



Universidade de Brasília

Instituto de Química

Programa de Pós-Graduação em Química

DISSERTAÇÃO DE MESTRADO

***Desenvolvimento de método para discriminação entre
marcas de uísque e identificação de adulteração por
espectroscopia UV-Vis e PLS-DA***

Angélica Rocha Martins

Orientador

Prof. Dr. Jez Willian Batista Braga

Coorientador

Dr. Márcio Talhavini

Brasília, 2015.



Universidade de Brasília

Instituto de Química

Programa de Pós-Graduação em Química

DISSERTAÇÃO DE MESTRADO

***Desenvolvimento de método para discriminação entre
marcas de uísque e identificação de adulteração por
espectroscopia UV-Vis e PLS-DA***

Angélica Rocha Martins

*Dissertação apresentada ao Instituto
de Química da Universidade de
Brasília como parte dos requisitos
exigidos para a obtenção do Título de
Mestre em Química.*

Orientador

Prof. Dr. Jez Willian Batista Braga

Coorientador

Dr. Márcio Talhavini

Brasília, 2015.

COMUNICADO

Comunicamos a aprovação da Defesa de Dissertação de Mestrado do (a) aluno (a) **Angélica Rocha Martins**, matrícula nº **13/0165565**, intitulada “**Desenvolvimento de método para discriminação entre marcas de uísque e identificação de adulteração por espectroscopia UV-Vis e PLS-DA**”, apresentada no (a) Auditório Verde do Instituto de Química (IQ) da Universidade de Brasília (UnB) em 7 de agosto de 2015.

Prof. Dr. Jez Willian Batista Braga
Presidente de Banca (IQ/UnB)

Profa. Dra. Ana Cristi Basile Dias
Membro Titular (IQ/UnB)

Prof. Dr. Fábio Moreira da Silva
Membro Titular (IQ/UnB)

Profa. Dra. Simone Monteiro e Silva
Membro Suplente (IQ/UnB)

Em 7 de agosto de 2015.

DEDICATÓRIA

Dedico este trabalho ao meu marido Henrique, pessoa de maior influência nesta conquista, meu melhor amigo, quem mais me apoia e incentiva.

AGRADECIMENTOS

À minha mãe, por todo o esforço que me permitiu chegar até aqui.

Ao meu marido, por uma década de companheirismo, amizade, confiança, apoio e incentivo.

Ao meu orientador prof. Dr. Jez Willian Batista Braga pela oportunidade de realizar este trabalho, pelos ensinamentos, acompanhamento, confiança e paciência.

Ao meu coorientador, Dr. Márcio Talhavini, pelos conselhos, ensinamentos, atenção, apoio, incentivo, motivação, orientação e amizade.

Ao Instituto de Química da Universidade de Brasília pela oportunidade de realizar o mestrado.

À Polícia Federal, pela parceria no desenvolvimento do projeto, pela disponibilização das amostras e do espaço para realização das análises.

Ao Lanagro-MG e a Diageo pela doação de amostras.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida.

À Fernanda Lintomen, pelas amostras doadas.

Ao Maurício e ao Zacca, pela contribuição com o trabalho.

À Camila Letti, pela grande amizade, pelos ótimos momentos que passamos juntas na UnB, pelo acolhimento e por toda a ajuda que me deu e que ainda dará.

À Nancy, pela grande amizade, descontração e momentos agradáveis desde nosso primeiro semestre na UnB.

À Carla, pela amizade, atenção a qualquer hora do dia, companheirismo, prestatividade, conselhos e ajuda com as figuras desta dissertação.

À Karen, pela amizade, simpatia, conselhos, prestatividade e pelas palavras de conforto no momento em que precisei.

À Rosy, pelo carinho, simpatia e amizade.

Aos colegas e amigos do Grupo de Automação, Quimiometria e Química Ambiental (AQQUA) do IQ/UnB, por proporcionarem um ambiente familiar e acolhedor, pelos ótimos momentos que passamos juntos, pela descontração e risadas.

À Gabriele, pela amizade, confiança, torcida pelo meu sucesso e pelos momentos agradáveis no horário do café.

À Marianna Brandão, pela amizade e ajuda.

Aos colegas do INC, por proporcionarem um ambiente muito amigável e acolhedor, que fizeram com que minha passagem por lá se torne inesquecível.

À prof.^a Elaine Rose Maia, por toda ajuda durante minha graduação, o meu sincero agradecimento e admiração.

Aos professores, colegas e funcionários do IQ/UnB.

A todos que colaboraram direta ou indiretamente com este trabalho, o meu muito obrigada.

RESUMO

O uísque é uma bebida alcoólica de grande aceitação mundial e alto valor comercial. Devido a estes fatores, esta bebida tem sido alvo de falsificações de diversos tipos: outro tipo de bebida ser colorida com caramelo e comercializada como uísque, a bebida ser diluída com água ou misturada com outro uísque de qualidade inferior, o conteúdo de uma garrafa de uísque de maior valor comercial ser totalmente substituído por outro de menor valor. As análises de rotina realizadas pelas polícias Federal e Civil em amostras de uísques suspeitas não são eficientes para identificar este último tipo de falsificação, pois elas discriminam diferentes tipos de bebidas e não diferentes marcas de uísque. Com base nesta demanda, este trabalho de pesquisa foi desenvolvido em parceria com a Polícia Federal com o intuito de contribuir para a solução deste tipo de falsificação. Foram desenvolvidos modelos de discriminação de marcas de uísque através de espectroscopia UV-Vis aliada a análise discriminante por mínimos quadrados parciais (PLS-DA), cujos espectros foram obtidos sem nenhum preparo de amostra. Os modelos foram construídos com sete marcas de uísque: Red Label, Black Label, White Horse, Chivas Regal (12 anos), Ballantine's Finest, Old Parr e Natu Nobilis. A predição de amostras autênticas e falsas apresentaram taxa de classificação correta superior a 98% e 87%, respectivamente. Contudo, a aplicação dos modelos em casos reais de apreensão da Polícia Federal não apresentou a mesma eficiência. Suspeita-se que esse resultado se deve ao pequeno número de lotes disponível para o desenvolvimento dos modelos. Sugere-se que aumento da representatividade das amostras pela inclusão de mais lotes seja capaz de aumentar a eficiência do método em novas amostras.

ABSTRACT

Whisky is an alcoholic beverage with great worldwide acceptance and high commercial value. Due to these factors, this beverage has been used as a subject of various types of counterfeits: another type of beverage colored with caramel pigment and sold as whisky, dilution with water or mixing with other lower quality whisky, or even the replacement of the content in the bottle of a high commercial value whisky by another of lower value. Routine analysis carried out by the Federal and Civil Police on suspicious whisky samples are not efficient to identify the latter type of falsification, because they can discriminate different types of beverages instead of different brands of whisky. Based on this demand, this research was developed in partnership with the Federal Police in order to solution of this type of counterfeiting. Discrimination models of whisky brands were developed by UV-Vis spectroscopy combined with partial least squares for discriminant analysis (PLS-DA), whose spectra were obtained with no sample preparation. The models were built with seven whisky brands: Red Label, Black Label, White Horse, Chivas Regal (12 years), Ballantine's Finest, Old Parr and Natu Nobilis. The prediction of genuine and false samples showed correct classification rates over 98% and 93%, respectively. However, the application of the models in real cases of apprehension of the Federal Police did not provide the same efficiency, leading to inconclusive result rates ranging from 37.5% to 56.5%. This result was associated with variations in the beverage composition in different batches of the same brand, originated in production steps. It is suspected that this result is due the small number of batch available for models development. It is suggested that the samples representativeness increasing by the inclusion of more batches be able to increase the method's efficiency in new samples.

ÍNDICE

LISTA DE ABREVIATURAS E ACRÔNIMOS	xii
LISTA DE TABELAS	xiv
LISTA DE FIGURAS	xv
1. INTRODUÇÃO E OBJETIVOS	2
2. REVISÃO BIBLIOGRÁFICA	6
2.1 UÍSQUE	6
2.1.1 Definição.....	6
2.1.2 Produção	8
2.1.3 Composição.....	10
2.1.4. Análise de amostras de uísque.....	11
2.2 ESPECTROSCOPIA DE UV-VIS.....	15
2.3 QUIMIOMETRIA	17
2.3.1 Calibração Multivariada	18
2.3.1.1 PCA.....	19
2.3.1.2 PLS.....	21
2.3.1.2.1 Pré-processamentos	23
2.3.1.1.2 PLS-DA	24
2.3.1.1.3 Determinação do número de variáveis latentes do modelo por validação cruzada	25
2.3.1.1.4 Detecção de amostras anômalas	27
2.3.1.1.5 Figuras de mérito dos modelos de discriminação	29

3. DISCRIMINAÇÃO DE MARCAS DE UÍSQUE POR ESPECTROSCOPIA UV-VIS	34
3.1 MATERIAIS E MÉTODOS	34
3.1.1.1 Amostras autênticas	34
3.1.1.2 Amostras falsas	36
3.1.1.3 Amostras da apreensão.....	36
3.1.1.4 Reagentes	37
3.1.2 Obtenção dos espectros	37
3.2 ANÁLISE DE DADOS	37
3.3 RESULTADOS E DISCUSSÃO	41
3.3.1 Validação do método pela análise de amostras autênticas	44
3.3.1.1 Validação com amostras autênticas de Red Label	45
3.3.1.2 Validação com amostras autênticas de Black Label.....	48
3.3.1.3 Validação com amostras autênticas de White Horse.....	49
3.3.1.4 Validação com amostras autênticas de Chivas Regal	50
3.3.1.5 Validação com amostras autênticas de Ballantine's Finest	51
3.3.1.6 Validação com amostras autênticas de Natu Nobilis	52
3.3.1.7 Validação com amostras autênticas de Old Parr	53
3.3.2 Predição de amostras falsas.....	54
3.3.2.1 Análise de amostras falsas com o modelo de discriminação de Red Label	56
3.3.2.2 Análise de amostras falsas com o modelo de discriminação de Black Label	58
3.3.2.3 Análise de amostras falsas com o modelo de discriminação de White Horse.....	60
3.3.2.4 Análise de amostras falsas com o modelo de discriminação de Chivas Regal	61
3.3.2.5 Análise de amostras falsas com o modelo de discriminação de Ballantine's Finest	62
3.3.2.6 Análise de amostras falsas com o modelo de discriminação de Natu Nobilis	63

3.3.2.7 Análise de amostras falsas com o modelo de discriminação de Old Parr	64
3.3.3 Aplicação dos modelos em caso real.....	65
3.3.3.1 Análise de amostras reais a partir de modelo de discriminação de Red Label	68
3.3.3.2 Análise de amostras reais a partir de modelo de discriminação de Black Label	70
3.3.3.3 Análise de amostras reais a partir de modelo de discriminação de White Horse.....	72
3.3.3.4 Análise de amostras reais a partir de modelo de discriminação de Old Parr	73
3.4 CONCLUSÕES DO CAPÍTULO.....	75
4. CONCLUSÕES GERAIS.....	78
5. REFERÊNCIAS BIBLIOGRÁFICAS	80

LISTA DE ABREVIATURAS E ACRÔNIMOS

ABRABE	Associação Brasileira de Bebidas
DITEC	Diretoria Técnico Científica
DPER	Divisão de Perícias
DPF	Departamento de Polícia Federal
FN	Falso Negativo
FP	Falso Positivo
FTIR	Infravermelho com Transformada de Fourier (<i>Fourier Transform Infrared</i>)
INC	Instituto Nacional de Criminalística
LANAGRO	Laboratório Nacional Agropecuário
LDA	Análise Discriminante Linear (<i>Linear Discriminant Analysis</i>)
MLR	Regressão Linear Múltipla (<i>Multiple Linear Regression</i>)
NIR	Infravermelho Próximo (<i>Near Infrared</i>)
PCA	Análise por components Principais (<i>Principal Component Analysis</i>)
PCR	Regressão por Componentes Principais (<i>Principal Component Regression</i>)
PLS	Mínimos Quadrados Parciais (<i>Partial Least Squares</i>)
PLS-DA	Análise Discriminante por Mínimos Quadrados Parciais (<i>Partial Least Squares for Discriminant Analysis</i>)

PLSR	Regressão por Mínimos Quadrados Parciais (<i>Partial Least Squares Regression</i>)
RMSEC	Raiz quadrada do erro médio quadrático de calibração (<i>Root Mean Squared Error of Calibration</i>)
RMSECV	Raiz quadrada do erro médio quadrático de validação cruzada (<i>Root Mean Squared Error of Cross-Validation</i>)
SEPLAB	Setor de Perícias de Laboratório
SIMPLS	Modificação Estatisticamente Inspirada do PLS (<i>Statistically Inspired Modification of PLS</i>)
TCC	Taxa de Classificação Correta
TEF	Taxa de Eficiência
TFN	Taxa de Falso Negativo
TFP	Taxa de Falso Positivo
TSP	Taxa de Especificidade
TST	Taxa de Seletividade
UV-VIS	Ultravioleta-Visível
VL	Variáveis Latentes
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

LISTA DE TABELAS

Tabela 1. Compostos que compõem o padrão de identidade dos uísques ⁵	12
Tabela 2: Legendas para os parâmetros utilizados no cálculo das figuras de mérito	31
Tabela 3: Relação das amostras autênticas utilizadas no trabalho	35
Tabela 4: Relação das amostras falsas utilizadas no trabalho	36
Tabela 5. Relação das amostras provenientes da apreensão	36
Tabela 6: quantidades de amostras utilizadas nos conjuntos de calibração e validação	38
Tabela 7: Parâmetros dos modelos de calibração	44
Tabela 8. Parâmetros e figuras de mérito da predição de amostras autênticas	46
Tabela 9. Critérios de identificação de falsos positivo e negativo e verdadeiros positivo e negativo de predições de amostras falsas	55
Tabela 10. Parâmetros e figuras de mérito dos modelos de predição de amostras falsas.....	57
Tabela 11. Figuras de mérito de modelos de discriminação de amostras provenientes da apreensão da Polícia Federal	67
Tabela 12. Figuras de mérito de modelos de discriminação de amostras provenientes da apreensão da Polícia Federal	68

