



UNIVERSIDADE ESTADUAL DO SUDOESTE DA BAHIA - UESB
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA E CIÊNCIA DE ALIMENTOS



Áreas de Concentração: Engenharia de Alimentos e Ciência de Alimentos

PROSPECÇÃO E IDENTIFICAÇÃO DE CACHAÇAS ENVELHECIDAS EM
DIFERENTES REGIÕES E MADEIRAS POR QUIMIOMETRIA.

Autora: Roberta Magalhães Dias Cardozo

Orientador: Prof. Dr. Leandro Soares Santos

ITAPETINGA
BAHIA – BRASIL
Dezembro de 2015

ROBERTA MAGALHÃES DIAS CARDOZO

**PROSPECÇÃO E IDENTIFICAÇÃO DE CACHAÇAS ENVELHECIDAS EM
DIFERENTES REGIÕES E MADEIRAS POR QUIMIOMETRIA.**

Dissertação apresentada como parte das exigências para obtenção do título de Mestre em Engenharia e Ciência de Alimentos, no Programa de Pós-Graduação em Engenharia e Ciência de Alimentos da Universidade Estadual do Sudoeste da Bahia.

Orientador:

Prof. Dr. Leandro Soares Santos

Co-Orientadores:

Prof. Dr. Paulo Bonomo

Prof.^a Dr.^a Silmara Almeida de Carvalho

ITAPETINGA

BAHIA – BRASIL

Dezembro de 2015

663.16 C269p	<p>Cardozo, Roberta Magalhães Dias Prospecção e identificação de cachaças envelhecidas em diferentes regiões e madeiras por quimiometria. / Roberta Magalhães Dias Cardozo. - Itapetinga: UESB, 2016. 111f.</p> <p>Dissertação apresentada como parte das exigências para obtenção do título de Mestre em Engenharia e Ciência de Alimentos, no Programa de Pós-Graduação em Engenharia e Ciência de Alimentos da Universidade Estadual do Sudoeste da Bahia. Sob a orientação do Prof. D.Sc. Leandro Soares Santos e co-orientação do Prof. D.Sc. Paulo Bonomo e Profª. D.Sc. Silmara Almeida de Carvalho.</p> <p>1. Cachaça. 2. Cachaça - Análises Quimiométricas. 3. Cachaça - Análises Físico-Químicas. I. Universidade Estadual do Sudoeste da Bahia. Programa de Pós-Graduação em Engenharia e Ciência de Alimentos. II. Santos, Leandro Soares. III. Bonomo, Paulo. IV. Carvalho, Silmara Almeida de. V. Título.</p> <p style="text-align: center;">CDD(21): 663.16</p>
-----------------	---

Catalogação na fonte:
 Adalice Gustavo da Silva – CRB/5-535
 Bibliotecária – UESB – Campus de Itapetinga-BA

Índice Sistemático para Desdobramento por Assunto:

1. Cachaça
2. Cachaça - Análises Quimiométricas
3. Cachaça - Análises Físico-Químicas



Universidade Estadual do Sudoeste da Bahia
Programa de Pós-Graduação
Mestrado em Engenharia e Ciência de Alimentos



Áreas de Concentração: Engenharia de Alimentos
Ciência de Alimentos

DECLARAÇÃO DE APROVAÇÃO

Título: PROSPECÇÃO DE CACHAÇAS ENVELHECIDAS EM DIFERENTES
REGIÕES E MADEIRAS POR QUIMIOMETRIA

Autor (a): ROBERTA MAGALHÃES DIAS CARDOZO

Orientador (a): Prof.^o Dr.^o Leandro Soares Santos

Co-orientador (a): Prof.^o Dr. Paulo Bonomo

Prof.^a Dr.^a Silmara Almeida de Carvalho

Aprovado como parte das exigências para obtenção do Título de **MESTRE
EM ENGENHARIA E CIÊNCIA DE ALIMENTOS, ÁREA DE
CONCENTRAÇÃO: CIÊNCIA DE ALIMENTOS**, pela Banca
Examinadora.

Leandro Soares Santos
Prof.^o Dr. Leandro Soares Santos (UESB)

Renata Cristina Ferreira Bonomo
Prof.^a Dr.^a Renata Cristina Ferreira Bonomo (UESB)

Anderson Santos Souza
Prof.^o Dr. Anderson Santos Souza (UFBA)

Itapetinga-BA, 04 de dezembro de 2015.

“Deus quer, o homem sonha e a obra nasce”

Fernando Pessoa

A Deus.

Aos meus pais.

Aos meus irmãos.

A minha família.

DEDICO

Agradecimentos

A DEUS por cuidar de cada passo dado ao longo deste curso, me orientando, iluminando e me dando força para superar as dificuldades.

Aos meus pais, Alberto e Enilda, por serem meus maiores incentivadores e por, muitas vezes, abdicarem dos seus sonhos para que eu concretizasse os meus.

Aos meus irmãos, Ricardo e Rayssa, pela torcida e incentivo.

A Felipe, pelas contribuições, companheirismo, paciência e carinho.

Aos meus familiares, que sempre me apoiaram, orientaram e ajudaram.

Ao meu orientador, Leandro, pela orientação, confiança, incentivo, ensinamentos e amizade.

Aos professores Paulo e Silmara pela co-orientação, disponibilidade e amizade.

Ao programa de pós-graduação em Engenharia e Ciência de Alimentos pela oportunidade de realização desse curso.

Aos professores da UESB pela amizade e ensinamentos.

Aos amigos e companheiros do mestrado Edilson, Olga, Gabriel, Joyce, Mylena, Marília, Mariana, Alexandre, Jorge, Carilan, Márcio, Hudson e Carol.

Aos secretários da pós-graduação, Jamille e Jaime pela dedicação no atendimento, pela amizade e estarem sempre dispostos a colaborar com nosso trabalho.

Aos amigos William, Thiara e Cristina pelo acolhimento e cuidado;

Ao IFNMG – Campus Salinas pela oportunidade da realização deste curso.

Aos servidores do IFNMG – Campus Salinas, em especial a Profª. Dra. Maria Araci Magalhães e o Prof. Dr. Leonardo Lara e Lanna, pelo apoio, pela atenção e presteza com as quais sempre fui muito bem atendida.

E a todos que direta ou indiretamente contribuíram para a conclusão deste trabalho.

MUITO OBRIGADA!

SUMÁRIO

LISTA DE FIGURAS	11
LISTA DE TABELAS	14
LISTA DE ABREVIATURAS	15
RESUMO	16
ABSTRACT	17
CAPÍTULO 1	18
1 INTRODUÇÃO GERAL	18
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Cachaça: Definição e Principais Regiões Produtoras	19
2.2 Produção da Cachaça	21
2.2.1 Fermentação	22
2.2.2 Destilação	23
2.2.3 Envelhecimento	24
2.3 Quimiometria e suas aplicações em pesquisas envolvendo cachaça	25
2.3.1 Análise de Componentes Principais (ACP)	27
2.3.2 Análise de Fatores (AF)	30
2.3.3 Análise de Correlações Canônicas (ACC)	31
2.3.4 Análise Discriminante (AD)	32
2.4 Redes Neurais Artificiais (RNA)	34
3 OBJETIVOS	36
3.1 OBJETIVO GERAL	36
3.2 OBJETIVOS ESPECÍFICOS	36
4 REFERÊNCIAS	37
CAPÍTULO 2 – PROSPECÇÃO E IDENTIFICAÇÃO DE CACHAÇAS PRODUZIDAS EM SALINAS E PARATY POR QUIMIOMETRIA	44
RESUMO	44
ABSTRACT	44
1 INTRODUÇÃO	45
2 MATERIAL E MÉTODOS	46
2.1 Coleta dos Dados	46
2.2 Delineamento Experimental dos Dados	47
2.3 Medidas de Dispersão e Padronização da Matriz de Dados	48

2.4	Análises Estatísticas Multivariadas dos Dados	49
2.4.1	Análise de Componentes Principais	49
2.4.2	Análise de Fatores.....	50
2.4.3	Análise de Correlações Canônicas.....	50
2.4.4	Análise Discriminante.....	51
2.5	Redes Neurais Artificiais	52
2.5.1	Treinamento e Validação dos Dados	54
3	RESULTADOS E DISCUSSÃO.....	55
3.1	Análise de Componentes Principais	55
3.2	Análise de Fatores.....	58
3.3	Análise de Correlações Canônicas.....	63
3.4	Análise Discriminante.....	67
3.5	Rede Neural.....	69
4	CONCLUSÃO.....	74
5	REFERÊNCIAS	74
CAPÍTULO 3 – INVESTIGAÇÃO DE CACHAÇAS ENVELHECIDAS EM DIFERENTES MADEIRAS POR ANÁLISE QUIMIOMÉTRICA.....		77
RESUMO		77
ABSTRACT		77
1	INTRODUÇÃO.....	78
2	MATERIAL E MÉTODOS.....	79
2.1	Coleta dos Dados.....	79
2.2	Delineamento Experimental dos Dados	80
2.3	Medidas de Dispersão e Padronização da Matriz de Dados.....	80
2.4	Análises Estatísticas Multivariadas dos Dados.....	81
2.4.1	Análise de Componentes Principais	81
2.4.2	Análise de Fatores.....	82
2.4.3	Análise de Correlações Canônicas.....	82
2.4.4	Análise Discriminante.....	83
2.5	Redes Neurais Artificiais.....	84
2.5.1	Treinamento e Validação dos Dados	86
3	RESULTADOS E DISCUSSÕES.....	87
3.1	Análise de Componentes Principais.....	87
3.2	Análise de Fatores.....	92

3.3	Análise de Correlação Canônica	96
3.4	Análise Discriminante	98
3.5	Redes Neurais Artificiais	101
4	CONCLUSÕES	104
5	REFERÊNCIAS	105
	CONSIDERAÇÕES FINAIS	109
	APÊNDICES	110

LISTA DE FIGURAS

CAPÍTULO 1

Figura 1. Fluxograma da Cachaça de Alambique.....	22
Figura 2. <i>Scree plot</i> dos autovalores associados com os componentes principais.....	29
Figura 3. Representação e fronteiras de decisão, não-linear (esquerda) e linear (direita).....	32
Figura 4. Funcionamento de um neurônio computacional.....	35

CAPÍTULO 2

Figura 1. Construção das Matrizes de Dados A_1 e A_2	47
Figura 2. Algoritmo <i>Backpropagation</i>	52
Figura 3. Gráfico de escores das cachaças produzidas nas regiões de Salinas e Paraty em relação aos três primeiros componentes principais (CP1, CP2 e CP3) da matriz A_1	55
Figura 4. Gráfico de pesos dos coeficientes de ponderação (Autovetores) da matriz A_1 contendo as cachaças produzidas nas regiões de Salinas e Paraty.....	56
Figura 5. Gráfico de escores das cachaças produzidas nas regiões de Salinas e Paraty em relação aos três primeiros componentes principais (CP1, CP2 e CP3) da matriz A_2	57
Figura 6. Gráfico de pesos dos coeficientes de ponderação (Autovetores) para matriz A_2 contendo as cachaças produzidas nas regiões de Salinas e Paraty.....	58
Figura 7. Gráfico de pesos das cargas fatoriais finais após a rotação da matriz A_1 para as cachaças produzidas nas regiões de Salinas e Paraty.....	59
Figura 8. Gráfico de escores das cachaças produzidas nas regiões de Salinas e Paraty em relação aos dois primeiros fatores (Fator 1 e Fator 2) da matriz A_1	61
Figura 9. Gráfico de pesos das cargas fatoriais finais após a rotação da matriz A_2 para as cachaças produzidas nas regiões de Salinas e Paraty.....	61
Figura 10. Gráfico de escores das cachaças produzidas nas regiões de Salinas e Paraty em relação aos três primeiros fatores (Fator 1, Fator 2 e Fator 3) da matriz A_2	63
Figura 11. Correlação do primeiro par de variáveis canônicas obtida para a matriz A_1	64

Figura 12. Representação canônica das observações obtidas pelas variáveis da fração álcool da matriz A_1	65
Figura 13. Representação canônica das observações obtidas pelas variáveis da fração não álcool da matriz A_2	65
Figura 14. Correlação do primeiro par de variáveis canônicas obtida para a matriz A_2	66
Figura 15. Representação canônica das observações obtidas pelas variáveis da fração álcool da matriz A_2	66
Figura 16. Representação canônica das observações obtidas pelas variáveis da fração álcool da matriz A_2	67
Figura 17. RMSError x Número de iterações para dados de treinamento e validação da matriz A_1	69
Figura 18. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de treinamento da matriz A_1	70
Figura 19. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de validação da matriz A_1	70
Figura 20. Importância relativa (%) de cada variável na classificação de amostras de cachaça da matriz A_1	71
Figura 21. RMSError x Número de iterações para dados de treinamento e validação da matriz A_2	72
Figura 22. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de treinamento da matriz A_2	73
Figura 23. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de treinamento da matriz A_2	73
Figura 24. Importância relativa (%) de cada variável na classificação de amostras de cachaça da matriz A_2	74

CAPÍTULO 3

Figura 1. Algoritmo <i>Backpropagation</i>	84
Figura 2. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana, Bálsamo e Carvalho em relação aos três primeiros componentes principais (CP1, CP2 e CP3).....	88

Figura 3. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana e Bálamo em relação aos três primeiros componentes principais (CP1, CP2 e CP3).....	89
Figura 4. Gráfico de pesos dos coeficientes de ponderação (Autovetores) para as cachaças envelhecidas nas madeiras Amburana e Bálamo.....	90
Figura 5. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana e Carvalho em relação aos dois primeiros componentes principais (CP1 e CP2).....	91
Figura 6. Gráfico de pesos dos coeficientes de ponderação (Autovetores) para as cachaças envelhecidas nas madeiras Amburana e Carvalho.....	91
Figura 7. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana, Bálamo e Carvalho em relação aos três primeiros fatores (Fator 1, Fator 2 e Fator 3).....	92
Figura 8. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana e Bálamo em relação aos três primeiros fatores (Fator 1, Fator 2 e Fator 3).....	93
Figura 9. Gráfico de pesos das cargas fatoriais finais após a rotação para as cachaças envelhecidas nas madeiras Amburana e Bálamo).....	94
Figura 10. Gráfico de pesos das cargas fatoriais finais após a rotação para as cachaças envelhecidas nas madeiras Amburana e Carvalho.....	95
Figura 11. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana e Carvalho em relação aos dois primeiros fatores (Fator 1 e Fator 2).....	96
Figura 12. Correlação do primeiro par de variáveis canônicas obtida para a matriz dois.....	97
Figura 13. Representação canônica das observações obtidas pelas variáveis da fração álcool.....	98
Figura 14. Representação canônica das observações obtidas pelas variáveis da fração não álcool.....	98
Figura 15. RMSError x Número de iterações para dados de treinamento e validação da matriz A_2	102
Figura 16. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de treinamento.	103
Figura 17. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de validação.....	103
Figura 18. Importância relativa (%) de cada variável na classificação de amostras de cachaça da matriz A_2	104

LISTA DE TABELAS

CAPÍTULO 1

Tabela 1. Teores máximo e mínimo para os componentes secundários e contaminantes presentes na aguardente de cana e cachaça.....	20
--	----

CAPÍTULO 2

Tabela 1. Número de amostras por região e amplitude das variáveis utilizadas para construção da matriz A_1	47
Tabela 2. Número de amostras por região e amplitude das variáveis utilizadas para construção da matriz A_2	48
Tabela 3. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaças da matriz A_1	68
Tabela 4. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaças da matriz A_2	69

CAPÍTULO 3

Tabela 1. Número de amostras por madeira e amplitude das variáveis utilizadas para construção da matriz A_2	80
Tabela 2. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaça envelhecidas em Amburana, Bálsamo e Carvalho...99	99
Tabela 3. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaça envelhecidas em Amburana e Bálsamo.....	100
Tabela 4. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaça envelhecidas em Amburana e Carvalho.....	101

LISTA DE ABREVIATURAS

ACC: Análise de Correlações Canônicas

ACP: Análise de Componentes Principais

AD: Análise Discriminante

ADL: Análise Discriminante Linear

AF: Análise de Fatores

AFE: Análise de Fatores

Amet: Álcool Metílico

ASIsoA: Álcool Superior Isoamílico

ASIsoB: Álcool Superior Isobutílico

ASnP: Álcool Superior n-propílico

ASTot: Alcoóis Superiores Totais

CP: Componentes Principais

GL: Grau Alcoólico

SCV: Soma dos Componentes Voláteis

CARDOZO, Roberta Magalhães Dias. **Prospecção e identificação de cachaças envelhecidas em diferentes regiões e madeiras por quimiometria**. Itapetinga, BA: UESB, 2015. 111p. Dissertação. (Mestrado em Engenharia e Ciências de Alimentos, Área de Concentração em Ciência de Alimentos). *

RESUMO

A expansão do comércio da cachaça e os valores envolvidos na cadeia produtiva vem estimulando as pesquisas na área. Ainda por conta disto, ao longo dos anos, regiões ganharam fama por causa da bebida, que se tornou símbolo que reflete a economia e cultura local, como ocorre em Salinas e Paraty. Como consequência receberam o selo de Indicação de Procedência, mas, em função da popularidade, são frequentes os problemas relacionados com falsificação da bebida. Outra questão relevante para a expansão do comércio da cachaça é o envelhecimento destas bebidas em barris de madeiras que tornam a cachaça mais agradável ao paladar, aumenta o seu valor comercial e caracteriza a comercialização por região. Desta forma, uma alternativa para identificar as madeiras utilizadas no envelhecimento da cachaça e as regiões de produção, a fim de gerar conhecimento sobre variáveis importantes no processamento da bebida e, por consequência ter subsídio para evitar possíveis fraudes, é a quimiometria. Neste trabalho objetivou-se a identificação das variáveis que apresentam maior relevância e a identificação de cachaças produzidas nas regiões de Salinas e Paraty em diferentes madeiras, através da utilização de técnicas multivariadas e dos dados de composição físico-químicos exigidos pela legislação Brasileira. Foram coletados, por meio de visitas *in loco*, certificados oficiais de análises físico-químicas que estruturaram as matrizes de dados com n medidas de diferentes propriedades físico-químicas (variáveis) realizadas sobre m amostras (cachaças). Para o tratamento das matrizes de dados foram utilizadas as seguintes técnicas estatísticas: análise de componentes principais (ACP), fatores (AF), correlações canônicas (ACC), discriminante (AD) e a técnica não estatística, rede neural artificial (RNA). Na prospecção de cachaças produzidas em Salinas e Paraty os resultados indicaram que as técnicas não supervisionadas (ACP, AF e ACC) foram eficientes na formação de agrupamentos com registro de alguns pontos distribuídos desuniformemente em função das singularidades dos processos de produção destas regiões. Quanto às variáveis que possuem maior influência na identificação das regiões, pode se destacar a acidez, ésteres e os alcoóis superiores. Já na aplicação das técnicas supervisionadas obteve-se uma excelente classificação na etapa de validação com valores entre 83,87 e 100% com ênfase para a rede neural que apresentou 100% de separação na validação para as duas matrizes avaliadas. Posteriormente na investigação de cachaças envelhecidas em diferentes madeiras constatou-se que o processo de classificação utilizando a ACP e AF das cachaças envelhecidas nas madeiras, Amburana, Bálsamo e Carvalho resultou na escolha das melhores separações que ocorreram com as combinações Amburana e Carvalho e Amburana e Bálsamo. No entanto devido a questões relacionadas ao processamento não houve separação total dos grupos. A Análise de Correlações Canônicas obteve alta correlação, mas a separação por grupo obteve resultados semelhantes aos observados na ACP e AF. O resultado da ACP e AF foi utilizado, ainda, para realização da Análise Discriminante (AD). A AD e Redes Neurais Artificiais (RNA) apresentaram, na validação, uma porcentagem de acerto que variou entre 10,0 e 80,0%. As variáveis que possuem maior importância na classificação das cachaças segundo ACP, AF e RNA são acidez, GL e alcoóis superiores.

Palavras-chave: Cachaça, Análises Quimiométricas, Análises Físico-Químicas.

* Orientador: Prof. Leandro Soares Santos, Dr. UESB e Co-orientadores: Prof. Paulo Bonomo, Dr. UESB e Prof.^a Silmara Almeida de Carvalho, Dr.^a UESB.

ABSTRACT

Expanding trade of cachaça and the amounts involved in the production chain has been stimulating research in the area. Yet because of this, over the years, regions have gained fame because of drinking, which has become a symbol which reflects the economy and local culture, as in Salinas and Paraty. As a result of receiving the indication of origin seal, but depending on the popularity, there are frequent problems associated with the beverage forgery. Thus, an alternative to identify the wood used in the aging of cachaça and production regions to generate knowledge about important variables in the processing of drink and therefore have benefits to prevent possible fraud, is chemometrics. This study aimed to identify the variables that have greater relevance and cachaças identification produced in the regions of Salinas and Paraty in different woods, through the use of multivariate techniques and the physical and chemical composition data required by Brazilian law. Were collected through site visits, official certificates of physic-chemical analysis that structured arrays of data with n measurements of different physicochemical properties (variables) performed on samples m (cachaça). For the treatment of matrix data the following statistical techniques were used: principal component analysis (PCA), factors (FA), canonical correlation (CCA), cluster analysis (CA), discriminant (DA) and non-statistical technique, neural network artificial (ANN). The prospect of cachaça produced in Salinas and Paraty the results indicated that the unsupervised techniques (PCA, FA and CCA) were efficient in the formation of clusters with log some points distributed non-uniformly depending on the particularities of these regions production processes. As for the variables that have the greatest influence on identification of regions, can stand out the acidity, esters and higher alcohols. In the application of supervised techniques we obtained an excellent rating in the validation stage with values between 83.87 and 100% focusing on the neural network showed 100% separation validation for both evaluated arrays. Later in the investigation of cachaça aged in different woods was found that the classification process using the PCA and FA of aged cachaça in the woods, Amburana, Balsam and Oak resulted in choosing the best separations that occurred with Amburana and Balsam combinations and Amburana and Oak the matrix A_1 . In the matrix A_2 satisfactory results occurred Amburana and Balsam and Amburana and oak. However due to issues related to the processing no total separation of the groups. The Correlation Analysis Canonical got high correlation, but the separation group obtained results similar to those seen in the PCA and the FA arrays A_1 and A_2 . The result of the PCA and FA was used also for carrying out the discriminant analysis (AD). The AD and Artificial Neural Networks (ANN) had, in the validation, a percentage of settlement ranging between 20.0 and 100.0%. The variables that have the greatest importance in the classification of cachaça second PCA, FA and ANN are acidity, GL and higher alcohols

Keywords: Cachaça, Chemometric Analysis, Physic-Chemical Analysis

CAPÍTULO 1

1 INTRODUÇÃO GERAL

A cachaça é uma bebida genuinamente brasileira, com 1,3 bilhões de litros anuais que representa 87% da produção nacional de bebidas destiladas, movimentando a cifra de R\$ 7 bilhões por ano em toda a sua cadeia produtiva. De acordo com o Programa Brasileiro de Desenvolvimento da Cachaça (PBDAC), o setor gera 600 mil empregos diretos e indiretos (CBRC, 2012). No ano de 2014 a Cachaça foi exportada para 66 países, por mais de 60 empresas exportadoras e um volume de 10,18 milhões de litros, gerando receita de US\$ 18,33 milhões. Esses números indicam uma expansão de, aproximadamente, 10% comparada ao ano anterior, tornando-se, assim, o 3º destilado mais consumido no mundo (IBRAC, 2015a). Ainda é possível observar esta valorização da bebida por meio da participação e premiação da cachaça brasileira em grandes concursos mundiais de bebidas destiladas, a exemplo do *Concours Mondial Spirits Selection* que ocorreu em 2014 onde 59 cachaças artesanais foram reconhecidas por sua qualidade recebendo prêmios nas categorias ouro e prata (IBRAC, 2015b).

Desde a criação do programa PROCACHAÇA, na década de 90, tem-se observado crescimento na demanda interna e, especialmente da externa, por meio do aumento das exportações, da cachaça. Essa expansão requer dos produtores uma maior preocupação e atendimento aos padrões de qualidade estabelecidos pela legislação brasileira e por órgãos internacionais. No entanto, qualidade não está associada somente aos limites mínimos e máximos destes documentos legislativos, e, portanto, se faz necessário conhecer as variáveis que possuem maior influência na qualidade físico-química e sensorial desta bebida. A partir destas informações é possível, utilizar recursos estatísticos e computacionais para classificação e identificação das cachaças, agregando valor e facilitando o processo de reconhecimento das regiões produtoras e das madeiras utilizadas no envelhecimento, a fim de atribuir selos de identidade de origem e identificar possíveis fraudes ou falsificações.

O crescimento da produção, do consumo e da exigência em relação à qualidade da cachaça, tem aumentado o estímulo da pesquisa sobre a bebida. No entanto ainda são incipientes os trabalhos na área de cachaça se comparado com outros destilados. Neste contexto o presente trabalho foi realizado com intuito de avaliar se a análise estatística multivariada e as redes neurais artificiais são capazes de gerar informações sobre as principais variáveis físico-químicas envolvidas na qualidade do processo produtivo e promover a classificação da cachaça, de acordo com bebidas de diferentes regiões e madeiras, utilizando

parâmetros físico-químicos estabelecidos pela Instrução Normativa (IN) nº 13, de 29/6/2005 e a Instrução Normativa nº 24, de 08/09/2005 do Ministério da Agricultura, Pecuária e Abastecimento.

O estudo foi organizado em três capítulos. A revisão bibliográfica do capítulo 1, de forma concisa, aborda a temática escolhida incluindo os seguintes tópicos, cachaça: definição e principais regiões produtoras; produção da cachaça; fermentação, destilação, envelhecimento, quimiometria e suas aplicações em pesquisas envolvendo cachaça, análise de componentes principais, análise de fatores, análise de correlações canônicas, análise discriminante e redes neurais artificiais.

No segundo capítulo estão descritas, de forma detalhada, as etapas das análises quimiométricas para seleção das variáveis que possuem maior associação com as regiões de Salinas e Paraty, mostrando a relação destas com as fases de produção da bebida. Apresenta-se ao leitor, ainda, os resultados de classificação das amostras de acordo com a região de produção.

O terceiro capítulo explorou as mesmas análises do capítulo supracitado para triagem de variáveis relacionadas com o envelhecimento da cachaça em diferentes tipos de madeira e apresentou a conexão destes dados com as reações que acontecem durante esta etapa do processo produtivo. Os resultados de identificação das amostras por madeira também foram apresentados.

2 REVISÃO BIBLIOGRÁFICA

2.1 Cachaça: Definição e Principais Regiões Produtoras

A aguardente é obtida do caldo da cana de açúcar, e recebe várias denominações, variando de acordo com a região do país. Aguardente de cana é definida como a bebida com graduação alcoólica entre 38% e 54% em volume, a 20 °C, obtida de destilado alcoólico simples de cana-de-açúcar ou pela destilação do suco fermentado de caldo de cana-de-açúcar, podendo ser adicionada de açúcares. Já a cachaça só pode receber essa denominação se seu teor alcoólico ficar entre 38% e 48% em volume nas mesmas condições da aguardente (BRASIL, 2005a).

A Comissão de Desenvolvimento Econômico, Indústria e Comércio aprovou a proposta que regulamenta a produção, a comercialização e a fiscalização de aguardente de cana e cachaça, entrando em vigor a Instrução Normativa nº 13, de 29 de Junho de 2005 que

estabelece os limites para os compostos secundários e contaminantes em aguardente de cana e cachaça como apresentado na Tabela 1: (BRASIL, 2005a).

Tabela 1. Teores máximo e mínimo para os componentes secundários e contaminantes presentes na aguardente de cana e cachaça

Componente	Unidade	Limite	
		Mínimo	Máximo
Graduação alcoólica (aguardente)	% v.v ⁻¹ de álcool etílico a 20 °C	38,0	54,0
Graduação alcoólica (cachaça)	% v.v ⁻¹ de álcool etílico a 20 °C	38,0	48,0
Sacarose, em açúcar refinado, invertido ou glicose	g L ⁻¹	6,0	30,0
Acidez volátil, em ácido acético	mg 100 mL a a ⁻¹	-	150,0
Ésteres, em acetato de etila	mg 100 mL a a ⁻¹	-	200,0
Aldeídos, em aldeído acético	mg 100 mL a a ⁻¹	-	30,0
Furfural	mg 100 mL a a ⁻¹	-	5,0
Alcoóis superiores*	mg 100 mL a a ⁻¹	-	360,0
Álcool metílico	mg 100 mL a a ⁻¹	-	20,0
Álcool sec-butílico	mg 100 mL a a ⁻¹	-	10,00
Álcool n-butílico	mg 100 mL a a ⁻¹	-	3,00
Congêneres**	mg 100 mL a a ⁻¹	200,0	650,0
Acroleína	mg 100 mL a a ⁻¹	-	5
Cobre	mg L ⁻¹	-	5,0
Chumbo	µg L ⁻¹	-	200
Arsênio	µg L ⁻¹	-	100
Carbamato de etila	µg L ⁻¹	-	150
Extrato seco	g L ⁻¹	-	6,0

*Alcoóis superiores = isobutílico + isoamílico + propílico

**Congêneres = acidez volátil + ésteres + aldeídos + furfural + alcoóis superiores.

Fonte: BRASIL, (2005a).

A cachaça, certamente, possui grande influência na cultura brasileira, especialmente nas cidades de Paraty e Salinas. Acredita-se que, as condições edafoclimáticas e a facilidade na construção de rodas d'água, foram elementos que transformaram Paraty no maior e melhor centro produtor da bebida durante os períodos do Brasil Colonial e Imperial. Atualmente, a cachaça de Paraty é caracterizada pela produção artesanal em alambiques com equipamentos modernos, porém mantendo as práticas e a forma de produção da tradicional cachaça de Paraty, que resultam em um destilado especial, encorpado, com um buquê que lembra o bagaço de cana e sabor agradável que promove o ardor característico da cachaça sem agredir o paladar. Os sete alambiques em atividade em Paraty são abertos à visitaç o e a produç o anual gira em torno de 300.000 litros, rotulados em quinze marcas. (C MARA, 2014)

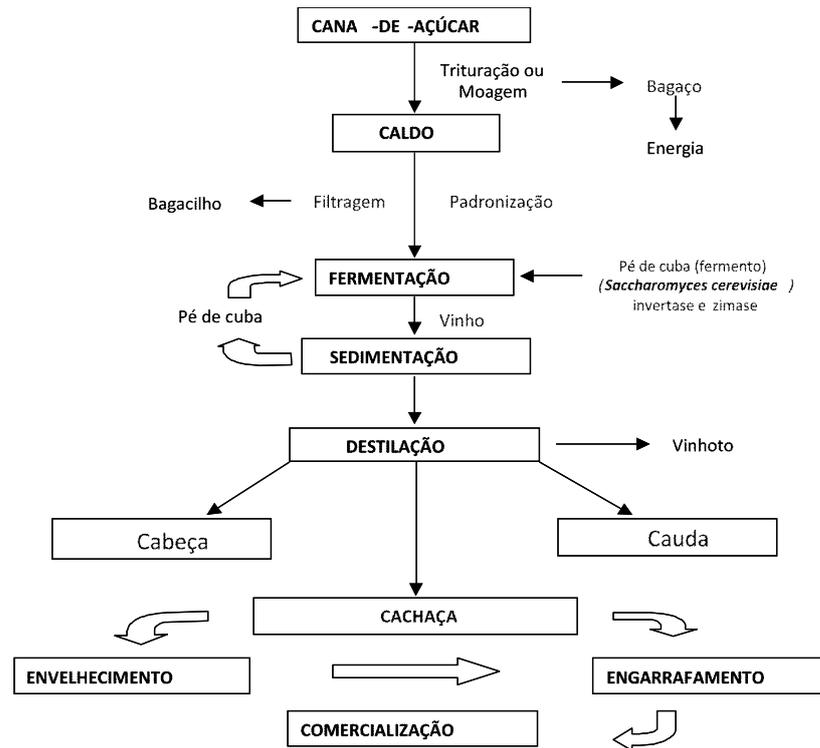
Por sua vez, o município de Salinas é conhecido como o mais importante centro de produção de cachaça de alambique em Minas Gerais. A cachaça começou a ser produzida no município com a vinda dos primeiros fazendeiros para a região, seguindo os rastros da pecuária, cujos primeiros rebanhos bovinos vieram da Bahia, porém era produzida como atividade complementar à atividade pecuária. As boas perspectivas para a cachaça de Salinas tiveram início a partir das décadas de 40 e 50. Nessa época, algumas marcas de cachaça começaram a ser produzidas em Salinas, como a Piragybana, Havana entre outras. Tais marcas foram prenunciadoras de outras, que foram produzidas a partir do lançamento do PROCACHAÇA (Programa Mineiro de Incentivo de Produção à Aguardente), pelo governo do Estado, em 1992. A cachaça de Salinas é comercializada com aproximadamente 60 marcas em todo o país e no exterior (SEBRAE, 2001). O setor produtivo da cachaça em Salinas responde por 1/3 da arrecadação do ICMS. A microrregião de Salinas possui 45 produtores de cachaça formalizados (inscritos na Receita Federal, Receita Estadual e Ministério da Agricultura) e em 2012 faturaram 26,8 milhões de reais, 9,24% a mais em relação a 2011. (R\$ 26.680.176,36) (SANTIAGO, 2006; SANTIAGO, 2013).

