores e coeficientes de ponderação (autovetores normalizados) associados aos atributos padronizados aroma alcoólico (AALCp), aroma de madeira (AMADp), ..., viscosidade (VISCp). Atributos, estes, avaliados nas 14 cachaças envelhecidas.

| i        | Componentes<br>principiais (CP <sub>i</sub> ) | Autovalores<br>λ <sub>i</sub> de R | Proporção da<br>variância (%)* | Proporção<br>acumulada (%) | Coeficientes de ponderação (autovetores normalizados) |       |     |       |
|----------|-----------------------------------------------|------------------------------------|--------------------------------|----------------------------|-------------------------------------------------------|-------|-----|-------|
|          |                                               |                                    |                                |                            | AALCp                                                 | AMADp | ... | VISCp |
| <b>1</b> | <b>CP<sub>1</sub></b>                         | <b>6,4332</b>                      | <b>49,49</b>                   | <b>49,49</b>               | -0,29                                                 | 0,39  | ... | 0,32  |
| <b>2</b> | <b>CP<sub>2</sub></b>                         | <b>2,7143</b>                      | <b>20,88</b>                   | <b>70,37</b>               | 0,31                                                  | 0,03  | ... | 0,16  |
| <b>3</b> | <b>CP<sub>3</sub></b>                         | <b>1,3724</b>                      | <b>10,56</b>                   | <b>80,92</b>               | 0,30                                                  | 0,10  | ... | 0,03  |
| <b>4</b> | <b>CP<sub>4</sub></b>                         | <b>1,0102</b>                      | <b>7,77</b>                    | <b>88,69</b>               | 0,20                                                  | -0,05 | ... | -0,15 |
| 5        | CP <sub>5</sub>                               | 0,5602                             | 4,31                           | 93,00                      | -0,03                                                 | -0,05 | ... | 0,61  |
| 6        | CP <sub>6</sub>                               | 0,4530                             | 3,48                           | 96,49                      | 0,11                                                  | -0,01 | ... | -0,11 |
| 7        | CP <sub>7</sub>                               | 0,2791                             | 2,15                           | 98,63                      | -0,38                                                 | -0,05 | ... | -0,04 |
| 8        | CP <sub>8</sub>                               | 0,1270                             | 0,98                           | 99,61                      | -0,08                                                 | 0,20  | ... | -0,66 |
| 9        | CP <sub>9</sub>                               | 0,0241                             | 0,19                           | 99,80                      | 0,69                                                  | -0,06 | ... | 0,01  |
| 10       | CP <sub>10</sub>                              | 0,0133                             | 0,10                           | 99,90                      | 0,20                                                  | 0,42  | ... | -0,03 |
| 11       | CP <sub>11</sub>                              | 0,0065                             | 0,05                           | 99,95                      | 0,09                                                  | 0,71  | ... | 0,10  |
| 12       | CP <sub>12</sub>                              | 0,0060                             | 0,05                           | 100,00                     | 0,12                                                  | -0,22 | ... | 0,14  |
| 13       | CP <sub>13</sub>                              | 0,0006                             | 0,00                           | 100,00                     | 0,00                                                  | -0,25 | ... | 0,03  |

$$CP_1 = -0,2856 * AALC_p + 0,3881 * AMAD_p + \Lambda + 0,3153 * VISC_p$$

$$CP_2 = 0,3052 * AALC_p + 0,0273 * AMAD_p + \Lambda + 0,1570 * VISC_p$$

...