LISTA DE FIGURAS

Figura 1. À esquerda: selos de IPI autêntico (acima) e falso (abaixo). À direita: os mesmos selos de IPI iluminados com luz UV.	12
Figura 2. Extratos secos de uísque autêntico (esquerda) e falso (direita).....	13
Figura 3. Tipos de transição eletrônica decorrentes da absorção de radiação ultravioleta e visível. Adaptada da referência [32].....	17
Figura 4. Organização da matriz de dados X a partir de dados de UV-Vis, na qual as amostras são dispostas nas linhas e as variáveis nas colunas.	20
Figura 5. Decomposição da matriz de dados $X_{n,m}$ por PCA. Adaptada da referência [46]	21
Figura 6. Decomposição das matrizes X e Y em variáveis latentes por PLS. Adaptada da referência [46]	22
Figura 7. Representação esquemática do PLS-DA. Adaptada da referência [58]	24
Figura 8. exemplo de gráfico de dispersão de classe.....	25
Figura 9. a) importância da escolha correta do número de VL; b) gráfico do valor de RMSECV em função da quantidade de variáveis latentes.	27
Figura 10. Identificação de <i>outliers</i> através do gráfico de influência e resíduos	29
Figura 11. Organograma da rotina de exclusão de <i>outliers</i> na construção dos modelos de calibração.	40
Figura 12. Fluxograma para análise de amostras desconhecidas nos modelos de discriminação desenvolvidos.....	41
Figura 13. Espectros UV-Vis das amostras autênticas de uísque 12 e 8 anos ou menos, utilizadas no conjunto de calibração. Black Label, Old Parr, Chivas, Red Label, White Horse, Ballantine's, Natu Nobilis.	42
Figura 14. A) Gráfico de influência (T^2) e resíduos (Q) do modelo de discriminação entre amostras de Red Label (\blacktriangledown) e as demais marcas. B) Ampliação da região dentro do limite de 99,9%.	46
Figura 15. Resultados obtidos para modelo de discriminação entre amostras autênticas de Red Label (\blacktriangledown) e as amostras de Black Label (*), White Horse (\blacksquare), Chivas Regal (+), Ballantine's (\diamond), Natu Nobilis (\blacktriangle), Grand Old Parr (\star) e as marcas não incluídas na calibração (demais símbolos).	47

Figura 16. Resultados obtidos para modelo de discriminação entre amostras autênticas de Black Label (*) e as demais marcas.....	48
Figura 17. Resultados obtidos para modelo de discriminação entre amostras autênticas de White Horse (■) e as demais marcas.....	49
Figura 18. Resultados obtidos para modelo de discriminação entre amostras autênticas de Chivas Regal (+) e as demais marcas.	50
Figura 19. Resultados obtidos para modelo de discriminação entre amostras autênticas de Ballantine’s Finest (◇) e as demais marcas.	52
Figura 20. Resultados obtidos para modelo de discriminação entre amostras autênticas de Natu Nobilis (▲) e as demais marcas.	53
Figura 21. Resultados obtidos para modelo de discriminação entre amostras autênticas de Grand Old Parr (★) e as demais marcas.	54
Figura 22. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas Red Label (▼) e as amostras de Black Label (*), White Horse (■), Chivas Regal (+), Ballantine’s (◇), Natu Nobilis (▲), Grand Old Parr (★). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.	56
Figura 23. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Red Label (▼). As demais marcas do modelo são: Black Label (*), White Horse (■), Chivas Regal (+), Ballantine’s (◇), Natu Nobilis (▲), Grand Old Parr (★)	57
Figura 24. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Black Label (*). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.	58
Figura 25. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Black Label (*).....	59
Figura 26. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de White Horse (■). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.	60

Figura 27. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de White Horse (■).	60
Figura 28. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Chivas Regal (+). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.	61
Figura 29. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Chivas Regal (+).	61
Figura 30. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Ballantine's (◇). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.	62
Figura 31. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Ballantine's (◇);	62
Figura 32. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Natu Nobilis (▲). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.	63
Figura 33. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Natu Nobilis (▲).	63
Figura 34. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Old Parr (★). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.	64
Figura 35. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Old Parr (★).	64
Figura 36. Espectros de absorção de UV-Vis de amostras autênticas da apreensão (vermelho), amostras falsas (verde) e amostras do conjunto de calibração (azul).	66

Figura 37. Resultados obtidos para modelo de discriminação entre amostras de Red Label (▼) e as demais marcas provenientes da apreensão da Polícia Federal. As demais marcas do modelo são: Black Label (*), White Horse (■) e Grand Old Parr (★)	70
Figura 38. Resultados obtidos para modelo de discriminação entre amostras de Black Label (*) provenientes da apreensão da Polícia Federal.....	71
Figura 39. Resultados obtidos para modelo de discriminação entre amostras de White Horse (■) provenientes da apreensão da Polícia Federal.....	73
Figura 40. Resultados obtidos para modelo de discriminação entre amostras de Old Parr (★) provenientes da apreensão da Polícia Federal.	74

CAPÍTULO 1
Introdução e Objetivos

1. INTRODUÇÃO E OBJETIVOS

A expansão da pirataria atingiu tal dimensão que se tornou um problema mundial, e se trata, de forma geral, da apropriação e reprodução da ideia ou produto sem a permissão do autor, fabricante ou detentor da patente.¹

No Brasil, o combate a essa prática é uma realidade como política pública e está presente em diferentes frentes de atuação, como legislativa, institucional, repressiva, educacional e internacional². Dentre as perdas sofridas por tal prática, podem ser citadas a não arrecadação tributária e tudo o que dela decorre, empobrecimento da nação, prejuízo da livre concorrência, riscos à saúde do consumidor, péssimas condições de trabalho dos produtores das mercadorias piratas, além de poder estar relacionada a outros crimes, como receptação, roubo, falsidade de selos públicos, crimes contra a saúde pública, contrabando, homicídios, tráfico de drogas e armas, dentre outros.²

A falsificação de um produto ocorre quando é feita uma cópia não autorizada do produto com a intenção de ser comercializado como legítimo.³ Dentre os produtos pirateados e falsificados no Brasil estão as bebidas alcoólicas, especialmente os uísques.

As falsificações de uísque podem ocorrer de diversas formas, desde as mais grosseiras, como álcool de cereais misturado com água de torneira e caramelo, à falsificações mais sofisticadas, como quando se comercializa uísques de menor qualidade envasados em frascos de marcas de maior valor comercial.

Para a correta tipificação do crime de adulteração ou falsificação de bebidas é necessária a realização de análises físico-químicas do material questionado.⁴

Entretanto, as análises realizadas atualmente visam determinar se a bebida em questão obedece aos padrões de identidade definidos pela Instrução Normativa nº 15 de 2011, do Ministério da Agricultura, Pecuária e Abastecimento⁵, que delimita os teores dos componentes que definem o uísque. Através destes parâmetros, é possível distinguir o uísque das demais bebidas alcoólicas. A execução de análises que atestem que determinada bebida é de fato uísque pode responder questões de interesse comercial ou de saúde

pública, entretanto, na esfera criminal, estes dados muitas vezes não são suficientes.⁴

Os dados gerados por estas análises podem detectar vários tipos de adulterações, principalmente as mais grosseiras, em que o líquido em questão perdeu as características definidas pela referida Instrução Normativa⁴. Entretanto, não permite identificar se o uísque contido na garrafa é o mesmo referente à marca do rótulo, tipos e tempo de envelhecimento da bebida, o que dificulta a identificação de alguns tipos de adulteração, como, por exemplo, a substituição do conteúdo da garrafa por um uísque de menor tempo de envelhecimento e, portanto, de menor valor comercial. De tal forma, percebe-se que há a necessidade de metodologias de análise que permitam tal discriminação.

Tendo em vista a relevância e a necessidade de um método capaz de detectar adulterações/falsificações sutis, o objetivo da presente dissertação de mestrado consiste no desenvolvimento de métodos baseados em espectroscopia e análise multivariada que permitam identificar e discriminar uísques com relação à marca e tempo de envelhecimento, assim como detectar adulterações, com o intuito de contribuir com o trabalho das polícias Federal e Civil. O método espectroscópico de análise selecionado foi de absorção molecular na região de ultravioleta e visível devido ao fato da grande maioria dos laboratórios de perícia do país possuírem tal equipamento, da sua facilidade de operação e capacidade de detectar compostos presentes em baixas concentrações nas amostras.

Esta dissertação foi dividida em cinco capítulos, nos quais são expostos conteúdos bibliográficos e teóricos, resultados de diferentes predições de amostras de uísque e as referências consultadas pertinentes ao estudo realizado. Neste Capítulo 1 consta uma breve introdução e o objetivo geral desta dissertação.

No capítulo 2 é apresentada a revisão bibliográfica dos conteúdos abordados neste trabalho, dividida em três seções. A seção 1 fala sobre o uísque, sua definição, produção e composição. A seção 2 trata brevemente da espectroscopia de UV-Vis e a seção 3 trata da quimiometria.

O capítulo 3 descreve o desenvolvimento dos modelos PLS-DA e sua aplicação na predição de amostras autênticas, falsas e aplicação em um caso real de apreensão da Polícia Federal.

Por fim, o capítulo 4 apresenta as conclusões gerais e o capítulo 5 apresenta as referências bibliográficas desta dissertação.

CAPÍTULO 2
Revisão Bibliográfica

2. REVISÃO BIBLIOGRÁFICA

2.1 UÍSQUE

O nome uísque surgiu da expressão *uisge beatha*, equivalente gaélico do latim *aquae vitae* – “água da vida”. Bebida de elevado teor alcoólico e sabor marcante, atraiu diversos apreciadores ao longo do tempo e vem movimentando um enorme mercado mundial de produção de uísque. Só no Brasil, em 2012 foram vendidos cerca de 24 milhões de litros.⁶

Sua origem é incerta, pois não há registros de quem realmente deu origem à bebida, assim como exatamente em qual país e ano. O que se pode afirmar, de acordo com Marco de La Roche no livro *um brinde à vida: a história das bebidas* (lançado pela associação brasileira de bebidas em comemoração aos seus 40 anos) é que a primeira menção ao uísque foi feita em 1494, em um documento tributário escocês, que fala na venda de barris de malte a frei John Corr, o que demonstra que a bebida pode ter tido ligação com monges e mosteiros. Por esse motivo, especula-se que o uísque tenha sido produzido na Irlanda antes mesmo que na Escócia, pois aquela ilha sempre foi centro religioso e, na Idade Média, os monges eram os principais produtores de bebidas alcoólicas com fins medicinais.⁷

A seguir, será apresentada uma breve descrição do uísque, incluindo sua definição de acordo com a legislação brasileira, assim como produção e composição.

2.1.1 Definição

A definição de uísque varia de país para país, devido às suas regulamentações legais. Entretanto, o que é internacionalmente aceito é que a bebida se trata de um destilado feito de cereais, levedura e água, maturado em barris de madeira por determinado período de tempo.⁸ De acordo com a legislação brasileira, a definição de uísque, whisky ou whiskey “é a bebida com graduação alcoólica de trinta e oito a cinquenta e quatro por cento em volume, a vinte graus Celsius, obtida do destilado alcoólico simples de cereais envelhecido, parcial ou totalmente maltados, podendo ser adicionado de álcool etílico potável de origem agrícola, ou de destilado alcoólico simples de cereais, bem como de

água para redução da graduação alcoólica e caramelo para correção da cor” e deve ser envelhecido em barril de madeira por no mínimo 2 anos.⁹

Há diversos tipos diferentes de uísques, baseados no cereal utilizado como matéria-prima, sendo dada a cada um deles uma denominação, de acordo com a legislação brasileira:

- **Puro malte:** quando elaborado exclusivamente com destilado alcoólico de malte envelhecido, com coeficiente de congêneres (voláteis não-álcool) não inferior a $350 \text{ mg} \cdot 100 \text{ mL}^{-1}$, em álcool anidro.⁹
- **Blended ou cortado:** há diversos tipos de *blends*, que obedecem à legislação do país produtor. Entretanto, a legislação brasileira define o uísque *blended* como obtido pela mistura de, no mínimo, trinta por cento de destilado alcoólico de malte envelhecido com destilados alcoólicos de cereais, álcool etílico potável de origem agrícola ou ambos, envelhecidos ou não, com coeficientes de congêneres não inferior a $100 \text{ mg} \cdot 100 \text{ mL}^{-1}$, em álcool anidro.⁹
- **Grain whisky ou uísque de cereais:** bebida obtida de cereais reconhecidos internacionalmente na produção de uísque, sacarificados, total ou parcialmente, por diastases (enzimas que atuam sobre o amido) da cevada maltada, adicionada ou não de outras enzimas naturais e destilada em alambique ou coluna, envelhecido por um período mínimo de dois anos, com coeficiente de congêneres não inferior a $100 \text{ mg} \cdot 100 \text{ mL}^{-1}$ em álcool anidro.⁹
- **Bourbon whisky:** produzido nos Estados Unidos da América, de acordo com sua legislação, sem prejuízo à definição de uísque estabelecida pela legislação brasileira. Elaborado com no mínimo 50% de destilado alcoólico simples de milho, sacarificado com cevada maltada, envelhecido por um período mínimo de dois anos, adicionado ou não de álcool etílico potável de origem agrícola, podendo ser envelhecido ou não, com o coeficiente de congêneres não inferior a $150 \text{ mg} \cdot 100 \text{ mL}^{-1}$ em álcool anidro.⁹
- **Tennessee whisky:** quando for produzido nos Estados Unidos da América, de acordo com sua legislação, sem prejuízo à definição de uísque estabelecida pela legislação brasileira.⁹ Similar ao bourbon,

difere em um passo adicional de filtração por uma coluna de carvão de 3 m de espessura, denominado “*charcoal mellowing*”, antes de ser colocado no barril para maturação.⁸

De acordo com o Decreto nº 6871 de 04 de junho de 2009 do MAPA “o uísque engarrafado no território nacional somente poderá fazer uso das denominações de origem, ou seja, scotch whisky, canadian whisky, irish whisky, bourbon whisky, tennessee whisky e outras reconhecidas internacionalmente, quando elaborado, exclusivamente, com matérias-primas importadas a granel, cujos destilados sejam produzidos e envelhecidos em seus respectivos países de origem e que mantenham as características determinadas por suas legislações, podendo apenas ser adicionado de água para redução da graduação alcoólica e de caramelo para a correção da cor”.⁹

Os principais cereais utilizados na produção de uísque são milho, centeio, cevada e trigo. A cevada é utilizada principalmente na forma de malte, devido ao sabor que proporciona ao destilado e também pelo fato de ser fonte de enzimas para converter o amido do malte e dos grãos não maltados em açúcares fermentáveis.¹⁰

2.1.2 Produção

Há diversos detalhes que proporcionam as peculiaridades do sabor e odor do uísque, como, por exemplo, o cereal utilizado ou a mistura de cereais e a levedura utilizada, uma vez que existem vários tipos e geralmente as destilarias tem sua própria cultura.^{8,11}

Neste trabalho foi utilizado apenas uísque de cevada maltada, portanto todo o processo de produção será referente apenas a este cereal.