Os produtores consideram que as características do clima, solo e localização geográfica favorecem o desenvolvimento de uma variedade de leveduras selvagens especiais, responsáveis pela qualidade e sabor das cachaças produzidas nesta região e o estudo conduzido por Barbosa (2013), mostrou diferenças entre as linhagens da microrregião de salinas de outras regiões de Minas Gerais, o que, segundo a autora constitui-se numa estratégia adicional para o requerimento de Indicação Geográfica do tipo Denominação de Origem.

2.2 Produção da Cachaça

O processo de produção da cachaça de alambique tem como única matéria prima a cana de açúcar, que resumidamente passa pelo processo de moagem para obtenção de caldo, fermentação e destilação, como mostra a descrição detalhada no fluxograma da figura 1. Na sequência estão descritas considerações sobre as etapas que mais influenciam na qualidade da bebida, a saber, fermentação, destilação e envelhecimento

Figura 1. Fluxograma da Cachaça de Alambique



Fonte: DUARTE, 2013

2.2.1 Fermentação

A fermentação alcoólica é o processo químico desencadeado por leveduras as quais convertem os açúcares do caldo de cana em álcool etílico, gás carbônico e componentes secundários, tais como: ácidos carboxílicos, metanol, ésteres, aldeídos e alcoóis superiores que contribuem para a determinação das propriedades sensoriais da cachaça. Para que a fermentação ocorra em sua total plenitude é necessário controlar algumas condições como preparo do mosto, temperatura, pH, vitaminas e minerais, entre outros fatores (SCHWAN e CASTRO, 2006).

No início da fermentação, quando o açúcar disponível para as leveduras é a sacarose, esta é quase imediatamente hidrolisada à glicose e frutose, pela ação da enzima invertase. A maioria das linhagens de leveduras são glicolíticas, ou seja, fermentam a glicose em taxas mais rápidas que a frutose e outros monossacarídeos (SCHWAN e CASTRO, 2006).

Vale ressaltar que a molécula de glicose passa pelo processo anaeróbico mediante doze etapas intermediárias antes de ser transformada em etanol e gás carbônico. Ao lado do etanol e do gás carbônico formam o acetaldeído glicerol, 2,3 butilenoglicol, ácido succínico e ácido cítrico como produtos constantes da fermentação alcoólica, o que contribui para o sabor e para o aroma da cachaça. Os ésteres, como o acetato de etila, são também formados pela esterificação de ácidos orgânicos fixos como o ácido tartárico e málico também catalisados pelas enzimas da levedura. O álcool etílico e o dióxido de carbono são produtos de excreção da célula e não são utilizados pela levedura em condições anaeróbicas (DUARTE, 2013; SCHWAN; CASTRO, 2006).

O processo de fermentação artesanal pode ocorrer de forma espontânea ou, mais raramente, por meio do uso de leveduras selecionadas. O uso destas linhagens tem aumentado a produtividade e melhorado a qualidade da bebida em muitos alambiques, pois diminui os riscos de contaminações e, assim, reduz os componentes indesejáveis no produto final, além de promover a melhoria dos parâmetros físico-químicos e da qualidade do vinho obtido, principalmente em relação aos teores de acidez e concentração de alcoóis superiores (GOMES et al, 2009).

2.2.2 Destilação

O vinho contém, além da água e do álcool etílico, muitas substâncias, como aldeídos, ácidos, ésteres e alcoóis superiores. Também estão presentes alguns sólidos na sua composição, como os bagacilhos, leveduras, bactérias e, até mesmo, algum gás carbônico que não saiu para a atmosfera (DIAS, 2006).

Na aguardente é desejada uma concentração alcoólica, fixada por lei, entre 38 e 54% v.v⁻¹. Porém, na prática comercial estabelece 49 a 54% v.v⁻¹. Essa elevação do teor alcoólico é conseguida pela destilação. O processo baseia-se no aquecimento do vinho, cujos vapores serão mais ricos de substâncias voláteis. O equipamento usado para isso é a coluna de destilação nas grandes indústrias e do alambique de cobre utilizado pelos pequenos produtores artesanais. Nessa etapa de destilação, para a cachaça, o vinho é descarregado diretamente na panela do alambique (DIAS, 2006). Nesse processo, há o fracionamento das partes, separando a cabeça, o corpo e a cauda.

Estas frações são separadas em função do grau alcoólico: o primeiro corte é destinado a separar a cabeça (grau alcoólico de 70 a 50 % v.v⁻¹), o segundo corte separa a fração coração (50 a 38 % v.v⁻¹) e o restante é a cauda (38 a 14 % v.v⁻¹), sendo o coração a fração principal (aguardente ou cachaça), a depender do grau alcoólico.

Para definir o ponto de corte ideal entre as três frações do destilado, deve-se equacionar o perfil qualitativo e quantitativo dos alcoóis, ésteres, ácidos graxos e aldeídos, com especial atenção para o conteúdo de metanol (devido à sua toxicidade). No entanto, o corte é realizado, ainda, de forma totalmente empírica, baseado na experiência do mestre alambiqueiro e nas concentrações de alcoóis, avaliada por medidas de densidade. Esta falta de conhecimento sobre o perfil da bebida no momento do corte resulta, muitas vezes, em uma fração coração rica em metanol que pode provocar a intoxicação dos consumidores (MAIA, 2000; SILVA et al, 2000; RECHE et al, 2007).

2.2.3 Envelhecimento

A cachaça recém-destilada, apresenta qualidade sensorial para o consumo imediato, no entanto, o envelhecimento em barris de madeira é uma etapa fundamental no processo de fabricação dessa bebida, em função da efetiva contribuição que representa para sua qualidade sensorial (BORRAGINI, 2009). O envelhecimento é, portanto, uma etapa indispensável, quando se deseja agregar qualidade e, conseqüentemente, valor a uma bebida destilada.

A aguardente de cana recém destilada é incolor e apresenta sabor picante, áspero e seco por melhor que tenha sido conduzido o processo de fermentação e por mais cuidadosa que tenha sido a destilação. O produto final, mesmo apresentando as características em conformidade com as especificações legais, tem problemas de aceitação sensorial, devido ao forte sabor alcoólico e ao perfil químico do conjunto de componentes secundários recém destilados (MAÇATELLI, 2006; CARDELLO e FARIA, 1998).

Durante o envelhecimento ocorrem inúmeras transformações, incluindo as reações entre os compostos secundários provenientes da destilação, a extração direta de componentes da madeira, a decomposição de algumas macromoléculas da madeira (lignina, celulose e hemicelulose) e sua incorporação à bebida, as reações de compostos da madeira com os componentes originais do destilado e, ainda, a evaporação de compostos voláteis (PIGGOTT E CONNER, 2003; MOSEDALE e PUECH, 1998). Como resultado tem-se a diminuição do pH e das concentrações de cobre e de álcool metílico, enquanto se observam aumentos da acidez, da cor e das concentrações de ésteres, aldeídos, furfural, álcoois superiores, coeficiente de congêneres e compostos fenólicos (PARAZZI et al., 2008; MIRANDA et al., 2008).

No estudo de Chatonnet (2003) as análises sensoriais e físico-químicas mostraram diferenças importantes entre as espécies de madeira de diferentes regiões geográficas, que tipificam as notas de madeira desenvolvidas na cachaça, devido à extração de compostos

peculiares. Portanto, espécies de madeira, a sua região geográfica, idade madeira e manejo florestal são parâmetros relevantes na escolha do barril, uma vez que eles interagem para definir a qualidade da madeira e, conseqüentemente, o perfil sensorial e químico da cachaça envelhecida.

2.3 Quimiometria e suas aplicações em pesquisas envolvendo cachaça

Após a obtenção da cachaça e a realização das análises físico-química se faz necessário a análise dos dados obtidos. Esta é uma parte essencial em todo experimento, sendo univariada quando somente uma variável é medida sistematicamente para várias amostras. Há muito tempo a estatística univariada vem sendo aplicada a problemas químicos, mas a sua utilização tornou-se limitada em estudos que pretendiam avaliar, simultaneamente, muitas variáveis. Assim, a análise multivariada foi introduzida no tratamento de dados químicos, ganhando rapidamente popularidade e dando origem a uma nova disciplina, batizada de Quimiometria (SENA et al, 2000).

Assim, a tendência atual das pesquisas científicas é a utilização do raciocínio multivariado, principalmente devido ao desenvolvimento tecnológico atingido nos últimos tempos. Neste sentido, estas técnicas estão sendo muito utilizadas para determinar a origem geográfica e, ainda, caracterizar e controlar produtos alimentícios. Outro aspecto importante é a representação relativamente simples de conjuntos de dados complexos. Esta aparente simplicidade é importante para compreensão adequada das relações existentes entre as inúmeras variáveis que podem ser analisadas em um alimento (PATRAS et al, 2011)

Quando utilizado para analisar dados instrumentais, a quimiometria muitas vezes resulta em uma avaliação mais precisa e mais rápida da composição de um produto alimentar, ou até mesmo das propriedades físicas e sensoriais. Por exemplo, a composição (gordura, fibra, umidade, carboidratos) dos produtos lácteos ou de grãos pode ser rapidamente medida usando espectroscopia no infravermelho próximo e quimiometria. As propriedades dos alimentos (por exemplo, o gosto, o cheiro, a adstringência) também podem ser monitoradas continuamente. Em todos os casos, são utilizados padrões dos dados para desenvolver um modelo com o intuito de prever parâmetros de qualidade em dados futuros (INFOMETRIX, 1996).

Alguns autores têm utilizado estas ferramentas para classificação de cachaças, a exemplo de Fernandes e colaboradores (2005) que propuseram a identificação da origem geográfica e do modo de fabricação através da aplicação das análises de componentes principais (ACP) e agrupamento hierárquico (AH) para reconhecimento de padrões utilizando

dados de composição mineral. Os resultados mostraram que a categorização entre cachaça industrial e artesanal não é fácil, apesar de existir diferenças em relação a alguns minerais. Os compostos Al, Ca, Cu, Fe, K, Mg, Mn, Na, Pb, S, Se, Si, Sn, Sr, e Zn foram utilizados para classificação dos grupos de acordo com a procedência de origem de produção e a análise preliminar não foi bem-sucedida. No entanto, quando o número de classes foi reduzido para três, de acordo com as principais regiões geográficas (Nordeste, Centro-Oeste e Sul do Brasil), uma discriminação razoável entre os grupos foi obtida. A separação por classes relacionou os minerais com o tipo de solo em que a cana-de-açúcar foi cultivada, e o Nordeste foi o grupo que obteve melhor separação.

Outro estudo foi realizado para investigar a correlação entre compostos químicos e propriedades sensoriais de cachaças através do uso de ACP e análise discriminante (AD). Os resultados foram satisfatórios para AD, apresentando-se como uma ferramenta alternativa potencial para o reconhecimento das qualidades sensoriais de cachaças (SERAFIM et al, 2013). Baseado apenas no perfil sensorial por Análise Descritiva Quantitativa (ADQ), Cardello e Faria (1998) reportaram uma separação efetiva entre amostras de cachaças envelhecidas e não envelhecidas, por análise de componentes principais.

Adicionalmente, De Souza et al (1997) reportaram a possibilidade de separação de cachaças não apenas pelo tempo de envelhecimento, mas também pelo tipo de madeira utilizada na confecção dos barris de envelhecimento. Os autores reportaram uma separação bastante efetiva, por análise de componentes principais, de cachaças envelhecidas em barris de umburana e jequitibá, através de dados de espectrofotometria de massa.

Outra explanação de técnicas multivariadas usando os parâmetros analíticos foi abordada por Granato e colaboradores (2014) que avaliou a qualidade de cachaças comerciais brasileiras. A ACP não mostrou resultados positivos, no entanto a análise de agrupamento foi eficiente na separação de amostras produzidas por alambiques de diferentes regiões e a AD foi adequada para algumas amostras.

Nesta mesma linha de pesquisa há estudos que foram feitos para avaliar a distinção entre cachaças destiladas em alambique e coluna. A comparação feita entre análise discriminante com calibração multivariada por mínimos quadrados parciais e análise de fatores evidenciou que a AF conduziu a melhores resultados de explicação da variância do conjunto de dados. Já a AD permitiu a elaboração de um modelo com porcentagem de confiança na predição de amostras desconhecidas de 93,9% após a validação cruzada (RECHE e FRANCO, 2009).

2.3.1 Análise de Componentes Principais (ACP)

A análise de componentes principais é uma das técnicas multivariadas mais importantes e utilizadas, uma vez que, dentre outros fatores, constrói a base onde se fundamentam a maioria dos outros métodos multivariados de análise de dados (LYRA et al., 2010).

Essa técnica fundamenta-se essencialmente na transformação do conjunto original de variáveis em outro, os componentes principais (CP), por meio das combinações lineares das variáveis originais (PEREIRA e PEREIRA, 2004). Esse novo conjunto formado apresenta dimensão equivalente ao conjunto original, ou seja, se tivermos analisando um conjunto de p variáveis relacionadas, teremos p componentes principais que devem ser mutuamente não correlacionados, por consequência da ortogonalidade (DA SILVA e PADOVANI, 2006; FERREIRA, 2011).

Cada componente principal gerado contém uma quantidade máxima de variação em relação aos dados originais, que não são explicadas pelos demais componentes, uma vez que estes não apresentam correlação. Assim, as informações totais dos dados originais estão contidas nos componentes principais gerados, onde cada componente retém uma porcentagem da variância original (BORGOGNONE et al., 2001; PACHECO et al., 2013).

Os componentes principais são obtidos em ordem decrescente de máxima variância. Dessa forma, o componente principal 1 é a combinação linear de maior variância, detendo mais informação dos dados originais que o componente principal 2, que por sua vez tem mais informação que o componente principal 3, e assim sucessivamente (NETO e MOITA, 1998; BORGOGNONE et al., 2001; FERREIRA, 2011).

Dessa forma, pode-se sumarizar as principais propriedades dos componentes principais como: cada componente principal é uma combinação linear de todas as variáveis originais, são independentes entre si e estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos da variação total contida nos dados (MORRISON, 1976; JOHNSON e WICHERN, 1999).

Existem duas formas de obtenção dos componentes principais: por matriz de covariância (S) ou por matriz de correlação (R) (FERREIRA, 2011; BORGOGNONE et al., 2001). Assim, para o cálculo dos componentes principais a primeira etapa consiste em transformar a matriz de dados originais de n amostras (linhas) por p variáveis (colunas), em uma matriz de covariância ou de correlação. Qualquer uma das duas matrizes é de ordem $p \times p$ (BORGOGNONE et al., 2001).

Ao se realizar uma análise de componentes principais, espera-se que quase a totalidade das informações das variáveis originais esteja guardada nos primeiros componentes principais. Assim, promovendo uma simplificação estrutural dos dados, ao mesmo tempo em que preserva ao máximo as informações originais (PEREIRA e PEREIRA, 2004; PACHECO et al., 2013). Esse tipo de simplificação facilita a realização de inferências a respeito de algum fenômeno, sobre todo o grupo de variáveis (FERREIRA, 2011).

No entanto, um dos principais problemas na análise de componentes principais (ACP) é a estimativa do número correto de componentes. A escolha desta dimensionalidade é um passo crucial para a interpretação dos resultados ou análises posteriores, uma vez que, poderá levar a uma perda de informação (subestimação) ou a introdução de ruído aleatório (superestimação) (DRAY, 2008).

Segundo Da Silva e Padovani (2006), dentre os principais critérios encontrados na literatura para escolha do número de componentes principais, destacam-se: Diagrama de Autovalores, Critério de Kaiser e Fatores Interpretáveis. A escolha por um destes critérios é uma decisão do pesquisador e, portanto, pode levar a resultados diferentes, o que torna a análise, em partes, subjetiva.

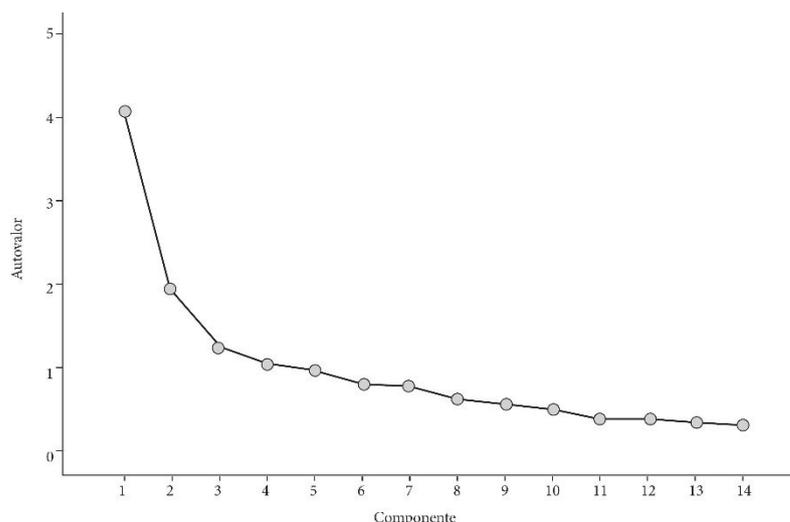
O diagrama de autovalores consiste na utilização de um gráfico denominado *scree plot* (Figura 2), em que é postada a ordem dos k componentes principais na abscissa e o seu autovalor estimado na ordenada (FERREIRA, 2011).

A interpretação do gráfico é baseada nas alterações das variâncias provocadas pelos componentes principais, ou seja, o gráfico é uma representação da porcentagem da variância explicada por componente.

Quando um componente principal não provoca alterações expressivas na variância, o segmento de reta atinge um patamar onde o coeficiente angular se aproxima de zero, ponto de estabilização, não havendo alterações significativas nos valores de suas variâncias (FERREIRA, 2011). Assim, o número de componentes principais escolhidos seria igual ao número de autovalores anteriores ao ponto de estabilização (MINGOTI, 2005).

Na figura 2 a estabilização da curva ocorre a partir do quarto autovalor. Dessa forma, os autovalores anteriores ao ponto de estabilização, 1, 2 e 3, são os escolhidos para representar a variância total dos dados, pelo gráfico de *scree plot* (OGILVIE et al., 2008).

Figura 2. *Scree plot* dos autovalores associados com os componentes principais.



Fonte: OGILVIE et al. (2008).

No critério de Kaiser, quando utilizado uma matriz de correlação, o número de componentes principais retidos deve ser igual ao número de autovalores maiores que 1 (MINGOTI, 2005; LEDESMA e VALERO-MORA, 2007).

Quando a análise é realizada com matriz de covariâncias a determinação do número de componentes pode ser baseada na média aritmética dos autovalores estimados da matriz de covariância. Assim, selecionam-se os componentes relacionados aos autovalores que são maiores ou iguais à média aritmética (FERREIRA, 2011 e MINGOTI, 2005).

Por sua vez, a técnica de fatores interpretáveis é um critério empírico, que considera conservar os componentes que podem representar com maior fidelidade às respostas do problema. Trata-se de um critério subjetivo, onde observa-se a quantidade de variância acumulada e um nível de corte é estabelecido, representando o total da variância contabilizada pelos componentes principais (DA SILVA e PADOVANI, 2006). Segundo autores citados por Da Silva e Padovani (2006) e Ferreira (2011), em geral, adota-se um valor mínimo como ponto de corte, entre 70% e 90% de explicação da variação dos dados.

Ainda é importante salientar que boa parte das pesquisas é fundamentada na avaliação de um grande número de p variáveis, muitas vezes de difícil interpretação. Todavia, um número menor de variáveis pode deter a mesma quantidade de informação contida no conjunto total (FERREIRA, 2011; PACHECO et al., 2013). Dessa forma, a técnica de componentes principais, além proporcionar o resumo das informações contidas no complexo de variáveis, permite a eliminação de variáveis com pouca contribuição, em termos de variação, no grupo avaliado (PEREIRA e PEREIRA, 2004).

2.3.2 Análise de Fatores (AF)

A análise de fatores e de componentes principais, apesar de algumas similaridades, são duas abordagens diferentes para analisar a estrutura de correlação de um conjunto de valores reais de variáveis aleatórias. Na prática, o resultado da análise de fatores torna-se, comumente, muito próximo daqueles da análise de componentes principais (SCHNEEWEISS e MATHES, 1995).

No entanto, existem diferenças básicas entre essas duas técnicas multivariadas (RENCHEER, 2002):

- A análise fatorial expressa as variáveis originais como uma combinação linear dos fatores, enquanto que, os componentes principais são uma combinação linear das variáveis originais e os seus componentes não são correlacionados.
- Na análise fatorial busca-se explicar as covariâncias ou correlações entre as variáveis. Todavia, com a análise de componentes principais propõe-se explicar grande parte da variância total das variáveis originais.

A análise fatorial é uma técnica estatística multivariada para simplificação de um conjunto de dados complexos, pela representação do conjunto de variáveis, em termos de um pequeno número de variáveis subjacentes (hipotéticas ou não observáveis), conhecida como fatores ou variáveis latentes (JAIN e SHANDLIYA, 2013).

Em geral, espera-se que o número de variáveis latentes (fatores) seja menor que o número das variáveis originais, e que estas novas variáveis criadas representem totalmente ou parcialmente a variação das observações originais (DUNCAN, 1968).

A base estabelecida para uma Análise Fatorial é a correlação (ou covariância) entre as medidas observadas. Assim, as medidas que sejam altamente correlacionadas (positiva ou negativamente) são, provavelmente, influenciadas pelos mesmos fatores, enquanto que, aquelas não correlacionadas, influenciadas por fatores diferentes (DECOSTER, 1998).

Nesse sentido, a análise fatorial pode ser compreendida como uma técnica que busca analisar as relações entre as variáveis observadas e, a partir destas estimar um modelo fatorial subjacente que reproduzam essas relações (ARANHA e ZAMBALBI, 2008).

Outra característica da análise é sua capacidade de determinar o grau de influência de uma variável na explicação de um fator (posteriormente descoberto). Nesse sentido, os coeficientes (*loadings*) relacionados com cada fator para cada variável, representam a correlação entre o fator e a variável (MINGOTI, 2005).

Existem duas principais modalidades de análise fatorial: exploratória e confirmatória. Na análise exploratória o pesquisador não tem noção do número ou da natureza das variáveis,

ou seja, busca encontrar os fatores relacionados às variáveis originais. Por sua vez, na análise confirmatória o pesquisador parte de um modelo fatorial pré-definido e deseja verificar se é aplicável (DECOSTER, 1998; MINGOTI, 2005; WILLIAMS et al., 2010).

Willians et al. (2010) reporta em seu artigo de revisão que os principais critérios utilizados para escolha do número de fatores, são: 1. Número de autovalores maior ou igual a um (Critério de Kaiser); 2. O gráfico *scree plot* (Teste *scree*) e 3. Variância Percentual Acumulada. Todavia, os autores relatam que a utilização simultânea de vários critérios para escolha dos fatores é mais adequada e muitas vezes desejável, dada a natureza complexa da análise de fatores.

2.3.3 Análise de Correlações Canônicas (ACC)

A análise de correlações canônicas é uma técnica exploratória que resume grandes vetores de variáveis em um número menor delas através da generalização da análise de regressão múltipla. Pode-se dizer que esta análise tem como objetivo principal o estudo das relações lineares existentes entre dois conjuntos de variáveis. A ideia básica é resumir a informação de cada conjunto de variáveis respostas em combinações lineares, sendo que a escolha dos coeficientes dessas combinações é feita tendo-se como critério a maximização da correlação entre os conjuntos de variáveis respostas. As combinações lineares que podem ser construídas são chamadas de variáveis canônicas, enquanto que a correlação entre elas é chamada de correlação canônica. Essa correlação mede basicamente o grau de associação existente entre os dois conjuntos de variáveis para estabelecer poucos pares de variáveis canônicas (FERREIRA, 2011; MINGOTI, 2005; VAINIONPAA et al., 2000).

O primeiro par de variáveis canônicas são as combinações lineares, com variância unitária que maximiza a correlação. Por sua vez, o segundo par de variáveis canônicas são as combinações lineares, com variância unitária, que maximiza a correlação entre todas as escolhas que não são correlacionadas com o primeiro par de variáveis canônicas. E assim, até a k-ésima variável canônica (LIRA, 2004).

Este conhecimento da inter-relação entre os conjuntos de variáveis pode propiciar ao pesquisador identificar possíveis estruturas de dependência entre os diferentes conjuntos que não seriam facilmente identificáveis pela simples inspeção das correlações entre os pares de variáveis originais (FERREIRA, 2011).

2.3.4 Análise Discriminante (AD)

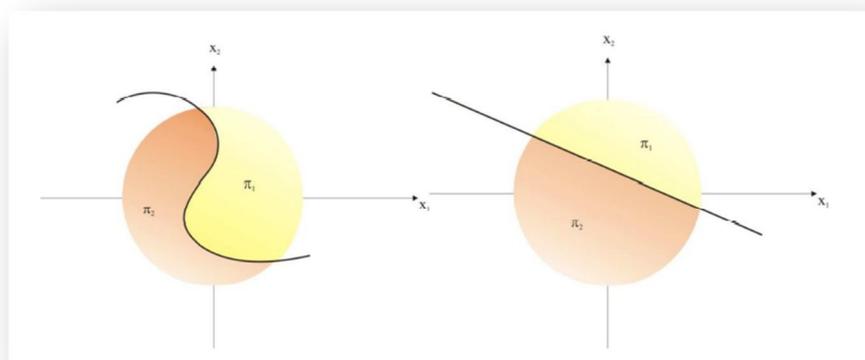
A análise discriminante (AD) é uma técnica estatística para classificar um conjunto de observações em classes pré-definidas (JIN e AN, 2011).

Este método é utilizado quando se quer estabelecer relações entre uma variável dependente e variáveis independentes e tem como objetivo criar índices que diferenciam ou discriminam os grupos, permitindo assim, a identificação de grupos similares e a classificação de novos casos, onde a inserção é feita no grupo que o caso tem maior probabilidade de pertencer, com base nas características do novo objeto. Assim, é possível obter uma função discriminadora, que apresente maior poder de discriminação entre os grupos, por meio de combinações lineares de variáveis independentes, a partir das quais seja possível classificar os elementos em cada uma das categorias da variável dependente (HAIR JR et al., 2010; MAROCO, 2003; PESTANA e GAGEIRO, 2000).

Essa função tem, ainda, a propriedade de minimizar as probabilidades de má classificação, quando as populações são normalmente distribuídas com média μ e variância Σ conhecidas e os grupos podem ser visualizados em um espaço multidimensional, muitas vezes menor que o original. Os coeficientes das funções discriminantes indicam a contribuição das variáveis originais para cada função discriminante (REGAZZI, 2000; SCHEEREN, 2000).

A função discriminante obtida é definida como uma fronteira que separa regiões de alocação, onde o conjunto de valores é separado em grupos. A Figura 3 ilustra regiões de alocação para o caso de duas populações (REGAZZI, 2000).

Figura 3. Representação e fronteiras de decisão, não-linear (esquerda) e linear (direita).



Fonte: REGAZZI, 2000

A extração das características representa uma importante etapa de pré-processamento para a análise de dados estatísticos multivariados. A extração de características para

classificação alcança a redução de dimensionalidade por meio da maximização de uma função objetivo adequada, preservando ou aumentando a separabilidade da classe. Este processo é geralmente referido como extração da característica discriminante. Deste modo, a seleção das variáveis a ser utilizada é um aspecto muito importante para a obtenção de uma boa classificação. A utilização de toda a informação disponível conduz em muitas situações problemáticas, como por exemplo, uma maior sobreposição nos grupos (MAROCO,2003).

A AD envolve a determinação de uma equação de regressão linear, que é utilizada para prever o grupo que a amostra em estudo pertence. A forma da equação é:

$$D = v_1X_1 + v_2X_2 + \dots + v_iX_i + a \quad (1)$$

Onde: D é a função discriminante; v é o coeficiente discriminante ou o peso atribuído a cada variável; X é o escore das variáveis estudadas; a é uma constante; i é o número de variáveis preditoras.

A discriminação é conseguida estabelecendo-se os pesos da variável dependente para cada variável independente (v), com o objetivo de maximizar a variância entre os grupos em relação à variância dentro dos grupos.

Definida a função discriminante procede-se a análise e classificação dos dados nos grupos. O procedimento da classificação consiste em delimitar um valor crítico de um índice Z (linha ou plano separatriz), de tal forma que, se um valor situa-se abaixo desse índice é classificado em um grupo, e se acima do mesmo, em outro. Porém, pode-se encontrar uma observação fora da região correspondente ao grupo definitivo podendo induzir a erros de classificação, os quais serão menores quanto mais diferentes forem os grupos (FLORENTINO, 2000).

A análise discriminante linear, também conhecido como o critério de Fisher, é a técnica mais conhecida desta classe (DAS e NENADIC, 2008) e se caracteriza por maximizar a razão entre as somas dos quadrados entre grupos e a soma de quadrados dentro dos grupos (MINGOTI, 2005). A função discriminante linear de Fisher é, portanto, uma combinação linear das variáveis originais e se caracteriza por produzir separação máxima entre duas populações.

Assim define-se a função discriminante de Fischer pela expressão:

$$D(x) = L' \cdot x = [x_1 - \bar{x}_2]' \cdot S_c^{-1} \cdot x \quad (2)$$

Onde,

D(x) = função discriminante linear amostral de Fischer;

L' = estimativa do vetor discriminante;

\bar{x}_1 = média amostral da população π_1 ;

\bar{x}_2 = média amostral da população π_2 ;

Adotando-se que as populações π_1 e π_2 têm a mesma matriz de covariância Σ podemos então estimar uma matriz comum de covariâncias denominada S_c . Assim, a função discriminante de Fisher é uma combinação linear do conjunto de caracteres que possibilita classificar um novo indivíduo, com um vetor de observações em certa população com máxima probabilidade de acerto.