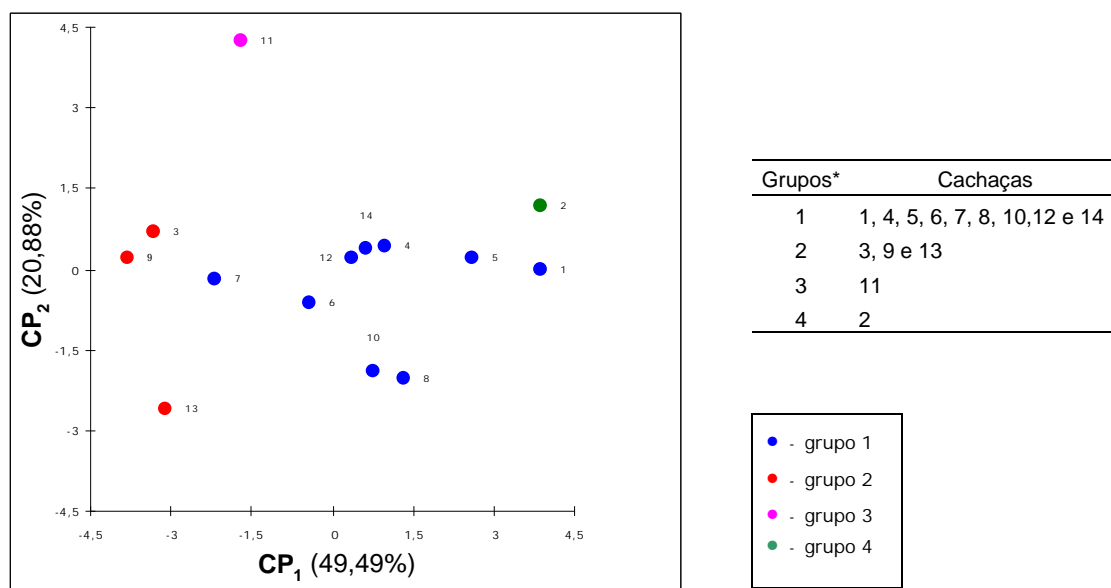
$$CP_{13} = 0,0049 * AALC_p - 0,2484 * AMAD_p + \Lambda + 0,0347 * VISC_p$$

$$* \text{ Proporção da variância (\%)} \text{ para } CP_i = \frac{\lambda_i}{\sum_{i=1}^{13} \lambda_i} \cdot 100\% = \frac{\lambda_i}{13} \cdot 100\% .$$

R: Matriz de correlação entre as médias originais dos atributos em estudo.

Quadro 5: Escores dos primeiros quatro componentes principais ( $CP_1, CP_2, CP_3$  e  $CP_4$ ) em relação às 14 cachaças envelhecidas.

| CACHAÇAS | $CP_1$ | $CP_2$ | $CP_3$ | $CP_4$ |
|----------|--------|--------|--------|--------|
| 1        | 3,87   | -0,04  | 1,06   | -0,41  |
| 2        | 3,91   | 1,15   | -2,91  | -1,01  |
| 3        | -3,31  | 0,69   | 0,10   | -1,05  |
| 4        | 1,00   | 0,40   | -0,69  | 0,72   |
| 5        | 2,62   | 0,19   | 1,08   | -0,81  |
| 6        | -0,40  | -0,62  | 1,38   | -1,52  |
| 7        | -2,16  | -0,20  | 0,01   | -1,28  |
| 8        | 1,33   | -2,06  | -0,93  | 1,16   |
| 9        | -3,80  | 0,21   | -0,84  | -0,25  |
| 10       | 0,75   | -1,91  | 0,80   | 0,61   |
| 11       | -1,68  | 4,23   | 0,24   | 1,21   |
| 12       | 0,38   | 0,19   | 0,89   | 1,10   |
| 13       | -3,12  | -2,62  | -0,99  | 0,59   |
| 14       | 0,61   | 0,36   | 0,80   | 0,96   |



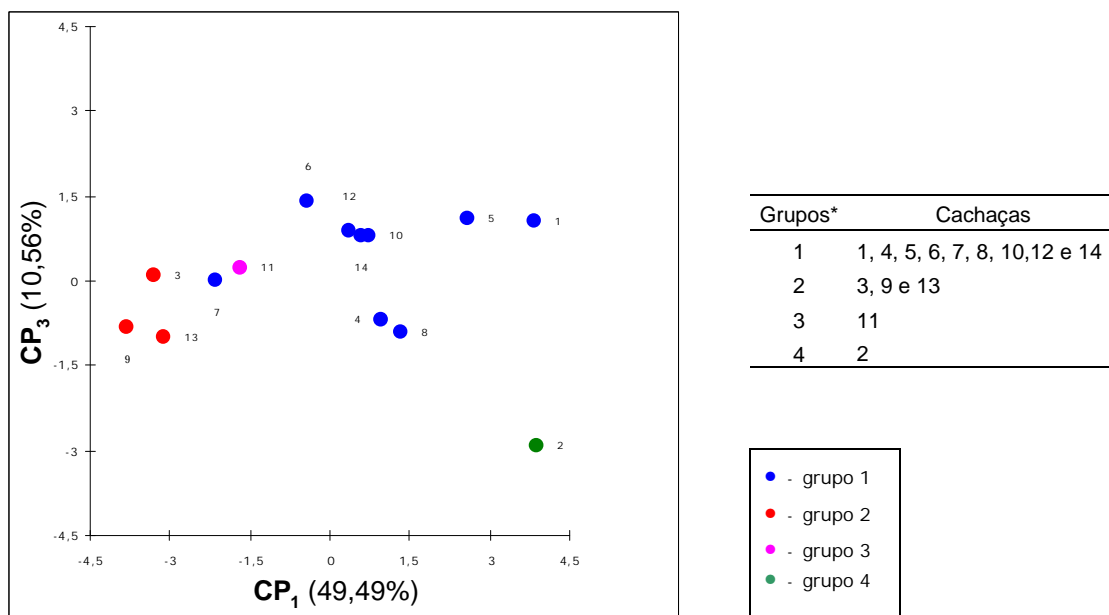
\*Observação: Os grupos foram estabelecidos pelo método de agrupamento de Tocher.