A produção de uísque maltado é bastante complexa e ocorre em diversas etapas, iniciando pela maltagem, processo que tem por finalidade elevar o teor enzimático dos grãos de cevada por meio da síntese de amilases, proteases, glucanases e de outras enzimas.¹⁰

A maltagem é iniciada fornecendo-se a umidade necessária à cevada para dar início à germinação, momento em que se ativa o sistema enzimático do grão. Após atingidos os desejados nível enzimático e grau de germinação, a etapa seguinte é a secagem, cujo objetivo é obter o produto seco e armazenável e

impedir posterior desenvolvimento da atividade biológica. Os grãos são secos até umidade de aproximadamente 5% em estufas, onde também ocorre a queima da turfa, material vegetal, que proporciona ao uísque um sabor defumado, provenientes de compostos fenólicos. A queima da turfa não é obrigatória, sendo utilizada apenas quando se deseja que o uísque tenha seu sabor característico, porém é importante na determinação das características do uísque escocês.^{8,10,11}

A etapa seguinte é a de maceração. O malte é moído e misturado com água quente em grandes tonéis, geralmente feitos de cobre ou aço inoxidável, com cobertura de cobre para melhor manter a temperatura. Nesta etapa ocorre a liberação dos açúcares fermentáveis e o líquido resultante é chamado de mosto.⁸

Na etapa de fermentação do mosto os açúcares obtidos a partir da hidrólise do amido são metabolizados pelas leveduras e convertidos em etanol, CO₂ e subprodutos.¹⁰ Para a fermentação de uísque pode-se utilizar misturas de leveduras, escolhida de acordo com o sabor que se deseja atribuir à bebida, inclusive a levedura de cerveja por proporcionar sabor mais agradável.¹²

Após a fermentação, inicia-se a destilação do mosto, sem a necessidade de separação das leveduras presentes no fermentado. A destilação pode ser realizada de duas formas: em alambique ou contínua em coluna. Na primeira, geralmente são feitas duas destilações (eventualmente três), gerando uma bebida de sabor mais forte.^{10,12} Os alambiques normalmente são feitos de cobre por ser bom condutor de calor e por sua resistência ao desgaste. O produto da primeira destilação, chamado de “*low wine*” possui teor alcóolico de aproximadamente 21-23% (v/v). Ao final da segunda destilação, a bebida possui teor alcóolico de aproximadamente 70% (v/v). A destilação contínua em coluna produz bebida de sabor mais suave, normalmente usada na produção de uísques de cereais (não maltado) para formação do *blended*.^{8,10,12}

A etapa seguinte é a de maturação em barris de madeira, geralmente carvalho, importante passo no desenvolvimento das propriedades sensoriais da bebida, uma vez que uísques recentemente destilados geralmente apresentam características sensoriais inaceitáveis.¹⁰ Uma grande quantidade de fatores pode influenciar na qualidade final do produto, como o tipo de madeira utilizada, a forma como foi produzido e até mesmo as condições climáticas durante o armazenamento podem afetar o processo de maturação.¹³ Cor, pH, ácidos,

ésteres e açúcares são muito influenciados pelo tipo do barril, sendo o carvalho americano (*Quercus alba*) e carvalho europeu (*Quercus petraea*) as espécies mais utilizadas.^{10,12} O tempo de envelhecimento mínimo depende das especificações legais de cada país, por exemplo uísques *bourbon* e de grãos americanos são armazenados por no mínimo 1 ano, enquanto que uísques escoceses, irlandeses e canadenses devem ser envelhecidos por no mínimo 3 anos.¹⁰ No rótulo da garrafa deve constar o tempo de envelhecimento mínimo. No rótulo dos *blended* deve constar o do uísque de menor tempo de envelhecimento utilizado, incluindo os de cereais.⁸

Após a maturação desejada é feito o *blend* ou corte misturando-se bebidas mais encorpadas com menos encorpadas, na proporção que convier à destilaria, obedecendo às legislações do país de origem, no intuito de elaborar um uísque com sabor único. Nesta etapa a bebida apresenta elevado teor alcoólico, cerca de 70% (v/v). Após a mistura, a bebida é diluída com água para ajustar o teor alcoólico a uma faixa de 40%-45% (v/v).¹² Finalizada esta etapa, o uísque é filtrado e já está pronto para ser engarrafado.

2.1.3 Composição

O uísque é uma bebida de composição química complexa, possuindo centenas de compostos, porém sua composição química era pouco conhecida até final do século XIX. Os primeiros trabalhos a respeito de uísques escoceses descreviam métodos para a detecção de álcoois, aminas, aldeídos, ésteres e furfural. Com o avanço de novas técnicas cromatográficas e seu acoplamento com espectrometria de massas, o número de compostos conhecidos aumentou significativamente.¹⁰ Atualmente são conhecidos centenas de congêneres no uísque, incluindo álcoois, ácidos, ésteres, compostos carbonilados, fenóis, hidrocarbonetos e compostos com nitrogênio e enxofre.¹⁰

A maior parte dos compostos presentes no uísque é adquirida pela absorção de componentes da madeira do barril onde ocorre a maturação. O carvalho consiste de aproximadamente 10% (m/m) de óleos voláteis, ácidos voláteis e não voláteis, açúcares, taninos, pigmentos e compostos inorgânicos, porém a composição do extrato depende de alguns fatores, dentre eles a espécie do carvalho utilizado, o pré-tratamento que a madeira sofreu antes da maturação.

Um exemplo da influência da espécie da madeira é o fato de o carvalho europeu possuir maior concentração de taninos e menor concentração de lactonas que o carvalho americano. Portanto, a composição do uísque é bastante influenciada por inúmeros fatores.¹²

Além de etanol, outros álcoois são encontrados, como isobutanol, n-propanol e 2 e 3-metilbutanol. Outros compostos também encontrados em maior quantidade são os acetato de etila e acetaldeído, que, juntamente com os álcoois citados, formam o conjunto dos chamados componentes majoritários. Estes compostos são produzidos na etapa de fermentação e sua concentração é influenciada pela destilação e pelo *blending*.¹⁴

Além dos componentes majoritários, também estão presentes diversos componentes minoritários provenientes do processo de maturação da matéria-prima utilizada. Como visto na seção anterior, esses componentes minoritários são responsáveis pelas características sensoriais do uísque e compreendem álcoois, ácidos graxos, aldeídos, ésteres, fenóis e compostos contendo nitrogênio e enxofre. A presença e concentração dos componentes minoritários são características de cada processo de fabricação, sendo parâmetros de grande utilidade em análises de autenticidade¹⁵

A presença de compostos fenólicos voláteis é proveniente da queima da turfa na etapa de secagem da cevada maltada. Estes compostos fenólicos são responsáveis pela cor marrom-dourada do uísque¹⁶. Em alguns países, inclusive aqui no Brasil, esta cor pode ser ajustada com a adição de caramelo, porém o teor máximo de caramelo que pode ser adicionado à bebida é definido pela legislação.

2.1.4. Análise de amostras de uísque

Como citado anteriormente na Introdução, o Ministério da Agricultura sugere análises de uísque baseadas na determinação dos teores dos componentes que constituem seu padrão de identidade, delimitados pela Instrução Normativa nº 15 de 2011.⁵ Estes compostos são mostrados na Tabela 1.

Tabela 1. Compostos que compõem o padrão de identidade dos uísques⁵.

Item		Limite Mínimo	Limite Máximo	Classificação
1	Acidez volátil, em ácido acético, em mg/100 mL de álcool anidro	---	150	---
2	Álcool superior (somatório do álcool n-propílico, álcool iso-butílico e álcoois iso-amílicos) em mg/100 mL de álcool anidro	---	300	---
3	Aldeídos, em aldeído acético, mg/ 100 mL de álcool anidro	---	20	---
4	Coeficiente de congêneres, em mg/100 mL de álcool	> 350	---	Uísque malte puro
		> 100	---	Uísque cortado ou uísque de cereais
		> 150	---	Bourbon Whisky
5	Ésteres, em acetato de etila, em mg/100 mL de álcool anidro	---	150	---
6	Gradação alcoólica, em % v/v a 20 °C	38	54	---
7	Somatório de Furfural e hidroximetilfurfural, em mg/100 mL de álcool anidro	---	5	---

As análises realizadas pela Polícia Federal são descritas a seguir. Todas foram retiradas da referência 4 por ter sido escrita por peritos desta instituição.

a) Análise visual das características do material

É parte da rotina forense o exame de embalagens, lacres, selos de segurança e todos os demais elementos que possam ser indicativos de fraudes ou alteração do produto original.



Figura 1. À esquerda: selos de IPI autêntico (acima) e falso (abaixo). À direita: os mesmos selos de IPI iluminados com luz UV.

b) Teor alcoólico

Análise muito realizada, uma vez que é comum encontrar bebidas falsas com teor alcoólico menor que o esperado para bebidas autênticas.

c) Resíduo do evaporado ou “extrato seco”

Uma alíquota da bebida é seca em estufa, a 100 °C em cápsula de porcelana. Uísques autênticos podem possuir corante caramelo em pequenas quantidades, de forma que o extrato seco não ultrapassa 0,3% m/v. Por outro lado, uísques falsos costumam ter valores de extrato seco da ordem de 1,0% a 1,5% m/v. Nota-se também uma diferença visual entre extrato seco de uísque falso e autêntico, pois a bebida falsa com alto teor de caramelo resulta em um extrato viscoso, como pode ser visto na Figura 2. Este teste tem aplicação limitada, porque nem todo uísque falso possui grande quantidade de caramelo, da mesma forma, misturas de uísques falsos com autênticos geram resultados intermediários.

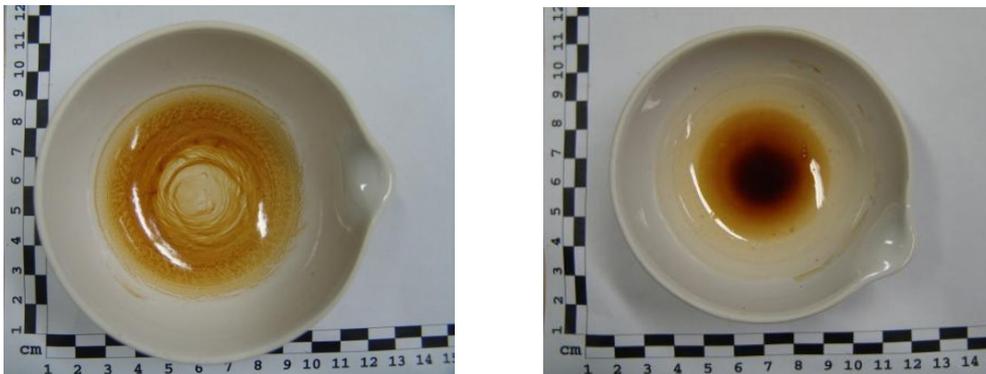


Figura 2. Extratos secos de uísque autêntico (esquerda) e falso (direita)

d) Determinação de composição de voláteis por CG-EM-Headspace

Os uísques autênticos possuem dezenas de componentes voláteis em sua composição. Caso a bebida em análise seja uma mistura de água, etanol e corante, ou caso se trate de uma mistura de pequena quantidade de uísque com cachaça e água, a análise de voláteis não mostrará a presença dos voláteis característicos ou mostrará em proporções diferentes das esperadas para o uísque autêntico.

e) Condutividade elétrica

O uísque autêntico possui baixa concentração de íons, sendo a condutividade elétrica esperada entre 20 e 35 mS.cm⁻¹. Uma vez que a bebida adulterada muitas vezes tem adição de água de torneira, espera-se nestes casos que a concentração iônica seja mais elevada que em uísques autênticos. De fato, a condutividade de uísques adulterados situa-se em geral acima de 70 mS.cm⁻¹, sendo que valores acima de 125 mS.cm⁻¹ já foram obtidos nos laboratórios da PF.

Vale ressaltar que os resultados destes testes podem ser inconclusivos dependendo do tipo de falsificação. Portanto, para se ter uma maior confiança no resultado, muitas vezes é necessário combinar a informação de diferentes métodos para se afirmar com segurança que a amostra é falsa, tendo em vista ser muito grave condenar um inocente.

A partir de pesquisas pela literatura foram encontrados trabalhos que relatam a identificação de adulteração de uísques, porém utilizam cromatografia gasosa e microextração em fase sólida.^{17,18,19,20,21} Também foram encontrados trabalhos que utilizam espectroscopia FTIR, NIR, Raman e UV-Vis na identificação de adulteração destas bebidas através da determinação do teor de etanol e/ou metanol. Entretanto, estes trabalhos utilizaram quantidade restrita de amostras e de diferentes marcas e na maioria das vezes não utilizaram amostras falsas, provenientes de apreensões policiais.^{22,23,24,25,26,27,28}

De acordo com a revisão da literatura realizada, poucos trabalhos foram encontrados utilizando espectroscopia de UV-Vis na análise de amostras de uísque, pela discriminação entre marcas e entre regiões de produção e na identificação de adulteração.^{25,27,28,29} Dois destes trabalhos merecem destaque por se assemelharem ao trabalho apresentado nesta dissertação, que foram realizados utilizando espectroscopia na região de UV-Vis. Mackenzie e Ayllot²⁸ utilizaram a mesma metodologia de obtenção dos espectros, porém com diferente análise de dados. Monitoraram 10 comprimentos de onda (220, 230, 240, 250, 260, 280, 300, 320, 340 e 360 nm) em cada espectro e construíram um algoritmo baseado na soma dos quadrados dos escores (z), definidos pela seguinte equação, para comparações entre amostras de referências e questionadas:

$$z(\lambda) = \frac{x-X}{\sigma} \quad (1)$$

onde x é a absorvância medida para a amostra questionada, X é a absorvância média para a marca de referência e σ é o desvio padrão da absorvância da marca de referência. O trabalho consistiu apenas na identificação de adulteração, não discriminaram entre marcas.

Por outro lado, Cantarelli et al²⁷ também utilizaram análise multivariada na análise dos dados e fizeram discriminação entre marcas, através de PCA, PLS-DA e LDA (Análise Discriminante Linear, do inglês *Linear Discriminant Analysis*) e PLS-DA, porém utilizaram uma diferente metodologia de obtenção dos espectros, pois fizeram preparo de amostra. Fizeram diluição a 25 mL de uma alíquota de 2 mL de amostra e 10 mL de solução tampão de *Clark and Lubs* a pH 12 (KCl e NaOH).

O fato de haverem poucos trabalhos em discriminação entre marcas indica que existe uma demanda por métodos que são capazes de detectar esse tipo de fraude/adulteração em uísques e que a metodologia utilizada ainda não foi realizada por outros grupos de pesquisa.

2.2 ESPECTROSCOPIA DE UV-VIS

A espectroscopia de absorção molecular na região de ultravioleta e visível é uma técnica analítica que compreende faixa do espectro eletromagnético entre 160 e 780 nm. Suas medidas tem ampla aplicação na determinação quantitativa de uma grande variedade de espécies orgânicas e inorgânicas.³⁰

A espectroscopia de absorção molecular por transmissão está baseada na medida de transmitância T ou absorvância A de soluções contidas em células transparentes, tendo um caminho óptico de b cm. A concentração c de um analito absorvente está relacionada linearmente com a absorvância, conforme a equação 2:

$$A = -\log T = \log \frac{P_0}{P} = \epsilon bc \quad (2)$$

sendo ϵ a absorvidade molar ($\text{L mol}^{-1} \text{ cm}^{-1}$).³⁰

A absorção da radiação ultravioleta ou visível resulta da excitação de elétrons de ligação. Como consequência, os comprimentos de onda dos picos

de absorção podem ser correlacionados com os tipos de ligações nas espécies em estudo. Contudo, como o espectro nessa região é constituído de bandas largas, sua aplicação para estudos qualitativos é limitada, mas apresenta grande potencial para determinações quantitativas de compostos contendo grupos absorventes devido a sua sensibilidade e baixo erro associado às medidas instrumentais.³⁰

Três tipos de transições podem ocorrer em absorção molecular na região de UV-Vis:

- a) Envolvendo elétrons π , σ e n , ocorrem em moléculas e íons orgânicos e alguns ânions inorgânicos
- b) Envolvendo elétrons d e f , ocorrem em íons de metais de transição e de metais da série dos lantanídeos e actinídeos
- c) Envolvendo transferência de carga, ocorrem em alguns tipos específicos de complexos inorgânicos.³⁰

Neste trabalho serão abordadas somente as transições envolvendo elétrons π , σ e n por serem as mais relevantes para o estudo. Vale também ressaltar que a técnica não será profundamente abordada, devido ao fato de já estar muito bem estabelecida e ser bastante tratada em livros-texto.