A classificação de um novo indivíduo baseia-se na seguinte regra:

Classificar o novo x_0 em π_1 se $D(x_0) = [\bar{x}_1 - \bar{x}_2]' \cdot S_c^{-1} \cdot x_0 \geq m$ ou

Classificar o novo x_0 em π_2 se $D(x_0) = [\bar{x}_1 - \bar{x}_2]' \cdot S_c^{-1} \cdot x_0 \leq m$

Sendo m o ponto médio entre as duas médias amostrais das populações \bar{x}_1 e \bar{x}_2

Atualmente a técnica de análise discriminante é abrangente em todas as áreas de estudos, sendo uma dessas a área de Ciência e Tecnologia de Alimentos, visando a discriminação e classificação de produtos com relação as suas variáveis analisadas.

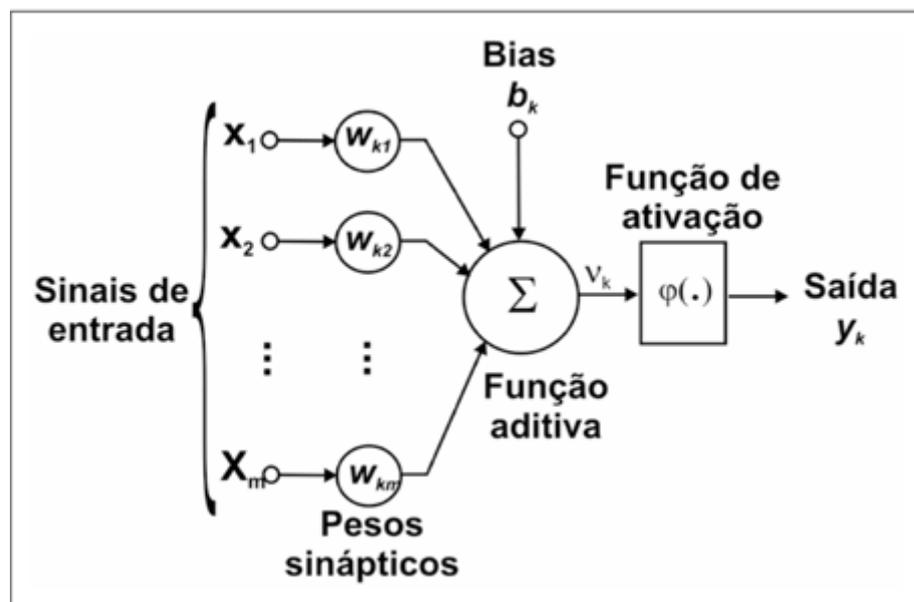
2.4 Redes Neurais Artificiais (RNA)

As Redes Neurais Artificiais (RNA) é um método capaz de capturar relações funcionais entre variáveis de entrada e saída a partir de uma base de dados, representando uma ferramenta valiosa para os sistemas de processamento de alimentos, envolvendo estimativa, previsão, classificação e otimização (EIM et al., 2013). Uma RNA é uma estrutura interconectada composta por muitos elementos de processamento simples, os neurônios, capazes de realizar cálculos em paralelo para o processamento de dados (BAS e BOYACI, 2007). Com efeito, a RNA pode lidar com múltiplas variáveis dependentes e independentes ao mesmo tempo, sem o conhecimento prévio da relação funcional que possa existir entre elas (ZHOU e THERDTHAI, 2010).

Inspiradas no funcionamento do neurônio biológico típico, as RNA são técnicas computacionais que utilizam modelos matemáticos para a classificação de padrões, simulação de atividades humanas, agrupamento de dados, previsões temporais, entre outras aplicações (ZUBEN, 2003 e KOVÁCS, 2002). Uma importante característica das RNA é a habilidade de aprender a partir de casos reais conhecidos, o que proporciona o aprimoramento de seu desempenho. Esse aprendizado é feito por um processo iterativo e ocorre quando a RNA atinge uma solução generalizada para uma classe de problemas previamente apresentados a ela (BRAGA et al., 2007).

A RNA é composta de camadas de elementos de processamento, os quais são denominados de neurônios artificiais (ARTEAGA, 1993). Os neurônios artificiais recebem os sinais e os passam às camadas intermediárias, por meio de conexões matemáticas. Cada neurônio processa uma parte das informações, o que ocorre em paralelo, e automaticamente as distribui através das diferentes unidades que compõem a rede. As unidades de processamento transferem as somas ponderadas de informações por uma função de ativação, geralmente uma função sigmoideal, que contém um grande número de sinais informativos, os quais são convertidos a valores definidos durante a operação de treinamento e teste (Figura 4) (EERIKAINEN, et al., 1993; HORIMOTO, et al., 1995; KATZ, et al., 1992; LAWRENCE e PETERSON, 1993).

Figura 4. Funcionamento de um neurônio computacional



Fonte: MARTINS-FILHO et al, 2005.

Há uma série de algoritmos de aprendizagem utilizados na composição de redes neurais, os quais operam de maneira diferente. Além disso, as redes diferem também em sua topologia, ou organização dos neurônios. Redes de multicamadas que se utilizam do algoritmo *Backpropagation*, no qual o aprendizado se dá por um processo de retropropagação do erro, são atualmente as mais difundidas (BLOCK, 1997 e MENDES, 1999).

Durante a fase de treinamento, os valores previstos são subsequentemente comparados com o valor real e, as conexões entre as unidades de processamento, com seus devidos pesos, são modificadas para minimizar o desvio entre os valores. A rede mede o erro entre o valor obtido e o valor desejado, usando o algoritmo para ajustar os pesos das

interconexões, reduzindo o erro nas determinações subsequentes. Pela utilização desta estrutura de rede, o programa continua a fazer alterações nas conexões internas, através do processo de tentativa e erro, até atingir um determinado nível de certeza (KATZ, 1992 e WYTHOFF, 1990).

3 OBJETIVOS

3.1 OBJETIVO GERAL

Avaliar se a análise estatística multivariada e as redes neurais artificiais são capazes de gerar informações sobre as principais variáveis físico-químicas envolvidas na qualidade do processo produtivo e promover a classificação da cachaça, de acordo com bebidas de diferentes regiões e madeiras, utilizando parâmetros físico-químicos estabelecidos pelas Instruções Normativas (IN) nº 13, de 29/6/2005 e a nº 24, de 08/09/2005 do Ministério da Agricultura, Pecuária e Abastecimento.

3.2 OBJETIVOS ESPECÍFICOS

- 3.2.1 Obter dados físico-químicos de amostras de cachaça das regiões de Paraty e Salinas;
- 3.2.2 Aplicar a técnica de ACP, para escolha dos componentes principais que melhor represente o conjunto de dados, com a consequente separação das cachaças através do agrupamento das amostras que possuem maior similaridade;
- 3.2.3 Utilizar a AF, para determinação dos fatores que expresse adequadamente o conjunto de dados, e como consequência separar as cachaças, por meio de dispersão gráfica que represente os escores dos fatores selecionados.
- 3.2.4 Empregar a análise de correlações canônicas, para o estudo das relações lineares entre as variáveis da fração alcoólica e as variáveis da fração não alcoólica da cachaça;
- 3.2.5 Realizar a análise discriminante para classificação das cachaças e obtenção de um modelo que permita prever a separação de novas amostras;
- 3.2.6 Criar uma Rede Neural Artificial para classificar as cachaças de acordo com a origem e o tipo de madeira de envelhecimento.

4 REFERÊNCIAS

- ARANHA, F.; ZAMBALDI, F. **Análise Fatorial em administração**. São Paulo: Cengage Learning, 2008.
- ARTEAGA, G.E.; NAKAI, S. Predicting protein functionality with artificial neural networks: foaming and emulsifying properties. **Journal of Food Science**, v. 58, p. 1152-1156, 1993.
- BAŞ, D.; BOYACI, İ.H. Modeling and optimization II: Comparison of estimation capabilities of response surface methodology with artificial neural networks in a biochemical reaction. **Journal of Food Engineering**, v. 78, n. 3, p. 846–854. 2007.
- BLOCK, J.M. **Formulação de Gorduras Hidrogenadas através de Redes Neurais**. 1997. 115f. Tese (Doutorado em Tecnologia de Alimentos) – Universidade Estadual de Campinas, Campinas, SP.
- BORGOGNONE, M. G.; BUSSI, J.; HOUGH, G. Principal component analysis in sensory analysis: covariance or correlation matrix? **Food Quality and Preference**, v.12, p. 323–326, 2001.
- BORRAGINI, M. C. C. **Envelhecimento da cachaça com circulação forçada e aeração**. 2009. 90f. Dissertação (Mestrado em Ciência dos Alimentos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”. Faculdade de Ciências Farmacêuticas, Araraquara, SP.
- BRAGA, A. P, CARVALHO APLF, LUDEMIR TB. **Redes neurais artificiais: teoria e aplicações**. 2ª ed. Rio de Janeiro: Livro Técnico e Científico; 2007.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa nº 13, de 29 de junho de 2005a**. Aprova o regulamento técnico para fixação dos padrões de identidade e qualidade para aguardente de cana e para cachaça. Diário Oficial da República Federativa do Brasil. Brasília, DF, 30 Jul. 2005 Disponível em: <<http://www.agricultura.gov.br/>>. Acesso em: 04 de mai. 2015.
- CÂMARA, M. **A Indicação de Procedência da Cachaça de Paraty**. 2014. Disponível em: <<http://www.ilhaverde.net/paraty.htm>> Acesso em: 11. Jun. 2015
- CARDELLO, H. M. A. B.; FARIA J. B. Análise descritiva quantitativa da aguardente de cana durante o envelhecimento em tonel de carvalho (*Quercus alba* l.). **Ciência e Tecnologia de Alimentos**, v. 18, n. 2, p. 169-175, 1998.

- CENTRO BRASILEIRO DE REFERÊNCIA DA CACHAÇA. **Oportunidades para a cachaça no mercado interno e externo.** Disponível em: <<http://www.sebraemercados.com.br/oportunidades-para-a-cachaca-no-mercado-interno-e-externo/>>. Acesso em: 18 Mai. 2015.
- DAS, K. NENADIC, Z. Approximate information discriminant analysis: A computationally simple heteroscedastic feature extraction technique. **Pattern Recognition**, v.41, n.5, p. 1548-1557, 2008.
- DA SILVA, N. R.; PADOVANI, C. R. Utilização de componentes principais em experimentação agrônômica. **Energia na Agricultura**, Botucatu, v.21, n.4, p.98-113, 2006.
- DECOSTER, J. **Overview of Factor Analysis**, 1998. Disponível em: <http://www.stat-help.com/notes.html>. Acesso em: 08 de março de 2015>
- DE SOUZA, P. P. de; AUGUSTI, D. V.; CATHARINO, R. R.; SIEBALD H G. L.; EBERLIN, M. N.; AUGUSTI R. Differentiation of rum and Brazilian artisan cachaça via electrospray ionization mass spectrometry fingerprinting. **Journal of Mass Spectrometry**, v. 42, p. 1294–1299, 2007.
- DIAS, S. M. B. C. O processo da destilação. In: CARDOSO, M.G. **Produção de aguardente de cana**. 2. ed. Lavras: UFLA, 2006. p. 137-202.
- DRAY, S. On the number of principal components: A test of dimensionality based on measurements of similarity between matrices. **Computational Statistics e Data Analysis**, v.52, p.2228-2237, 2008.
- DUARTE, F. C. **Argilas como adsorvente para remoção de cobre em cachaças**. 2013. 116f. Tese (Doutorado em Ciência dos Alimentos) – Universidade Federal de Lavras, Lavras, MG.
- DUNCAN, R. C. Factor Analysis, A Farm Management Research Tool. **Review of Marketing and Agricultural Economics**, v.36, n.1, p.37-46, 1968.
- EIM, V. S. SIMAL, S. ROSSELLÓ, C. JOSÉ, B. A. F. Optimisation of the addition of carrot dietary fibre to a dry fermented sausage (sobrassada) using artificial neural networks. **Meat Science**, v.94, n.3, p. 341–348, 2013.
- EERIKAINEN, T.; LINKO, P.; LINKO, S.; SIIMES, T.; ZHU, Y-H. FUZZY Logic and Neural Network Applications in Food Science and Technology. **Trends Food Science and Technology**, v.4, p. 237-242, 1993.

- FERNANDES A. P.; SANTOS M. C.; LEMOS S. G.; FERREIRA M. M. C.; NOGUEIRA A. R. A.; NÓBREGA J. A. Pattern recognition applied to mineral characterization of Brazilian coffees and sugar-cane spirits, **Spectrochimica Acta B**, v. 60, n. 5, p. 717-724, 2005.
- FERREIRA, D.F. **Estatística multivariada**. Lavras: Editora UFLA, 2011. 675p.
- FLORENTINO, M. **Aplicação de uma metodologia multivariada ao controle interno de qualidade dos laboratórios de análises clínicas**. 2000. 63f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, SC.
- GOMES, F.C.; ARAÚJO, R.A.C.; CISALPINO, P.S.; MOREIRA, E.S.A.; ZANI, C.L.; ROSA, C.A. Comparison between two selected *Saccharomyces cerevisiae* strains as fermentation starters in the production cachaça. **Brazilian archives of Biology and Technology**, v. 52, n. 2, p. 449-455, 2009.
- GRANATO, D.; OLIVEIRA, C. C.; CARUSO, M. S. F.; NAGATO, L. A. F.; ALABURDA, J. Feasibility of different chemometric techniques to differentiate commercial Brazilian sugarcane spirits based on chemical markers. **Food Research International**, v. 60, p. 212-217, 2014.
- HAIR JR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate Data Analysis**. 7.ed. Pearson Prentice Hall, 2010.
- HORIMOTO, Y.; DURANCE, T.; NAKAI, S.; LUKOW, O.M. Neural Networks vs Principal Component Regression for Prediction of Wheat Flour Loaf Volume in Baking Tests. **Journal of Food Science**, v.60, p.429-433, 1995.
- INFOMETRIX. **Chemometrics in Food and Beverage**. 10634 E. Riverside Dr., Bothell, WA. 1996.
- INSTITUTO BRASILEIRO DA CACHAÇA. **Mercado Externo da Cachaça - a**. Disponível em: <<http://www.ibrac.net/index.php/servicos/estatisticas/mercado-externo>>. Acesso em: 21 Mar. 2015.
- INSTITUTO BRASILEIRO DA CACHAÇA. **Premiação no Concours Mondial Spirits Selection 2014 - b**. Disponível em: < <http://www.ibrac.net/index.php/noticias/noticias-do-ibrac/359-mais-de-200-destilados-de-todo-mundo-sao-premiados-no-concours-mondial-spirits-selection-2014>>. Acesso em: 20 Out. 2015.

- JAIN, P. M.; SHANDLIYA, V. K. A survey paper on comparative study between Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA). **International Journal Of Computer Science And Applications**, v. 6, n. 2, p.373-375, 2013
- JIN, J. AN, J. Robust discriminant analysis and its application to identify protein coding regions of rice genes. **Biosciences Matemática**, v. 232, n. 2, p.96-100, 2011
- JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. New Jersey: Prentice-Hall, 1999.816p.
- KATZ, W.T.; SNELL, J.W.; MERICKEL, M.B. Artificial Neural Networks. **Methods in Enzymology**, v. 210, p. 610-636, 1992.
- KOVÁCS, Z. L. **Redes neurais artificiais: fundamentos e aplicações**. 3ª ed. São Paulo: Livraria da Física Editora; 2002.
- LAWRENCE, J.; PETERSON, A. **Brainmaker. User's Guide and Reference Manual**. 4thed., Nevada City, CA. 1993.
- LEDESMA, R. D.; VALERO-MORA. Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel. **Analysis, Practical Assessment, Research e Evaluation**, v. 12, n. 2, p. 1-11, 2007.
- LIRA, S. A. **Análise de correlação: abordagem teórica e de construção dos coeficientes com aplicações**. 2004. 196f. Dissertação (Mestrado em Ciências) – Universidade Federal do Paraná, Curitiba, PR.
- LYRA, W. S.; DA SILVA, E. C.; DE ARAÚJO, M. C. U.; FRAGOSO, W. D.; VERAS, G. Classificação periódica: um exemplo didático para ensinar análise de componentes principais. **Química Nova**, v. 33, n. 7, p. 1594-1597, 2010.
- MAÇATELLI, M. **Determinação do perfil sensorial de marcas comerciais de cachaça**. 2006. 128f. Dissertação (Mestrado em Alimentos e Nutrição) - Faculdade de Ciências Farmacêuticas, UNESP, Araraquara, SP.
- MAIA, A. B. **Apostila destilação da cachaça**, Vassouras, RJ, 2000, 10p.
- MAROCO, J. **Análise Estatística, com utilização do SPSS**, Edições Sílabo, Lisboa, Portugal, 2003.
- MARTINS-FILHO, L. S.; MOL, A. A.; ROCHA, R. Desenvolvimento de ferramenta computacional para auxílio ao projeto de gemas lapidadas. **Revista Escola de Minas**, v. 58, n. 4, p. 367-373, 2005.

- MENDES, E.F.F. **Redes Neurais: introdução a redes neurais artificiais, algoritmos genéticos e aplicações**, 1999. Disponível em: <http://www.icmc.sc.usp.br/~prico/neural1.html>. Acesso em: 20 de Mai. 2015
- MINAI, A. A., WILLIAMS, R. D. **Back-propagation heuristics: A study of the extended delta-bar-delta algorithm,**” Proc. IEEE Int. Joint Conf. Neural Networks, San Diego, USA, 595–600, 1990.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005. 297p.
- MOSEDALE, J. R.; PUECH, J. L. Wood maturation of distilled beverages. **Trends in Food Science e Technology**, v. 9, n. 3, p. 95-101, 1998.
- MORRISON, D.F. **Multivariate statistical method**. McGraw - Hill Book Co., 1976. 415 p.
- NETO, J. M. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. **Química Nova**, v.21, n.4, p.467-469, 1998.
- OGILVIE, D.; MITCHELL, R.; MUTRIE, N.; PETTICREW, M.; PLATT, S. Perceived characteristics of the environment associated with active travel: development and testing of a new scale. **International Journal of Behavioral Nutrition and Physical Activity**, v. 5, n. 32, p. 1-10, 2008.
- PACHECO, J.; CASADO, S.; PORRAS, S. Exact methods for variable selection in principal component analysis: Guide functions and pre-selection. **Computational Statistics and Data Analysis**, v.57, p. 95–111, 2013.
- PATRAS, A.; BRUNTON N. P.; DOWNEY, G., RAWSON, A.; WARRINER, K.; GERNIGON, G. Application of principal component and hierarchical cluster analysis to classify fruits and vegetables commonly consumed in Ireland based on in vitro antioxidant activity, **Journal of Food Composition and Analysis**, v. 24, p. 250-256, 2011.
- PEREIRA, E. B. B.; PEREIRA, M. B. Um critério para o descarte de variáveis na análise de componentes principais. **Revista Universidade Rural, Série Ciências Exatas e da Terra**. v.23, n. 1-2, p. 1-7, 2004.
- PESTANA, M. H.; GAGEIRO, J. N. **Análise de Dados para Ciências Sociais**. 2. ed. Lisboa: Edições Sílabo, 2000.
- RECHE, R. V.; FRANCO, D. W. Distinção entre cachaças destiladas em alambiques e em colunas usando quimiometria. **Química Nova**, v. 32, n. 2, p. 332-336, 2009

- RECHE, R. V.; LEITE-NETO, A. F.; DA SILVA, A. A.; GALINARO, C. A.; DE OSTI, R. Z.; FRANCO, D. W. Influence of type of distillation apparatus on chemical profiles of Brazilian cachaças **Journal of agricultural and food chemistry**, v. 55, n. 16, p. 6603, 2007.
- REGAZZI, A.J. **Analise multivariada, notas de aula INF 766**, Departamento de Informática da Universidade Federal de Viçosa, v.2, 2000.
- RENCHER, A. C. **Methods of multivariate analysis**. 2.ed. New York: John Wiley e Sons, 2002. 708 p.
- SANTIAGO, R. C. M. **A força da Cachaça de Salinas**, 2013. Disponível em: <<http://partes.com.br/wp-content/uploads/2013/05/Salinas.pdf>> Acesso em: 08. Mar. 2015
- SANTIAGO, R. C. M. **O mito da Cachaça Havana-Anísio Santiago**. Belo Horizonte: Cuatiara, 2006.
- SENA, M. M. DE; POPPI, R. J.; FRIGHETTO, R. T. S.; VALARINI, P. J. Avaliação do uso de métodos quimiométricos em análise de solos. **Química Nova**, v. 23, n. 4, p. 547-556, 2000.
- SERAFIM, F. A. T.; SEIXAS, F. R. F.; SILVA, A. A.; GALINARO, C. A.; NASCIMENTO, E. S. P.; BUCHVISER, S. F.; ODELLO, L.; FRANCO, D. W. Correlation between Chemical Composition and Sensory Properties of Brazilian Sugarcane Spirits (Cachaças). **Journal of the Brazilian Chemical Society**, v. 24, n. 6, p. 973-982, 2013.
- SCHNEEWEISS, H.; MATHES, H. Factor analysis and principal components. **Journal of Multivariate Analysis**, v.55, n.1, p.105-124, 1995.
- SCHWAN, R. F.; CASTRO, H. A. Fermentação alcoólica. In: CARDOSO, M.G. **Produção de cachaça de cana-de-açúcar**. 2. ed. Lavras: UFLA, p.101-135, 2006.
- SCHEEREN, L. W. et al. Agrupamento de unidades amostrais de *Araucaria angustifolia* (Bert.) O. Ktze. em função de variáveis do solo, da serapilheira e das acículas, na região de Canela, RS. **Ciência Florestal**, Santa Maria, v.10, n.2, p.39 - 57, 2000.
- SILVA, M. L.; MACEDO, A. C.; MALCATA, F. X. Review: Steam distilled spirits from fermented grape pomace. **Food Science and Technology International**, v.6, n.4, p. 285-300, 2000.
- VAINIONPAA, J.; KERVINEN, R.; DE PRADO, M.; LAURILA, E.; KARI, M.; MUSTONEN, L.; AHVANEAINEN, R. Exploration of storage and process tolerance

of different potato cultivars using principal component and canonical correlation analyses. **Journal Food Engineering**, v. 44, p. 47-61, 2000.

WILLIAMS, B.; BROWN, T.; ONSMAN, A. Exploratory factor analysis: A five-step guide for novices. **Australasian Journal of Paramedicine**, v.8, n.3, p.1-13, 2010.

WYTHOFF, B.J. *ET AL.* Spectral Peak Verification and Recognition Using a Multi Layered Neural Network. **Analytical Chemistry**, v. 62, p. 702-2709, 1990.

ZHOU, W. THERDTHAI, N. **Artificial neural networks in food processing**. M.M. Farid (Ed.), **Mathematical modeling of food processing**, CRC Press, Boca Raton, p. 901–920. 2010.

ZUBEN, F. J. V. Uma caricatura funcional de redes neurais artificiais. **Rev SBRN**. v. 1, n. 2, p.66-76, 2003.

CAPÍTULO 2 – PROSPECÇÃO E IDENTIFICAÇÃO DE CACHAÇAS PRODUZIDAS EM SALINAS E PARATY POR QUIMIOMETRIA

RESUMO

A comercialização da cachaça no Brasil vem aumentando a cada ano e com isto é possível observar a notoriedade de algumas regiões, a exemplo, de Salinas e Paraty no que diz respeito a qualidade da bebida. Neste sentido, os dados analíticos obtidos das cachaças podem ser utilizados para caracterizar amostras com base em um parâmetro previamente estabelecido. Uma opção viável para extrair o máximo de informações de dados físico-químicos obtidos nas regiões supracitadas é o uso de métodos matemáticos e estatísticos compilados na quimiometria. Objetivou-se neste trabalho a identificação das variáveis que apresentam maior relevância e a identificação de cachaças produzidas nas regiões de Salinas e Paraty, através da utilização de técnicas multivariadas e dos dados de composição físico-químicos exigidos pela legislação Brasileira. Foram coletados, por meio de visitas *in loco*, certificados oficiais de análises físico-químicas que estruturaram duas novas matrizes (A_1 e A_2) de dados com n medidas de diferentes propriedades físico-químicas (variáveis) realizadas sobre m amostras (cachaças) das regiões de Salinas e Paraty. Para o tratamento das matrizes de dados foram utilizadas as seguintes técnicas estatísticas: análise de componentes principais (ACP), fatores (AF), correlações canônicas (ACC), agrupamento (AA), discriminante (AD) e a técnica não estatística, rede neural artificial (RNA). Os resultados indicaram que as técnicas não supervisionadas (ACP, AF e ACC) foram eficientes na formação de agrupamentos com registro de alguns pontos distribuídos desuniformemente em função das singularidades dos processos de produção de Salinas e Paraty. Quanto às variáveis que possuem maior influência na identificação das regiões, pode se destacar a acidez, ésteres e os alcoóis superiores. Já na aplicação das técnicas supervisionadas obteve-se uma excelente classificação na etapa de validação com valores entre 83,87 e 100% com ênfase para a rede neural que apresentou 100% de separação na validação para as duas matrizes.

ABSTRACT

The commercialization of the cachaça in Brazil is increasing every year and it is possible to observe the notoriety of some regions, like in Salinas and Paraty as regards the quality of the beverage. In this regard, the analytical data obtained from cachaças can be used to characterize samples based on a predetermined parameter. A viable option to extract the maximum physical-chemical data information obtained in the above areas is the use of mathematical and statistical methods collected in chemometrics. The objective of this study was to identify the variables that have greater relevance and cachaça identification produced in the regions of Salinas and Paraty, through the use of multivariate techniques and the physical and chemical composition data required by Brazilian law. Were collected through site visits, official certificates of physical-chemical analysis, structured two new arrays (A_1 and A_2) of data with n measurements of different physical and chemical properties (variables) performed on m samples (cachaça) regions Salinas and Paraty. For the treatment of matrix data the following statistical techniques were used: principal component analysis (PCA), factors (FA), canonical correlation (CCA), discriminant (DA) and non-statistical technique, neural network artificial (RNA). The results indicated that the unsupervised techniques (PCA, FA and CCA) were efficient in the formation of clusters with log some points distributed non-uniformly depending on the peculiarities of Salinas and Paraty production processes. As for

the variables that have the greatest influence on the identification of regions, can stand out the acidity, esters and higher alcohols. In the application of supervised techniques gave an excellent rating in the validation stage with values between 83.87 and 100% focusing on the neural network showed 100% separation validation for the two arrays.

1 INTRODUÇÃO

Ao longo dos anos, algumas cidades ou regiões ganharam fama por causa de seus produtos ou serviços que são símbolos que refletem a economia e cultura local. Salinas e Paraty alcançaram notoriedade com a produção e comercialização da cachaça. Para evitar a descaracterização da autêntica cachaça de Salinas e de Paraty foi concedido, pelo INPI (Instituto Nacional de Propriedade Intelectual), a Indicação de Procedência (uma das modalidades de Indicação Geográfica - IG). Com a concessão do selo, apenas as cachaças produzidas nestas regiões, usando as práticas agrícolas e modos de fabricação específicos de cada cidade, terão o selo apostado em suas garrafas (APACS, 2013).

No entanto esta notabilidade traz consigo algumas consequências negativas, pois são recorrentes as notificações de alterações, adulterações, falsificações e fraudes em algumas etapas da produção e, especialmente, na comercialização da bebida. Uma alternativa para minimizar estes episódios é a utilização da investigação científica para obtenção de informações acerca das principais variáveis que influenciam na produção da bebida. Este processo facilita o reconhecimento do produto e contribui para a consolidação de uma boa imagem do destilado símbolo do Brasil.

Desta forma, os dados analíticos obtidos das cachaças podem ser utilizados para caracterizar as amostras de acordo com um parâmetro previamente estabelecido. Esta caracterização seria relativamente simples se o número de variáveis fosse pequeno, mas com o desenvolvimento tecnológico das ferramentas e dos computadores, é possível processar uma ampla e complexa quantidade de informação de maneira a gerar conhecimento acerca dos diferentes processos. Na posse de uma grande quantidade de dados, o problema passa a ser como tratá-los (HELFER et al, 2006). Uma opção viável para extrair o máximo de informações de dados físico-químicos obtidos nas regiões supracitadas é o uso de métodos matemáticos e estatísticos compilados na quimiometria. Esta é uma ciência que usa os métodos supracitados para a obtenção de informações relevantes nas diversas medidas realizadas na área de química e, portanto, é uma estratégia contra a falsificação.

Ao pesquisar nas principais bases de dados é possível encontrar trabalhos envolvendo cachaça e quimiometria, no entanto a grande maioria utiliza métodos onerosos e análises que não são realizadas com frequência pelo produtor. Sendo assim, é possível utilizar

análise quimiométrica aplicada aos dados disponibilizados pelos produtores, que realizam este diagnóstico da bebida a cada lote produzido, com a finalidade de buscar padrões de agrupamento por região através das análises de componentes principais, fatores e correlação canônica. Esta última permite, ainda, a determinação da correlação entre as variáveis. Já a análise discriminante e a rede neural artificial são utilizadas na classificação supervisionada das amostras por região.

Faz-se necessário a adoção destes tipos de ferramentas, que permitem a discriminação da bebida pela região de produção, pois o novo perfil consumidor procura cachaças com alto padrão de qualidade, inclusive para suprir a demanda do mercado externo, mas que tenham a garantia da sua origem.

Diante do mencionado este trabalho teve como objetivo o reconhecimento das variáveis que estão relacionadas com a qualidade do processo produtivo e a identificação de cachaças produzidas nas regiões de Salinas e Paraty, através da utilização de técnicas multivariadas e dos dados de composição físico-químicos exigidos pela legislação Brasileira

2 MATERIAL E MÉTODOS

2.1 Coleta dos Dados

Para esta pesquisa optou-se pelas regiões de Salinas e Paraty que possuem o selo de Indicação Geográfica. Os dados foram obtidos com os produtores através de visitas *in loco*. Foram visitados 17 produtores que comercializam 36 marcas em Salinas e cinco produtores responsáveis por 15 marcas em Paraty. É importante salientar que existem apenas seis produtores formalizados na região de Paraty.

O período de coleta foi de Junho de 2014 a Fevereiro de 2015 e os produtores disponibilizaram as informações através de 171 certificados oficiais de análises físico-químicas que são arquivados para apresentação aos fiscais do Ministério da Agricultura, Pecuária e Abastecimento (MAPA). As amostras de cachaça utilizadas para elaboração dos laudos foram analisadas quanto a sua composição físico-química, seguindo os parâmetros estabelecidos pela IN nº 13, de 29/6/2005 e a IN nº 24, de 08/09/2005 do MAPA por laboratórios credenciados (BRASIL, 2005ab).

2.2 Delineamento Experimental dos Dados

As amostras foram organizadas em um conjunto de dados e a partir deste, foram constituídas duas matrizes (A_1 e A_2) diferentes a fim de avaliar o efeito do tipo de variável (Figura 1). As matrizes de dados foram construídas com n medidas de diferentes propriedades físico-químicas (variáveis) realizadas sobre m amostras (cachaças), de modo que foram formadas por $m \times n$ elementos (m linhas correspondentes as amostras e n colunas correspondentes as variáveis).

Figura 1. Construção das Matrizes de Dados A_1 e A_2

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

A primeira matriz de dados (A_1) apresentou sete medidas de diferentes propriedades físico-química ($n=7$), a saber, grau alcoólico, acidez volátil, ésteres totais, aldeído, alcoóis superiores totais, álcool metílico e soma dos componentes voláteis, e em relação ao número de amostras de cachaças foram avaliadas 142 ($m=142$) (Tabela 1). Esta primeira matriz de dados foi estruturada para avaliar apenas as variáveis que representam os parâmetros globais exigidos pelo MAPA com níveis restritos de presença nas cachaças.