Figura 1: Dispersão das 14 cachaças em relação aos dois primeiros componentes principais ( $CP_1$  e  $CP_2$ ).

O quarto componente principal (eixo vertical da Figura 3) apresentou menor variabilidade entre os grupos, pois dos quatro, ele foi o que reteve a menor proporção de variância dos dados (7,77%).

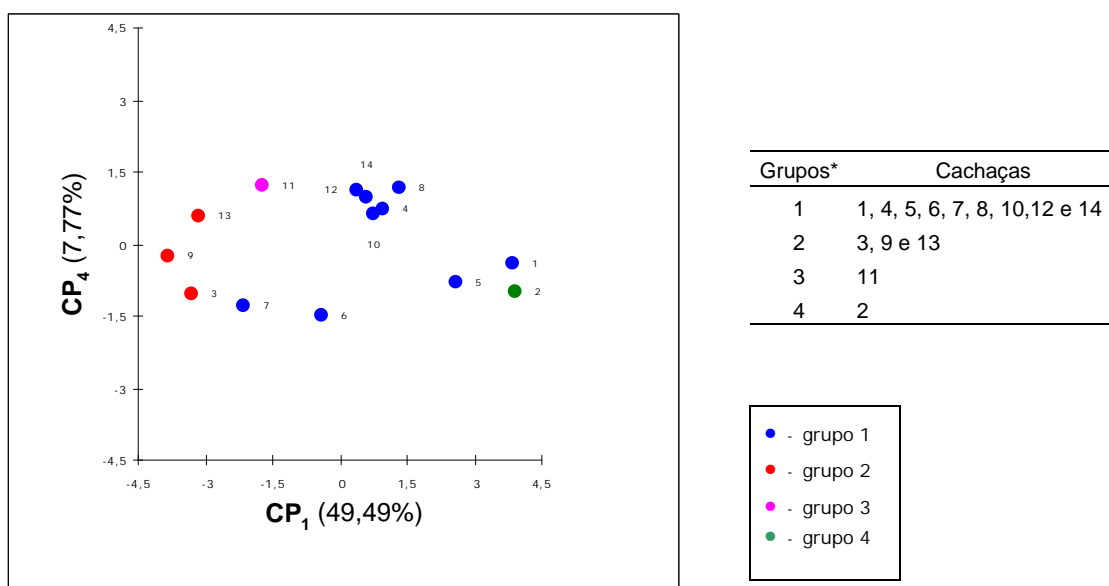
As cargas apresentadas no Quadro 6 permitiram a interpretação dos componentes, ou seja, permitiram verificar a importância relativa dos atributos ou a influência de cada um deles sobre os componentes. Cada um dos atributos destacados em negrito (aroma de madeira e de baunilha, sabor de madeira residual, coloração amarela, viscosidade, aroma alcoólico, sabor

alcoólico inicial e gosto amargo) apresentou a maior carga em um dos últimos nove componentes principais. Estes atributos estão relacionados a componentes que retêm uma pequena parte da variância dos dados e, portanto, foram considerados como os atributos de menor importância relativa.



\*Observação: Os grupos foram estabelecidos pelo método de agrupamento de Tocher.

Figura 2: Dispersão das 14 cachças em relação ao primeiro e terceiro componente principal ( $CP_1$  e  $CP_3$ ).



\*Observação: Os grupos foram estabelecidos pelo método de agrupamento de Tocher.

Figura 3: Dispersão das 14 cachças em relação ao primeiro e quarto componente principal ( $CP_1$  e  $CP_4$ ).