Os tipos de transições envolvendo elétrons π , σ e n são:

a) Transições $\sigma \rightarrow \sigma^*$

Transição que envolve mais alta energia de excitação, correspondendo à comprimentos de onda na região ultravioleta de vácuo (<185 nm), sendo observada em alcanos.^{30,31}

b) Transições $n \rightarrow \sigma^*$

Observada em compostos saturados contendo pares de elétrons não ligantes. Podem ser originadas por radiação na região entre 150 e 250 nm e na presença de solventes polares os máximos de absorção sofrem deslocamento para comprimentos de ondas menores. Como pode ser visto na Figura 3, este tipo de transição requer menor energia de excitação que a transição $\sigma \rightarrow \sigma^*$.^{30,31}

c) Transições $n \rightarrow \pi^*$ e $\pi \rightarrow \pi^*$

Observadas em compostos insaturados. A maioria das aplicações da espectroscopia de absorção molecular UV-Vis a compostos orgânicos envolve estes tipos de transição devido ao fato das energias necessárias para promover excitação dos elétrons n e π ao nível π^* situarem-se em região espectral experimentalmente conveniente (200 a 700 nm).³⁰

A mudança de polaridade do solvente também influencia nos máximos de absorção destas transições. Ao se aumentar a polaridade do solvente, os picos de absorção das transições $n \rightarrow \pi^*$ geralmente sofrem deslocamento para menores comprimentos de onda. Normalmente, uma tendência oposta é observada nas transições $\pi \rightarrow \pi^*$.³¹

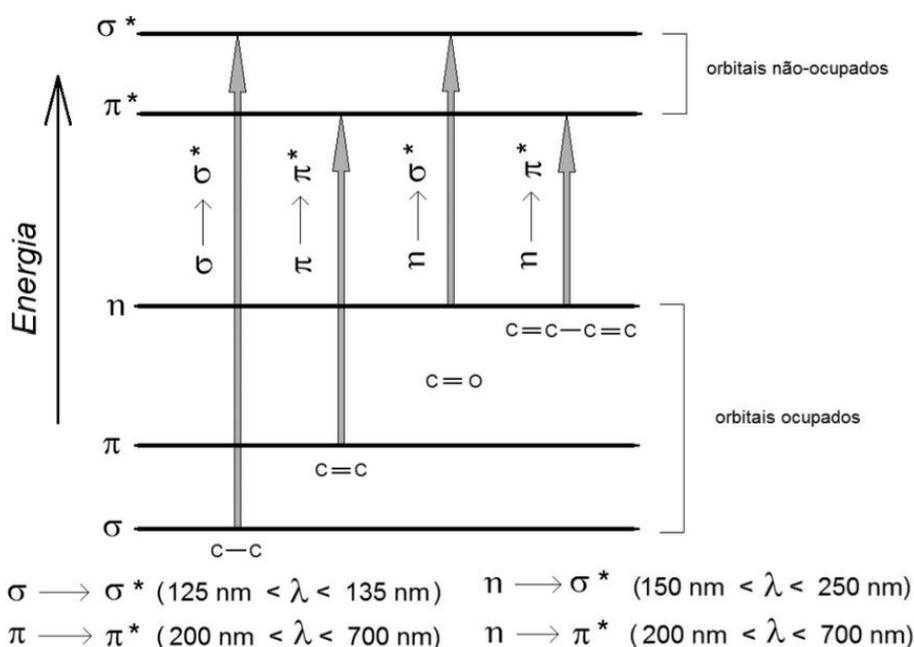


Figura 3. Tipos de transição eletrônica decorrentes da absorção de radiação ultravioleta e visível. Adaptada da referência [32]

2.3 QUIMIOMETRIA

A quimiometria pode ser definida como uma disciplina da Química que utiliza métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada, assim como extrair o máximo de informação relevante a partir da análise dos dados obtidos.^{33,34}

O termo quimiometria foi introduzido por Svante Wold e Bruce R. Kowalsky no início dos anos 1970 e se desenvolveu rapidamente.³⁵ Uma explicação para

tal fato é que neste período se observou a evolução da informática, uma vez que a aplicação de métodos de processamento de sinal, análise multivariada e interpretação de grande volume de dados químicos se desenvolveu com o aumento da capacidade computacional.^{36,35}

O primeiro artigo sobre quimiometria (com esta nomenclatura) foi publicado por Wold em 1972³⁷ e, no Brasil, a primeira dissertação de mestrado envolvendo métodos quimiométricos foi defendida em 1981 por Ieda Scarmínio.^{36,38}

A quimiometria possui ampla aplicação não apenas na química analítica, mas também em outras áreas da química, assim como em outras áreas da ciência, como por exemplo engenharia de produção e alimentos, biologia, geoquímica, ciências forenses, dentre outras.³⁹

Algumas das principais linhas de pesquisas na quimiometria envolvem planejamento de experimentos, reconhecimento de padrões, resolução de curvas multivariadas, calibração multivariada.³⁹ As ferramentas quimiométricas que foram utilizadas neste trabalho são detalhadas a seguir.

2.3.1 Calibração Multivariada

Em Química Analítica, calibração é o procedimento que relaciona a medida instrumental ao analito de interesse. A partir da relação matemática encontrada, é possível estimar a propriedade de interesse em outras amostras com base em suas medidas instrumentais⁴⁰.

Dependendo da necessidade, a calibração pode ser uni ou multivariada. Porém, com os avanços da instrumentação e da computação, houve também um aumento na capacidade de aquisição de dados que proporcionou a obtenção de inúmeras variáveis em um único ensaio, levando à necessidade da calibração multivariada.⁴⁰

A calibração univariada utiliza apenas uma variável por amostra na construção de um modelo que irá estimar a propriedade de interesse em outras amostras. Por outro lado, a multivariada utiliza mais de uma variável por amostra, podendo até mesmo usar, no caso de espectroscopia, o espectro inteiro. A grande vantagem em se utilizar um grande número de variáveis é que é possível contornar problemas com seletividade encontrados na utilização de uma variável. Em casos em que há interferência no sinal da propriedade de interesse,

a calibração univariada pode produzir resultados duvidosos, pois ela exige uma medida altamente seletiva para o analito de interesse. Nestes casos, a calibração multivariada se torna mais eficiente, podendo ser usada quando não há nenhuma variável seletiva, possibilitando previsões na presença de sobreposição de sinais e de interferentes.^{40,41}

O desenvolvimento de um método empregando calibração multivariada é composto por duas etapas. A primeira é a calibração, na qual se relacionam as medidas instrumentais (matriz \mathbf{X}) e a(s) propriedade(s) de interesse (matriz ou vetor \mathbf{y}) por meio de uma regressão. O modelo multivariado é o resultado desta regressão. O vetor \mathbf{y} consiste das variáveis dependentes enquanto que a matriz \mathbf{X} consiste das independentes.⁴²

Após a obtenção de um modelo de calibração otimizado, a etapa seguinte é aplicar esse modelo para futuras estimativas da propriedade de interesse em novas amostras, que são utilizadas para validação do método proposto.⁴²

Há diversos modelos de calibração multivariada, como por exemplo Regressão Linear Múltipla (MLR), Regressão por Componentes Principais (PCR) e Regressão por Mínimos Quadrados Parciais (PLSR).⁴⁰ Porém, tendo em vista os modelos aplicados nessa dissertação, apenas a Análise por Componentes Principais (PCA) e PLS serão detalhados.

2.3.1.1 PCA

A Análise por componentes Principais (PCA, do inglês *Principal Component Analysis*) é a base da análise de dados multivariada.⁴³ O princípio da PCA é construir um novo sistema de eixos, denominados componentes principais, a fim de reduzir o número de variáveis das matrizes de dados e tornar possível a visualização de sua natureza multivariada em poucas dimensões, mostrando uma visão geral dos dados obtidos de forma mais fácil de serem interpretados.³⁴

Para se realizar o cálculo das componentes principais, é necessário organizar os dados em uma matriz, de forma que as linhas sejam compostas pelas amostras e as colunas sejam compostas pelas variáveis, conforme Figura 4.

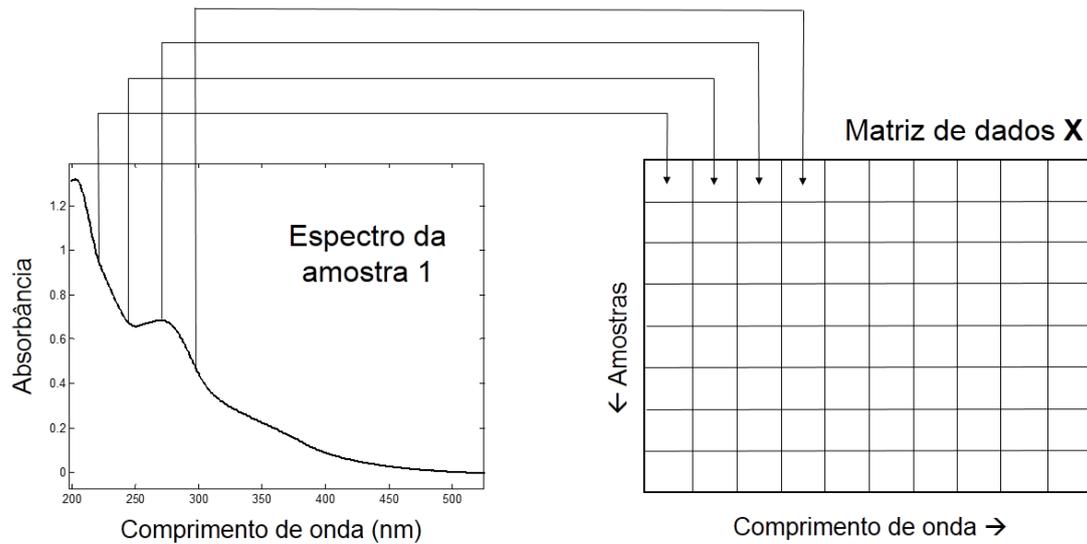


Figura 4. Organização da matriz de dados \mathbf{X} a partir de dados de UV-Vis, na qual as amostras são dispostas nas linhas e as variáveis nas colunas.

A PCA promove a decomposição da matriz de dados \mathbf{X} numa soma de matrizes⁴⁴:

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \dots + \mathbf{M}_r \quad (3)$$

Em que r é o número de componentes principais. As matrizes \mathbf{M} podem ser escritas como um produto de dois vetores, o de escores (\mathbf{t}_r) e o de pesos (“loadings”) (\mathbf{p}_r):

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \mathbf{t}_3\mathbf{p}_3^T + \dots + \mathbf{t}_r\mathbf{p}_r^T = \sum_{i=1}^r \mathbf{t}_i\mathbf{p}_i^T + \mathbf{E} \quad (4)$$

Ou na forma matricial $\mathbf{X} = \mathbf{TP}^T$, em que \mathbf{P}^T é composta por \mathbf{p}^T como linhas e \mathbf{T} é composta por \mathbf{t} como as colunas.

Os escores são as projeções das amostras nos novos eixos de coordenadas definidos pelas componentes principais. Portanto, a matriz \mathbf{T} tem tantas linhas quanto a matriz original \mathbf{X} . A matriz \mathbf{P} é composta pelos pesos, que são os cossenos dos ângulos entre as variáveis originais e os eixos das componentes principais. Nesse caso, a matriz \mathbf{P}^T tem tantas colunas quanto a matriz \mathbf{X} original. Outra propriedade é o fato do número de colunas na matriz \mathbf{T} ser igual ao número de linhas na matriz \mathbf{P}^T . Cada coluna de \mathbf{T} e cada linha de \mathbf{P}^T correspondem a uma componente principal. \mathbf{E} é composta pelos resíduos, a

parte de \mathbf{X} que não é explicada pelo modelo.^{16,20,45,27} A Figura 5 ilustra esta decomposição.

$$\begin{array}{c} m \\ \square \\ n \end{array} \mathbf{X} = \begin{array}{c} 1 \\ \square \\ n \end{array} \mathbf{t}_1 \begin{array}{c} m \\ \square \\ 1 \end{array} \mathbf{P}_1^T + \dots + \begin{array}{c} 1 \\ \square \\ n \end{array} \mathbf{t}_A \begin{array}{c} m \\ \square \\ 1 \end{array} \mathbf{P}_A^T + \begin{array}{c} m \\ \square \\ n \end{array} \mathbf{E}$$

Figura 5. Decomposição da matriz de dados $\mathbf{X}_{n,m}$ por PCA. Adaptada da referência [46]

2.3.1.2 PLS

A regressão por Mínimos Quadrados Parciais (PLS, do inglês Partial Least Squares) foi desenvolvida pelo estatístico Herman Wold na década de 1970, sendo introduzida, mais tarde, na quimiometria por seu filho, Svante Wold.^{47,48}

A regressão por PLS é baseada na PCA, porém além de se trabalhar com a matriz de dados \mathbf{X} , também é utilizado vetor ou matriz de variáveis dependentes (\mathbf{y} ou \mathbf{Y}), que possuem os dados de referência. No caso deste trabalho, foi utilizado um vetor \mathbf{y} que contém os valores de referência das classes de uísque, tendo em vista que foi desenvolvido um modelo PLS específico para cada marca de uísque, o que será detalhado mais adiante.