Por sua vez, a segunda matriz de dados (A_2) foi elaborada com dez medidas de diferentes propriedades físico-químicas ($n=10$), que são as variáveis sobreditas e os alcoóis n-propílico, isobutílico e isoamílico obtidos da estratificação dos alcoóis superiores totais. Estas foram realizadas em 91 amostras de cachaças (Tabela 2). A segunda matriz foi estruturada com objetivo de avaliar a importância dos alcoóis estratificados na identificação das cachaças produzidas em Salinas e Paraty.

Tabela 1. Número de amostras por região e amplitude das variáveis utilizadas para construção da matriz A_1

	Salinas	Paraty	Total
Dados	104	38	142
GL°	39,0-49,0	40,7-59,1	39,0-59,1
Acidez*	10,5-150,0	5,0-119,0	5,0-150,0

Ésteres*	16,0-169,7	15,0-73,0	15,0-169,7
Aldeído*	0,4-30,0	3,6-15,8	0,4-30,0
ASTot*	96,5-315,5	178,0-360,0	96,5-360,0
AMet*	0,0-13,5	0,48-6,0	0,0-13,5
SCV*	218,0-556,2	227,0-499,0	218,0-556,2

* mg/100 mL etanol

GL: Grau Alcoólico; **ASTot:** Alcoóis Superiores Totais; **Amet:** Álcool Metílico; **SCV:** Soma dos Componentes Voláteis.

Tabela 2. Número de amostras por região e amplitude das variáveis utilizadas para construção da matriz A_2

	Salinas	Paraty	Total
Dados	70	21	91
GL°	39,5-48,1	40,7-59,1	39,5-59,1
Acidez*	39,6-150,0	11,0-119,0	11,0-150,0
Ésteres*	16,0-137,3	15,0-59,0	15,0-137,3
Aldeído*	1,8-19,5	5,0-12,0	1,8-19,5
ASTot*	123,1-315,5	191,0-360,0	123,1-360,0
ASnP*	24,0-98,8	11,0-79,0	11,0-98,8
ASIsoB*	27,3-72,7	30,0-60,0	27,3-72,7
ASIsoA*	71,4-184,0	100-262,0	71,4-262,0
AMet*	0,0-12,9	1,0-5,0	0,0-12,9
SCV*	243,0-556,2	235,0-499,0	235,0-556,2

* mg/100 mL etanol

GL: Grau Alcoólico; **ASTot:** Alcoóis Superiores Totais; **ASnP:** Álcool Superior n-propílico; **ASIsoB:** Álcool Superior Isobutílico; **ASIsoA:** Álcool Superior Isoamílico **Amet:** Álcool Metílico; **SCV:** Soma dos Componentes Voláteis.

2.3 Medidas de Dispersão e Padronização da Matriz de Dados

Foram realizadas as determinações de algumas medidas de dispersão, a saber, desvio padrão e coeficiente de variação. A partir deste resultado, fez-se a padronização dos dados, que consistiu na implementação de uma nova variável (Z) com média zero e desvio padrão igual a 1, eliminando as diferenças entre as unidades de medidas das variáveis estudadas. (Equação 1).

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Onde Z : nova variável (forma padrão)

X : variável

μ : média

σ : desvio padrão

2.4 Análises Estatísticas Multivariadas dos Dados

Para realização das análises estatísticas multivariadas foram utilizados os procedimentos descritos em SAS, do pacote computacional SAS, versão 9.0 e, nos anexos estão apresentados os comandos utilizados para os conjuntos de dados.

2.4.1 Análise de Componentes Principais

A partir do conjunto de médias padronizadas obtido no item 2.3, foi realizada a transformação das variáveis Z em um novo sistema de coordenadas com eixos ortogonais, dando origem aos CP's, obtidos através de combinações lineares (Equação 2).

$$CP_i = a_{i,1}Z_1 + a_{i,2}Z_2 + \dots + a_{i,n}Z_n \quad (2)$$

com $i = 1, 2, \dots, n$, em que:

CP_i : i-ésimo componente principal;

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

Na sequência, fez-se a interpretação de cada componente através da avaliação das variáveis originais mais importantes em cada eixo. O programa utilizado forneceu as CP's em ordem decrescente de máxima variância, o que permitiu a redução da dimensionalidade por meio da seleção de três CP's, que é um número bem menor do que o de variáveis originais para representar o conjunto inicial dos dados que contém sete variáveis para a matriz A_1 e dez para matriz A_2 . Isto foi possível porque os últimos componentes são responsáveis pela explicação de uma pequena porção da variância disponível.

Para a escolha do número de componentes principais foi considerado o critério de fatores interpretáveis e o critério de Kaiser, que selecionaram os primeiros k componentes principais que retêm, juntos, a maior proporção de variância dos atributos originais (acima de 70% da variância) e autovalor maior que um. Já para a seleção de variáveis, avaliou-se no

componente de maior autovalor, as variáveis que apresentam, em valor absoluto, maior correlação. Posteriormente, deu-se prosseguimento a análise do segundo e terceiro componentes para escolha das variáveis que não foram previamente escolhidas (JOLLIFFE, 2002; FERREIRA, 2011).

2.4.2 Análise de Fatores

Nesse estudo, optou-se pelo uso do modelo fatorial e o número de fatores (m) foi escolhido a partir de autovalores maiores que um, conforme Johnson e Wichern (1999) e Guttman (1954). A fim de estabelecer o número de fatores, também utilizou-se como critério, aqueles cuja soma explicava no mínimo 70% da variação dos dados.

Quanto ao descarte de variáveis, foram retiradas, em primeiro lugar, as que apresentaram comunalidade menor que 0,5, pois estas representaram uma pequena proporção da variância na análise. Na sequência foram eliminadas as variáveis que apresentaram cargas (correlações entre as variáveis e os fatores) inferiores a 0,4 sendo, este um limite aceitável para contribuição da variável na criação do fator. Adicionalmente, foram excluídas variáveis que apresentem cargas superior a 0,4, em mais de um fator, uma vez que uma variável não pode contribuir para dois fatores distintos, pois dificulta a produção de grupos por ela estar associada aos dois fatores (HAIR JR et al, 2010). Utilizando os critérios citados procedeu-se, à análise definitiva, na qual se especificou o número de fatores final.

Para a rotação, foi escolhido, o método ortogonal Varimax para que cada fator incluísse a maior porcentagem possível da variância existente nos atributos originais por ele representados. Foram estimadas, ainda, com a rotação, as cargas fatoriais finais e estas foram usadas para identificar os atributos que compunham cada fator (CARNEIRO, 2005).

A partir dos fatores comuns foram estimados os escores para serem utilizados na separação das cachaças, por meio do agrupamento de amostras com o mais alto grau de similaridade, em um espaço gráfico bi ou tridimensional.

2.4.3 Análise de Correlações Canônicas

A análise de correlação canônica foi realizada com o intuito de observar as relações lineares entre dois vetores de variáveis p (fração álcool) e q (fração não álcool), a fim de avaliar a influência do álcool na formação de compostos secundários importantes para qualidade sensorial da cachaça.

Uma combinação linear foi desenvolvida em cada um dos pares de variáveis canônicas

dos conjuntos de amostras (Salinas e Paraty), para que a correlação entre os dois fosse maximizada e, desta forma, considerou-se as combinações lineares do primeiro par como U (fração álcool) e V (fração não álcool) expressas por:

$$\begin{aligned} U_k &= a_k' X \\ V_k &= b_k' Y \end{aligned} \tag{3}$$

tal que:

$$\text{Corr}(U_k, V_k) = \frac{a_k' \Sigma_{XY} b_k}{\sqrt{a_k' \Sigma_{XX} a_k} \sqrt{b_k' \Sigma_{YY} b_k}}$$

seja a maior possível, ou seja, $\max \text{Corr}(U, V)$, que expressa a máxima associação entre as funções lineares dos dois grupos de variáveis (JOHNSON e WICHERN, 1999).

Após a obtenção das correlações canônicas, foi realizado o teste estatístico F ($p < 0,05$), cuja hipótese de nulidade foi não há correlação entre os conjuntos de dados, para testar a significância dos resultados.

Já as cargas canônicas, que se referem às associações entre cada variável original e a correspondente variável canônica foram obtidas através das correlações estruturais.

Foi calculada, ainda, a proporção da variância total que é explicada por cada par de variáveis canônicas e isto permitiu avaliar a relevância deste par no conjunto total dos dados. Por fim, utilizou-se as observações obtidas pelas variáveis de cada fração (álcool e não álcool) para representar graficamente a distribuição das amostras por região.

2.4.4 Análise Discriminante

A análise discriminante consistiu em obter funções matemáticas capazes de classificar uma amostra de cachaça em uma das duas populações π_i , ($i=1$ - Paraty ou 2 - Salinas), com base em medidas de sete variáveis para a matriz A_1 e dez variáveis para a matriz A_2 , buscando minimizar a probabilidade de má classificação, isto é, minimizar a probabilidade de classificar erroneamente uma amostra em uma população π_1 , quando realmente pertence a população π_2 . Para isto a AD procurou maximizar a variância entre os diferentes grupos e minimizar a variação dentro de cada um deles, a fim de selecionar o limite que estabelece máxima separação entre os dois grupos. O treinamento foi a primeira etapa, sendo a parte exploratória da análise, que consistiu em procurar características capazes de serem utilizadas para alocar

amostras nos dois grupos previamente definidos, a partir do emprego de limites de decisão linear.

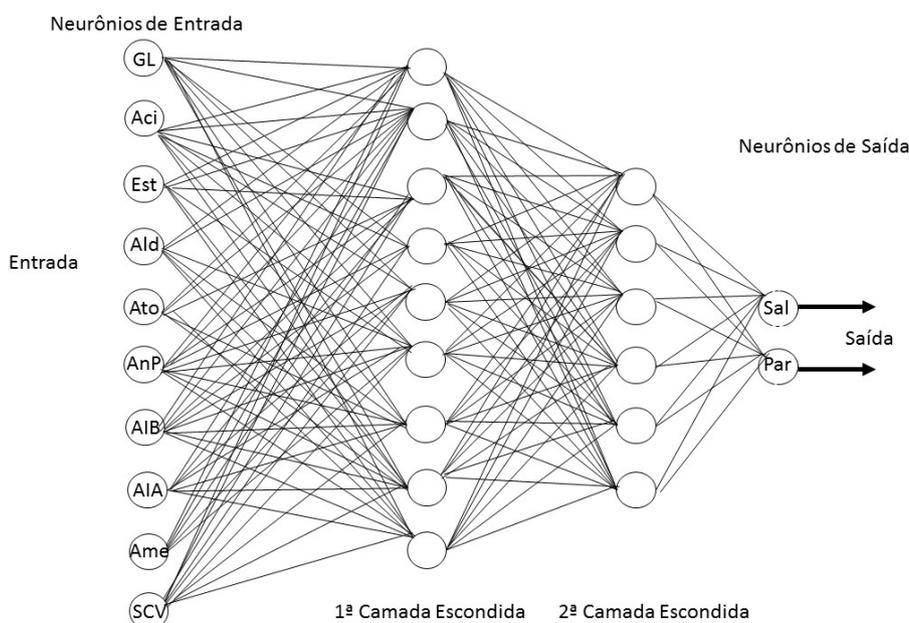
Os dados foram aleatorizados e divididos em dois grupos, um de teste e o outro de validação utilizando o algoritmo de Kenstone. Optou-se por usar a proporção 70% e 30% (Treinamento; Validação) e os dados obtidos foram analisados, para obter a função discriminante linear.

2.5 Redes Neurais Artificiais

Utilizou-se uma rede neural artificial (RNA) a fim obter a separação das cachaças em dois grupos, a saber, Salinas e Paraty. Nas RNA's, cada neurônio recebeu um sinal de entrada e enviou um sinal de saída como os neurônios do cérebro para geração da informação. A força das ligações depende dos pesos aplicados em cada neurônio durante a etapa de aprendizagem.

Para este trabalho foi escolhida a metodologia de redes neurais supervisionadas denominada *perceptrons* multicamadas ou *multilayers perceptrons* (MLPs) e o algoritmo de treinamento utilizado para tanto foi o algoritmo de retropropagação do erro, conhecido na literatura como *Backpropagation Algorithm* ou simplesmente *Backprop*.

Figura 2. Algoritmo *Backpropagation*



GL: Grau Alcoólico; **Aci:** Acidez; **Est:** Ésteres; **Ald:** Aldeído **Ato:** Alcoóis Superiores Totais; **AnP:** Álcool Superior n-propílico; **AIB:** Álcool Superior Isobutílico; **AIA:** Álcool Superior Isoamílico **Ame:** Álcool Metílico; **SCV:** Soma dos Componentes Voláteis.

Para classificar os diferentes tipos de cachaças em função das propriedades físico-químicas, foi utilizado o algoritmo supracitado com a regra de aprendizado Delta-Bar-Delta-expanded (MINAI E WILLIAMS, 1990). A regra de aprendizagem utilizou uma heurística com variação da taxa e do coeficiente de aprendizado com valores de 0,3 e 0,4, respectivamente, durante todo o treinamento. Para ativação do neurônio (transferência de sinais) foi utilizada uma função hiperbólica (φ), por ser a mais utilizada para sistemas biológicos ou mais complexos (RAI et al., 2005) (Equação 4).

$$\varphi = \tanh(\nu_k) \quad (4)$$

onde ν_k é o potencial de ativação do neurônio de ordem k (a soma dos sinais de entrada ponderados por pesos sinápticos e os desvios).

Com a finalidade de obter a melhor configuração da rede neural para classificação das amostras de cachaça, foi feito um rastreamento do número de iterações, e do valor do RMSE nas fases de treinamento e validação. Na sequência fez-se a análise de 10.000 até 160.000 e para conseguir a melhor arquitetura, foram feitos testes, dentro deste intervalo, com um total de 30 configurações que continha de 3 a 15 neurônios na primeira camada e de 0 a 15 na segunda camada.

Todas as redes testadas tiveram os neurônios de uma camada totalmente conectados com os neurônios de uma camada vizinha

A escolha do número de neurônios em cada camada escondida foi baseada na minimização do erro quadrático e maximização do coeficiente de correlação, associada à capacidade preditiva da rede (RAI et al., 2005).

A taxa de aprendizagem indicou a taxa de variação de pesos de conexão durante a formação. É válido salientar que valores baixos retardam o treinamento, enquanto grandes valores aceleram o processo. Os valores típicos para a taxa de aprendizagem variam entre 0,05 e 0,75 (RAI et al., 2005).

Para acelerar a etapa de aprendizagem, foram realizadas modificações no algoritmo *backpropagation* através da regra estendida delta-bar-delta (EDBD) que utilizou uma taxa de aprendizado para aumentar a velocidade de aprendizagem e conseqüentemente, reduzir os riscos de instabilidade (HAYKIN, 1999, BRAGA et al., 2000). A taxa de aprendizado foi adicionada aos pesos para ajustar a função e ajudar as mudanças de peso.

O número de iterações usadas também é de grande importância. Um pequeno número de iterações leva a grandes erros enquanto um grande número de iterações poderia exigir

muito tempo computacional. A determinação do melhor número de iteração melhor para a formação da rede foi efetuada por tentativa e erro, técnica que resultou em um menor valor da raiz do erro quadrado médio (RMSE) de validação (Equação 5).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_d - x_p)^2} \quad (5)$$

onde n é o número de dados, e x_d e x_p são os valores experimentais e previstos das propriedades físico-químicas, respectivamente.

2.5.1 Treinamento e Validação dos Dados

Os conjuntos de dados foram utilizados para treinamento da rede e ajuste das funções discriminantes. Os dados foram aleatorizados e divididos com auxílio do algoritmo de Kenstone em dois subconjuntos, usando 70% para o treinamento e 30% para validação de rede. Foram utilizados sete neurônios na camada de entrada, grau alcoólico (X1), acidez volátil (X2), ésteres totais (X3), aldeído (X4), alcoóis superiores totais (X5), álcool metílico (X6), soma dos componentes voláteis (X7). Já para matriz dois foram dez, que são as variáveis supracitadas e os alcoóis n-propílico (X8), isobutílico (X9) e isoamílico (X10). A saída do RNA foi representada por um vetor multidimensional com uma série de dimensões iguais ao número de tipos de cachaça e para cada dimensão vetorial foi atribuída um tipo. As cachaças foram classificadas como mostrado abaixo:

(1,0) = pertence a região de Salinas

(0,1) = pertence a região de Paraty

Nos arquivos de treinamento e teste/validação, o conjunto de dados foi criado em formato numérico, atribuindo 1 quando a cachaça foi corretamente classificada e 0 caso contrário.

Os dados de entrada e saída foram normalizados usando através das equações 6 e 7, respectivamente

$$x_{i,norm} = \frac{x_i - \bar{x}_i}{R_{i,max}} \quad (6)$$

$$x_{i,norm} = \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}} \quad (7)$$

onde $x_{i,norm}$ são os dados normalizados, x_i são os dados experimentais, \bar{x}_i é a média dos dados, $R_{i,max}$ é dada por $R_{i,max} = \max [x_{i,max} - x_i, x_i - x_{i,min}]$, $x_{i,min}$ e $x_{i,max}$ são os valores mínimos e máximos de x_i , respectivamente.

As performances das diferentes formações das redes neurais foram comparadas utilizando a taxa de classificação, matriz de confusão e a raiz do erro quadrático médio (RMSE), expresso pela equação 5.

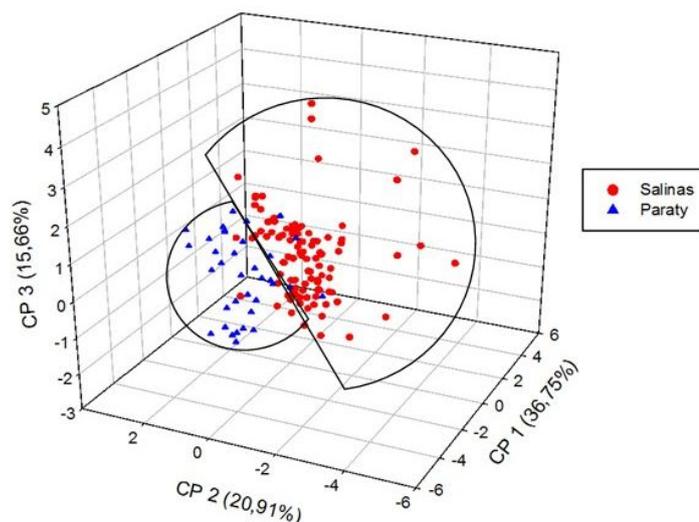
3 RESULTADOS E DISCUSSÃO

3.1 Análise de Componentes Principais

Para a matriz A_1 , composta por 142 linhas relacionadas as amostras de cachaça fabricadas nas regiões de Salinas e Paraty ($m=142$) e 07 colunas relativas as variáveis físico-químicas da qualidade da cachaça ($n=7$) obteve-se como resultado três componentes principais (CP), sendo que os percentuais da variação dos dados explicados pela CP1, CP2 e CP3 foram 36,75%, 20,91% e 15,66%, respectivamente, totalizando 73,32% da variância.

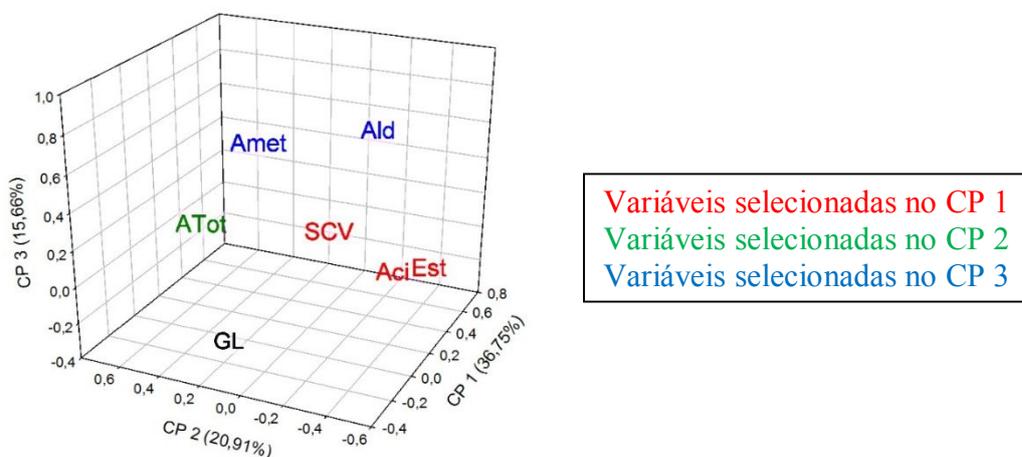
Avaliando o gráfico de escores das cachaças produzidas em Salinas e Paraty (Figura 3) é possível observar a formação de um padrão de agrupamento por região, no entanto, não há separação total das amostras. Como esta análise identifica a relação entre as características extraídas dos dados pode-se dizer que algumas amostras de localidades diferentes possuem propriedades muito semelhante e este fato dificulta a separação das mesmas. Isto provavelmente ocorre em função da ampla faixa apresentada por algumas variáveis analisadas neste estudo. A acidez, por exemplo, foi quantificada entre os seguintes valores: 5,0 a 150 mg/100 mL (Tabela 1 e 2).

Figura 3. Gráfico de escores das cachaças produzidas nas regiões de Salinas e Paraty em relação aos três primeiros componentes principais (CP1, CP2 e CP3) da matriz A_1 .



Na sequência, tem-se que as variáveis associadas foram acidez, ésteres e soma dos componentes voláteis para o CP 1, álcool superior total para o CP 2 e álcool metílico e aldeído para o CP 3 baseado nos maiores coeficientes (Figura 4). Como a CP 1 representa 36,7% da variação, e é o componente que melhor representa a distribuição dos dados, as variáveis que estão relacionadas a este componente, apresentam maior importância relativa e podem determinar a característica de cada grupo estudado (Figura 4). É importante salientar que são justamente estas variáveis que apresentam maior amplitude (Tabela 1 e 2), reiterando a justificativa para a dificuldade de separação de algumas amostras.

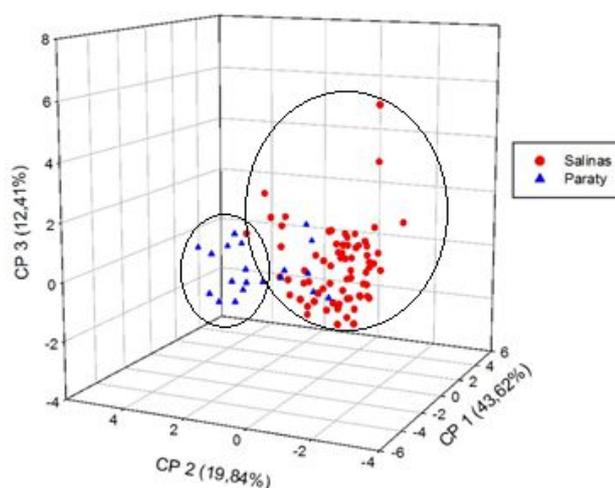
Figura 4. Gráfico de pesos dos coeficientes de ponderação (Autovetores) da matriz A_1 contendo as cachaças produzidas nas regiões de Salinas e Paraty.



GL: Grau Alcoólico; **Aci:** Acidez; **Est:** Ésteres; **Ald:** Aldeído; **ATot:** Alcoóis Superiores Totais; **Amet:** Álcool Metílico; **SCV:** Soma dos Componentes Voláteis.

Por sua vez, para o segundo conjunto de dados que constituiu a matriz A_2 com 91 linhas referentes as amostras de cachaças produzidas em Salinas e Paraty ($m=91$) e 10 variáveis físico-químicas ($n=10$), tem-se que três componentes principais resultaram em uma explicação de 75,88% da variação dos dados. O gráfico com a dispersão dos escores dos componentes principais 1, 2 e 3 (Figura 5) manteve o padrão da matriz A_1 , ou seja, os perfis por região foram formados, mas a presença de amostras alocadas incorretamente corrobora a ideia sobre as peculiaridades do processo produtivo apresentada anteriormente para a ocorrência de sobreposições.

Figura 5. Gráfico de escores das cachaças produzidas nas regiões de Salinas e Paraty em relação aos três primeiros componentes principais (CP1, CP2 e CP3) da matriz A_2 .

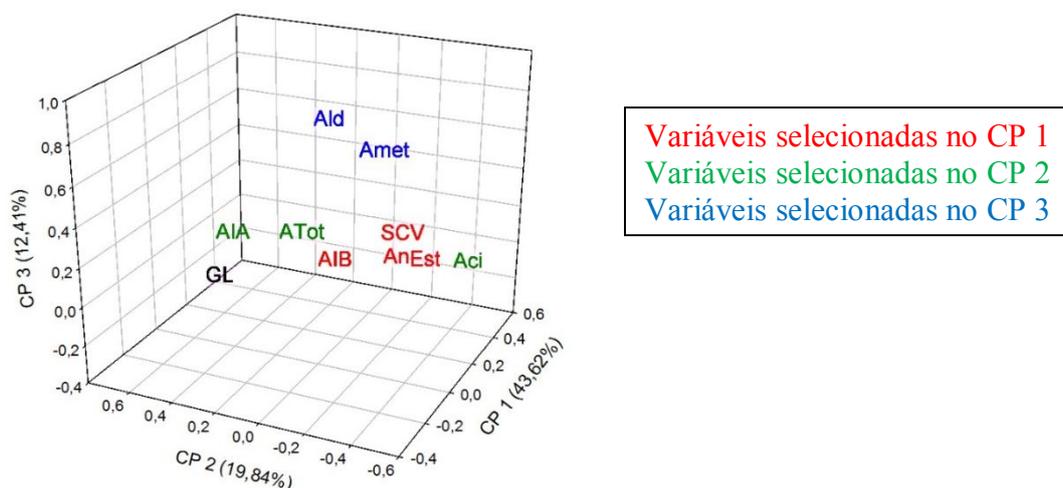


A triagem das variáveis, de acordo com os pesos, resultou na escolha dos ésteres, os alcoóis superiores n-propílico e isobutilíco e a soma dos componentes voláteis para o CP 1, acidez e álcool superior total e isoamílico para o CP 2, aldeído e álcool metílico para o CP 3, uma vez que estes possuem os maiores coeficientes (Figura 6). A CP 1 explicou 43,62% da variação, ou seja, as variáveis a ele relacionadas são muito importantes e representam, através das combinações lineares das variáveis originais, a maior parte da variância de todo o conjunto.

A análise agrupou as cachaças segundo seu comportamento dentro da população, representado pela variação do conjunto de características que define a amostra. Assim, com

um número maior de variáveis é notória a importância da estratificação dos alcoóis superiores na determinação do resultado.

Figura 6. Gráfico de pesos dos coeficientes de ponderação (Autovetores) para matriz A_2 contendo as cachaças produzidas nas regiões de Salinas e Paraty.



GL: Grau Alcoólico; **Aci:** Acidez; **Est:** Ésteres; **Ald:** Aldeído; **ATot:** Alcoóis Superiores Totais; **AnP:** Álcool Superior n-propílico; **AIB:** Álcool Superior Isobutílico; **AIA:** Álcool Superior Isoamílico **Amet:** Álcool Metílico; **SCV:** Soma dos Componentes Voláteis.

É válido salientar que o tipo de variável estudada altera os resultados encontrados, uma vez que após com a introdução dos alcoóis, observou-se uma modificação nas variáveis que determinam a qualidade e, isto, na prática mostra que mesmo mantendo todas as outras vistas na matriz A_1 estas foram as que tiveram maior importância durante todo o processo produtivo, e, portanto, podem ser utilizadas como referências para cada região. Ademais o resultado corrobora a ideia apresentada para estratificação das variáveis com maior relevância, pois a estratificação dos alcoóis alterou a ordem de importância e o mesmo pode ser feito com os ácidos, o éster, entre outros.

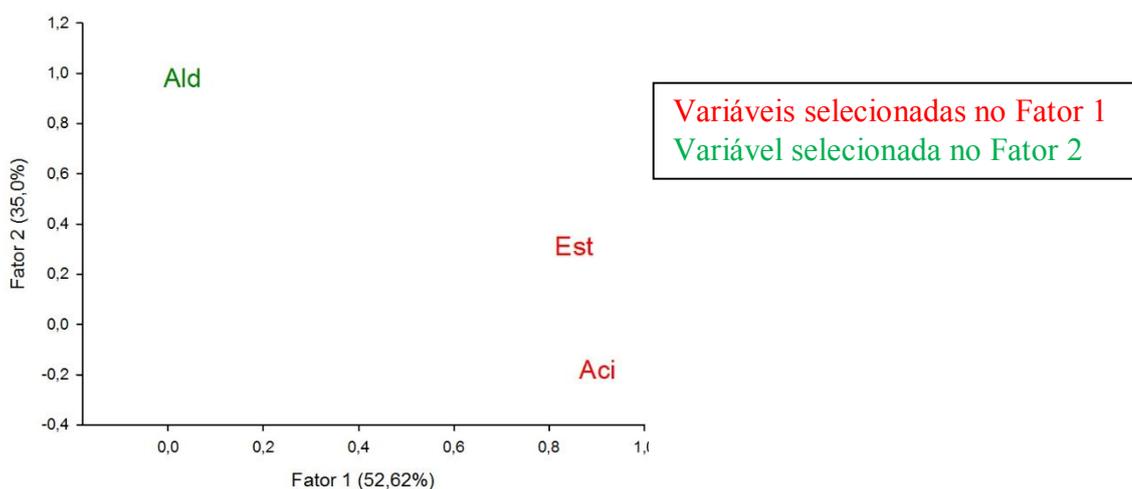
3.2 Análise de Fatores

A análise da matriz A_1 composta por 142 amostras de cachaça das regiões de Salinas e Paraty ($m=142$) e 7 variáveis físico-químicas ($n=7$) resultou na escolha de dois fatores que representaram 87,67% da variação dos dados, dos quais 52,62% foram representados pelo fator 1 e 35,05% pelo fator 2.

Em primeiro lugar, foi feita uma análise preliminar, para definir, com base na análise de componentes principais, qual o número inicial de fatores e este método resultou na escolha de três fatores. Posteriormente, fez-se a análise definitiva seguida da rotação dos dados. É importante salientar que esta etapa é fundamental para a seleção de variáveis, já que os dados rotacionados fornecem uma melhor interpretação das cargas e da comunalidade, facilitando a leitura dos resultados

A análise foi repetida várias vezes até que todos os pressupostos relatados por Hair Jr. e colaboradores (2010) fossem alcançados e de acordo com a figura 7, com base nos valores das cargas fatoriais, as variáveis acidez e ésteres estão altamente associadas ao fator 1 e a variável aldeído com o fator 2.

Figura 7. Gráfico de pesos das cargas fatoriais finais após a rotação da matriz A_1 para as cachaças produzidas nas regiões de Salinas e Paraty.



Aci: Acidez; **Est:** Ésteres; **Ald:** Aldeído.