Quadro 6: Estimativas das cargas associadas aos atributos em cada componente principal (correlações entre as médias padronizadas dos atributos e os componentes principais).

| Componentes principais | Atributos   |             |             |       |       |             |              |       |       |              |       |              |              |
|------------------------|-------------|-------------|-------------|-------|-------|-------------|--------------|-------|-------|--------------|-------|--------------|--------------|
|                        | AALC        | AMAD        | ABAU        | DOCE  | ACID  | AMAR        | SALI         | SALR  | SMAI  | SMAR         | ADST  | COLA         | VISC         |
| CP <sub>1</sub>        | -0,72       | 0,98        | 0,77        | -0,31 | -0,18 | 0,49        | -0,18        | -0,33 | 0,98  | 0,98         | 0,59  | 0,96         | 0,80         |
| CP <sub>2</sub>        | 0,50        | 0,04        | -0,11       | -0,40 | 0,48  | 0,54        | 0,90         | 0,85  | 0,04  | 0,06         | 0,39  | 0,01         | 0,26         |
| CP <sub>3</sub>        | 0,35        | 0,12        | 0,35        | 0,54  | -0,61 | -0,37       | 0,35         | 0,34  | 0,11  | 0,10         | -0,17 | 0,16         | 0,03         |
| CP <sub>4</sub>        | 0,20        | -0,05       | 0,25        | 0,54  | 0,48  | -0,10       | -0,07        | -0,18 | 0,01  | 0,06         | 0,56  | -0,05        | -0,15        |
| CP <sub>5</sub>        | -0,02       | -0,03       | -0,20       | 0,37  | 0,22  | 0,13        | 0,02         | -0,06 | 0,02  | -0,01        | -0,32 | 0,06         | <b>0,46</b>  |
| CP <sub>6</sub>        | 0,07        | -0,01       | 0,16        | 0,15  | -0,21 | <b>0,55</b> | -0,12        | -0,02 | -0,09 | -0,07        | -0,01 | -0,14        | -0,07        |
| CP <sub>7</sub>        | -0,20       | -0,02       | <b>0,37</b> | -0,03 | 0,19  | -0,03       | 0,05         | 0,11  | -0,05 | -0,09        | -0,19 | -0,06        | -0,02        |
| CP <sub>8</sub>        | -0,03       | 0,07        | -0,08       | 0,07  | 0,08  | 0,06        | 0,00         | 0,07  | 0,06  | 0,07         | -0,11 | 0,13         | <b>-0,23</b> |
| CP <sub>9</sub>        | <b>0,11</b> | -0,01       | 0,04        | -0,05 | 0,03  | 0,00        | -0,02        | -0,05 | 0,00  | 0,00         | -0,05 | 0,05         | 0,00         |
| CP <sub>10</sub>       | 0,02        | 0,05        | 0,00        | 0,00  | 0,01  | -0,01       | -0,01        | 0,00  | 0,06  | 0,01         | -0,02 | <b>-0,08</b> | 0,00         |
| CP <sub>11</sub>       | 0,01        | <b>0,06</b> | 0,00        | 0,00  | 0,01  | -0,01       | -0,02        | 0,01  | -0,03 | -0,04        | 0,01  | 0,01         | 0,01         |
| CP <sub>12</sub>       | 0,01        | -0,02       | 0,00        | 0,00  | 0,00  | 0,00        | <b>-0,05</b> | 0,04  | -0,01 | 0,03         | 0,00  | 0,00         | 0,01         |
| CP <sub>13</sub>       | 0,00        | -0,01       | 0,00        | 0,00  | 0,00  | 0,00        | -0,01        | 0,01  | 0,02  | <b>-0,02</b> | 0,00  | 0,01         | 0,00         |

Observação: Cada linha representa as cargas, em um componente principal, associadas aos atributos padronizados.

AALC: aroma alcoólico, AMAD: aroma de madeira, ABAU: aroma de baunilha, DOCE: gosto doce, ACID: gosto ácido, AMAR: gosto amargo, SALI: sabor alcoólico inicial, SALR: sabor alcoólico residual, SMAI: sabor de madeira inicial, SMAR: sabor de madeira residual, ADST: adstringência, COLA: coloração amarela e VISC: viscosidade.