PLS é um método que relaciona os dados \mathbf{X} e \mathbf{y} por um modelo linear multivariado, com o intuito de obter uma melhor relação linear entre esses dados.^{42,49} Como em PCA, PLS também realiza a decomposição das matrizes e cria um novo sistema de eixos de coordenadas, a partir de componentes principais, agora denominadas de variáveis latentes. Em PLS essas novas variáveis, as variáveis latentes, são obtidas de forma a maximizar a covariância entre \mathbf{X} e \mathbf{y} .^{41,50}

Em PLS, a relação entre \mathbf{X} e \mathbf{y} é obtida por sua decomposição de forma similar a que ocorre com PCA. Contudo, na regressão por PLS ocorre a decomposição da matriz de dados \mathbf{X} e do vetor \mathbf{y} simultaneamente em uma soma de matrizes,^{39,51,47,52,53} como mostram as equações a seguir⁵³:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{i=1}^A \mathbf{t}_i \mathbf{p}_i^T + \mathbf{E} \quad (5)$$

$$\mathbf{y} = \mathbf{Tq}^T + \mathbf{f} = \sum_{i=1}^A \mathbf{t}_i \mathbf{q}_i^T + \mathbf{f} \quad (6)$$

Em que \mathbf{E} (n,m) e \mathbf{f} ($n,1$) contém os resíduos de \mathbf{X} e \mathbf{y} , respectivamente; o vetor \mathbf{t}_i , que compõe as colunas de \mathbf{T} (n, A) é o vetor de scores; \mathbf{P} (A, m) é a matriz de loadings, composta pelos vetores \mathbf{p}_i ; \mathbf{q} ($A, 1$) é o vetor loading de \mathbf{y} , composto de \mathbf{q}_i valores e A é o número de variáveis latentes usadas no modelo. A Figura 6 ilustra esta decomposição.

$$\begin{array}{c}
 \begin{array}{c} m \\ \square \\ n \end{array} \mathbf{X} = \begin{array}{c} 1 \\ \square \\ n \end{array} \mathbf{t}_1 \begin{array}{c} m \\ \square \\ 1 \end{array} \mathbf{P}_1^T + \dots + \begin{array}{c} 1 \\ \square \\ n \end{array} \mathbf{t}_A \begin{array}{c} m \\ \square \\ 1 \end{array} \mathbf{P}_A^T + \begin{array}{c} m \\ \square \\ n \end{array} \mathbf{E} \\
 \\
 \begin{array}{c} 1 \\ \square \\ n \end{array} \mathbf{y} = \begin{array}{c} 1 \\ \square \\ n \end{array} \mathbf{t}_1 \begin{array}{c} 1 \\ \square \\ 1 \end{array} \mathbf{q}_1^T + \dots + \begin{array}{c} 1 \\ \square \\ n \end{array} \mathbf{t}_A \begin{array}{c} 1 \\ \square \\ 1 \end{array} \mathbf{q}_A^T + \begin{array}{c} 1 \\ \square \\ n \end{array} \mathbf{f}
 \end{array}$$

Figura 6. Decomposição das matrizes \mathbf{X} e \mathbf{Y} em variáveis latentes por PLS. Adaptada da referência [46]

As estimativas feitas para as propriedades de interesse ($\hat{\mathbf{y}}$) de um conjunto de amostras é obtida através da multiplicação dos dados pré-processados de \mathbf{X} pelo vetor regressão \mathbf{b} ($m,1$)⁵³:

$$\hat{\mathbf{y}} = \mathbf{T}\mathbf{q}^T = \mathbf{X}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{q}^T = \mathbf{X}\mathbf{b} \quad (7)$$

Em que \mathbf{W} é uma segunda matriz pesos (*weights*) determinada pelo algoritmo utilizado na regressão por PLS. Vale ressaltar que este peso é diferente dos pesos *loadings*.

Há diversos algoritmos para se calcular os parâmetros do modelo PLS. Atualmente um dos mais utilizados é o SIMPLS (do inglês *Straightforward Implementation of a Statistically Inspired Modification of the PLS*), que calcula scores (\mathbf{T}), loadings (\mathbf{P}) e weight loadings (\mathbf{W}).⁵⁴ O algoritmo repete o cálculo para cada parâmetro ($\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A$; $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_A$; $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_A$) até que o número de variáveis latentes (A) seja atingido.

Deve-se destacar que há dois diferentes modelos de regressão por PLS: PLS1 e PLS2. Em PLS1 é construído um modelo de calibração distinto para cada propriedade de interesse e neste caso sempre é utilizado um vetor \mathbf{y} . Já em PLS2, um único modelo de calibração é construído para todas as propriedades de interesse e nesse caso as várias propriedades são agrupadas em uma matriz \mathbf{Y} e modeladas simultaneamente.^{51,52} Para propósitos de interpretação, o PLS2 pode ser vantajoso por construir apenas um modelo, porém, para propósitos de predição, calibrar cada propriedade de interesse separadamente geralmente fornece melhores resultados.⁴¹ Neste trabalho foi utilizado o PLS1.

2.3.1.2.1 Pré-processamentos

Os dados experimentais originais podem não ter uma distribuição adequada para a análise, o que pode prejudicar sua interpretação ou modelagem.³⁴ Medidas em diferentes unidades e variáveis com diferentes variâncias são algumas razões que podem levar a estes problemas.³⁴ Nestes casos, é essencial a aplicação de pré-processamentos nos dados.

Há diversos pré-processamentos disponíveis e a escolha mais adequada de qual utilizar depende dos dados e do objetivo da análise.³⁹ Dois pré-processamentos foram utilizados neste trabalho. São eles:

a) Centrar na Média (*Mean Centring*):

Consiste em calcular a média das colunas dos dados e subtrair essa média de sua respectiva coluna.^{34,39} A equação a seguir ilustra o cálculo:

$$x_{ij}^* = x_{ij} - \bar{x}_j \quad (8)$$

Em que x_{ij}^* são os dados centrados, x_{ij} representa cada valor da matriz \mathbf{X} da amostra i e variável j e \bar{x}_j é o valor da média de todas as amostras com relação à variável j . Ao se centrar os dados na média, o centro dos dados é deslocado para a origem do sistema de coordenadas. Esse pré-processamento destaca as diferenças/variações existentes nos dados e geralmente fornece valores de estimativas melhores.

b) Primeira derivada e alisamento

Derivação e alisamento são métodos comumente usados para tratar espectros por reduzirem o ruído, corrigirem efeitos de variação de linha de base e realçarem as diferenças existentes nos dados.^{47,52,54} O método mais utilizado de derivação e alisamento polinomial é o método de Savitzky-Golay, no qual deve-se selecionar o tamanho da janela do alisamento e o grau do polinômio desejados.^{47,52,54} Um importante aspecto da derivação é que ela não afeta nenhuma relação linear entre os dados. Desta forma, derivar espectros de absorvância não muda a relação linear entre os dados e as concentrações das espécies químicas presentes.⁵⁴

2.3.1.1.2 PLS-DA

o PLS foi desenvolvido originariamente como uma ferramenta para regressão/calibração multivariada, porém também encontrou aplicação em classificação e discriminação.^{55,56} O PLS-DA (*Partial Least Squares for Discriminant Analysis*) é uma ferramenta de classificação capaz de reconhecer diferenças entre grupos conhecidos.⁵⁷

Assim como nas aplicações de regressão, no PLS1-DA, ao invés de se ter uma matriz Y , se tem vetores y .

A principal diferença entre o PLS utilizado para regressão e o PLS-DA está no vetor y . Neste caso, se atribui o valor 1 para as amostras que pertencem à classe de interesse e 0 para as demais classes, como mostra a Figura 7.^{55,57} Neste método, é construído um modelo para cada classe, de forma que o modelo 1 discrimina a classe 1 das demais, o modelo 2 discrimina a classe 2 das demais e assim por diante.

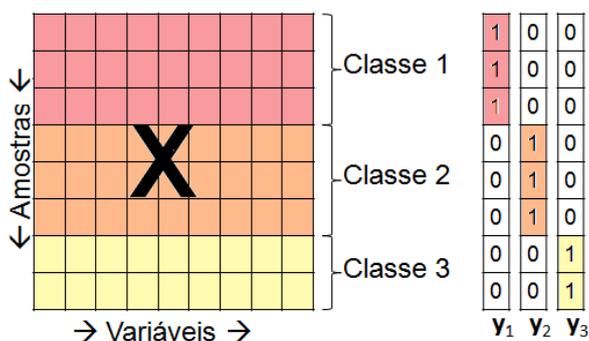


Figura 7. Representação esquemática do PLS-DA. Adaptada da referência [58]

Desta forma, no modelo 1, as amostras que pertencerem à classe 1 apresentarão dispersão no gráfico de estimativa do valor de classe em torno de 1, e as amostras que pertencerem às demais classes apresentarão dispersão em torno de 0. Um exemplo do gráfico de dispersão é mostrado na Figura 8.

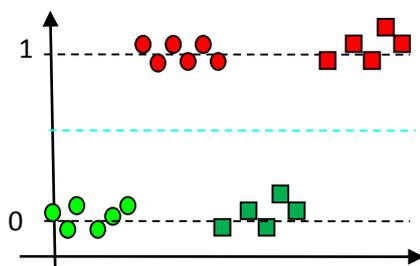


Figura 8. exemplo de gráfico de dispersão de classe

A linha tracejada em azul (----) na Figura 8 representa o limite de discriminação entre as classes. Amostras que apresentarem valor de classe acima do valor do limite de discriminação é classificada como pertencente à classe 1, por outro lado, amostras que apresentarem valor de classe menor que do limite de discriminação é classificada como pertencente à classe 0.

O limite de discriminação possui um valor entre 0 e 1. Caso as dispersões das classes 0 e 1 tiverem variâncias similares e não houver erros sistemáticos, esse limite terá valor próximo de 0,5.⁵⁴ É calculado com base no teorema de Bayes, assumindo que os valores estimados de y seguem uma distribuição similar ao que será observado em amostras futuras.⁵⁴ Seu objetivo é identificar o limite entre as classes para evitar erros nas classificações.⁵⁴

2.3.1.1.3 Determinação do número de variáveis latentes do modelo por validação cruzada

De acordo com Ferreira et. Al.,³⁴ a validação cruzada consiste em “uma metodologia utilizada para a escolha do número de variáveis latentes baseada na avaliação da magnitude dos erros de previsão de um dado modelo”.

O princípio da validação cruzada consiste em dividir os dados em S_i subconjuntos e construir um modelo contendo todos os subconjuntos exceto um (S_{i-1}). Este modelo será então usado para prever a propriedade de interesse do i -ésimo subconjunto. Este processo é repetido até que todos os subconjuntos

sejam analisados. Desta forma, uma comparação das predições obtidas com os valores conhecidos das propriedades de interesse fornece uma avaliação dos erros associados ao modelo em amostras que a rigor não foram utilizadas no seu desenvolvimento.⁴⁰ O método utilizado neste trabalho foi o método por blocos contíguos, em que as amostras do conjunto de calibração são subdivididas em blocos (subconjuntos) construídos de forma que todas as replicatas de uma mesma amostra estejam presentes no mesmo bloco.

Uma estimativa do erro médio de predição é feita através da raiz quadrada do erro médio quadrático da validação cruzada (RMSECV, do inglês *root mean square error of cross validation*) que é obtido de todas as repetições da validação cruzada.⁵⁴ A equação 9 define o RMSECV⁴¹:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_{CV,i} - y_i)^2}{n}} \quad (9)$$

Em que $\hat{y}_{CV,i}$ é a estimativa para y_i através do método da validação cruzada, ou seja, da predição através do modelo de calibração sem a amostra i .

Um importante problema que ocorre com todos os métodos de compressão de dados é a seleção da quantidade certa de variáveis ou componentes. Se muitas componentes são usadas, então um excesso de variáveis é inserido e a modelagem se torna sobreajustada, gerando predições ruins. Por outro lado, se poucas variáveis são usadas, ocorre o chamado subajuste, que significa que o modelo não possui variáveis suficientes para fazer uma adequada modelagem dos dados.⁴¹ A Figura 9a ilustra este fenômeno. A escolha da quantidade de variáveis latentes deve ser feita cuidadosamente para que não ocorra problemas de ajustes dos dados. Uma forma de se escolher adequadamente o número de variáveis latentes é escolhendo-se a quantidade que gera o menor valor de RMSECV ou a partir da qual se observa uma estabilização desse parâmetro, como mostra a Figura 9b. Neste caso, o número de variáveis latentes seria 11.

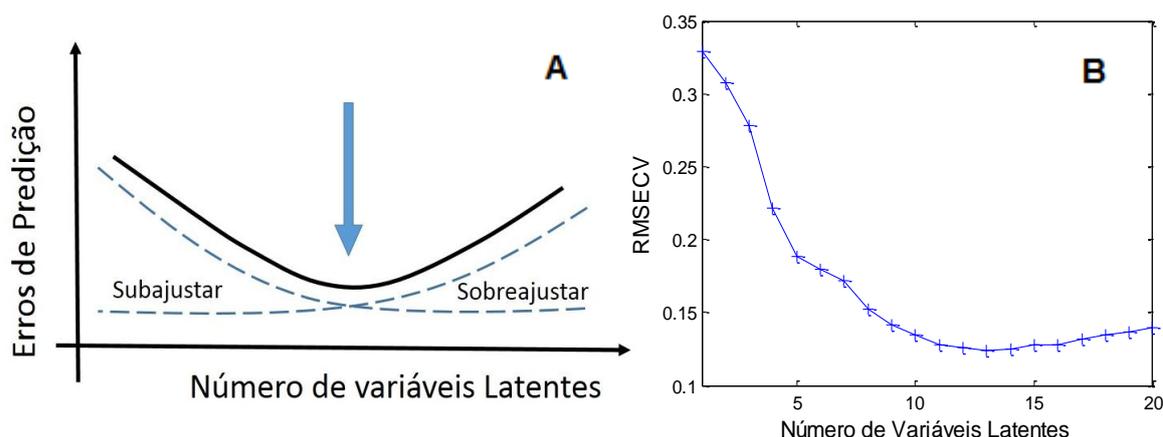


Figura 9. a) importância da escolha correta do número de VL; b) gráfico do valor de RMSECV em função da quantidade de variáveis latentes.

2.3.1.1.4 Detecção de amostras anômalas

Amostras anômalas, do inglês *outliers* (como serão chamadas de agora em diante neste trabalho), são amostras que por alguma razão apresentaram resultados experimentais diferentes das demais amostras do conjunto de amostras de calibração.⁴¹ A presença dessas amostras no conjunto de calibração pode prejudicar o desenvolvimento do modelo devido ao fato de alterarem os coeficientes da regressão.⁵³ Portanto, um passo muito importante no desenvolvimento dos modelos é a identificação e exclusão de *outliers*.

O método utilizado neste trabalho foi aplicado por Borin e Poppi⁴⁴ e é baseado em dados que apresentam altos valores de influência, definida pela estatística de Hotelling (T^2) e altos valores de resíduos não modelados em \mathbf{X} , definidos pelo parâmetro estatístico Q e resíduos não modelados em \mathbf{Y} , levando em consideração um intervalo de confiança de 99,9%. A seguir são apresentadas as equações principais do método de identificação de outliers que se baseia na abordagem de cartas de controle multivariadas. Maiores detalhes podem ser obtidos na referência [44].

Este método é executado na etapa de decomposição dos dados, na qual é definido o número de variáveis latentes, que por sua vez são usadas para a construção da carta de T^2 e a carta Q .

A carta de T^2 foi originada do trabalho de Hotelling e é estimada dos escores em um espaço formado pelas variáveis latentes mais significativas. T^2 é a medida da variação em cada amostra dentro do modelo PLS e é definida como:

$$\mathbf{T}_i^2 = \mathbf{t}_i \boldsymbol{\lambda}^{-1} \mathbf{t}_i^T = \mathbf{x}_i \mathbf{P}_k \boldsymbol{\lambda}^{-1} \mathbf{P}_k^T \mathbf{x}_i^T \quad (10)$$

em que:

- \mathbf{t}_i refere-se à i-ésima linha de \mathbf{T}_k (a matriz de k vetores scores do modelo PLS);
- $\boldsymbol{\lambda}$ é a matriz diagonal contendo os autovalores associados aos autovetores incluídos nos modelos;
- \mathbf{x}_i é a i-ésima linha de \mathbf{X} ;
- \mathbf{P}_k é a matriz de k vetores *loadings* retidos no modelo PCA.