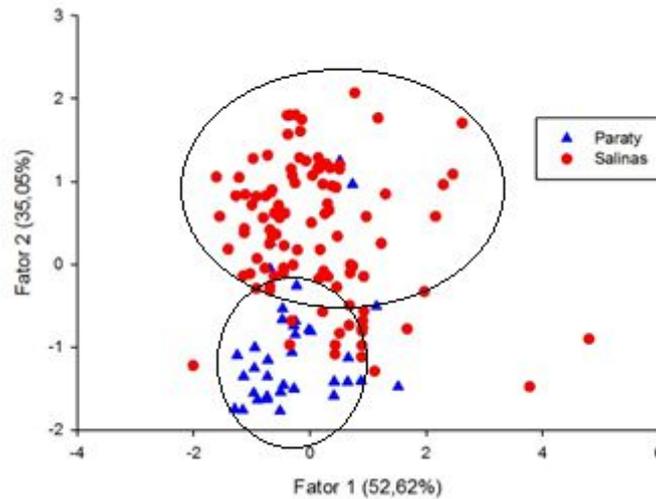
A acidez da cachaça vai depender do controle no processo de fermentação, em relação a fatores como: qualidade microbiológica da cana e do caldo; estirpe da levedura predominante no pé de cuba, pureza da fermentação, o tempo e a temperatura de fermentação e o manejo do mosto. Deve-se evitar durante a fermentação a aeração do mosto, já que o aumento de oxigênio faz com que o levedo transforme o açúcar em ácido acético em vez de etanol. (CARDOSO, 2006). Outra questão relacionada é que ácidos em contato com oxigênio originam ésteres. O acetato de etila é o principal éster encontrado na bebida e é obtido pela reação entre etanol e ácido acético, provenientes do processo de fermentação, porém outros ésteres estão presentes em concentrações que variam em função da extensão das reações de transesterificação (NYKÄNEN e NYKÄNEN, 1983).

A proporção de ésteres no vinho varia acentuadamente de uma região para outra, sendo um dos itens mais importantes para a identificação da origem da cachaça, já que diferentes cepas de leveduras produzem proporções de ésteres bem variadas. A quantidade deste composto gerada durante o processo de fermentação é fortemente influenciada por micronutrientes do solo que são incorporados à cana (como o zinco), pois esses atuam como cofatores enzimáticos na reação intracelular de esterificação (MAIA e CAMPELO, 2005). Aproximadamente 80% da quantidade de ésteres presentes no mosto é representado pelo acetato de etila, podendo sua formação ser estimulada através do resfriamento do mosto, logo após o término da fermentação (MAIA, 1994). Assim pode-se dizer que o fator 1 está altamente relacionado com a etapa de fermentação do processo produtivo, e como os produtores, normalmente, utilizam leveduras endêmicas, estas estão associadas com a caracterização de cada região estudada,

Por sua vez, o fator 2 é representado pelo aldeído que é produzido pelas leveduras principalmente durante os primeiros estágios da fermentação alcoólica, diminuindo sua presença gradativamente até o estágio final da fermentação, pela oxidação a ácido acético (CARDOSO, 2006). A presença de oxigênio favorece também a formação de outros aldeídos, provenientes da oxidação de álcoois superiores. Os aldeídos oxidam-se com facilidade, convertendo-se nos ácidos correspondentes, como por exemplo, o acetaldeído, que se converte em ácido acético (MAIA e CAMPELO, 2005). Neste caso a fase de produção com maior relevância também é a fermentação o que, possivelmente, justifica o fato de não se conseguir 100% de agrupamento correto nestas análises e corrobora a importância das leveduras no processo de identificação das regiões.

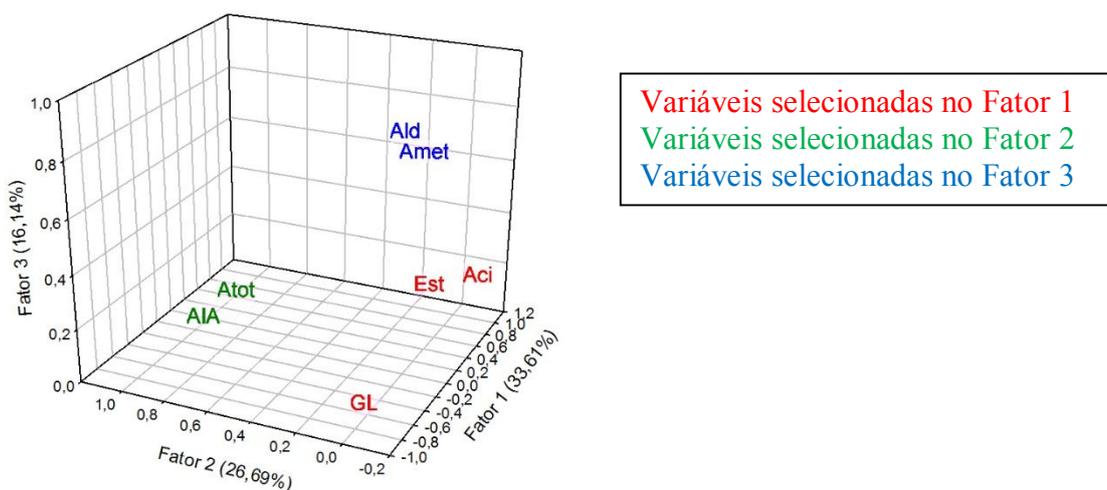
Finalmente, na Figura 8 está a representação no espaço bidimensional das cachaças produzidas em Salinas e Paraty. É válido lembrar que existem amostras que não estão alocadas adequadamente em razão da especificidade do processamento de cada cachaça e das especificidades das regiões conforme discussão anterior.

Figura 8. Gráfico de escores das cachaças produzidas nas regiões de Salinas e Paraty em relação aos dois primeiros fatores (Fator 1 e Fator 2) da matriz A_1 .



No que diz respeito à matriz A_2 com $m=91$ e $n=10$, utilizando as mesmas pressuposições, obteve-se como resultado três fatores que representam 76,44% da variabilidade dos dados. Já as informações das cargas fatoriais e da comunalidade resultaram na escolha das variáveis grau alcoólico, acidez e ésteres para o fator 1, álcool superior total e isoamílico para o fator 2 e o aldeído e álcool metílico para o fator 3 (Figura 9).

Figura 9. Gráfico de pesos das cargas fatoriais finais após a rotação da matriz A_2 para as cachaças produzidas nas regiões de Salinas e Paraty



GL: Grau Alcoólico; **Aci:** Acidez; **Est:** Ésteres; **Ald:** Aldeído **Atot:** Alcoóis Superiores Totais; **AIA:** Álcool Superior Isoamílico **Amet:** Álcool Metílico

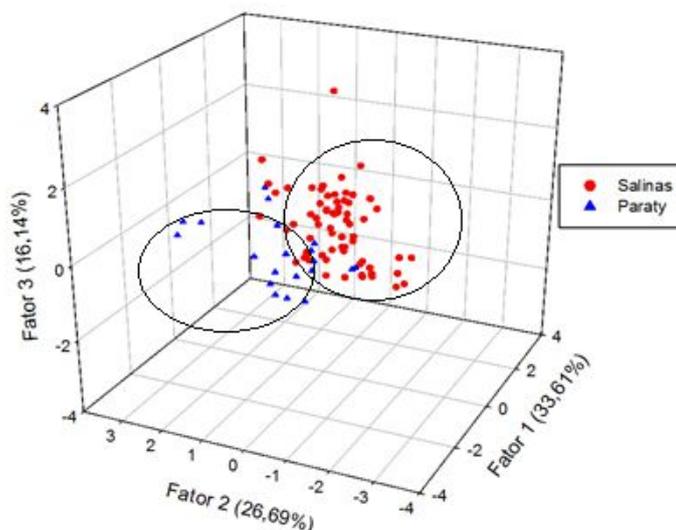
A seleção do fator 1 corrobora a fundamentação apresentada na matriz A_1 que relatou à importância dos ácidos e ésteres na caracterização das regiões. Quanto ao grau alcoólico tem-se que esta variável está associada com a formação de novos compostos, e, portanto, é muito importante no processo de agrupamento das amostras.

Para o fator 2, foram escolhidos os alcoóis superiores que são oriundos, em sua grande maioria, das reações de degradação de aminoácidos, que ocorrem durante o processo de fermentação. Esse tipo de reação química que ocorre com alguns aminoácidos, pode explicar a formação dos principais alcoóis superiores, como o álcool d-amílico a partir da d-leucina, o álcool isoamílico a partir da leucina e o álcool isobutílico a partir da valina, os quais apresentam odores característicos frequentemente encontrados em bebidas (YOKOYA, 1995). Este fator também está associado a etapa de fermentação.

Já para o fator 3, tem-se, além do aldeído citado anteriormente, o álcool metílico. Este composto em vinho e por consequência em destilados tem sua origem na hidrólise de matérias pécticas. Como o teor de matérias pécticas em cana-de-açúcar é muito baixo, é, portanto, um dos menores componentes orgânicos (HORII, 2004). A molécula de pectina é um polímero, formado pela associação de centenas de moléculas de ácido galacturônico, que possuem fragmentos de moléculas de metanol, as quais são liberadas durante o processo de fermentação, através das reações de hidrólise ácidas ou enzimáticas (MAIA, 1994). Assim, o fator 3 tem alta relação com plantação, colheita e higienização da cana-de-açúcar e deve ser controlado já que é um parâmetro relacionado a intoxicação dos consumidores, uma vez que no organismo humano o metanol é oxidado a ácido fórmico e posteriormente a gás carbônico, provocando acidose grave (diminuição do pH sanguíneo), afetando o sistema respiratório, podendo levar ao coma e até mesmo à morte (MAIA, 1994).

Na figura 10 está a dispersão das amostras e também há pontos distribuídos desuniformemente em função das peculiaridades de cada região, como solo, clima, leveduras endêmicas, entre outros.

Figura 10. Gráfico de escores das cachaças produzidas nas regiões de Salinas e Paraty em relação aos três primeiros fatores (Fator 1, Fator 2 e Fator 3) da matriz A_2 .



3.3 Análise de Correlações Canônicas

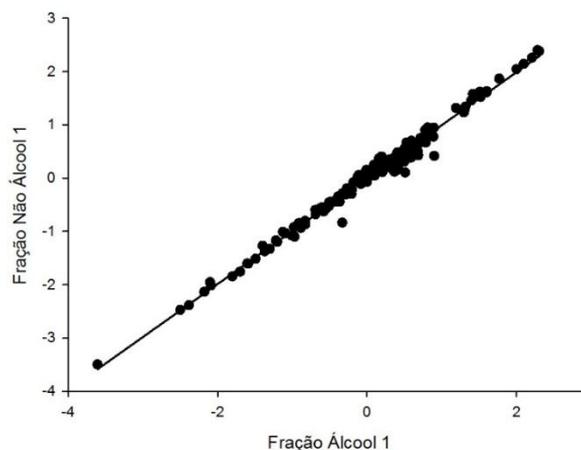
Para realização desta análise, foi feita a divisão das variáveis da matriz A_1 ($m=142$; $n=7$) em dois grupos. Optou-se pela separação em fração álcool e fração não álcool, para avaliar a influência dos alcoóis na formação dos principais compostos que conferem aroma e gosto à cachaça. Estes produtos secundários da segunda fração são as substâncias formadas durante a fermentação alcoólica e o envelhecimento da cachaça. Eles são encontrados em pequenas quantidades e muitas vezes são difíceis de serem determinados ou quantificados (YOKOYA, 1995). A natureza e proporção desses compostos são determinadas, em diferentes níveis, pelas características da matéria-prima, da fermentação, da destilação e do envelhecimento. Os principais compostos secundários formados durante a fermentação alcoólica pertencem às classes funcionais dos ácidos, ésteres e aldeídos, e em proporções menores formam-se também cetonas, compostos fenólicos, aminas e mercaptanas (MAIA, 1994).

Na sequência foi feita a avaliação da estatística das correlações canônicas ($p<0,05$), sendo que duas das três correlações foram significativas com destaque para a primeira correlação que explica 99,45% da variabilidade dos dados e apresentou uma correlação de 0,9936. Abaixo estão descritas as equações que foram utilizadas para determinar a correlação canônica do primeiro par de variáveis (Equação 8 e 9) e o gráfico que representa esta correlação. (Figura 11).

$$Y1_{\text{Fração álcool}} = -0,0130 * GL + 1,0050 * ASTot - 0,0194 * Amet \quad (8)$$

$$Y1_{\text{Fração não álcool}} = -0,6722 * aci - 0,5598 * Est - 0,0832 * Ald + 1,4183 * SCV \quad (9)$$

Figura 11. Correlação do primeiro par de variáveis canônicas obtida para a matriz A_1 .



Como resultado obteve-se uma correlação próxima de 100%, indicando que há durante o processo de produção, uma associação entre estas variáveis que resulta em uma dependência linear entre os compostos dos pares escolhidos, ou seja, quanto maior a fração álcool, maior a concentração dos compostos secundários que são fundamentais para a qualidade da cachaça. Isto, provavelmente, ocorre em função das inúmeras reações químicas que utilizam a molécula de álcool para produção dos congêneres durante a fermentação, destilação e envelhecimento. A título de exemplo tem-se na fermentação, as reações realizadas pela ação de micro-organismos sobre os açúcares presentes no caldo; na destilação, por meio do calor e da ação catalítica do cobre, as reações de formação e transformação de compostos, além das que acontecem no envelhecimento.

Foi possível, ainda, através dos escores dos dois primeiros pares de variáveis canônicas das frações álcool e não álcool, representar graficamente a distribuição das amostras. O resultado foi semelhante ao encontrado nas análises de componentes principais e de fatores, pois nas figuras 12 e 13 há a formação de perfis com pontos sobrepostos. Estes podem ser explicados pelas especificidades de cada região conforme discussão já apresentada.

Figura 12. Representação canônica das observações obtidas pelas variáveis da fração álcool da matriz A_1

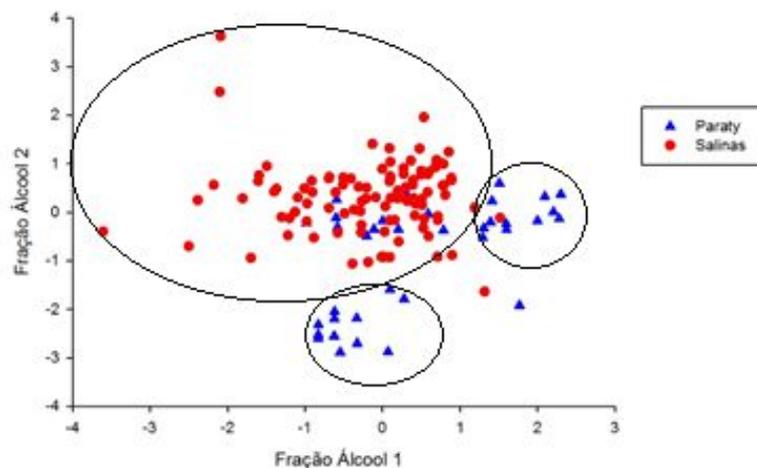
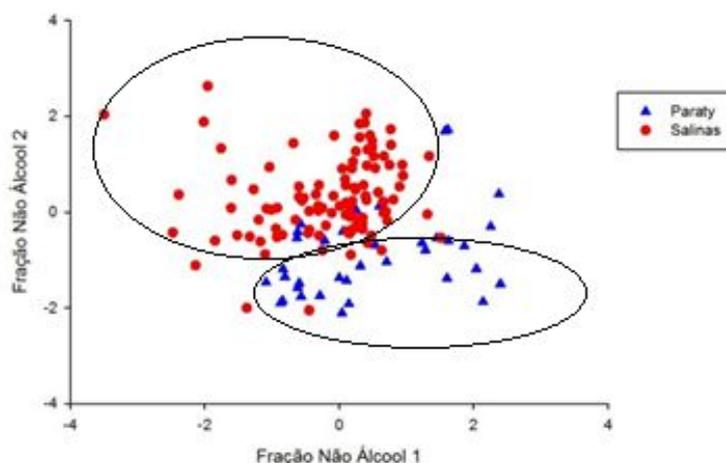


Figura 13. Representação canônica das observações obtidas pelas variáveis da fração não álcool da matriz A_2

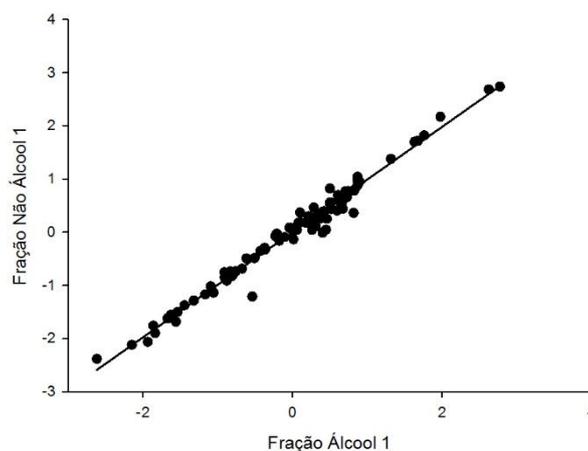


A mesma premissa para separação de variáveis foi utilizada para o tratamento da matriz A_2 construída com $m=91$ e $n=10$. Das quatro correlações obtidas tem-se que três foram significativas, ($p<0,05$), com destaque para a primeira que retém 95,84% da variabilidade dos dados e tem uma correlação de 0,9882. A seguir estão apresentadas as equações utilizadas para geração dos resultados.

$$Y1_{\text{Fração álcool}} = -0,0689 * GL + 0,7958 * ASTot + 0,0618 ASnP + 0,0045 * ASIsOB + 0,1850 ASIsOA - 0,0228 * Amet \quad (10)$$

$$Y1_{\text{Fração não álcool}} = -0,7137 * aci - 0,6243 * Est - 0,0604 * Ald + 1,7427 SCV \quad (11)$$

Figura 14. Correlação do primeiro par de variáveis canônicas obtida para a matriz A_2 .



Estes dados também resultaram em uma correlação alta, mostrando que, há uma dependência entre as variáveis pertencentes as duas frações (Figura 14). Nas figuras 15 e 16 estão representadas as distribuições das amostras e verificou-se a formação de agrupamentos naturais, com divisão de subgrupos dentro da mesma região, indicando que as peculiaridades do processo produtivo podem interferir na separação das amostras.

Figura 15. Representação canônica das observações obtidas pelas variáveis da fração álcool da matriz A_2

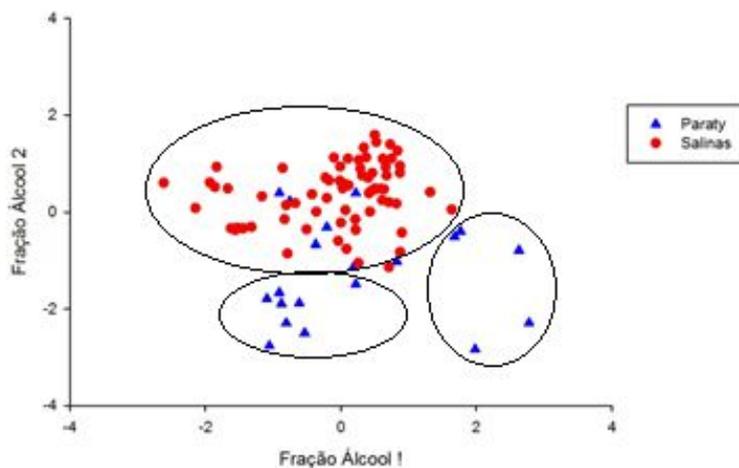
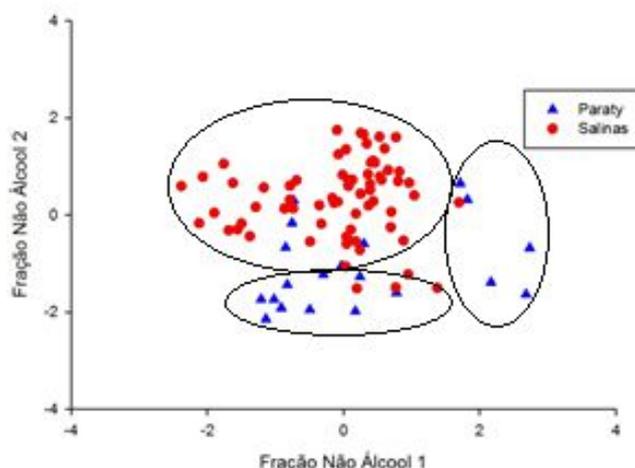


Figura 16. Representação canônica das observações obtidas pelas variáveis da fração álcool da matriz A_2



3.4 Análise Discriminante

Esta técnica teve a finalidade de separar as amostras da população em dois grupos, a saber, Salinas e Paraty a partir das características de cada região. Em razão disto foram determinadas duas funções discriminatórias como combinações lineares das variáveis utilizadas no estudo, a fim de maximizar a razão entre as classes e minimizar a dispersão dentro de cada grupo (Equações 12 e 13). Estas funções fornecem, ainda, informações sobre a relação das variáveis com as regiões, sendo assim, de acordo com os maiores coeficientes, a soma dos componentes voláteis, acidez, ésteres e o álcool superior total estão associadas com Paraty, enquanto que a soma dos componentes voláteis e o álcool superior total estão relacionadas com Salinas.

$$Y_{Paraty} = -1,1718 + 1,0170 * GL - 2,9385 * aci - 2,4120 * Est + -0,47882 * Ald - 2,1457 * ASTot - 0,0599 * Amet + 4,3442 * SCV \quad (12)$$

$$Y_{Salinas} = -0,2360 - 0,4370 * GL - 0,2564 * aci - 0,7138 * Est - 0,1120 * Ald - 1,5848 * ASTot - 0,1253 * Amet + 1,7781 * SCV \quad (13)$$

No que diz respeito à capacidade de treinamento, de acordo com a tabela 3, a taxa de classificação foi de 76,92% e 87,67% para as cachaças de Paraty e Salinas, respectivamente. Em relação ao procedimento de validação, que foi utilizado para avaliar o desempenho da

classificação, a capacidade de previsão das duas regiões foi de 100% e 83,87%, respectivamente.

Esta classificação satisfatória, é muito importante, uma vez que trata-se de análises realizadas frequentemente pelos produtores, pois são exigidas pelo MAPA. Após o tratamento dos dados obtidos pode-se conseguir, a determinação da origem, ou seja, a definição da região de produção. Este processo além de agregar valor, impulsiona a comercialização das cachaças e inibe a ocorrência de fraudes.

Tabela 3. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaças da matriz A_1 .

Grupo	Paraty T/V	Salinas T/V	Total T/V
Paraty	20/12 76,92/ 100,00	6/0 23,08/0,00	26/12 100,00
Salinas	9/5 12,33/16,13	64/26 87,67/ 83,87	73/31 100,00

Na tabela 4 estão apresentados os resultados para matriz A_2 com $m=91$ e $n=10$ e constatou-se uma taxa de classificação de 76,47% no treinamento e 93,62% na validação para a região de Paraty e 100,00% e 95,65% para a região de Salinas. Para a previsão dos dados foram utilizadas as equações 14 e 15. Segundo os maiores coeficientes destas equações é possível dizer que o álcool superior total e o isobutílico estão correlacionados com Paraty, ao passo que o álcool superior total e isoamílico estão vinculados à Salinas.

$$Y_{Paraty} = -2,5811 + 1,5408 * GL - 0,9164 * aci - 1,9441 * Est - 0,9798 * Ald + 5,9432 * ASTot - 2,2171 * ASnP - 3,4955 * ASIsoB - 2,0222 * ASIsoA + 0,2418 * Amet + 1,9992 * SCV \quad (14)$$

$$Y_{Salinas} = -0,2096 - 0,5144 * GL + 0,5680 * aci + 0,8753 * Est + 0,30747 * Ald - 4,0678 * ASTot + 2,0051 * ASnP + 1,342 * ASIsoB + 2,6646 * ASIsoA + 0,0327 * Amet - 1,3745 * SCV \quad (15)$$

A introdução das medidas dos alcoóis superiores n-propílico, isobutílico e isoamílico proporcionaram o aumento, na etapa de validação, do número de amostras alocadas adequadamente. Este fato concorda com a ideia de que estes compostos devem ser usados para caracterizar e diferenciar as cachaças produzidas nas diferentes regiões, tendo em vista

que os alcoóis superiores totais e n-propílico estão associados a região de Salinas enquanto o álcool superior total, o isobutílico e o metílico estão associados com a região de Paraty, como atesta as análises de ACP.

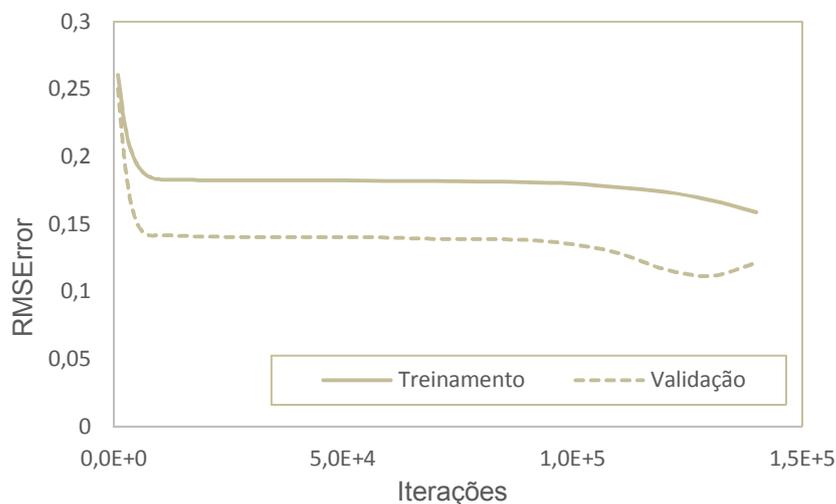
Tabela 4. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaças da matriz A_2 .

Grupo	Paraty T/V	Salinas T/V	Total
Paraty	13/4 76,47/ 100,00	4/0 23,53/0,00	100,00
Salinas	3/1 6,38/4,35	44/22 93,62/ 95,65	100,00

3.5 Rede Neural

Para a matriz A_1 com ($m=142$; $n=7$), foi feita a avaliação de 100 até 100.000 iterações em intervalos de 100 (Figura 17). Este procedimento teve como finalidade acompanhar o erro de validação que diminuiu um pouco durante a fase inicial de treinamento, assim como o erro do conjunto de treinamento. No entanto, quando a rede começou a ajustar os dados, o erro de validação definido começou a subir, embora o erro de treinamento continuasse diminuindo. Quando o erro de validação aumentou, após 140.000 iterações, o treinamento foi interrompido, pois neste a rede decorou e já não aprende mais ou absorveu os ruídos

Figura 17. RMSError x Número de iterações para dados de treinamento e validação da matriz A_1



O desenvolvimento do modelo de RNA envolve duas etapas básicas, aprendizagem e validação. O preceito para a seleção da melhor topologia foi baseado na escolha do número de neurônios e camadas ocultas que apresentaram menor valor de RMSE na validação e uma taxa máxima de classificação (Figuras 18 e 19).

Obteve-se a melhor configuração da RNA com nove neurônios na primeira camada e seis na segunda com um coeficiente de RMSE de 0,1589 e 0,1120 e uma taxa de classificação para Salinas e Paraty de 0,9863 e 0,7308, 1,0000 e 1,0000 para o conjunto de treinamento e validação, respectivamente.

Figura 18. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de treinamento da matriz A_I .

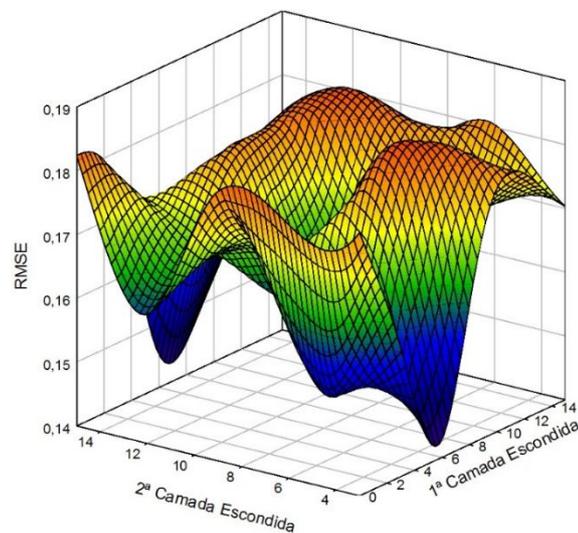
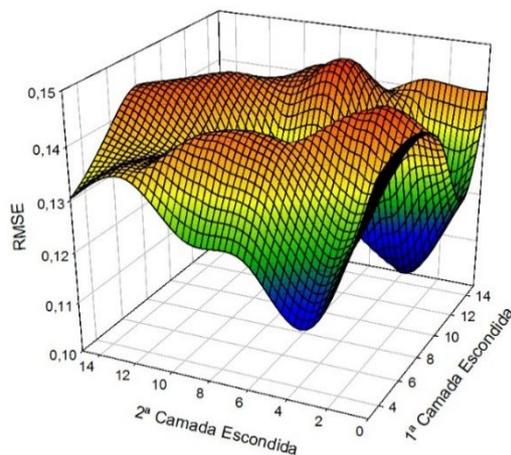


Figura 19. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de validação da matriz A_I .

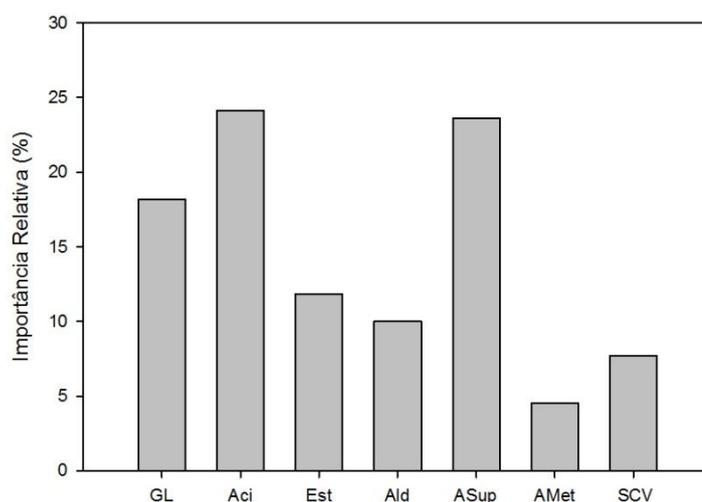


De acordo com Cybenko (1989), uma camada oculta é suficiente para aproximar qualquer função contínua, enquanto duas camadas escondidas são suficientes para aproximar qualquer função matemática. Pode ser observado que, em geral, quando apenas uma camada escondida é utilizada, há um valor de RMSE mais elevado, o que implica que a função utilizada para classificar as cachaças é provavelmente uma função descontínua.

A importância relativa de cada variável na classificação das amostras de cachaça das regiões de Salinas e Paraty está apresentada na figura 20. Esta ferramenta apresentou uma análise de pesos entre a camada de entrada e outras camadas. A altura de cada barra do gráfico mostra facilmente a contribuição relativa efetuada por cada elemento de processamento da camada de entrada (CE), com a barra mais alta que indica o maior contribuinte.

Sendo assim, constatou-se que a acidez e o álcool superior total foram as variáveis com maior contribuição, seguida do grau alcoólico e ésteres. Este resultado corrobora os obtidos nas análises não supervisionadas (ACP, AF e ACC), pois, mostrou a relevância destes mesmos compostos na identificação das regiões de produção da bebida conforme discussão apresentada anteriormente.

Figura 20. Importância relativa (%) de cada variável na classificação de amostras de cachaça da matriz A_1 .



Para a matriz A_2 com $m=91$ e $n=10$, foi feito o mesmo monitoramento do número de iterações e do valor do RMSE visto na matriz A_1 (Figura 21). Neste caso específico, a análise final foi realizada com 120.000 iterações e após testar as configurações, optou-se pela topologia que continha nove neurônios na primeira camada oculta e seis na segunda uma vez

que esta apresentou maior taxa de classificação e menor RMSE. Para o conjunto de dados de treinamento e validação, os coeficientes de RMSE foram 0,1036, 0,0978 a taxa de classificação para esta configuração ideal foram 0,96, 1,0 e 1,0 e 1,0, para Salinas e Paraty, respectivamente (Figura 22 e 23).

Esta metodologia apresentou melhores resultados que a análise discriminante, e tal fato pode ser explicado pela função utilizada, já que a AD usa uma função linear para descrever os resultados e a rede utiliza uma função que melhor se adequa aos dados. Desta forma, esta metodologia resultou em uma melhor separação das amostras de acordo com as regiões estudadas.