Vale ressaltar que quando em um componente principal de menor variância o maior coeficiente de ponderação está associado a um caráter já previamente descartado, tem-se optado por não fazer nenhum outro descarte com base nos coeficientes daquele componente, mas prosseguir a identificação da importância relativa dos caracteres no outro componente de variância imediatamente superior (CRUZ e REGAZZI, 1997). Isto aconteceu para o atributo viscosidade (Quadro 6).

Os atributos de menor importância relativa apresentaram baixa variabilidade e/ou correlacionaram-se com outros que fazem parte do estudo. A baixa variabilidade de alguns atributos indica que eles contribuíram pouco para a discriminação entre as cachaças. Já os atributos correlacionados com outro(s) tiveram seu comportamento explicado por este(s). Portanto, os atributos de menor importância relativa são passíveis de descarte em experimentos futuros de mesma natureza, ou até podem ser desconsiderados neste estudo. O gosto amargo está entre os atributos de menor variabilidade, ou seja, está entre aqueles que apresentaram os menores valores da estatística F (Quadro 3). Já o aroma de madeira e de baunilha, sabor de madeira residual, coloração amarela, viscosidade, aroma alcoólico e sabor alcoólico inicial estão correlacionados com pelo menos um dos outros atributos (Quadro 7), e portanto, têm seu comportamento explicado por estes.

Os demais atributos avaliados: sabor de madeira inicial, sabor alcoólico residual, adstringência, gosto doce e gosto ácido, considerados de maior importância relativa por se correlacionarem com componentes que retêm considerável parte da variância total dos dados (Quadro 6), apresentaram variabilidade entre as cachaças ou não se correlacionaram com outros atributos em estudo. Houve variabilidade entre cachaças para os atributos sabor de madeira inicial, sabor alcoólico residual e adstringência. Isto pode ser constatado pelo efeito significativo para cachaça, em relação a estes atributos (Quadro 3). O gosto doce e o gosto ácido não se correlacionaram ( $P > 0,05$ ) com nenhum outro atributo, assim eles não foram indicados para descarte, apesar de estarem entre os atributos com menor variabilidade (menores valores de F, Quadro 3).

As cargas (correlações entre atributos e componentes) relativas aos atributos de maior importância relativa estão apresentadas no Quadro 8. Observa-se que o sabor de madeira inicial se correlacionou com o primeiro componente e o sabor alcoólico residual com o segundo, assim, estes atributos foram os que mais contribuíram para a variabilidade observada no primeiro e segundo componente, respectivamente (Figura 1). Já a adstringência e os

gostos doce e ácido correlacionaram-se fracamente com mais de um componente, sobre os quais exerceram influência.

Quadro 7: Correlações (coeficientes de correlação de Pearson) entre os atributos avaliados nas cachaças envelhecidas.

| Atributos de menor importância relativa | Atributos de maior importância relativa |                          |               |            |             |
|-----------------------------------------|-----------------------------------------|--------------------------|---------------|------------|-------------|
|                                         | Sabor de madeira inicial                | Sabor alcoólico residual | Adstringência | Gosto doce | Gosto ácido |
| Aroma de madeira                        | 0,99<br><0.0001                         | —                        | 0,56<br>0,04  | —          | —           |
| Aroma de baunilha                       | 0,75<br>0,00                            | —                        | —             | —          | —           |
| Sabor de madeira residual               | 0,99<br><0.0001                         | —                        | 0,63<br>0,01  | —          | —           |
| Coloração amarela                       | 0,98<br><0.0001                         | —                        | —             | —          | —           |
| Viscosidade                             | 0,80<br>0,00                            | —                        | —             | —          | —           |
| Aroma alcoólico                         | -0,65<br>0,01                           | 0,72<br>0,00             | —             | —          | —           |
| Sabor alcoólico inicial                 | —                                       | 0,96<br><0.0001          | —             | —          | —           |
| Gosto amargo                            | —                                       | —                        | —             | —          | —           |

Observação: Estão apresentadas apenas as correlações significativas (P = 0,05).