Há apenas um limite de confiança para T^2 que é calculado pela distribuição F e expresso como:

$$T_{lim}^2 = \frac{(m-1)k}{m-K} F_{k,m-k,\alpha} \quad (11)$$

em que:

- m é o número de amostras utilizadas no desenvolvimento do modelo;
- k é o número de variáveis latentes escolhido;
- α corresponde ao limite de confiança.

A carta Q corresponde à falta de ajuste dos dados espectrais pelo modelo PLS e é definida como:

$$Q_i = \mathbf{e}_i \mathbf{e}_i^T = \mathbf{x}_i (\mathbf{I} - \mathbf{P}_k \mathbf{P}_k^T) \mathbf{x}_i^T \quad (12)$$

Em que \mathbf{e}_i é a i-ésima linha de \mathbf{E} e \mathbf{I} é a matriz identidade.

O valor de Q (que é um escalar) mede a variação que não foi modelada pelo modelo PLS. Se as amostras de calibração são todas de uma mesma população e sem presença de *outliers*, os valores de Q devem ser pequenos. O limite de confiança é definido como:

$$Q_{\alpha \text{ lim}} = \Theta_1 \left[\frac{c_{\alpha} \sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right]^{\frac{1}{h_0}} \quad (13)$$

$$\Theta_i = \sum_{j=k+1}^A \lambda_j^i \text{ para } i = 1, 2, 3 \quad (14)$$

$$h_0 = \frac{2\Theta_1 \Theta_3}{3\Theta_2^2} \quad (15)$$

em que c_α na equação 13 é o desvio padrão normal correspondente à porcentagem superior $(1-\alpha)$. Na equação 14, k é o número de variáveis latentes usadas no modelo, A é o número total de variáveis latentes e λ_j é o autovalor associado à j -ésima variável latente.

Com base nessas duas cartas, diferentes tipos de outliers podem ser identificados. A Figura 10 mostra o gráfico de T^2 e Q . As linhas tracejadas representam os limites de 99,9% de confiança tanto para T^2 quanto para Q . Com o objetivo de excluir apenas amostras estritamente necessárias, foram consideradas como *outliers* apenas as amostras que se encontravam no quadrante correspondente à intersecção entre estes dois parâmetros, que são amostras que apresentam simultaneamente elevados influência e resíduos de modelagem.

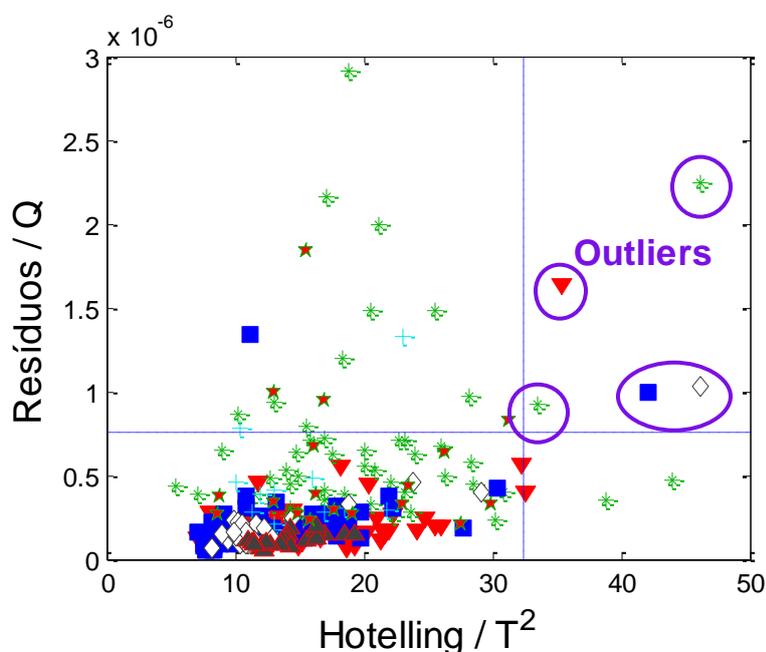


Figura 10. Identificação de *outliers* através do gráfico de influência e resíduos

2.3.1.1.5 Figuras de mérito dos modelos de discriminação

Sempre que um procedimento analítico é proposto ou desenvolvido, há a necessidade de se averiguar se o método atende ao seu propósito, nas condições em que será aplicado, e às normas estabelecidas por agências reguladoras e órgãos de fiscalização nacionais. Esse processo de averiguação é conhecido como validação.⁵⁹

Antes da aplicação do modelo construído, este deve ser validado com o objetivo de testar sua capacidade preditiva.³⁴ A validação consiste em testar o modelo a fim de estimar as propriedades de interesse de amostras que não foram usadas em sua construção, cujos valores sejam conhecidos, para estabelecer se ele irá de fato prever o comportamento do analito de interesse através da comparação entre o valor estimado e o conhecido.^{34,40}

A forma de validação dos modelos utilizada neste trabalho foi a validação externa, que utiliza um conjunto de amostras diferente do utilizado na etapa de calibração. Através dos resultados obtidos para essas amostras externas ou de validação são calculados parâmetros que demonstram a eficiência do método proposto. Atualmente esses parâmetros são conhecidos como figuras de mérito. A forma como são determinadas geralmente é estabelecida por órgãos de fiscalização e encontra-se descrita em normas específicas, guias de validação e trabalhos científicos.⁵⁹

A validação dos modelos de discriminação empregados nesse trabalho vem sendo abordada em diversos estudos recentes.^{59,60} Devido a isso, algumas figuras de mérito utilizadas neste trabalho foram baseadas em trabalhos científicos ao invés de normas técnicas.

Para atestar a eficiência dos modelos desenvolvidos foram determinadas as seguintes figuras de mérito:

a) Taxa de falso positivo – TFP

Para compreender esta e as próximas figuras de mérito, faz-se necessário a construção da tabela 2, que detalha cada fator das equações.

TFP é definida como a porcentagem baseada na razão entre o número de erros falso positivo (FP) e a soma entre os erros falso positivo e verdadeiro negativo (VN), de acordo com a equação:

$$TFP = \frac{FP}{FP+VN} 100 \quad (16)$$

b) Taxa de falso negativo -TFN

A taxa de falso negativo (TFN) é definida pela razão entre o número de erros falso negativo e a soma entre os erros falso negativo e verdadeiro positivo, de acordo com a equação:

$$TFN = \frac{FN}{FN+VP} 100 \quad (17)$$

Tabela 2: Legendas para os parâmetros utilizados no cálculo das figuras de mérito

Fatores	Abreviação	Significado
Falso Positivo	FP	Amostras pertencentes às demais classes (valor de classe 0) classificadas como amostras da classe 1
Falso Negativo	FN	Amostras da classe 1 classificadas como pertencentes à classe 0
Verdadeiro Positivo	VP	Amostras da classe 1 classificadas corretamente como 1
Verdadeiro Negativo	VN	Amostras da classe 0 classificadas corretamente como 0

c) Taxa de seletividade – TST

A taxa de seletividade (TST) é definida pela razão entre o número de verdadeiros positivo e a soma dos erros falso negativo e verdadeiro positivo, de acordo com a equação:

$$TST = \frac{VP}{FN+VP} 100 \quad (18)$$

d) Taxa de especificidade – TSP

A taxa de especificidade (TSP) é definida pela razão entre o número de verdadeiros positivo e a soma entre o número de verdadeiro negativo e os erros falso positivo, de acordo com a equação:

$$TSP = \frac{VN}{VN+FP} 100 \quad (19)$$

e) Taxa de eficiência – TEF

A taxa de eficiência (TEF) é definida pela equação:

$$TEF = 100 - (TFP + TFN) \quad (20)$$

É importante destacar que, para a análise de amostras de uísque falsas ou de amostras autênticas de marcas não modeladas, o resultado esperado do método para essas amostras é indicar que as mesmas apresentam uma condição distinta daquelas utilizadas para a calibração do modelo, isto é, a amostra deve ser identificada como um *outlier* pelos testes de T^2 e Q. Entretanto, decidiu-se usar a terminologia *outlier* apenas para amostras anômalas excluídas do conjunto de calibração. Como o fato de amostras falsas e não modeladas apresentarem as mesmas características de *outliers* ser um resultado esperado e elas não serem excluídas, estas amostras serão classificadas como “não autênticas ou não pertencentes às classes modeladas”.

Uma consequência deste tipo de classificação é que para essas amostras não se tem formalmente verdadeiros positivos, verdadeiros negativos e falsos negativos. Logo, para a análise dessas amostras apenas os valores de erros falsos positivos podem ser calculados, de forma que a eficiência do método foi medida em termos da taxa de classificações corretas (TCC), calculado como:

$$TCC = \frac{(N_T - FP - FN)}{N_T} 100 \quad (21)$$

em que N_T é o número total de amostras do conjunto a ser analisado, seja autênticas, falsas ou pertencentes a outras classes. Deve-se notar que o resultado da equação 21 expressa a eficiência do método em termos do número de classificações corretas e não utilizando os erros de falso positivo e negativo.

CAPÍTULO 3
Discriminação de marcas de uísque por espectroscopia UV/Vis

3. DISCRIMINAÇÃO DE MARCAS DE UÍSQUE POR ESPECTROSCOPIA UV-VIS

A espectroscopia de UV-Vis é uma técnica analítica bastante simples e amplamente utilizada em diversas áreas da ciência. Esta técnica foi selecionada para este trabalho devido ao fato de ser um equipamento encontrado na grande maioria dos laboratórios e por sua simplicidade de operação, tornando possível que, caso haja necessidade, a discriminação de marcas de uísque seja feita em laboratórios mais simples, assim como pela ausência da necessidade de treinamento de pessoal para operar o equipamento.

3.1 MATERIAIS E MÉTODOS

A parte experimental deste trabalho foi realizada no laboratório de Química do Setor de Perícias de Laboratório do Instituto Nacional de Criminalística da Polícia Federal (SEPLAB/DPER/INC/DITEC/DPF).

3.1.1 Materiais

Sete marcas de uísque foram selecionadas para a execução do trabalho. O critério de escolha das marcas foi o histórico de apreensões pela Polícia Federal. Foi realizado um levantamento das bebidas apreendidas que constavam no depósito do Instituto Nacional de Criminalística da Polícia Federal e as marcas que apresentavam maior quantidade de unidades foram escolhidas, sob a conclusão de que se tratavam dos maiores alvos de falsificação.

3.1.1.1 Amostras autênticas

Um total de 164 amostras, de diversas marcas, foram obtidas por diversas fontes, através de doação e compras no mercado local. São elas:

- 1) Doação do Laboratório Nacional Agropecuário de Minas Gerais (LANAGRO - MG): 14 unidades de diversas marcas, amostradas em frascos âmbar de 100 mL.

- 2) Doação da Diageo^a: 36 unidades, sendo 9 de Johnnie Walker Red Label, 9 unidades de Johnnie Walker Black Label, 9 unidades de Grand Old Parr e 9 unidades de White Horse. Para cada marca, foram recebidas 3 amostras de 3 lotes diferentes, totalizando 9 amostras para cada marca.
- 3) Doação da Polícia Federal: A Polícia Federal doou 22 miniaturas (50 mL) de Johnnie Walker Red Label, 22 miniaturas (50 mL) de Johnnie Walker Black Label e 23 unidades (500 mL) de White Horse.
- 4) Demais amostras: foram adquiridas com recursos próprios dos membros envolvidos no projeto ou por meio de doações de terceiros, totalizando 11 amostras de Chivas Regal (50 mL), 18 amostras de Ballantine's finest (50 mL), 1 amostra de Black Label (100 mL), 1 amostra de Grant's (200 mL), 1 amostra de Glenfiddich (50 mL) e 15 amostras de Natu Nobilis (1 L).

Tabela 3: Relação das amostras autênticas utilizadas no trabalho

Marca	Tempo de envelhecimento	Unidades
Johnnie Walker – Red Label	8 anos	32
Johnnie Walker – Black Label	12 anos	33
Johnnie Walker – Gold Label	18 anos	1
Johnnie Walker – Green Label	15 anos	1
Logan	12 anos	1
Grant's Royal	12 anos	1
Grant's	8 anos	1
Grand Old Parr	12 anos	10
Dimple	12 anos	1
Dimple	15 anos	1
House of Lords	Não especificado	1
Jonnie Walker - Swing	15 anos	1
Glenfiddich	12 anos	2
Ballantine's Finest	8 anos	19
Chivas Regal	12 anos	11

^a A Diageo é uma empresa do ramo de bebidas alcoólicas, que conta com cervejas, destilados e vinhos. Possui representação em 80 países.

Marca	Tempo de envelhecimento	Unidades
Buchanan's	Não especificado	1
White Horse	8 anos	32
Natu Nobilis	3 anos	15
Total		164

3.1.1.2 Amostras falsas

Um total de 73 amostras provenientes de apreensões foram cedidas pela Polícia Federal, atestada como falsas pelos seus métodos de análises de rotina.

Tabela 4: Relação das amostras falsas utilizadas no trabalho

Amostra	Unidades
Johnnie Walker Red Label	23
Johnnie Walker Black Label	15
Ballantine's	8
White Horse	9
Natu Nobilis	9
Chivas Regal	7
Grand Old Parr	2
Total	73

3.1.1.3 Amostras da apreensão

Um total de 16 amostras foram cedidas pelo perito responsável pelo laudo pericial das bebidas apreendidas. Dentre elas, havia amostras de referência, com o intuito de comparar seu comportamento com das amostras apreendidas.

Tabela 5. Relação das amostras provenientes da apreensão

Marcas	Unidades - apreensão	Unidades - referência
Johnnie Walker	4 autênticas	1 autêntica
Red Label	2 falsas	1 falsa
Johnnie Walker	2 autênticas	1 autêntica
Black Label		
White Horse	2 autênticas	-
Grand Old Parr	2 autênticas	1 autêntica
Total	12 amostras	4 amostras

3.1.1.4 Reagentes

Foram utilizados etanol absoluto da marca Dinâmica com pureza de 99,8% e água ultrapura obtida do sistema de purificação Millipore Direct-Q 5.

3.1.2 Obtenção dos espectros

As análises foram realizadas em triplicata em espectrofotômetro UV-Vis, da marca Agilent Technologies, modelo 8453, utilizando cubeta de fluxo de 1 mm de caminho óptico e volume de 62 μ L, da marca Hellma, confeccionada em quartzo, acoplada a uma bomba peristáltica. Os dados foram adquiridos e exportados utilizando o software UV-VisibleChemstation, no qual a região espectral foi selecionada de 190 a 1100 nm.