Figura 21. RMSError x Número de iterações para dados de treinamento e validação da matriz A_2

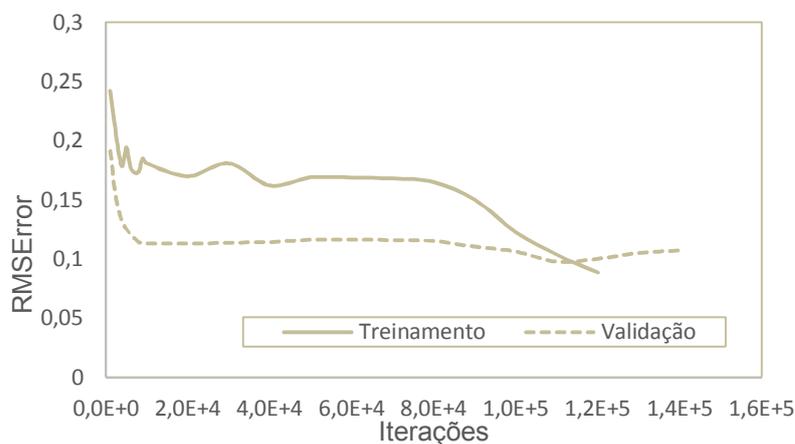


Figura 22. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de treinamento da matriz A_2 .

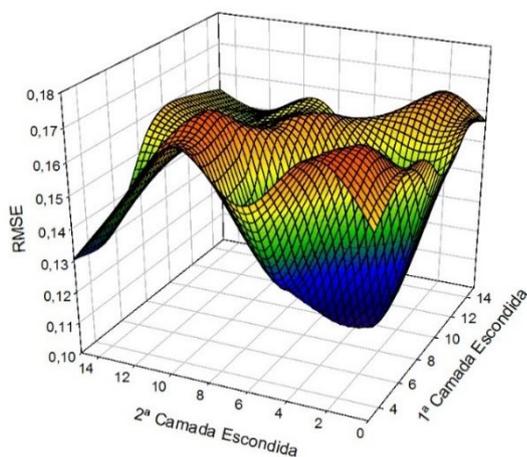
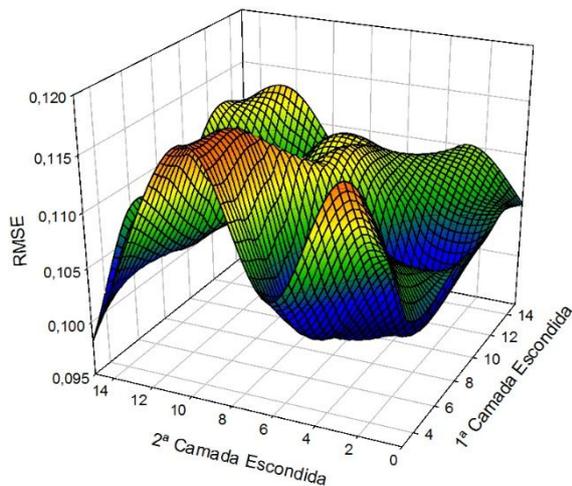
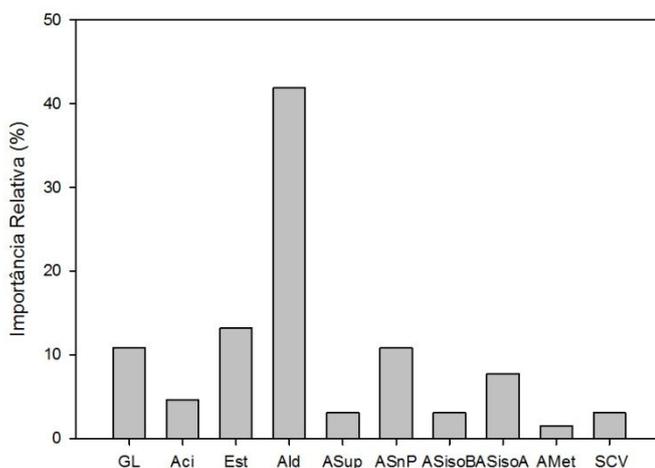


Figura 23. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de treinamento da matriz A_2 .



A participação efetiva de cada variável no processo de identificação das amostras de cachaça é apresentada na figura 24. É notório o destaque da contribuição do aldeído neste processo, uma vez que é a barra mais alta, logo em seguida estão o éster e o grau alcoólico. Estas variáveis foram citadas na discussão anterior que trata das peculiaridades de cada região, mas neste caso os alcoóis superiores tiveram pequena participação, indicando que a introdução de novas variáveis nesta técnica não gerou o mesmo tipo de informação vista nos métodos anteriores.

Figura 24. Importância relativa (%) de cada variável na classificação de amostras de cachaça da matriz A_2 .



4 CONCLUSÃO

Os resultados indicaram que as técnicas não supervisionadas foram eficientes na formação de agrupamentos com registro de alguns pontos distribuídos desuniformemente em função das singularidades dos processos de produção de Salinas e Paraty. Quanto às variáveis que possuem maior influência na identificação das regiões, pode se destacar a acidez, ésteres para matriz A_1 e os alcoóis superiores na matriz A_2 . Considerando que as análises físico-químicas utilizadas são relativamente simples e realizadas com frequência pelos produtores, obteve-se uma excelente classificação na etapa de validação das técnicas supervisionadas com valores entre 83,87 e 100% com ênfase para a rede neural que apresentou 100% de separação na validação para as duas matrizes.

Sendo assim, considerando a complexidade das amostras analisadas pode-se concluir que as técnicas utilizadas foram adequadas para separação das cachaças de acordo com as regiões de produção, com ênfase para análise de fatores e a rede neural artificial que foram as metodologias com melhores resultados e, portanto, são as mais indicadas para este tipo de conjunto de dados.

5 REFERÊNCIAS

- ASSOCIAÇÃO DOS PRODUTORES ARTESANAIS DE CACHAÇA DE SALINAS (APACS). **Projeto XII festival mundial da cachaça de Salinas/MG 12 a 14 julho / 2013**. Salinas.
- BRAGA, A.P., CARVALHO, A.C.P.L.F., LUDERMIR, T.B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC Editora S.A, 2000. 262p.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. **Indicação Geográfica valoriza produtos agropecuários**. Brasília, DF, 10 jan. 2011. Disponível em: < <https://www.fao.org.br/igvpa.asp>>. Acesso em: 21 de out. 2015.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa nº 13, de 29 de junho de 2005a**. Aprova o regulamento técnico para fixação dos padrões de identidade e qualidade para aguardente de cana e para cachaça. Diário Oficial da República Federativa do Brasil. Brasília, DF, 30 Jul. 2005 Disponível em: <<http://www.agricultura.gov.br/>>. Acesso em: 04 de mai. 2015.

- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. **Instrução Normativa nº 24, de 08 de Setembro de 2005b**. Manual Operacional de Bebidas e Vinagres. Diário Oficial da União, Brasília, DF, 20 set.2005. Seção 1, p.11. Disponível em: <<http://www.agricultura.gov.br/>>. Acesso em: 04 de mai. 2015.
- CARDOSO, M. das G. **Produção de aguardente de cana-de-açúcar**. Lavras: Editora UFLA, 2006. 444p
- CARNEIRO, J. C. S. **Análise estatística multivariada aplicada à avaliação sensorial de alimentos**. 2005. 89f. Tese (Doutorado em Ciência e Tecnologia de Alimentos) – Universidade Federal de Viçosa, Viçosa, MG.
- CYBENKO, G. Approximation by superpositions of a sigmoid function. **Mathematics of Control, Signals and Systems**, v. 2, p. 303-314, 1989.
- FERREIRA, D.F. **Estatística multivariada**. Lavras: Editora UFLA, 2011. 675p.
- GUTTMAN, L. **A new approach to factor analysis: The radex**. In P.F. Lazarsfeld (Ed.) *Mathematical Thinking in the Social Sciences*. New York: Free Press, 1954
- HAYKIN, S. **Neural networks – a comprehensive foundation**, 2nd edition. Pearson Education, Inc., Delhi, 823p, 1999.
- HAIR JR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate Data Analysis**. 7.ed. Pearson Prentice Hall, 2010.
- HELPER, G. A.; FERRÃO, M. F.; FERREIRA, C. V; HERMES, N. Aplicação de métodos de análise multivariada no controle qualitativo de essências alimentícias empregando espectroscopia no infravermelho médio. **Ciência e Tecnologia dos Alimentos**, v. 26, n. 4, p. 779-786, 2006.
- HORII, J. **A cana de açúcar como matéria-prima**. Piracicaba: ESALQ, Departamento de Agroindústria, Alimentos e Nutrição, 2004, 27p.
- JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. New Jersey: Prentice-Hall, 1999.816p.
- JOLLIFFE, I. T. Discarding Variables in a Principal Component Analysis. I: Artificial Data. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, v.21, n.2, p.160-173, 1972.

- MAIA, A.B.R.A. Componentes secundários da aguardente. **STAB Açúcar Álcool e Subprodutos**, v.12, n.6, p. 29-34, 1994.
- MAIA, A.B.R.A.; CAMPELO, E.A.P. **Tecnologia da cachaça de alambique**. Belo Horizonte: SEBRAE/MG; SINDBEBIDAS, 2005. 129 p.
- MINAI, A. A., WILLIAMS, R. D. **Back-propagation heuristics: A study of the extended delta-bar-delta algorithm,**” Proc. IEEE Int. Joint Conf. Neural Networks, San Diego, USA, 595–600, 1990.
- NYKÄNEN, L.; NYKÄNEN, I. Rum Flavour. In: PIGGOTT, J. R. (Ed) Flavour of distilled beverages: origin and development. **Ellis Horwood Limited**, Cap. 3, p .49-63, 1983.
- RAI, P., MAJUMDAR, G.C., DASGUPTA, S., DE, S. Prediction of the viscosity of clarified fruit juice using artificial neural network: a combined effect of concentration and temperature. **Journal of Food Engineering**, v. 68, p. 527-533, 2005.
- YOKOYA, F. **Fabricação da cachaça de cana**. Campinas: Fundação Tropical de Pesquisas e Tecnologia André Tosello, 1995. 92 p.

CAPÍTULO 3 – INVESTIGAÇÃO DE CACHAÇAS ENVELHECIDAS EM DIFERENTES MADEIRAS POR ANÁLISE QUIMIOMÉTRICA

RESUMO

Diante do grande potencial social e econômico da cachaça e do fortalecimento do mercado consumidor da bebida, os produtores têm buscado ferramentas para melhoria da qualidade e opções para diminuição de falsificações. Uma alternativa para identificar as madeiras utilizadas no envelhecimento da cachaça, a fim de gerar conhecimento sobre variáveis importantes no processamento da bebida e, por consequência ter subsídio para evitar possíveis fraudes, é a quimiometria. No presente trabalho objetivou-se identificar as variáveis mais importantes no processamento de cachaças envelhecidas em diferentes madeiras utilizando técnicas multivariadas e os dados de composição físico-químicos exigidos pela legislação Brasileira. O tratamento das matrizes de dados foi feito utilizando as seguintes técnicas estatísticas: análise de componentes principais (ACP), fatores (AF), correlações canônicas (ACC), discriminante (AD) e a técnica não estatística, rede neural artificial (RNA). O processo de classificação utilizando a ACP e AF das cachaças envelhecidas nas madeiras, Amburana, Bálsamo e Carvalho mostrou que as melhores separações ocorreram com as combinações Amburana e Carvalho e Amburana e Bálsamo. No entanto devido a questões relacionadas ao processamento não houve separação total dos grupos. A Análise de Correlações Canônicas atingiu 94,84% de correlação, mas a separação por grupo obteve resultados semelhantes aos observados na ACP e AF. O resultado da ACP e AF foi utilizado, ainda, para realização da Análise Discriminante (AD). A AD e Redes Neurais Artificiais (RNA) apresentaram, na validação, uma porcentagem de acerto que variou entre 10,0 e 87,5%. As variáveis que possuem maior importância na classificação das cachaças segundo ACP, AF e RNA são acidez, GL e alcoóis superiores.

ABSTRACT

Before the great social and economic potential of cachaça and the strengthening of consumer beverage market, producers have sought tools to improve the quality and options for reduction of fakes. An alternative to identify woods used the aging cachaça, to generate knowledge about important variables in the processing of the beverage, and consequently have allowance to avoid possible fraud is chemometrics. In the present study aimed to identify the most important variables in the processing of cachaça aged in different woods using multivariate techniques and the physicochemical composition data required by Brazilian law. The treatment of data arrays was done using the following statistical techniques: principal components analysis (PCA), factors (FA), canonical correlation (CCA), discriminant (DA) and non-statistical technique, artificial neural network (ANN). The classification process using PCA and FA of aged cachaça in the woods, Amburana, Balm and Carvalho showed that the best separation occurred with Amburana and Carvalho and Amburana and Balm. However due to issues related to the processing no total separation of the groups. The canonical correlation analysis reached 94.84% correlation, but the separation group obtained similar results to those observed in PCA and FA. The result of the PCA and FA was used also to perform the discriminant analysis (DA). The DA and Artificial Neural Networks (ANN) presented at validation, a percentage of correct answers ranging between 10.0 and 87.5%. The

variables that have the greatest importance in the classification of cachaça second PCA, FA and ANN are acidity, AD and higher alcohols

1 INTRODUÇÃO

Os setores públicos e privados têm destinado grande atenção ao mercado produtor de cachaça, uma vez que vislumbram o seu grande potencial social e econômico e, portanto, investem no fortalecimento do meio e no aumento do mercado consumidor da bebida.

Neste sentido, a expansão do comércio da cachaça passa por uma melhor adequação diante das exigências do consumidor nacional e internacional, além de cumprir todos os requisitos das leis federais e estaduais. É possível adquirir equipamentos mais eficientes, embalagens mais atrativas e desenvolver novas tecnologias, mas a valorização do produto final está intimamente ligada à diminuição de fraudes e da garantia de procedência. Assim, uma forma de obter informações sobre a bebida, para facilitar este processo de identificação, é a geração de conhecimento através de pesquisas científicas.

Outra questão relevante para a expansão do comércio da cachaça é o envelhecimento destas bebidas em barris de madeiras que tornam a cachaça mais agradável ao paladar, aumenta o seu valor comercial e caracteriza a comercialização por região. Como exemplo, tem-se que a região de Salinas, tradicionalmente, distribui a maior parte envelhecida em Amburana e Bálsamo e Paraty tem comercializado a bebida envelhecida em tonéis de Amendoim e Jequitibá, além do envelhecimento no Carvalho. Sendo assim, devido à diversidade da flora brasileira, não só o carvalho, que é importado, (*Quercus sp*) é usado para o envelhecimento, mas também várias outras madeiras, como, amendoim (*Pterogyne nitens*), bálsamo (*Myroxylon peruiferum*), Jequitibá (*Cariniana estrellensis*) e Amburana (*Amburana cearensis*) que são extraídas da natureza segundo as leis ambientais vigentes (BERNARDES et al, 2014).

A aguardente de cana recém-destilada é incolor e apresenta sabor picante, áspero e seco, características que agradam sensorialmente apenas um pequeno grupo de consumidores (MAÇATELLI, 2006; CARDELLO e FARIA, 1998). Por meio do envelhecimento é possível observar a diminuição significativa do sabor alcoólico e da agressividade da bebida, com simultâneo aumento da doçura e do sabor de madeira, proporcionando a melhora sensorial do produto (CARDELLO e FARIA, 1999; ODELLO et al, 2009).

Uma alternativa para identificar as madeiras utilizadas no envelhecimento da cachaça, a fim de gerar conhecimento sobre variáveis importantes no processamento da bebida e, por consequência ter subsídio para evitar possíveis fraudes, é a quimiometria. O

modelo estatístico baseado em métodos multivariados, utilizado por esta disciplina, considera a correlação entre muitas variáveis analisadas simultaneamente, permitindo a extração de uma quantidade muito maior de informação (SENA et al, 2000). Esta técnica matemática é uma poderosa ferramenta e já foi utilizada para o reconhecimento de padrões e identificação de fraude alimentar (LAVINE e WORKMAN, 2002; CEVOLI et al., 2011, CIMPOIU et al., 2011) e para a classificação de alguns tipos de cachaça (RECHE e FRANCO, 2009; GRANATO et al, 2014; SERAFIM et al, 2013; SERAFIM et al, 2012; DE SOUZA et al, 2007, DE SOUZA et al, 2009, FERNANDES et al, 2005). Técnicas como a análise de fatores (AF), análise de correlações canônicas (ACC), análise discriminante (AD), e rede neural artificial (RNA) são comumente usadas para fins de autenticação (PILLONEL et al., 2005).

Objetivou-se no presente trabalho identificar as variáveis mais importantes no processamento e promover a classificação de cachaças envelhecidas em diferentes madeiras aplicando técnicas multivariadas aos dados de composição físico-químicos exigidos pela legislação Brasileira.

2 MATERIAL E MÉTODOS

2.1 Coleta dos Dados

Para esta pesquisa optou-se pelas regiões de Salinas e Paraty que possuem o selo de Indicação Geográfica. Os dados foram obtidos com os produtores através de visitas *in loco*. Foram visitados 17 produtores que comercializam 36 marcas em Salinas e cinco produtores responsáveis por 15 marcas em Paraty. É importante salientar que existem apenas seis produtores formalizados na região de Paraty.

O período de coleta foi de Junho de 2014 a Fevereiro de 2015 e os produtores disponibilizaram as informações através de 171 certificados oficiais de análises físico-químicas que são arquivados para apresentação aos fiscais do Ministério da Agricultura, Pecuária e Abastecimento (MAPA). As amostras de cachaça utilizadas para elaboração dos laudos foram analisadas quanto a sua composição físico-química, seguindo os parâmetros estabelecidos pela IN nº 13, de 29/6/2005 e a IN nº 24, de 08/09/2005 do MAPA por laboratórios credenciados (BRASIL, 2005ab).

2.2 Delineamento Experimental dos Dados

A matriz de dados apresentou dez medidas de diferentes propriedades físico-química ($n=10$), a saber, grau alcoólico, acidez volátil, ésteres totais, aldeído, alcoóis superiores totais, álcool metílico, soma dos componentes voláteis e os alcoóis n-propílico, isobutílico e isoamílico obtidos da estratificação dos alcoóis superiores totais. Estas foram realizadas em 70 amostras ($m=70$) de cachaças envelhecidas nas madeiras Amburana, Bálsamo e Carvalho (Tabela 1). Com o objetivo de avaliar o comportamento das madeiras por pares, a matriz original dos dados foi desmembrada nas seguintes matrizes, Amburana e Bálsamo (A_{11}), Amburana e Carvalho (A_{12}), Bálsamo e Carvalho (A_{13}).

Tabela 1. Número de amostras por madeira e amplitude das variáveis utilizadas para construção da matriz A_2

	Amburana	Bálsamo	Carvalho	Total
Dados	27	29	14	70
GL°	40,3-43,0	40,1-48,0	40,7-59,1	40,1-59,1
Acidez*	50,0-141,3	39,6-121,1	14,0-150,0	14,0-150,0
Ésteres*	19,0-128,8	16,0-137,3	15,0-112,1	15,0-137,3
Aldeído*	1,8-11,0	5,0-19,5	5,0-13,8	1,8-19,5
ASTot*	146,0-315,5	123,1-282,7	192,0-360,0	123,1-360,0
ASnP*	25,0-98,8	24,0-97,5	11,0-75,0	11,0-98,8
ASIsoB*	34,0-56,7	27,3-61,0	30,0-57,0	27,3-61,0
ASIsoA*	85,0-176,8	71,4-184,0	112,0-262,0	71,4-262,0
Amet*	0,0-7,2	0,0-12,9	1,0-11,4	0,0-12,9
SCV*	272,0-556,2	243,0-505,0	235,0-514,0	235,0-556,2

* mg/100 mL etanol

2.3 Medidas de Dispersão e Padronização da Matriz de Dados

Foram realizadas as determinações de algumas medidas de dispersão, a saber, desvio padrão e coeficiente de variação. A partir deste resultado, fez-se a padronização dos dados, que consistiu na implementação de uma nova variável (Z) com média zero e desvio padrão igual a 1, eliminando as diferenças entre as unidades de medidas das variáveis estudadas. (Equação 1).

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Onde Z: nova variável (forma padrão)

X: variável

μ : média

σ : desvio padrão

2.4 Análises Estatísticas Multivariadas dos Dados

Para realização das análises estatísticas multivariadas foram utilizados os procedimentos descritos em SAS, do pacote computacional SAS, versão 9.0 e, nos anexos estão apresentados os comandos utilizados para os conjuntos de dados.

2.4.1 Análise de Componentes Principais

A partir do conjunto de médias padronizadas obtido no item 2.3, foi realizada a transformação das variáveis Z em um novo sistema de coordenadas com eixos ortogonais, dando origem aos CP's, obtidos através de combinações lineares (Equação 2).

$$CP_i = a_{i,1}Z_1 + a_{i,2}Z_2 + \dots + a_{i,n}Z_n \quad (2)$$

com $i = 1, 2, \dots, n$, em que:

CP_i : i-ésimo componente principal;

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

Na sequência, fez-se a interpretação de cada componente através da avaliação das variáveis originais mais importantes em cada eixo. O programa utilizado forneceu as CP's em ordem decrescente de máxima variância, o que permitiu a redução da dimensionalidade por meio da seleção de três CP's, que é um número bem menor do que o de variáveis originais para representar o conjunto inicial dos dados que contém sete variáveis para a matriz A_1 e dez para matriz A_2 . Isto foi possível porque os últimos componentes são responsáveis pela explicação de uma pequena porção da variância disponível.

Para a escolha do número de componentes principais foi considerado o critério de fatores interpretáveis e o critério de Kaiser, que selecionaram os primeiros k componentes principais que retêm, juntos, a maior proporção de variância dos atributos originais (acima de 70% da variância) e autovalor maior que um. Já para a seleção de variáveis, avaliou-se no componente de maior autovalor, as variáveis que apresentam, em valor absoluto, maior correlação. Posteriormente, deu-se prosseguimento a análise do segundo e terceiro componentes para escolha das variáveis que não foram previamente escolhidas (JOLLIFFE, 2002; FERREIRA, 2011).

2.4.2 Análise de Fatores

Nesse estudo, optou-se pelo uso do modelo fatorial e o número de fatores (m) foi escolhido a partir de autovalores maiores que um, conforme Johnson e Wichern (1999) e Guttman (1954). A fim de estabelecer o número de fatores, também utilizou-se como critério, aqueles cuja soma explicava no mínimo 70% da variação dos dados.

Quanto ao descarte de variáveis, foram retiradas, em primeiro lugar, as que apresentaram comunalidade menor que 0,5, pois estas representaram uma pequena proporção da variância na análise. Na sequência foram eliminadas as variáveis que apresentaram cargas (correlações entre as variáveis e os fatores) inferiores a 0,4 sendo, este um limite aceitável para contribuição da variável na criação do fator. Adicionalmente, foram excluídas variáveis que apresentem cargas superior a 0,4, em mais de um fator, uma vez que uma variável não pode contribuir para dois fatores distintos, pois dificulta a produção de grupos por ela estar associada aos dois fatores (HAIR JR et al, 2010). Utilizando os critérios citados procedeu-se, à análise definitiva, na qual se especificou o número de fatores final.

Para a rotação, foi escolhido, o método ortogonal Varimax para que cada fator incluísse a maior porcentagem possível da variância existente nos atributos originais por ele representados. Foram estimadas, ainda, com a rotação, as cargas fatoriais finais e estas foram usadas para identificar os atributos que compunham cada fator (CARNEIRO, 2005).

A partir dos fatores comuns foram estimados os escores para serem utilizados na separação das cachaças, por meio do agrupamento de amostras com o mais alto grau de similaridade, em um espaço gráfico bi ou tridimensional.

2.4.3 Análise de Correlações Canônicas

A análise de correlação canônica foi realizada com o intuito de observar as relações

lineares entre dois vetores de variáveis p (fração álcool) e q (fração não álcool), a fim de avaliar a influência do álcool na formação de compostos secundários importantes para qualidade sensorial da cachaça.

Uma combinação linear foi desenvolvida em cada um dos pares de variáveis canônicas dos conjuntos de amostras (Amburana, Bálsamo e Carvalho), para que a correlação entre os três fosse maximizada e, desta forma, considerou-se as combinações lineares do primeiro par como U (fração álcool) e V (fração não álcool) expressas por:

$$\begin{aligned} U_k &= a_k' X \\ V_k &= b_k' Y \end{aligned} \tag{3}$$

tal que:

$$\text{Corr}(U_k, V_k) = \frac{a_k' \Sigma_{XY} b_k}{\sqrt{a_k' \Sigma_{XX} a_k} \sqrt{b_k' \Sigma_{YY} b_k}}$$

seja a maior possível, ou seja, $\max \text{Corr}(U, V)$, que expressa a máxima associação entre as funções lineares dos dois grupos de variáveis (JOHNSON e WICHERN, 1999).

Após a obtenção das correlações canônicas, foi realizado o teste estatístico F ($p < 0,05$), cuja hipótese de nulidade foi não há correlação entre os conjuntos de dados, para testar a significância dos resultados.

Já as cargas canônicas, que se referem às associações entre cada variável original e a correspondente variável canônica foram obtidas através das correlações estruturais.

Foi calculada, ainda, a proporção da variância total que é explicada por cada par de variáveis canônicas e isto permitiu avaliar a relevância deste par no conjunto total dos dados. Por fim, utilizou-se as observações obtidas pelas variáveis de cada fração (álcool e não álcool) para representar graficamente a distribuição das amostras por região.

2.4.4 Análise Discriminante

A análise discriminante consistiu em obter funções matemáticas capazes de classificar uma amostra de cachaça em uma das três populações π_i , ($i=1$ – Amburana; 2 – Bálsamo ou 3 – Carvalho), com base em medidas de dez variáveis, buscando minimizar a probabilidade de má classificação, isto é, minimizar a probabilidade de classificar erroneamente uma amostra em uma população π_1 , quando realmente pertence a população π_2 ou π_3 . Para isto a AD

procurou maximizar a variância entre os diferentes grupos e minimizar a variação dentro de cada um deles, a fim de selecionar o limite que estabelece máxima separação entre os três grupos. O treinamento foi a primeira etapa, sendo a parte exploratória da análise, que consistiu em procurar características capazes de serem utilizadas para alocar amostras nos três grupos previamente definidos, a partir do emprego de limites de decisão linear.

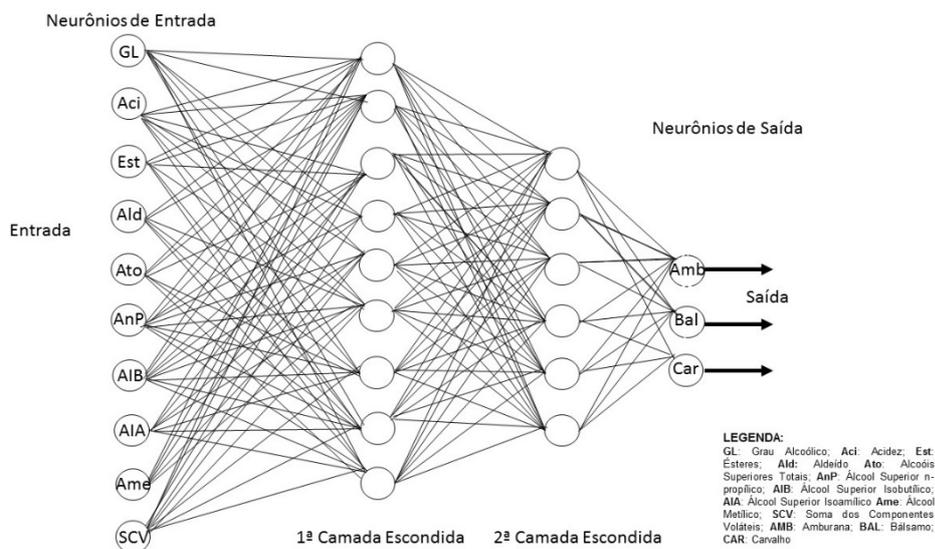
Os dados foram aleatorizados e divididos em dois grupos, um de teste e o outro de validação utilizando o algoritmo de Kenstone. Optou-se por usar a proporção 70% e 30% (Treinamento; Validação) e os dados obtidos foram analisados, para obter a função discriminante linear.

2.5 Redes Neurais Artificiais

Utilizou-se uma rede neural artificial (RNA) a fim obter a separação das cachaças em três grupos, a saber, Amburana, Bálsamo e Carvalho. Nas RNA's, cada neurônio recebeu um sinal de entrada e enviou um sinal de saída como os neurônios do cérebro para geração da informação. A força das ligações depende dos pesos aplicados em cada neurônio durante a etapa de aprendizagem.

Para este trabalho foi escolhida a metodologia de redes neurais supervisionadas denominada *perceptrons* multicamadas ou *multilayers perceptrons* (MLPs) e o algoritmo de treinamento utilizado para tanto foi o algoritmo de retropropagação do erro, conhecido na literatura como *Backpropagation Algorithm* ou simplesmente *Backprop*.

Figura 1. Algoritmo *Backpropagation*



GL: Grau Alcoólico; **Aci:** Acidez; **Est:** Ésteres; **Ald:** Aldeído **Ato:** Alcoóis Superiores Totais; **AnP:** Álcool Superior n-propílico; **AIB:** Álcool Superior Isobutílico; **AIA:** Álcool Superior Isoamílico **Ame:** Álcool Metílico; **SCV:** Soma dos Componentes Voláteis.

Para classificar os diferentes tipos de cachaças em função das propriedades físico-químicas, foi utilizado o algoritmo supracitado com a regra de aprendizado Delta-Bar-Delta-expanded (MINAI E WILLIAMS, 1990). A regra de aprendizagem utilizou uma heurística com variação da taxa e do coeficiente de aprendizado com valores de 0,3 e 0,4, respectivamente, durante todo o treinamento. Para ativação do neurônio (transferência de sinais) foi utilizada uma função hiperbólica (φ), por ser a mais utilizada para sistemas biológicos ou mais complexos (RAI et al., 2005) (Equação 4).

$$\varphi = \tanh(v_k) \quad (4)$$

onde v_k é o potencial de ativação do neurônio de ordem k (a soma dos sinais de entrada ponderados por pesos sinápticos e os desvios).

Com a finalidade de obter a melhor configuração da rede neural para classificação das amostras de cachaça, foi feito um rastreamento do número de iterações, e do valor do RMSE nas fases de treinamento e validação. Na sequência fez-se a análise de 10.000 até 160.000 e para conseguir a melhor arquitetura, foram feitos testes, dentro deste intervalo, com um total de 30 configurações que continha de 3 a 15 neurônios na primeira camada e de 0 a 15 na segunda camada.

Todas as redes testadas tiveram os neurônios de uma camada totalmente conectados com os neurônios de uma camada vizinha

A escolha do número de neurônios em cada camada escondida foi baseada na minimização do erro quadrático e maximização do coeficiente de correlação, associada à capacidade preditiva da rede (RAI et al., 2005).

A taxa de aprendizagem indicou a taxa de variação de pesos de conexão durante a formação. É válido salientar que valores baixos retardam o treinamento, enquanto grandes valores aceleram o processo. Os valores típicos para a taxa de aprendizagem variam entre 0,05 e 0,75 (RAI et al., 2005).