Quadro 8: Estimativas das cargas associadas aos atributos de maior importância relativa, nos quatro primeiros componentes principais (correlações entre as médias padronizadas dos atributos de maior importância relativa e os quatro primeiros componentes principais).

| Primeiros componentes | Proporção da variância (%) | Atributos de maior importância relativa |                          |               |             |             |
|-----------------------|----------------------------|-----------------------------------------|--------------------------|---------------|-------------|-------------|
|                       |                            | Sabor de madeira inicial                | Sabor alcoólico residual | Adstringência | Gosto doce  | Gosto ácido |
| CP <sub>1</sub>       | 49,49                      | <b>0,98</b>                             | -0,33                    | <b>0,59</b>   | 0,31        | 0,18        |
| CP <sub>2</sub>       | 20,88                      | 0,04                                    | <b>0,85</b>              | 0,39          | 0,40        | <b>0,48</b> |
| CP <sub>3</sub>       | 10,56                      | 0,11                                    | 0,34                     | -0,17         | <b>0,54</b> | <b>0,61</b> |
| CP <sub>4</sub>       | 7,77                       | 0,01                                    | -0,18                    | <b>0,56</b>   | <b>0,54</b> | <b>0,48</b> |

Observação: Cada linha representa as cargas, em um componente principal, associadas aos atributos padronizados.

A relação entre sabor de madeira inicial e o primeiro componente principal é direta, pois o coeficiente de correlação entre os dois é positivo (Quadro 8). Portanto, as cachaças ou grupos posicionados mais à direita das Figuras 1, 2 e 3 (maiores escores para o primeiro componente principal) foram caracterizadas por maior intensidade no sabor de madeira inicial.



Os atributos aroma de madeira e de baunilha, sabor de madeira residual, coloração amarela e viscosidade correlacionaram-se positivamente com o sabor de madeira inicial (Quadro 7) e, portanto, apresentaram comportamento semelhante ao deste atributo. Ou seja, houve variabilidade entre os grupos para tais atributos e os grupos posicionados mais à direita no eixo horizontal (Figuras 1, 2 e 3) apresentaram, além de um maior sabor de madeira inicial, maior intensidade nos atributos aroma de madeira e de baunilha, sabor de madeira residual, coloração amarela e viscosidade. Os grupos mais à direita apresentaram também, menor aroma alcoólico, pois este atributo correlacionou-se negativamente com o sabor de madeira inicial.

A relação entre sabor alcoólico residual e o segundo componente principal foi, também, direta (Quadro 8), assim, a cachaça número 11 (grupo 3), com maior escore no segundo componente (eixo vertical da Figura 1), caracterizou-se por um sabor alcoólico residual mais intenso. A cachaça 11 caracterizou-se, também, por maior aroma alcoólico e maior sabor alcoólico inicial, atributos estes correlacionados positivamente com o sabor alcoólico residual (Quadro 7).

A adstringência teve uma pequena contribuição para a variabilidade entre os grupos de cachaça observada ao longo do primeiro e quarto componente principal, pois a adstringência apresentou baixa correlação com estes componentes (coeficiente de correlação ( $r$ ) inferior a 0,60, Quadro 8). Como a relação foi direta, significa que os grupos posicionados mais à direita e acima na Figura 3 (grupos com maiores escores em  $CP_1$  e  $CP_4$ ) apresentaram adstringência ligeiramente maior. Comportamento semelhante foi apresentado pelo aroma de madeira e sabor de madeira residual, ambos correlacionados positivamente com a adstringência (Quadro 7).

As correlações entre os atributos gosto doce e gosto ácido com o segundo, terceiro e quarto componente foram todas baixas (coeficientes de correlação, em valor absoluto, inferiores a 0,61, Quadro 8), o que indica que estes são atributos que contribuíram pouco para a variabilidade entre os grupos nos referidos componentes. É preciso lembrar que o fato do gosto doce e o ácido estarem entre os atributos de maior importância relativa não implica, necessariamente, que eles apresentaram alta variabilidade. Provavelmente eles estão entre os de maior importância relativa por não estarem correlacionados ( $P > 0,05$ ) com nenhum dos outros atributos avaliados.

#### 4. CONCLUSÕES

A análise de componentes principais foi satisfatória, pois permitiu a redução no conjunto de dados de 13 atributos para quatro componentes principais. As 14 cachaças envelhecidas foram, então, caracterizadas e comparadas em relação aos componentes e, indiretamente, em relação aos 13 atributos. Porém, há de lembrar que os atributos sabor de madeira inicial, sabor alcoólico residual, adstringência, gosto doce e gosto ácido foram considerados de maior importância relativa, e, portanto, suficientes para a caracterização sensorial e comparação entre as cachaças. Os demais atributos apresentam baixa variabilidade ou foram redundantes por estarem correlacionados com os de maior importância. Assim, num próximo estudo de mesma natureza, por exemplo, se o produtor de cachaça quiser acompanhar, periodicamente, as características sensoriais de seu produto, basta avaliar apenas os cinco atributos de maior importância relativa.