A cubeta de fluxo foi utilizada com o intuito de consumir pequeno volume de amostra e de tornar o processo mais prático e rápido, uma vez que a utilização de bomba peristáltica para preenchimento da cubeta além de ser mais rápido que manualmente ainda minimiza a possibilidade de acidentes. Vale ressaltar que a análise não foi feita em fluxo. Quando a cubeta era preenchida com a amostra o fluxo era interrompido, feita a verificação da presença ou não de bolhas e só então o espectro era obtido.

A seleção das amostras para análise em cada dia de trabalho foi feita de forma aleatória para evitar a seleção preferencial de amostras para a calibração por outros critérios que poderiam introduzir tendência aos resultados.⁶¹ Desta forma, em cada dia de trabalho foram feitas análises de amostras de todas as marcas de uísque, para evitar que qualquer fator indesejável, alheio ao conhecimento do analista, contribuísse para a discriminação de uma marca de uísque caso todas as amostras fossem analisadas sob estas condições.

3.2 ANÁLISE DE DADOS

A análise de dados foi feita através do software MATLAB versão 7.12(R2011b) utilizando o pacote PLS Toolbox versão 6.5 para construção dos modelos de PLS-DA.

A construção dos modelos de calibração foi feita utilizando Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA). O critério de seleção de amostras para o conjunto de calibração foi destinar cerca de dois terços das

amostras de marcas com mais de 30 amostras para este conjunto. Marcas com menos de 20 amostras tiveram cerca de metade das amostras selecionadas para a calibração, conforme mostra a tabela 6. Marcas com menos de 10 amostras foram utilizadas apenas na etapa de validação. Sete meses depois que os espectros de todas as amostras foram obtidos, outras amostras foram selecionadas aleatoriamente e novos espectros foram obtidos, em triplicata, e foram inseridos no conjunto de calibração, de forma que este foi composto por 365 espectros. O intuito de se fazer novas análises foi inserir na calibração quaisquer variações que o equipamento tenha sofrido ao longo destes meses.

Tabela 6: quantidades de amostras utilizadas nos conjuntos de calibração e validação

Marcas	Nº de amostras	Amostras de calibração	Amostras de validação
Red Label	32	20	12
Black Label	33	20	13
White Horse	32	20	12
Chivas Regal	11	6	5
Ballantines Finest	19	10	9
Natu Nobilis	15	8	7
Old Parr	10	6	4

Para a construção dos modelos PLS-DA, valores de classe de referência foram atribuídos às classes (marcas) do conjunto de calibração, sendo o valor 1 às classes que deseja discriminar e o valor 0 às demais classes.

A etapa seguinte consistiu na regressão da matriz de dados X de calibração e do vetor y de cada classe, gerando-se, assim, um modelo para cada marca de uísque.

Após a regressão, foi feita uma rotina de otimização do modelo de calibração, avaliando-se os pré-processamentos mais adequados assim como o número de variáveis latentes. Para a construção dos modelos, os pré-processamentos foram selecionados com base na melhor separação entre as classes 0 e 1 no gráfico das estimativas dos valores de classes. Dentre os pré-processamentos testados, os que se mostraram mais adequados foram centrar na média e 1ª derivada, esta última baseada no algoritmo de Savitsky-Golay,

utilizando filtro com janela de 9 pontos e polinômio de segunda ordem. O número de variáveis latentes (VL) selecionados para a construção dos modelos foi determinado com base nos resultados de RMSECV das amostras de calibração, obtido por validação cruzada por blocos contínuos.

Outra etapa da otimização foi a realização da rotina para detecção e exclusão de *outliers*, que foram identificados de duas formas: espectros que apresentam elevados valores de resíduos e influência, os quais se localizavam no quadrante do gráfico de influência e resíduos correspondente à intersecção destes dois parâmetros, assim como espectros que apresentavam elevado erro de estimativa do valor de classe. Foi considerado erro de classificação de uma amostra apenas se pelo menos 2 replicatas da mesma amostra fossem classificadas incorretamente. Ou seja, se duas ou as três replicatas de uma mesma amostra apresentasse valor de classe compatível com de outra classe, esta amostra seria classificada incorretamente. Por outro lado, se apenas uma replicata de uma amostra apresentasse este comportamento, não caracterizaria um erro de classificação.

A rotina de detecção e exclusão de *outliers* foi feita em duas etapas. Um primeiro modelo foi construído e os *outliers* foram excluídos do conjunto de calibração. Em seguida o modelo foi reconstruído com as amostras restantes e foi feita uma nova identificação de *outliers*. Nos casos em que houveram, foram excluídos novamente e o modelo foi reconstruído. Esse procedimento de exclusão foi feito apenas duas vezes, ainda que no terceiro modelo houvesse novos *outliers*. A Figura 11 mostra os passos que foram seguidos para a construção dos modelos de calibração.

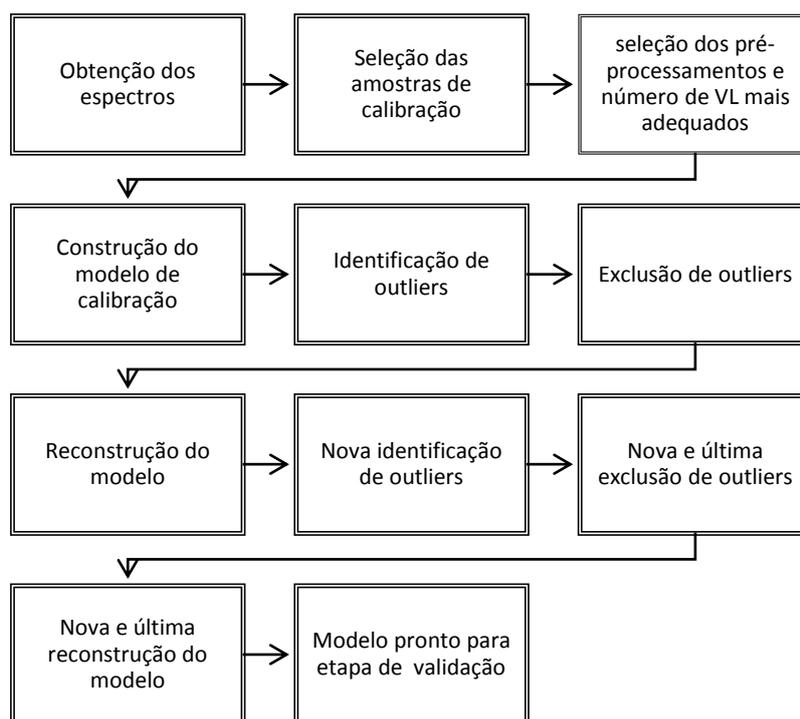


Figura 11. Organograma da rotina de exclusão de *outliers* na construção dos modelos de calibração.

Após a construção dos modelos de discriminação, sua eficiência de previsão foi testada utilizando-se o conjunto de dados de validação. Nesta etapa não há mais *outliers*, as amostras que apresentam valores de resíduo e influência acima do limite de 99,9% são denominadas amostras não autênticas ou não pertencentes às classes modeladas e, em uma situação real de rotina, são consideradas amostras que necessitariam ser analisadas novamente para confirmar sua condição ou analisadas por uma outra técnica independente. Após avaliada a eficiência de previsão com amostras autênticas, os modelos também foram utilizados na previsão de amostras reconhecidamente falsas e de amostras provenientes de uma apreensão da Polícia Federal. A Figura 12 apresenta os esquemas de análises adotados na etapa de previsão.

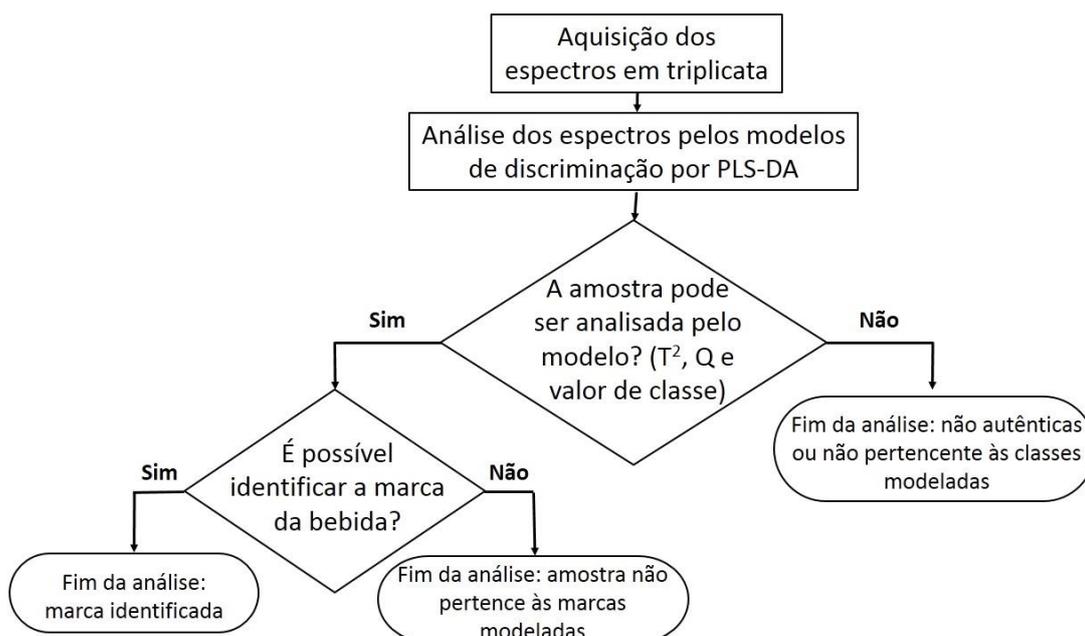


Figura 12. Fluxograma para análise de amostras desconhecidas nos modelos de discriminação desenvolvidos.

3.3 RESULTADOS E DISCUSSÃO

Embora a região espectral selecionada no espectrofotômetro tenha sido de 190 nm a 1100 nm, para a análise de dados foi selecionada a região entre 199 nm a 589 nm, pois é a região com a maior absorção. De 590 nm adiante a absorção é praticamente zero, restando apenas ruído instrumental. A Figura 13 apresenta os espectros das amostras utilizadas na etapa de calibração.

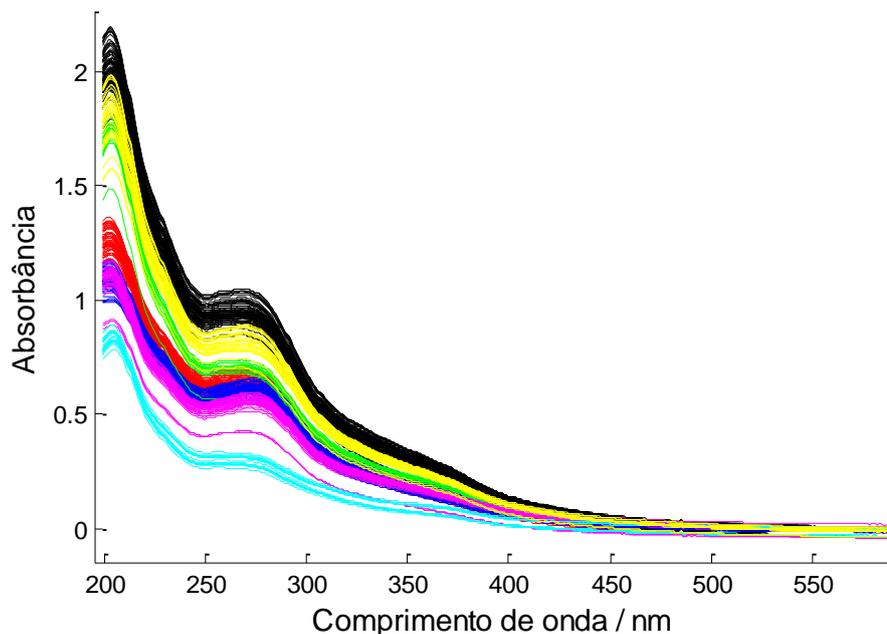


Figura 13. Espectros UV-Vis das amostras autênticas de uísque 12 e 8 anos ou menos, utilizadas no conjunto de calibração. Black Label, Old Parr, Chivas, Red Label, White Horse, Ballantine's, Natu Nobilis.

Apenas por uma análise visual dos espectros da Figura 13 é possível perceber que há diferenças na intensidade de absorbância entre as marcas, sendo o Black Label (identificada em preto) a que apresenta a maior intensidade e o Natu Nobilis (identificada em ciano) a menor. A absorbância na região UV-Vis pode ser explicada pela presença de congêneres resultantes do processo de maturação em barris de carvalho, presença de compostos fenólicos voláteis provenientes da queima da turfa na etapa de secagem da cevada maltada, assim como presença de caramelo para ajuste da cor da bebida.^{27,28} Portanto, através da Figura 13 conclui-se que estes fatores levam a uma relação entre o tempo de envelhecimento e o teor de congêneres. Boscolo et al¹⁶ associaram a absorção UV-vis na região próxima a 210 nm e 282 nm com a presença de furfural e hidroximetilfurfural em bebidas envelhecidas em barris de carvalho. De acordo com eles, o furfural é extraído do barril de carvalho e o hidroximetilfurfural é o principal composto do caramelo, sendo este último responsável por cerca de 60% da absorbância em 282 nm. Neste trabalho o grupo afirmou que cachaça não envelhecida e livre de caramelo não apresentou absorção significativa no

UV-Vis, o que corrobora com estas informações. Outros compostos que absorvem em 282 nm são os compostos fenólicos e o ácido elágico.⁶²

É possível também perceber na Figura 13 que algumas replicatas ou amostras de uma mesma marca apresentam certa variação com relação às outras da mesma marca, sobrepondo-se aos espectros de outra classe. Esta discrepância pode ser atribuída a alguma variação entre lotes ou variações instrumentais. É possível notar uma diferença significativa de absorção em aproximadamente 205 nm, notando-se uma clara separação entre os espectros das marcas com 12 anos de envelhecimento (Black Label, Old Parr e Chivas Regal) e as de 8 anos ou menos (Red Label, Ballantine's Finest, White Horse e Natu Nobilis), sendo as de 12 anos com maiores valores de absorbância. A região do espectro entre as bandas com máximo de absorção em aproximadamente 205 nm e 275 nm apresenta a menor sobreposição entre espectros de diferentes classes. Mesmo com muitas informações sendo extraídas destes espectros, ainda assim não é possível determinar a marca das bebidas com segurança apenas pela análise visual do espectro, o que justifica uma análise multivariada dos dados.

Os modelos de discriminação foram construídos através da rotina descrita na Figura 11. Os parâmetros estabelecidos para cada modelo são descritos na tabela 7. É importante ressaltar que o número de *outliers* descritos na tabela é referente ao número de espectros que apresentaram valores de T^2 e Q maiores que os limites de 99,9 % de confiança, independente do fato de ser apenas uma replicata de cada amostra, retirados apenas do conjunto de calibração.