Para acelerar a etapa de aprendizagem, foram realizadas modificações no algoritmo *backpropagation* através da regra estendida delta-bar-delta (EDBD) que utilizou uma taxa de aprendizado para aumentar a velocidade de aprendizagem e conseqüentemente, reduzir os

riscos de instabilidade (HAYKIN, 1999, BRAGA et al., 2000). A taxa de aprendizado foi adicionada aos pesos para ajustar a função e ajudar as mudanças de peso.

O número de iterações usadas também é de grande importância. Um pequeno número de iterações leva a grandes erros enquanto um grande número de iterações poderia exigir muito tempo computacional. A determinação do melhor número de iteração melhor para a formação da rede foi efetuada por tentativa e erro, técnica que resultou em um menor valor da raiz do erro quadrado médio (RMSE) de validação (Equação 5).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_d - x_p)^2} \quad (5)$$

onde n é o número de dados, e x_d e x_p são os valores experimentais e previstos das propriedades físico-químicas, respectivamente.

2.5.1 Treinamento e Validação dos Dados

Os conjuntos de dados foram utilizados para treinamento da rede e ajuste das funções discriminantes. Os dados foram aleatorizados e divididos com auxílio do algoritmo de Kenstone em dois subconjuntos, usando 70% para o treinamento e 30% para validação de rede. Foram utilizados dez neurônios na camada de entrada, grau alcoólico (X1), acidez volátil (X2), ésteres totais (X3), aldeído (X4), alcoóis superiores totais (X5), álcool metílico (X6), soma dos componentes voláteis (X7), álcool n-propílico (X8), álcool isobutílico (X9) e o álcool isoamílico (X10). A saída do RNA foi representada por um vetor multidimensional com uma série de dimensões iguais ao número de tipos de cachaça e para cada dimensão vetorial foi atribuída um tipo. As cachaças foram classificadas como mostrado abaixo:

(1,0,0) = envelhecida em Amburana

(0,1,0) = envelhecida em Bálamo

(0,0,1) = envelhecida em Carvalho

Nos arquivos de treinamento e teste/validação, o conjunto de dados foi criado em formato numérico, atribuindo 1 quando a cachaça foi corretamente classificada e 0 caso contrário.

Os dados de entrada e saída foram normalizados usando através das equações 6 e 7, respectivamente

$$x_{i,norm} = \frac{x_i - \bar{x}_i}{R_{i,max}} \quad (6)$$

$$x_{i,norm} = \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}} \quad (7)$$

onde $x_{i,norm}$ são os dados normalizados, x_i são os dados experimentais, \bar{x}_i é a média dos dados, $R_{i,max}$ é dada por $R_{i,max} = \max [x_{i,max} - x_i, x_i - x_{i,min}]$, $x_{i,min}$ e $x_{i,max}$ são os valores mínimos e máximos de x_i , respectivamente.

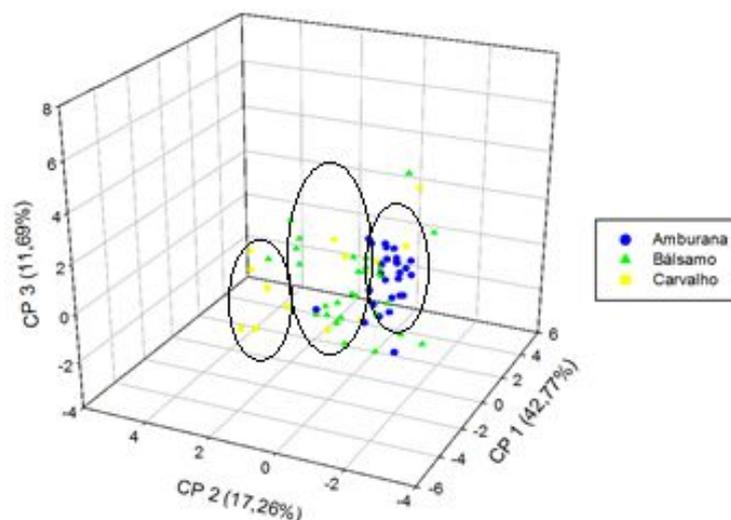
As performances das diferentes formações das redes neurais foram comparadas utilizando a taxa de classificação, matriz de confusão e a raiz do erro quadrático médio (RMSE), expresso pela equação 5.

3 RESULTADOS E DISCUSSÕES

3.1 Análise de Componentes Principais

Na matriz construída com 70 linhas relativas as amostras de cachaças envelhecidas nas madeiras Amburana, Bálsamo e Carvalho ($m=70$) e 10 variáveis físico-químicas ($n=10$), analisou-se, em primeiro lugar todas as madeiras, obtendo-se como resultado uma porcentagem de 71,72% de explicação da variação com os três componentes principais sendo que o CP1 representa 42,77%, o CP2, 17,26% e o CP3, 11,69%. Por sua vez a análise gráfica da dispersão das amostras (figura 2) mostra que há constituição de grupos naturais, porém observa-se, uma justaposição. Diante do grande número de amostras justapostas, optou-se por estratificar a matriz dois, com dez colunas, em três novas matrizes que são Amburana e Bálsamo, Amburana e Carvalho e Bálsamo e Carvalho e, neste caso, avaliando as combinações par a par, optou-se pelo estudo das duas primeiras combinações supracitadas, uma vez que na análise dos gráficos de escores houve uma melhor separação visual entre estes dois casos específicos.

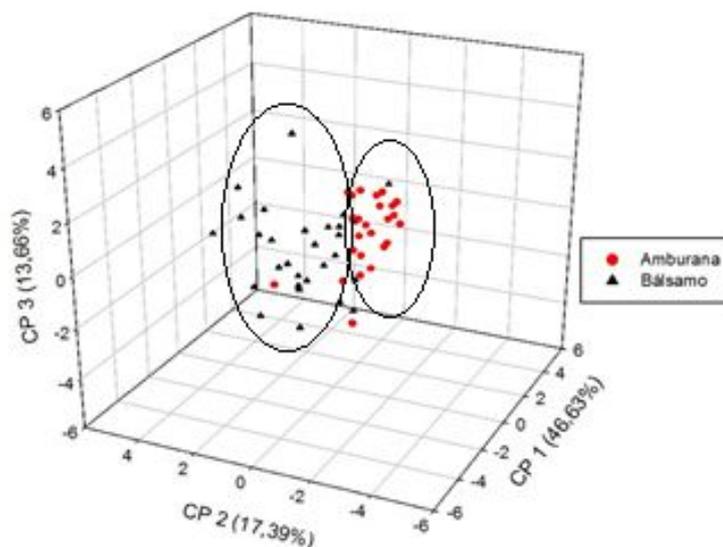
Figura 2. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana, Bálsamo e Carvalho em relação aos três primeiros componentes principais (CP1, CP2 e CP3).



Com relação as cachaças envelhecidas nas madeiras Amburana e Bálsamo que formaram a matriz A_{11} com ($m=56$ e $n=10$) os três primeiros componentes principais explicaram 77,67% da variação total, sete componentes apresentaram variância (autovalor) inferior a 1,0. Percebe-se que houve uma redução do número de variáveis, logo é necessário, uma quantidade menor de componentes principais para que os dados sejam bem representados.

A Figura 3 mostra a dispersão das amostras e, foi possível observar a formação de um perfil da Amburana e do Bálsamo. O gráfico que reúne os componentes principais 1, 2 e 3 fornece uma janela estatisticamente privilegiada para observação dos pontos no espaço tridimensional, no entanto, nota-se que não houve uma separação total dos grupos, tal fato pode ser explicado pelas condições de produção e armazenamento de cada produtor. Para o mesmo tipo de madeira empregada, a transferência de compostos, como aldeído e ácidos, para a cachaça depende de outros fatores, que não são controlados neste estudo, como, por exemplo, o tempo de envelhecimento e o número de ciclos de utilização dos tonéis no envelhecimento da cachaça. Outra característica importante são as características iniciais do destilado que podem influenciar a partição de diferentes compostos da madeira para a cachaça.

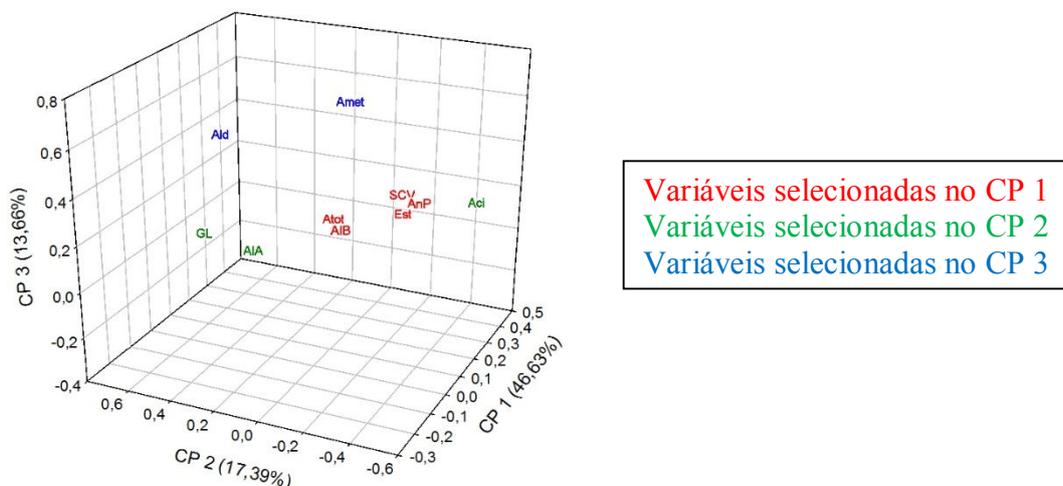
Figura 3. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana e Bálsamo em relação aos três primeiros componentes principais (CP1, CP2 e CP3).



Outro aspecto relevante é a diminuição da capacidade de transferência de compostos da madeira para a cachaça com as utilizações sucessivas dos tonéis, pois o processo de envelhecimento se dá até o ápice do estágio de saturação, mas para se atingir o mesmo padrão físico-químico numa reutilização, o tempo de envelhecimento será superior ao de uma utilização anterior (SILVA e VASCONCELOS, 2009).

A escolha das variáveis resultou na seleção dos ésteres, os alcoóis superiores totais, n-propílico e isobutilico e a soma dos componentes voláteis para o CP 1, grau alcoólico, acidez e álcool superior isoamílico para o CP 2, aldeído e álcool metílico para o CP 3, uma vez que estes possuem os maiores coeficientes (Figura 4). A CP 1 explicou 46,63% da variação, ou seja, as variáveis a ele relacionadas são muito importantes e representam, através das combinações lineares, quase a metade da variância de todo o conjunto e podem determinar a característica de cada grupo.

Figura 4. Gráfico de pesos dos coeficientes de ponderação (Autovetores) para as cachaças envelhecidas nas madeiras Amburana e Bálsamo.



* **GL:** Grau Alcoólico; **Aci:** Acidez; **Est:** Ésteres; **Ald:** Aldeído; **Atot:** Alcoóis Superiores Totais; **AnP:** Álcool Superior n-propílico; **AIB:** Álcool Superior Isobutílico; **AIA:** Álcool Superior Isoamílico; **AMet:** Álcool Metílico; **SCV:** Soma dos Componentes Voláteis.

A matriz A_{12} que compilou as variáveis estudadas para as amostras envelhecidas em Amburana e Carvalho ($m=41$; $n=10$), fazendo uso dos mesmos critérios utilizados para seleção dos CP's, obteve três componentes principais representando 79,04% da variabilidade total (Figura 5). A dispersão (figura 5) evidenciou a constituição dos perfis por madeira com amostras separadas de forma incorreta, conforme visto anteriormente e a figura 6 mostra que as variáveis que apresentam maior correlação com os componentes principais escolhidos são acidez, ésteres, álcool superior n-propílico, álcool superior isobutílico e a soma dos componentes voláteis para o CP 1, o álcool superior total e o álcool isoamílico para o CP 2, o grau alcoólico, aldeído e álcool metílico para o CP 3. Para este caso a CP 1 explica 50,08% da variabilidade, indicando, mais uma vez a influência significativa da acidez, dos ésteres, da SCV e, no caso da segunda matriz, da estratificação dos alcoóis.

Figura 5. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana e Carvalho em relação aos dois primeiros componentes principais (CP1 e CP2).

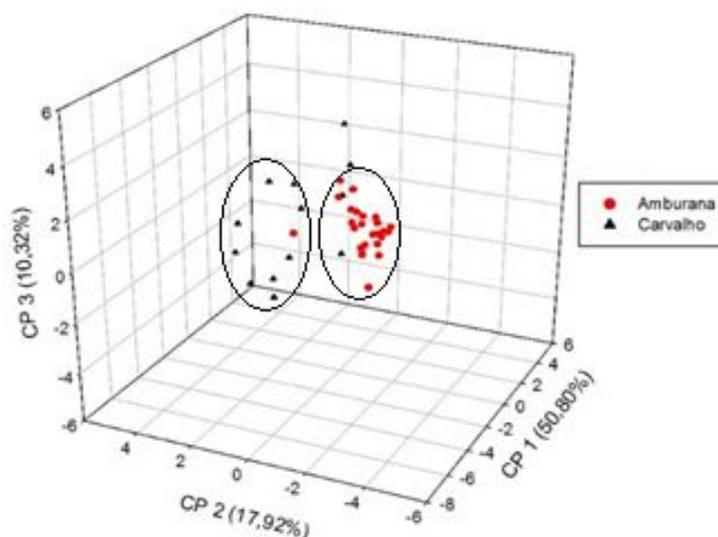
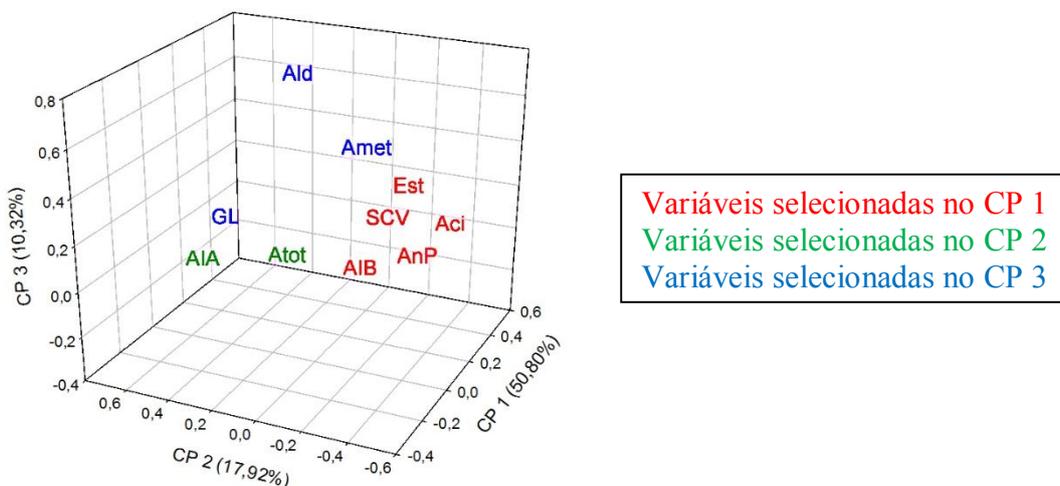


Figura 6. Gráfico de pesos dos coeficientes de ponderação (Autovetores) para as cachaças envelhecidas nas madeiras Amburana e Carvalho.



* **GL:** Grau Alcoólico; **Aci:** Acidez; **Est:** Ésteres; **Ald:** Aldeído; **Atot:** Alcoóis Superiores Totais; **AnP:** Álcool Superior n-propílico; **AIB:** Álcool Superior Isobutílico; **AIA:** Álcool Superior Isoamílico; **Amet:** Álcool Metílico; **SCV:** Soma dos Componentes Voláteis.

É importante dizer que as propriedades das madeiras interferem na composição final da bebida, sendo assim, a retirada de macromoléculas bem como a geração de novas, depende, por exemplo, da densidade e da permeabilidade da madeira. Este conhecimento é útil para o produtor já que o mesmo pode retirar as variáveis que não foram significativas e substituir

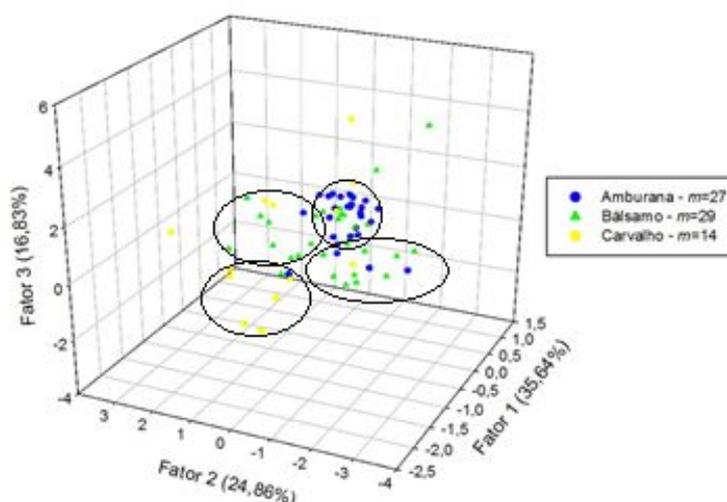
pela estratificação dos compostos mais importantes, obtendo informações que podem ser utilizadas na caracterização da bebida.

3.2 Análise de Fatores

Foram obtidos os resultados (fatores) da matriz estruturada com 10 colunas referentes as variáveis físico-químicas de qualidade de cachaça ($n=10$) e, 70 linhas referentes as amostras de cachaças envelhecidas nas madeiras Amburana, Bálsamo e Carvalho ($m=70$), produzidas nas cidades de Salinas e Paraty. As porcentagens de variância total, explicadas pelos três fatores, foram 35,64% para fator 1, 24,86% para fator 2 e 16,83% para fator 3, acumulando 77,33% do total da variância.

Analisando o gráfico com a dispersão dos escores dos fatores 3, 2 e 1 (Figura 7) para as cachaças envelhecidas foi possível observar que as amostras envelhecidas na mesma madeira estão próximas, mas houve sobreposição das cachaças envelhecidas nos diferentes tipos de madeira. Sendo assim, a matriz original dos dados foi estratificada em três novas matrizes, a saber Amburana e Bálsamo, Amburana e Carvalho e Bálsamo e Carvalho. Optou-se por utilizar as matrizes A_{11} e A_{12} , pois os resultados mostraram que houve melhor separação entre Amburana e Bálsamo (A_{11}) e Amburana e Carvalho (A_{12}).

Figura 7. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana, Bálsamo e Carvalho em relação aos três primeiros fatores (Fator 1, Fator 2 e Fator 3).



Para estimativa do número de fatores da matriz A_{11} com ($m=56$ e $n=10$) que foi constituída por dados das amostras envelhecidas em Amburana e Bálsamo, selecionou-se fatores com autovalores maiores que um. Assim, fez-se uma análise preliminar, utilizando o

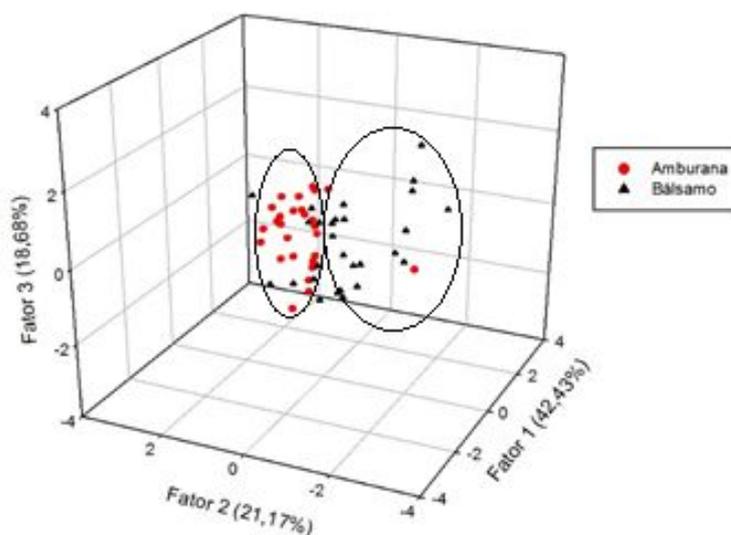
método exploratório de componentes principais, para determinar o número inicial, sendo escolhidos, a priori 3 fatores.

Com os fatores previamente escolhidos realizou-se a análise de fatores definitiva, e estes foram rotacionados, a fim de conseguir uma melhor interpretação dos resultados. A partir dos resultados rotacionados realizou-se a exclusão de variáveis.

Para atender as pressuposições descritas na metodologia de Hair Jr. e colaboradores (2010), a análise foi repetida várias vezes, até que todas as variáveis estivessem de acordo com os critérios estabelecidos. No final, restaram 3 fatores e 7 variáveis que representaram 82,28% da variabilidade dos dados.

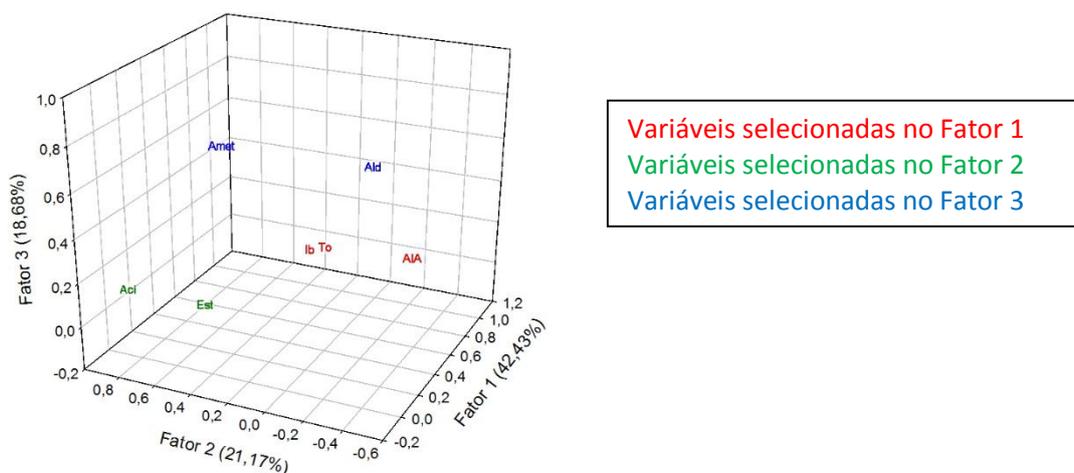
A dispersão das amostras, evidencia a formação dos perfis por madeira, como observado nos itens da ACP, com algumas amostras separadas incorretamente (Figura 8).

Figura 8. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana e Bálamo em relação aos três primeiros fatores (Fator 1, Fator 2 e Fator 3).



Os valores das cargas fatoriais finais, indicaram que as variáveis alcoóis superiores, total, isoamílico e isobutílico estão altamente relacionadas com o fator 1, a acidez e o teor de ésteres correlacionadas com o fator 2 e por fim o fator 3 está associado com as variáveis aldeído e álcool metílico (Figura 9).

Figura 9. Gráfico de pesos das cargas fatoriais finais após a rotação para as cachaças envelhecidas nas madeiras Amburana e Bálsamo



* **Aci:** Acidez; **Est:** Ésteres; **Ald:** Aldeído; **To:** Alcoóis Superiores Totais; **Ib:** Álcool Superior Isobutílico; **AIA:** Álcool Superior Isoamílico; **Amet:** Álcool Metílico.

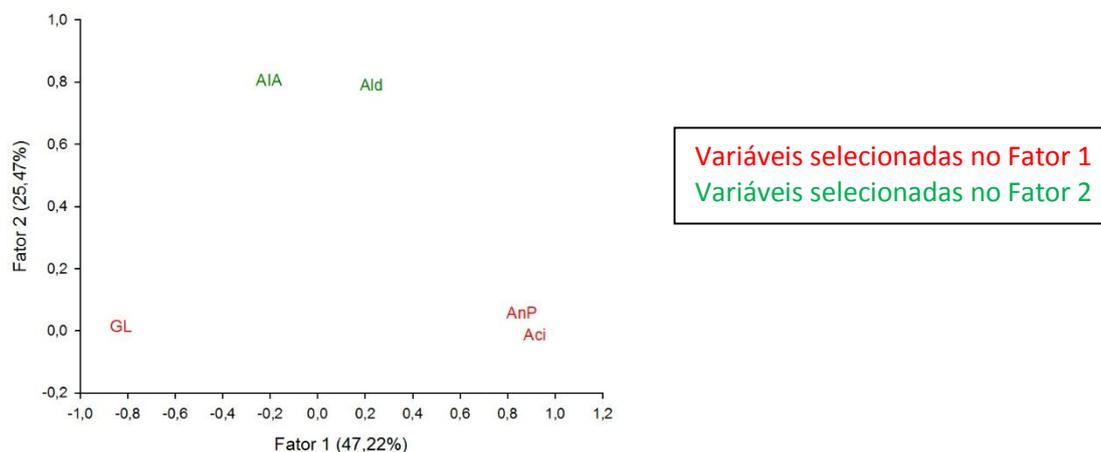
É válido ressaltar que os alcoóis superiores, que representam o fator 1, são produtos metabólicos decorrentes do crescimento de leveduras e do aproveitamento de aminoácidos como fonte de nutrientes amoniacais e dependem grandemente das condições do meio de fermentação, da quantidade e viabilidade do inóculo, da temperatura, do teor alcoólico final do vinho, entre outros fatores (CARDOSO, 2006; MAIA e CAMPELO, 2005, LÉAUTÉ, 1990). A relação deste fator com a etapa de fermentação pode ter sido um dificultador para uma efetiva separação das cachaças por madeira de envelhecimento. No fator 2, os valores das variáveis representam a quantificação dos compostos formados na fermentação e durante o envelhecimento através das reações de esterificação (ácidos + alcoóis = ésteres) e de oxidação (aldeídos + oxigênio = ácido) e contribuem de forma significativa para o buquê da cachaça. Além disso, nota-se que estas variáveis foram selecionadas nos primeiros fatores e, portanto, são muito importantes na representatividade das cachaças.

Já o fator 3 está relacionado com as etapas de plantio, fermentação e o envelhecimento pois os aldeídos são produzidos pelas leveduras principalmente durante os primeiros estágios da fermentação alcoólica, diminuindo sua presença gradativamente até o estágio final, pela oxidação. Exemplos de oxidação são a formação de acetaldeído e ácido acético a partir do álcool etílico e a formação de dimetilsulfóxido a partir de dimetilsulfeto (PIGGOTT e CONNER, 2003; CARDOSO, 2006; MAIA e CAMPELO, 2005). No envelhecimento, a madeira sofre degradação pela ação do álcool e da água, assim ocorre a hidrólise da hemicelulose e da lignina, sendo que os produtos dessa hidrólise passam para o destilado

(SHEREV e BRINK, 1980). A lignina, que é formada por monômeros aromáticos, tem estreita relação com o desenvolvimento de aroma e sabor nos destilados envelhecidos porque libera para a bebida aldeídos aromáticos durante o período de maturação. Porém, apenas pequena fração da lignina é solúvel e, conseqüentemente, extraída durante o período de maturação (SINGLETON, 1995). Por sua vez, o álcool metílico é um álcool indesejável na cachaça, originando-se da degradação de pectina, um polissacarídeo da cana-de-açúcar, presente em quantidades muito pequenas.

Os resultados observados para a matriz A_{12} elaborada a partir dos dados das cachaças envelhecidas em tonéis de Amburana e Carvalho mostraram que após o processo de rotação dos fatores iniciais são necessários dois fatores e cinco variáveis a fim de explicar 72,68% da variação total dos dados. As variáveis grau alcoólico, acidez e álcool superior n-propílico estão relacionadas com o fator um e, por sua vez, as variáveis aldeído e álcool superior isoamílico estão relacionadas com o fator dois. As variáveis selecionadas, com exceção do grau alcoólico, são as mesmas da outra análise reiterando a importância destas na qualidade final da bebida (Figura 10). O grau alcoólico é um fator importante na diferenciação, uma vez que é mais um parâmetro que caracteriza a volatilização das madeiras.

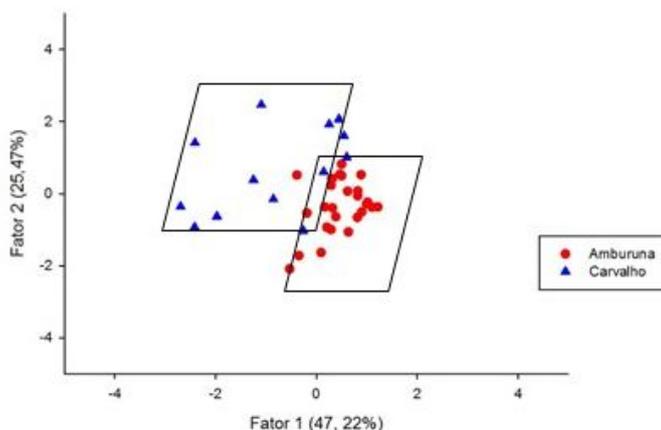
Figura 10. Gráfico de pesos das cargas fatoriais finais após a rotação para as cachaças envelhecidas nas madeiras Amburana e Carvalho



* **GL:** Grau Alcoólico; **Aci:** Acidez; **Ald:** Aldeído; **AnP:** Álcool Superior n-propílico; **AIA:** Álcool Superior Isoamílico

A figura 11 mostra a representação gráfica da dispersão dos grupos de acordo com o envelhecimento da bebida em diferentes tonéis, evidenciando que estas madeiras possuem características específicas e, portanto, promovem uma separação efetiva das amostras.

Figura 11. Gráfico de escores das cachaças envelhecidas nas madeiras Amburana e Carvalho em relação aos dois primeiros fatores (Fator 1 e Fator 2).



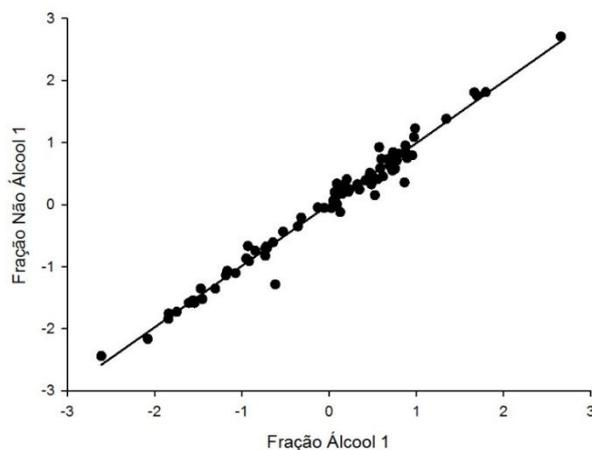
3.3 Análise de Correlação Canônica

Utilizando o conjunto de dados das amostras das madeiras Amburana, Bálamo e Carvalho, fez-se, primeiramente, a divisão das variáveis de acordo com o interesse da pesquisa, separando em um grupo os componentes alcoóis e em outro os que não são alcoóis com o objetivo de verificar se a composição alcóolica tem relação com a formação dos componentes secundários como os ésteres, os aldeídos, os ácidos e outros que são muito importantes para a qualidade da cachaça. Os resultados encontrados para as amostras de cachaça evidenciaram que, dos quatro pares de correlações encontrados, três foram significativos, ($p < 0,05$), com ênfase para a primeira correlação que retém 94,84% da variação dos dados e possui correlação igual a 0,9876. Tem-se, ainda, a representação gráfica da correlação (Figura 12) e as equações que geraram este resultado (Equação 8 e 9).

$$Y1_{\text{Fração álcool}} = -0,0909 * GL + 0,1159 * ASTot + 0,3942 ASnP + +0,1752 * ASIsoB + 0,5635 ASIsoA - 0,0394 * Amet \quad (8)$$

$$Y1_{\text{Fração não álcool}} = -0,6644 * aci - 0,6381 * Est - 0,0915 * Ald + 1,8028 SCV \quad (9)$$

Figura 12. Correlação do primeiro par de variáveis canônicas obtida para a matriz dois



Esta alta correlação indica que existe uma associação entre os componentes alcoóis e os não alcoóis da cachaça, ou seja, é possível encontrar uma associação que influencia as relações entre os mesmos, indicando que quanto maior a fração álcool, maior a formação de congêneres que modificam, de forma positiva as características sensoriais da bebida. Este resultado pode estar associado ao fato dos compostos serem formados através de reações químicas que incluem a molécula de álcool. Outro aspecto importante é sobre as análises físico-químicas realizadas pelo produtor, pois se o aumento da fração álcool é diretamente proporcional a fração não álcool, pode-se fazer apenas as análises da primeira fração e estimar a segunda, reduzindo assim os custos do processo.