Os atributos aroma de madeira e de baunilha, sabor de madeira residual, coloração amarela e viscosidade tiveram seu comportamento explicado pelo sabor de madeira inicial com a qual se correlacionaram. O comportamento do aroma alcoólico e sabor alcoólico inicial foi explicado pelo sabor alcoólico residual com o qual se correlacionaram. Portanto, na avaliação das cachaças envelhecidas, destacaram-se os atributos relacionados à madeira e ao teor alcoólico, que apresentaram variabilidade entre várias das cachaças.

Destaca-se, também, que os resultados obtidos foram equivalentes àqueles apresentados no primeiro artigo, em que os mesmos dados experimentais foram submetidos à análise por variáveis canônicas. Pois, em ambas as análises, a primeira e segunda dimensão de maior importância relativa (primeiro e segundo componente principal ou primeira e segunda variável canônica) foram influenciadas por atributos relacionados à madeira e ao teor alcoólico, respectivamente. Porém, a análise por variáveis canônicas apresentou algumas vantagens:

- as duas primeiras variáveis canônicas retiveram, juntas, 91,58% da variabilidade total disponível, contra 70,37% retida pelos dois primeiros componentes principais. Assim, houve menos perda de informação na redução do conjunto de dados ao usar variáveis canônicas;
- os grupos estabelecidos foram mais consistentes do que em componentes principais, ou seja, as cachaças alocadas em um mesmo grupo distam pouco entre si, o que não deixa dúvida quanto à sua alocação no grupo;

Portanto, recomenda-se que para avaliar dados obtidos por Análise Descritiva Quantitativa, em que as informações são obtidas com repetições, seja utilizada preferencialmente a análise por variáveis canônicas. Já a análise por componentes principais é mais apropriada para conjuntos de dados sem repetições, ou seja, em que há apenas uma observação por variável em cada indivíduo avaliado.

Vale ressaltar que a principal diferença entre as duas técnicas é que em variáveis canônicas são consideradas as variâncias e covariâncias residuais entre os dados, o que não acontece na análise de componentes principais, tornando variáveis canônicas mais apropriadas para dados com repetição.

Cabe, aqui, também, uma comparação dos resultados da análise de componentes principais com os da análise de fatores, apresentados no segundo artigo. Os atributos adstringência, gosto doce e gosto ácido correlacionaram-se fracamente com mais de um componente (coeficientes de correlação, em valor absoluto, inferiores a 0,61), o que dificulta identificar qual ou quais componentes foram influenciados por cada atributo. A rotação efetuada na análise de fatores minimizou este problema, pois algumas correlações se aproximaram dos limites extremos 1 ou -1, e outras de zero. Isto facilitou a interpretação dos resultados, pois cada fator passou a apresentar correlação relativamente forte com um ou mais atributos e correlação relativamente fraca com os demais. Porém, não é possível afirmar que uma ou outra análise seria melhor, pois tais análises não são concorrentes. Na análise de componentes principais a ênfase é explicar a variância, enquanto na análise de fatores o objetivo é tentar explicar as covariâncias, ou seja, enquanto na primeira o interesse foi identificar os atributos que mais contribuíram para a variabilidade observada entre as cachaças, na análise de fatores buscou-se separar os atributos em grupos (fatores), de forma que cada grupo reunisse os atributos correlacionados entre si.

Para finalizar, a análise por variáveis canônicas pode, também, ser caracterizada como uma análise em que a ênfase é a variância, a exemplo da análise de componentes principais, e estas sim, são consideradas como concorrentes, ou seja, usadas com o mesmo propósito.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

CARNEIRO, J. C. S. **Processamento industrial de feijão, avaliação sensorial descritiva e mapa de preferência**. 2001. 90f. Dissertação (Mestrado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa. Viçosa.

CARNEIRO, J. D. S.; REIS R. C.; MINIM, V. P. R.; CARNEIRO, J. C. S.; REGAZZI, A. J. Avaliação sensorial e mapa de preferência interno de marcas comerciais de refrigerante sabor guaraná. **Boletim do CEPPA**, Curitiba, v. 21, n.2, jun/dez. 2003.

CRUZ, C.D. **Programa Genes: versão Windows; aplicativo computacional em genética e estatística**. 1. ed. Viçosa: Editora UFV, 2001. 648 p.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. v. 2, 1. ed. Viçosa: Editora UFV, 2003. 585 p.

CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. v. 1, 1. ed. Viçosa: Editora UFV, 1997. 309p.

FREITAS, M. Q. de, **Características e aceitação sensorial de mortadelas produzidas com carne mecanicamente separada de frango**. 2002. 114f. Tese (Doutorado em Ciência e Tecnologia de Alimentos) – Universidade Federal de Viçosa. Viçosa.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4. ed. Upper Saddle River, New Jersey-USA: Prentice-Hall, 1998. 816 p.

POLIGNANO, L. A. C.; DRUMOND, F. B.; CHENG, L. C. Utilização dos mapas de percepção e preferência como técnicas auxiliares do QFD durante o desenvolvimento de produtos alimentícios. In: CONGRESSO BRASILEIRO DE GESTÃO DO DESENVOLVIMENTO DE PRODUTO. 1., 1999, Belo Horizonte. **Anais...** 1999. p. 274-284.

PONTES, L.V. **Avaliação sensorial e instrumental da cor de misturas em pó para refresco, bebida isotônica e gelatina utilizando corantes naturais**. 2004. 70f. Dissertação (Mestrado em Ciência e Tecnologia de Alimentos) – Universidade Federal de Viçosa. Viçosa.

RAO, R.C. **Advanced statistical methods in biometric research**. New York: John Wiley and Sons, 1952. 390 p.

REGAZZI, A. J. INF 766 - Análise Multivariada (notas manuscritas). DPI - UFV. Viçosa – MG, 1997.

SAS/STAT users' guide. Version 8. Cary, NC: **SAS Institute Inc.**, 1999.

YOKOTA, S.R.C. **Avaliação sensorial descritiva de cachaça envelhecida entre 18 e 24 meses por 4 painéis de julgadores**. 2005. 102f. Tese (Doutorado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa. Viçosa.

## CONCLUSÕES GERAIS

As três técnicas estatísticas multivariadas abordadas produziram resultados satisfatórios quando aplicadas aos dados descritivos obtidos pela avaliação sensorial das 14 cachaças envelhecidas. As análises permitiram a redução de 13 atributos para um número bem menor de dimensões que retiveram a maior proporção da variância total dos dados originais. Em variáveis canônicas foram necessárias apenas duas dimensões (duas variáveis canônicas) que retiveram 91,58% da variância. Em componentes principais e análise de fatores houve a redução para quatro dimensões (componentes ou fatores) que retiveram 88,69% da variância.

Foi possível, também, avaliar a importância ou influência dos atributos sobre as primeiras dimensões consideradas na caracterização e comparação entre as cachaças. Nas três análises, as duas primeiras dimensões foram influenciadas por atributos relacionados à madeira e ao teor alcoólico, respectivamente. Ficou, então, evidenciado que os resultados das três técnicas foram equivalentes e que o teor alcoólico e as características atribuídas pela madeira são importantes na avaliação de cachaça envelhecida.

A análise por variáveis canônicas apresentou algumas vantagens em relação aos componentes principais porque as primeiras variáveis canônicas retiveram maior proporção da variância do que os primeiros componentes e também porque os grupos estabelecidos foram mais consistentes. Portanto, recomenda-se que para avaliar dados obtidos por análise sensorial descritiva, em que as informações são obtidas com repetições, seja utilizada preferencialmente a análise por variáveis canônicas. Já a análise por componentes principais é mais apropriada para conjuntos de dados sem

repetições, ou seja, em que há apenas uma observação por variável em cada indivíduo avaliado.

A rotação efetuada na análise de fatores facilitou a interpretação dos resultados em comparação à análise de componentes principais. Porém, não se pode dizer que a primeira análise é a melhor, pois estas técnicas não são concorrentes. Na análise de componentes principais a ênfase é explicar a variância, enquanto que na análise de fatores o objetivo é tentar explicar as covariâncias.

Conclui-se, então, que as técnicas variáveis canônicas, componentes principais e análise de fatores podem ser aplicadas a dados obtidos por análise sensorial descritiva, desde que os resultados sejam satisfatórios como os aqui apresentados. Adicionalmente, recomenda-se que a análise por variáveis canônicas seja preferencial em relação à análise de componentes principais, pois os dados foram obtidos com repetição (cada uma das 14 cachaças foi avaliada por 8 provadores, sendo que cada um deles avaliou 5 amostras por cachaça).