A representação gráfica das duas primeiras variáveis canônicas da fração álcool e da fração não álcool estão apresentadas nas figuras 13 e 14, respectivamente. Estes gráficos mostram que há uma tendência para a formação dos agrupamentos naturais, mas em função das peculiaridades do processo de produção e envelhecimento, há uma sobreposição das amostras. Estes resultados corroboram as discussões apresentadas nos itens Análise de Componentes Principais e Análise de Fatores.

Figura 13. Representação canônica das observações obtidas pelas variáveis da fração álcool

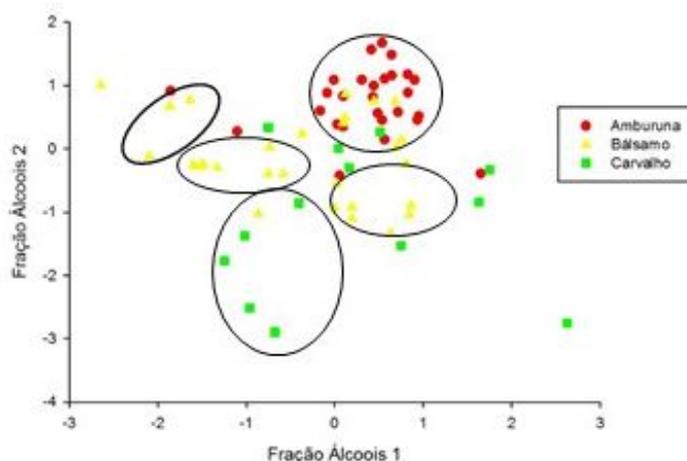
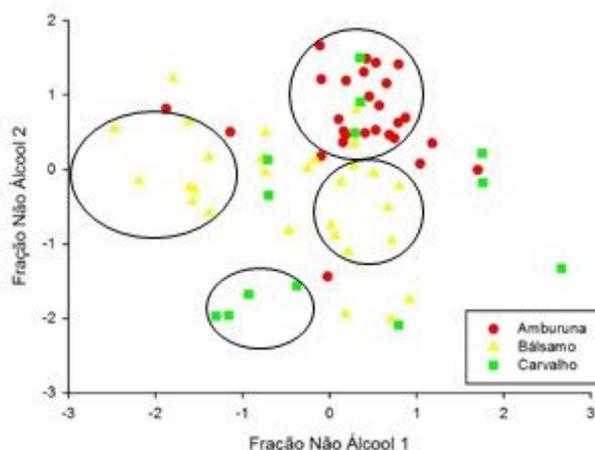


Figura 14. Representação canônica das observações obtidas pelas variáveis da fração não álcool



3.4 Análise Discriminante

Para a matriz A esta técnica da estatística multivariada foi utilizada com o propósito de estudar a separação das 70 amostras de cachaças de acordo com a madeira de envelhecimento, a saber, Amburana, Bálamo e Carvalho ($m=70$; $n=10$). O primeiro passo foi conseguir, em cada um dos grupos, atributos que diferenciassse cada um deles utilizando as amostras do treinamento. Posteriormente foi feita a validação de amostras desconhecidas com base nas equações 10 a 12. Outra função destes índices é relacionar as variáveis com cada madeira, sendo assim, de acordo com os valores dos coeficientes, pode-se afirmar que as três madeiras possuem maior associação com os alcoóis superiores.

$$Y_{Amburana} = -2,0053 + 0,0832 * GL - 0,7890 * aci - 3,1778 * Est - 1,1489 * Ald + 168,801 * ASTot - 77,7202 * ASnP - 31,028 * ASIsoB - 114,6330 * ASIsoA - 0,3786 * Amet + 5,38248 * SCV \quad (10)$$

$$Y_{Bálsamo} = -0,6721 - 0,1267 * GL - 0,1753 * aci + 1,1836 * Est + 1,1570 * Ald - 2,9925 * ASTot + 1,3457 * ASnP + 1,0924 * ASIsoB + 2,2673 * ASIsoA - 0,2761 * Amet - 2,2703 * SCV \quad (11)$$

$$Y_{Carvalho} = -3,2834 + 2,2476 * GL - 2,1282 * aci - 5,5185 * Est - 0,8221 * Ald - 97,820 * ASTot + 40,6239 * ASnP + 15,9615 * ASIsoB + 62,8261 * ASIsoA + 0,0767 * Amet + 13,1438 * SCV \quad (12)$$

Na etapa de treinamento o percentual de acerto variou de 75,0 a 93,75% (Tabela 2). Tem-se, ainda, a porcentagem de 63,7%, 33,4% e 10,0% de acerto na validação (Tabela 2). Diante disto pode-se dizer que as amostras apresentam uma faixa de intercessão em relação aos parâmetros analisados que não permitem a separação satisfatória das amostras.

Tabela 2. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaça envelhecidas em Amburana, Bálsamo e Carvalho.

Grupo	Amburana T/V	Bálsamo T/V	Carvalho T/V	Total T/V
Amburana	15/7 93,75/63,7	0/3 0,0/27,2	0/1 0,0/9,1	16/11 100,00
Bálsamo	1/7 5,8/58,3	14/4 82,4/33,4	0/0 0,0/0,0	17/12 100,00
Carvalho	0/1 0,0/16,7	1/2 12,5/33,4	6/0 75,0/10,0	8/6 100,00

*T=Treinamento e V=Validação

A partir destes resultados também foi feita as análises das matrizes estratificadas que tiveram melhores separações segundo a ACP e a AF. Para a matriz A_{11} com ($m=56$ e $n=10$) composta de dados das cachaças envelhecidas nas madeiras Amburana e Bálsamo, a separação das amostras foi feita com base nas funções discriminantes (Equações 13 e 14).

No treinamento e validação das amostras obteve-se como resultado 100,0 e 54,5% de acertos para as cachaças envelhecidas em Amburana e 94,4 e 63,6% de classificação correta para o Bálsamo (Tabela 3).

$$Y_{Amburana} = -2,57646 + 2,3450 * GL + 2,4088 * aci - 2,3011 * Est - 0,9487 * Ald + 3,0714 * ASTot - 3,3932 * ASnP + 2,4602 * ASIsoB - 5,6793 * ASIsoA + 0,0728 * Amet + 5,6638 * SCV \quad (13)$$

$$Y_{Bálsamo} = -1,5111 - 2,0856 * GL + 1,2211 * aci + 3,8401 * Est + 0,8223 * Ald + 4,02188 * ASTot - 0,3257 * ASnP - 2,2139 * ASIsoB + 3,4166 * ASIsoA - 0,2980 * Amet - 9,7032 * SCV \quad (14)$$

Tabela 3. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaça envelhecidas em Amburana e Bálsamo

Grupo	Amburana	Bálsamo	Total
Amburana	16/6 100,0/54,5	0/5 0,0/45,5	16/11 100,00
Bálsamo	1/4 5,5/36,3	17/7 94,4/63,6	18/11 100,00

Por fim para a matriz A_{12} ($m=41$; $n=10$), que reuniu dados das cachaças Amburana e Carvalho, a separação das amostras foi feita com base nas funções discriminantes (Equações 15 e 16).

Baseado nos resultados reportados na Tabela 4, observou-se 100,0% de acerto no treinamento, já na validação o percentual foi 63,6 e 80,0 para Amburana e Carvalho, respectivamente.

$$Y_{Amburana} = -2,4209 + 0,4950 * GL - 1,9794 * aci - 1,1037 * Est - 1,0815 * Ald + 31,6513 * ASTot - 17,088 * ASnP + 0,0791 * ASIsoB - 30,1517 * ASIsoA + 0,79538 * Amet + 1,5926 * SCV \quad (15)$$

$$Y_{Carvalho} = -4,3603 + 2,5276 * GL - 2,4934 * aci - 1,9738 * Est + 0,8481 * Ald - 9,1481 * ASTot + 3,4977 * ASnP - 5,6914 * ASIsoB + 11,0187 * ASIsoA + 0,4775 * Amet + 8,7731 * SCV \quad (16)$$

Tabela 4. Número de observações e o percentual de classificação por grupo no teste/validação das amostras de cachaça envelhecidas em Amburana e Carvalho

Grupo	Amburana	Carvalho	Total
Amburana	16/7 100,0/63,6	0/6 0,0/36,3	16/11 100,00
Carvalho	0/1 0,0/20,0	9/4 100,0/80,0	9/5 100,00

É importante salientar que a principal contribuição da madeira no envelhecimento de bebidas destiladas é relacionada com a extração de compostos a partir da madeira e a formação de novas moléculas aromáticas como resultado de hidrólise, oxidação e reação destes compostos com a cachaça. O perfil químico da cachaça envelhecida depende de vários fatores, tais como gênero e espécies, região geográfica, as operações da tanoaria (corte, envelhecimento e tratamento térmico), tempo de envelhecimento e armazém. Sendo assim, pode-se dizer que a complexidade deste processo e a impossibilidade de controle efetivo dificultou o processo de classificação.

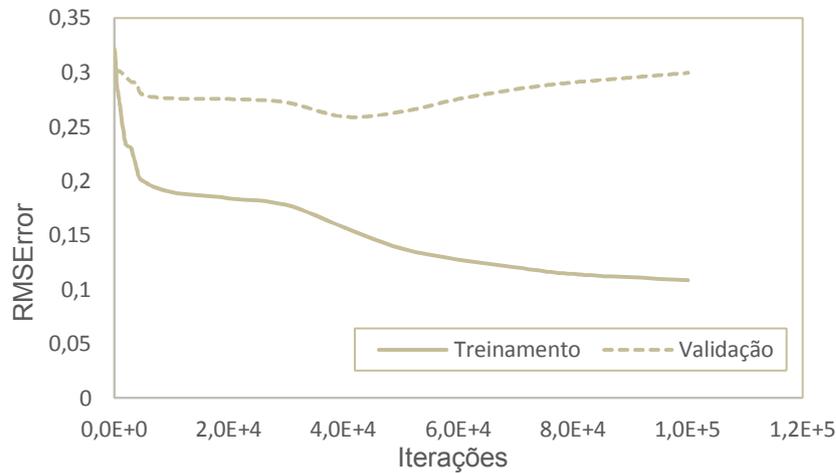
Outra questão que deve ser abordada é que os efeitos da maturação são influenciados principalmente pelo tipo de madeira do tonel (MOSEDALE e PUECH, 1998). Carvalho é a principal madeira utilizada para envelhecimento de destilados em todo o mundo porque participa ativamente do sabor das bebidas, devido à extração de moléculas aromáticas da madeira (RAMIREZ-RAMIREZ, 2002). No entanto, no Brasil, carvalho não é uma madeira nativa e é necessário importá-la da Europa ou da América do Norte, pagando-se preços altos. Muitas vezes os produtores adquirem barris antigos oriundos da produção de vinho ou de outros destilados e, portanto, a extração é mais difícil, dificultando a caracterização da bebida.

3.5 Redes Neurais Artificiais

Para obter o melhor número de iterações, o que proporcionou menor RMSE e o melhor desempenho da rede neural, fez-se o monitoramento, em função do número de iterações, do RMSE nas fases de treinamento e validação. A priori foi feita uma avaliação de 100 a 100.000 iterações em intervalos de 100 (Figura 15). Posteriormente foi feita a análise considerando que o erro da etapa de treinamento continua caindo, mesmo após os ajustes da rede, avaliou-se o erro da etapa de validação e após 35.000 iterações, este aumentou dentro do

intervalo estabelecido e interrompeu-se o treinamento, pois a partir deste momento considera-se que a rede decorou o procedimento ou passou a absorver os ruídos.

Figura 15. RMSError x Número de iterações para dados de treinamento e validação da matriz A_2



O critério utilizado para selecionar a topologia RNA adequada consistiu em selecionar o número de neurônios e as camadas ocultas que deram um RMSE mínimo de validação e uma taxa máxima de classificação (Figura 16 e 17). Os resultados mostraram que a configuração RNA composta de duas camadas ocultas com doze neurônios na primeira camada e três na segunda apresentou a melhor estrutura para a classificação de cachaças envelhecidas nas madeiras Amburana, Bálsamo e Carvalho. O coeficiente de RMSE e a taxa de classificação desta configuração para o conjunto de dados de treinamento e validação foram 0,1699, 0,9524, 0,8421, 0,6667 e 0,2669, 0,800, 0,8750, 0,6250, respectivamente.

Figura 16. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de treinamento.

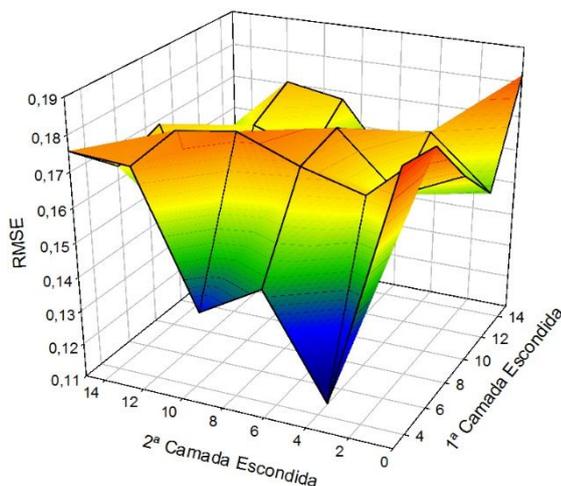
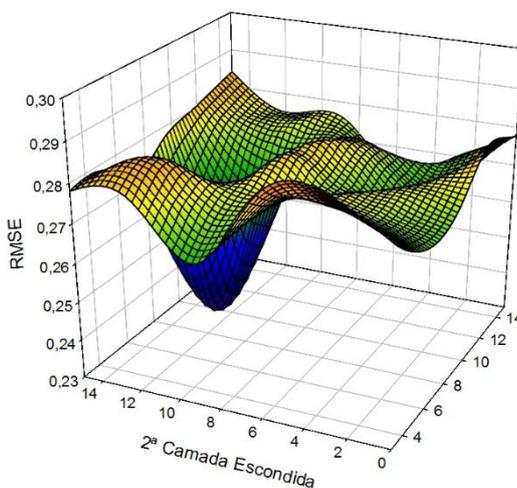


Figura 17. Representação das camadas ocultas das diferentes configurações da RNA em função do RMSE na etapa de validação.

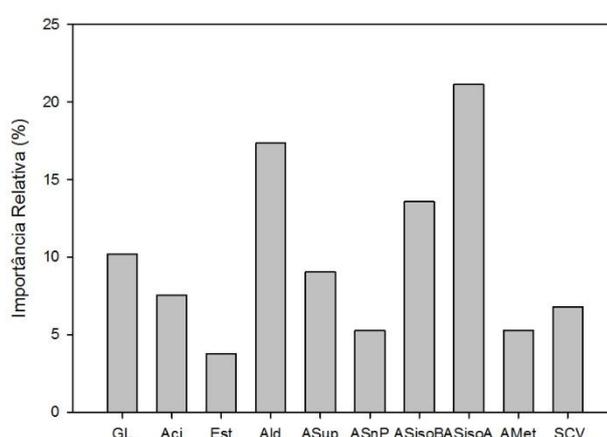


A participação de cada variável no processo de classificação é reportada na figura 18 e é possível afirmar, segundo esta análise, que o álcool superior isoamílico é a variável que apresenta maior contribuição, seguida do aldeído, fato que evidencia, mais uma vez a importância destas variáveis na classificação de cachaças. Estes parâmetros estão relacionados com o tempo de envelhecimento. Miranda e seus colaboradores (2008) observaram um acréscimo no teor de álcoois superiores em função do tempo de envelhecimento. O acréscimo

médio na concentração de álcoois superiores ao longo de 390 dias de envelhecimento foi de aproximadamente 13%, sendo que o álcool isobutílico apresentou um acréscimo de 12% e o isoamílico de 18% no mesmo período. A concentração do álcool n-propílico da aguardente praticamente não alterou durante o período de envelhecimento do destilado.

A causa da grande quantidade de aldeídos nas cachaças, que é a segunda variável mais importante, pode ser uma indicação de oxidação espontânea ou devido à atividade de bactérias contaminantes (YOKOYA, 1995).

Figura 18. Importância relativa (%) de cada variável na classificação de amostras de cachaça da matriz A_2 .



4 CONCLUSÕES

O processo de classificação utilizando a Análise de Componentes Principais (ACP) e Análise de Fatores (AF) das cachaças envelhecidas nas madeiras, Amburana, Bálsamo e Carvalho mostrou que as melhores separações ocorreram com as combinações Amburana e Bálsamo e Amburana e Carvalho. Foi possível observar a formação de perfis por madeira, no entanto devido a questões relacionadas ao processamento não houve separação total dos grupos. A Análise de Correlações Canônicas obteve correlação próxima de 100%, mas a separação por grupo obteve resultados semelhantes aos observados na ACP e AF.

A Análise de Correlações Canônicas obteve correlação próxima de 100%, mas a separação por grupo obteve resultados semelhantes aos observados na ACP e AF.

O resultado da ACP e AF foi utilizado, ainda, para realização da Análise Discriminante (AD). A AD e Redes Neurais Artificiais (RNA) apresentaram, na validação, uma porcentagem de acerto que variou entre 10,0 e 80,0%. As variáveis que possuem maior

importância na classificação das cachaças segundo ACP, AF e RNA são acidez, GL e alcoóis superiores.

5 REFERÊNCIAS

BRAGA, A.P., CARVALHO, A.C.P.L.F., LUDERMIR, T.B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC Editora S.A, 2000. 262p.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa nº 13, de 29 de junho de 2005a**. Aprova o regulamento técnico para fixação dos padrões de identidade e qualidade para aguardente de cana e para cachaça. Diário Oficial da República Federativa do Brasil. Brasília, DF, 30 Jul. 2005 Disponível em: <<http://www.agricultura.gov.br/>>. Acesso em: 04 de mai. 2015.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. **Instrução Normativa nº 24, de 08 de Setembro de 2005b**. Manual Operacional de Bebidas e Vinagres. Diário Oficial da União, Brasília, DF, 20 set.2005. Seção 1, p.11. Disponível em: <<http://www.agricultura.gov.br/>>. Acesso em: 04 de mai. 2015.

BERNARDES, C. D.; DE FIGUEIREDO, M. C. P.; BARBEIRA, P. J. S. Developing a PLS model for determination of total phenolic content in aged cachaças. **Microchemical Journal**, v. 116, p. 173-177, 2014

CARDELLO, H. M. A. B.; FARIA J. B. Análise descritiva quantitativa da aguardente de cana durante o envelhecimento em tonel de carvalho (*Quercus alba* L.). **Ciência e Tecnologia de Alimentos**, v. 18, n. 2, p. 169-175, 1998.

CARDELLO, H. M. A. B.; FARIA, J. B. Análise tempo-intensidade de características sensoriais de cachaça de cana durante o envelhecimento em tonel de carvalho (*Quercus* sp). **Boletim da Sociedade Brasileira de Ciência e Tecnologia de Alimentos**, v. 33, n. 1, p. 27-34, 1999.

CARDOSO, M. das G. **Produção de aguardente de cana-de-açúcar**. Lavras: Editora UFLA, 2006. 264p.

CARNEIRO, J. C. S. **Análise estatística multivariada aplicada à avaliação sensorial de alimentos**. 2005. 89f. Tese (Doutorado em Ciência e Tecnologia de Alimentos) – Universidade Federal de Viçosa, Viçosa, MG.

- CEVOLI, C., CERRETANI, L., GORI, A., CABONI, M.F., TOSCHI, T.G., FABBRI, A. Classification of Pecorino cheeses using electronic nose combined with artificial neural network and comparison with GC–MS analysis of volatile compounds. **Food Chemistry**, v.129, p.1315-1319, 2011.
- CIMPOIU, C., CRISTEA, V.M., HOSU, A., SANDRU, M., SESERMAN, L. Antioxidant activity prediction and classification of some teas using artificial neural networks. **Food Chemistry**, v.127, p.1323-1328, 2011.
- DE SOUZA, P. P. de; AUGUSTI, D. V.; CATHARINO, R. R.; SIEBALD H G. L.; EBERLIN, M. N.; AUGUSTI R. Differentiation of rum and Brazilian artisan cachaça via electrospray ionization mass spectrometry fingerprinting. **Journal of Mass Spectrometry**, v. 42, p. 1294–1299, 2007.
- DE SOUZA, P. P.; DE OLIVEIRA, L. C. A.; CATHARINO, R. R.; EBERLIN, M. N.; AUGUSTI, D. V.; SIEBALD, H. G. L.; AUGUSTI, R. Brazilian cachaça: “Single shot” typification of fresh alembic and industrial samples via electrospray, **Food Chemistry**, v. 115, p. 1064–1068, 2009.
- FERNANDES A. P.; SANTOS M. C.; LEMOS S. G.; FERREIRA M. M. C.; NOGUEIRA A. R. A.; NÓBREGA J. A. Pattern recognition applied to mineral characterization of Brazilian coffees and sugar-cane spirits, **Spectrochimica Acta B**, v. 60, n. 5, p. 717-724, 2005
- GRANATO, D.; OLIVEIRA, C. C.; CARUSO, M. S. F.; NAGATO, L. A. F.; ALABURDA, J. Feasibility of different chemometric techniques to differentiate commercial Brazilian sugarcane spirits based on chemical markers. **Food Research International**, v. 60, p. 212-217, 2014.
- GUTTMAN, L. **A new approach to factor analysis: The radex**. In P.F. Lazarsfeld (Ed.) *Mathematical Thinking in the Social Sciences*. New York: Free Press, 1954
- HAIR JR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate Data Analysis**. 7.ed. Pearson Prentice Hall, 2010.
- HAYKIN, S. **Neural networks – a comprehensive foundation**, 2nd edition. Pearson Education, Inc., Delhi, 823p, 1999.
- JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. New Jersey: Prentice-Hall, 1999.816p.

- LAVINE, B. K., e WORKMAN, J. **Chemometrics**. Analytical Chemistry, v. 74, p. 2763–2769. 2002.
- LÉAUTÉ, R. Distillation in alambic. **American Journal of Enology and Viticulture**, v. 41, p. 90-103, 1990.
- MAÇATELLI, M. **Determinação do perfil sensorial de marcas comerciais de cachaça**. 2006. 128f. Dissertação (Mestrado em Alimentos e Nutrição) - Faculdade de Ciências Farmacêuticas, UNESP, Araraquara, SP.
- MINAI, A. A., WILLIAMS, R. D. **Back-propagation heuristics: A study of the extended delta-bar-delta algorithm,”** Proc. IEEE Int. Joint Conf. Neural Networks, San Diego, USA, 595–600, 1990.
- MIRANDA, M. B. et al. Perfil físico-químico de aguardente durante envelhecimento em tonéis de carvalho. **Ciência e Tecnologia de Alimentos**, v. 28, n. 2, p. 313-319, 2008.
- MOSEDALE, J. R.; PUECH, J. L. Wood maturation of distilled beverages. **Trends in Food Science e Technology**, v. 9, n. 3, p. 95-101, 1998.
- ODELLO, L.; BRACESCHI, G. P.; SEIXAS, F. R. F.; SILVA, A. A.; GALINARO, C. A.; FRANCO, D. W. Avaliação sensorial de cachaça. **Química Nova**, v. 32, n. 7, p. 1839-1844, 2009.
- PIGGOTT, J. R.; CONNER, J. M. Whiskies. In: LEA, A. G. H.; PIGGOTT, J. R. (Ed.). **Fermented Beverage Production**. 2nd ed. New York: Klumer Academic/Plenum Publishers, 2003. cap.11, p. 239-262.
- PILLONEL, L., BÜTIKOFERA, U., SCHLICHTHERLE-CERNY, H., TABACCHI, R., BOSSET, J.O. Geographic origin of European Emmental. Use of discriminant analysis and artificial neural network for classification purposes. **International Dairy Journal**, v.15, p.557–562, 2005.
- RAI, P., MAJUMDAR, G.C., DASGUPTA, S., DE, S. Prediction of the viscosity of clarified fruit juice using artificial neural network: a combined effect of concentration and temperature. **Journal of Food Engineering**, v. 68, p. 527-533, 2005.
- RAMIREZ-RAMIREZ, G. **Étude de la sorption des composés d’arôme du vin par le bois de chêne em système modele d’élevage en fûts**. 2002. 135f. Tese (Doctorat en Oenologie) - Université de Bourgogne, Dijon, Bourgogne.
- RECHE, R. V.; FRANCO, D. W. Distinção entre cachaças destiladas em alambiques e em colunas usando quimiometria. **Química Nova**, v. 32, n. 2, p. 332-336, 2009.

- SILVA, B. V.; VASCONCELOS, J. N. Envelhecimento de Cachaça Artesanal em Tonéis de Diversos Tipos de Madeira com 20l de Capacidade. VIII **Congresso Brasileiro de Engenharia Química em Iniciação Científica**. 27 a 30 de julho de 2009, Uberlândia, Minas Gerais, Brasil
- SENA, M. M. DE; POPPI, R. J.; FRIGHETTO, R. T. S.; VALARINI, P. J. Avaliação do uso de métodos quimiométricos em análise de solos. **Química Nova**, v. 23, n. 4, p. 547-556, 2000.
- SERAFIM, F. A. T.; SEIXAS, F. R. F.; SILVA, A. A.; GALINARO, C. A.; NASCIMENTO, E. S. P.; BUCHVISER, S. F.; ODELLO, L.; FRANCO, D. W. Correlation between Chemical Composition and Sensory Properties of Brazilian Sugarcane Spirits (Cachaças). **Journal of the Brazilian Chemical Society**, v. 24, n. 6, p. 973-982, 2013.
- SERAFIM, F. A. T.; SILVA, A. A.; GALINARO, C. A.; FRANCO, D. W. Comparação do perfil químico entre cachaças de um mesmo vinho destiladas em alambiques e em colunas **Química Nova**, v. 35, n. 7, p. 1412-1416, 2012.
- SHEREV, R. N.; BRINK, J. A. Indústria de fermentação. In: SHEREV, R. N. (Ed.). **Indústrias de Processos Químicos**. 4 ed. Rio de Janeiro: Guanabara Dois, 1980, Cap. 31, p. 469-495.
- SINGLETON, V. L. Maturation of wines and spirits: comparison, facts and hypotheses. **American Journal of Enology and Viticulture**, v. 46, n. 1, p. 98-115, 1995.
- VARGAS, E.A.; GLÓRIA, M.B. Qualidade da aguardente de cana (*Saccharum officinarum* L.) produzida, comercializada e/ou engarrafada no Estado de Minas Gerais. **Ciência e Tecnologia de Alimentos**, Campinas, v.15, n.1, p.43-49, 1995.

CONSIDERAÇÕES FINAIS

O processo de prospecção e identificação das cachaças de acordo com a região de produção resultou em formação de agrupamentos com registro de alguns pontos sobrepostos em função das particularidades dos processos de produção de Salinas e Paraty. No que diz respeito às variáveis com maior participação na classificação das regiões, pode se destacar a acidez, ésteres para matriz A_1 e os alcoóis superiores na matriz A_2 . Como tratavam-se de amostras quimicamente complexas pode-se afirmar que as técnicas de separação foram eficientes, uma vez que os resultados foram de 83,87 a 100% de acerto na etapa de validação.

Por sua vez o processo de classificação utilizando os dados de envelhecimento da cachaça nas madeiras nas madeiras, Amburana, Bálsamo e Carvalho mostrou que as melhores separações ocorreram com as combinações Amburana e Bálsamo e Amburana e Carvalho. Neste caso também foi possível observar a formação de perfis por madeira, no entanto devido a peculiaridades do processamento não houve separação total dos grupos. A Análise de Correlações Canônicas obteve correlação próxima de 100%, mas a separação por grupo obteve resultados semelhantes aos observados na Análise de Componentes Principais (ACP) e Análise de Fatores (AF). A Análise Discriminante e as Redes Neurais Artificiais (RNA) apresentaram, na validação, uma porcentagem de acerto que variou entre 10,0 e 80,0%. As variáveis que possuem maior importância na classificação das cachaças segundo ACP, AF e RNA são acidez, GL e alcoóis superiores.

Todas as técnicas utilizadas neste estudo tiveram bom desempenho mas pode-se destacar análise de fatores e a rede neural artificial que foram as metodologias com melhores resultados e, portanto, são as mais indicadas para este tipo de conjunto de dados.

Outro aspecto positivo para este trabalho é que as informações utilizadas são obtidas com frequência pelos produtores pois são exigidas pela legislação vigente e, portanto, facilita o processo de reconhecimento da bebida.

APÊNDICES

APÊNDICE 1 - O algoritmo utilizado para obtenção da análise dos componentes principais encontra-se apresentado abaixo:

```
Title 'Matriz de Correlação';  
Proc Princomp data = ACPCachaca out=saida;  
Var GL Acidez Ésteres Aldeido ASupTot Amet SCV;  
Run;
```

```
Title 'Correlações dos yis com os xis';  
Proc Corr data=saida;  
Var GL Acidez Ésteres Aldeido ASupTot Amet SCV Prin1 Prin2 Prin3 Prin4 Prin5 Prin6  
Prin7;  
Run;
```

```
Title 'Escores dos componentes Principais';  
data saida; set saida;  
Proc Print;  
Var amostra Prin1 Prin2 Prin3 Prin4 Prin5 Prin6 Prin7 GL Acidez Ésteres Aldeido ASupTot  
Amet SCV;  
Run;
```

APÊNDICE 2 - Os procedimentos para obter as estimativas dos fatores encontram-se apresentado abaixo:

```
Title 'Análise de Correlação';  
Proc Corr;  
Var GL Acidez Ésteres Aldeido ASTot ASIsoA Amet;  
Run;
```

```
Title 'AF';  
Proc factor method = prin res scree;  
Var GL Acidez Ésteres Aldeido ASTot ASIsoA Amet;  
Run;
```

```
Proc factor n=4 method = prin score res scree r=v outstat = saida1 out = result;  
Var GL Acidez Ésteres Aldeido ASTot ASIsoA Amet;
```

```
Proc print data = result;  
var factor1 - factor4;
```

```
Proc print data = saida1;  
Run;
```

APÊNDICE 3 - A seguir estão descritos os procedimentos utilizados para obter os parâmetros da análise de correlações canônicas.

```
Title 'Análise de Correlações Canônicas';  
DATA ACCCachaca1; SET ACCCachaca;  
Proc Cancorr b int out=cscores outstat=canstat vprefix= alcooissup wprefix=outros;  
var ASTot ASnP ASIsoB ASIsoA;  
with GL Acidez Ésteres Aldeido SCV Amet;  
Run;
```

```
Title1 'Escores das variáveis canônicas';  
Proc Print Data=cscores;  
Run;
```

```
Title1 'Diversas estatísticas';  
Proc Print Data=canstat;  
Run;
```

APÊNDICE 4 - Segue abaixo o procedimento PROC discrim para análise discriminante:

```
Proc discrim data=saida testdata=validacao testlist;  
class classe;  
var GL Acidez Ésteres Aldeido ASTot ASnP ASIsoB ASIsoA Amet SCV;  
Run;
```