Na tabela é possível notar que os valores de RMSEC e RMSECV de cada marca são próximos, o que indica que não ocorreu sobreajuste do modelo. Outra informação importante fornecida pela tabela 7 é que os valores dos limites de discriminação foram próximos de 0,5. Como este valor depende da dispersão das amostras em torno do seu valor de classe esperado, 0 e 1, o limite de discriminação ser próximo de 0,5 indica que as dispersões destas amostras estão próximas do esperado, ou seja, indica que não houve erro sistemático. Entretanto, observa-se um comportamento diferente do modelo de Old Parr, pois apresenta o valor do limite de discriminação mais distante de 0,5, o que indica que houve erro sistemático, podendo ser confirmado através do gráfico de dispersão de classe (Figura 21) que será detalhado na seção a seguir.

Tabela 7: Parâmetros dos modelos de calibração

	Red Label	Black Label	White Horse	Old Parr	Chivas	Ball.	Natu Nobilis
PP	Centrado na média + 1ª derivada						
VL	16	15	14	19	16	11	16
Outliers	4	2	3	6	2	0	1
RMSEC	0,1688	0,1108	0,1274	0,1253	0,0931	0,1078	0,0247
RMSECV	0,2155	0,1429	0,1630	0,1894	0,1263	0,1276	0,0310
LD	0,4567	0,4682	0,4649	0,3848	0,5348	0,5393	0,4888

*número de espectros excluídos; PP = pré-processamentos; VL = número de variáveis latentes; LD = limite de discriminação; Ball = Ballantine's;

3.3.1 Validação do método pela análise de amostras autênticas

Sempre que um procedimento analítico é proposto ou desenvolvido, há a necessidade de se averiguar se o método atende ao seu propósito, nas condições em que será aplicado. Para tal averiguação, foi feita a predição de amostras autênticas a partir dos modelos de calibração criados.

Nos gráficos de dispersão mostrados a seguir, as amostras 1 a 365 são as amostras de calibração e as amostras 366 a 585 são as amostras de validação, sendo que destas, as amostras 550 a 585 são de marcas que não estão presentes no conjunto de calibração, ou seja, não foram modeladas.

O cálculo das Figuras de mérito foi feito a partir das fórmulas que constam na seção 2.3.1.1.5 e os erros falsos positivo e negativo, assim como os verdadeiros positivo e negativo, foram contabilizados de acordo com os critérios da tabela 2. A taxa de seletividade indica a taxa de classificação correta de amostras como classe 1, por outro lado, a taxa de especificidade indica a taxa de classificação correta de amostras como classe 0. Tanto a taxa de classificação correta quanto a taxa de eficiência indicam a eficiência do modelo em fazer classificações corretas das classes 1 e 0, porém calculadas de formas diferentes.

As amostras que apresentaram elevados valores de influência, resíduo, ou tiveram valores estimados de classe extremos (muito elevados ou muito baixos) foram consideradas como amostras não pertencentes às classes modeladas.

Este comportamento era esperado na análise de amostras pertencentes a marcas que não foram incluídas nas amostras de calibração, pois essas marcas devem conter composição diferente das marcas presentes neste conjunto. Quando essas amostras obtiveram tal classificação, foram consideradas como verdadeiros negativos.

Como citado anteriormente, para uma amostra ser classificada incorretamente, pelo menos duas replicatas desta amostra devem ser classificadas desta forma. Amostras autênticas foram consideradas como erro de classificação quando apresentaram valores de classe, T^2 e Q dentro dos limites aceitáveis e eram amostras de marcas não incluídas nas amostras de calibração que foram classificadas como pertencentes a classes 1 (classe discriminada).

3.3.1.1 Validação com amostras autênticas de Red Label

A Figura 14 apresenta o gráfico da estatística de Hotelling (T^2) e dos resíduos de modelagem (Q) para o modelo de discriminação do Red Label. Na Figura 14A pode-se observar que as amostras pertencentes a marcas diferentes das inseridas no modelo de discriminação apresentam influência e resíduo muito acima do limite de 99,9% estabelecido. Conforme relatado anteriormente, esse comportamento era esperado e representa um bom resultado, pois demonstra que o método não identificou amostras autênticas de outras marcas como pertencente a uma das marcas inseridas no modelo de calibração. A Tabela 8 mostra que todas as amostras autênticas pertencentes a marcas diferentes das incluídas na calibração foram corretamente identificadas como verdadeiros negativos.

A Figura 14B mostra uma ampliação do gráfico apresentado na Figura 14A. Nele é possível notar que duas amostras apresentaram uma replicata com altos valores de influência e resíduo e ultrapassaram os limites de 99,9% de confiança. Entretanto, como se trata apenas de uma replicata de cada amostra, não caracteriza erro de predição.

3. DISCRIMINAÇÃO DE MARCAS DE UÍSQUE POR ESPECTROSCOPIA UV-VIS

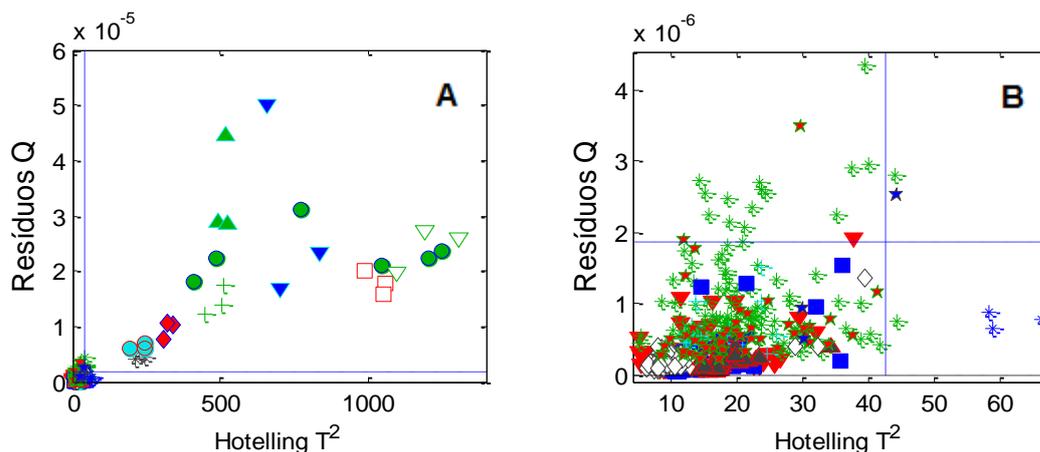


Figura 14. A) Gráfico de influência (T^2) e resíduos (Q) do modelo de discriminação entre amostras de Red Label (\blacktriangledown) e as demais marcas. **B)** Ampliação da região dentro do limite de 99,9%.

Tabela 8. Parâmetros e figuras de mérito da predição de amostras autênticas

	Red Label	Black Label	White Horse	Chivas Regal	Ball	Natu Nobilis	Old Parr	Outras marcas
N	12	13	12	5	9	7	4	11
NP	10	10	11	10	10	0	0	0
FP	0	0	0	0	1	0	1	0
FN	0	0	0	0	0	0	0	-
VP	12	13	12	5	9	7	4	-
VN	62	62	62	69	64	67	69	11
TFP*	0	0	0	0	1,5	0	1,4	-
TFN*	0	0	0	0	0	0	0	-
TST*	100	100	100	100	100	100	100	-
TSP*	100	100	100	100	98,5	100	98,6	-
TEF*	100	100	100	100	98,5	100	98,6	-
TCC	100	100	100	100	98,6	100	98,6	100

*Expresso em percentagem; N = número de amostras; NP = número de amostras que foram classificadas como não pertencentes às classes modeladas; FP = número de erros falsos positivos; FN = número de erros falsos negativos; VP = número de verdadeiros positivos; VN = número de verdadeiros negativos; TFP = Taxa de Falso positivo; TFN = Taxa de Falso Negativo; TST = taxa de seletividade; TSP = taxa de especificidade; TEF = taxa de eficiência; Ball = Ballantine's Finest;

A Figura 15 apresenta o gráfico de dispersão dos valores de classe estimados do modelo de discriminação entre amostras autênticas de Red Label e as demais marcas. Nesse modelo, o valor de classe 1 é esperado para as amostras de Red Label e 0 para as demais marcas. Nesse gráfico pode-se observar que todas as amostras de Red Label foram corretamente classificadas e que apenas uma replicata de Black Label apresentou erro de classificação, o que não foi suficiente para caracterizar um erro para essa amostra. Após esta análise, foi possível concluir que este modelo apresentou ótima separação entre as classes, visto que não houve erro de classificação e que apresentou taxas de seletividade, especificidade e eficiência de 100%. Isto significa que o modelo apresenta taxa de 100% de classificação correta de amostras tanto da classe 0 quanto da classe 1, conforme resultados apresentados na Tabela 8.

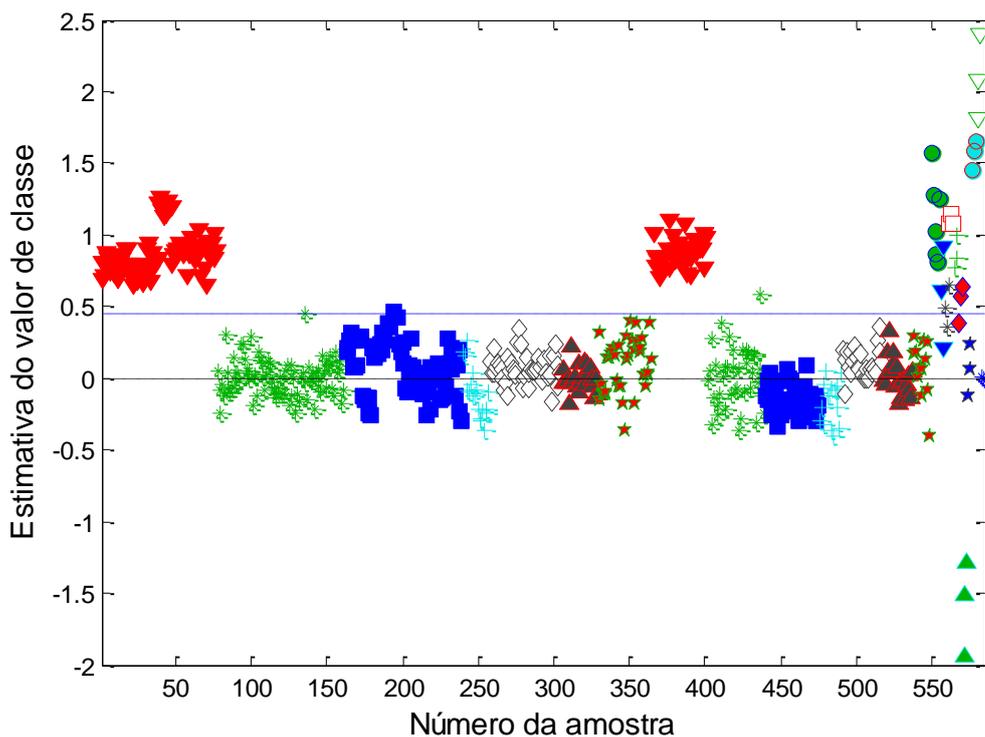


Figura 15. Resultados obtidos para modelo de discriminação entre amostras autênticas de Red Label (▼) e as amostras de Black Label (*), White Horse (■), Chivas Regal (+), Ballantine's (◇), Natu Nobilis (▲), Grand Old Parr (★) e as marcas não incluídas na calibração (demais símbolos).

3.3.1.3 Validação com amostras autênticas de White Horse

Para a classe White Horse não houve nenhuma amostra que apresentou valores de influência (T^2) e resíduos (Q) acima dos limites de 99,9% de confiança. Outro fato observado foi a ausência de erros falsos negativos e positivos. Ambos caracterizam a não ocorrência de erros de classificação.

A Figura 17 apresenta o gráfico de dispersão dos valores de classe estimados do modelo de discriminação entre amostras autênticas de White Horse e as demais marcas. Com base na Figura 17 é possível concluir que este modelo apresentou ótima separação entre as classes, visto não houve erro de classificação e apresentou taxas de seletividade, especificidade e eficiência de 100%. Isto significa que o modelo apresenta taxa de 100% de classificação correta de amostras tanto da classe 0 quanto da classe 1, conforme mostra a Tabela 8.

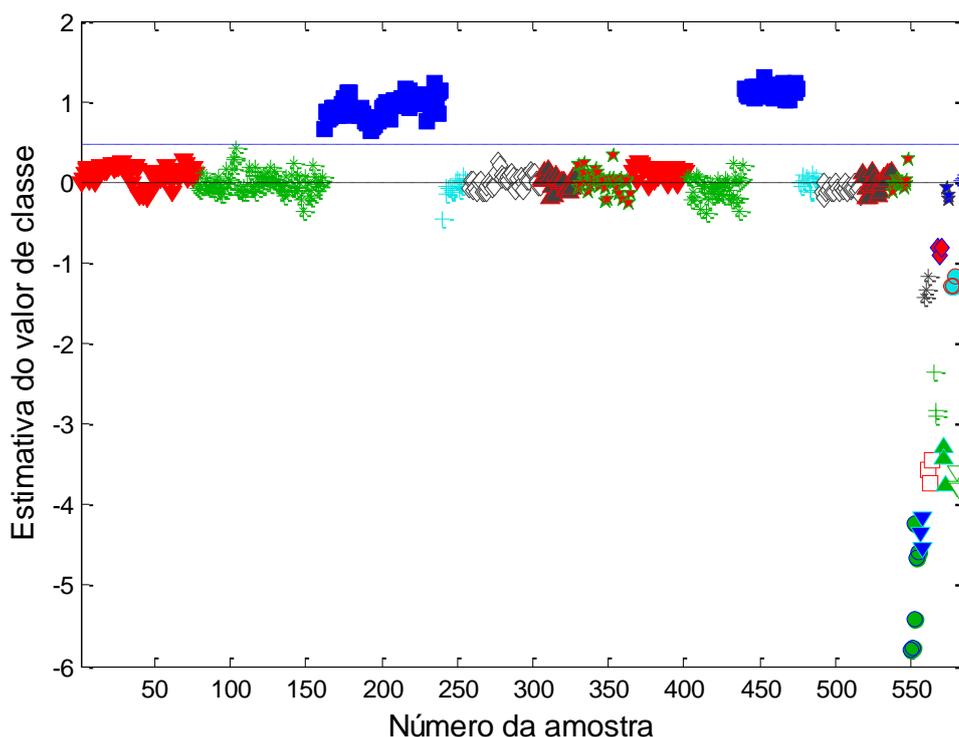


Figura 17. Resultados obtidos para modelo de discriminação entre amostras autênticas de White Horse (■) e as demais marcas.

3.3.1.4 Validação com amostras autênticas de Chivas Regal

Para a classe Chivas Regal, 2 replicatas apresentaram valores de influência (T^2) e resíduos (Q) acima do limite de 99,9% de confiança, entretanto, cada replicata correspondia a uma amostra diferente. Outro fato observado foi a ausência de falsos negativo e positivo. Ambos caracterizam a não ocorrência de erros de classificação.

A Figura 18 apresenta o gráfico de dispersão dos valores de classe estimados do modelo de discriminação entre amostras autênticas de Chivas Regal e as demais marcas. Com base na figura é possível concluir que este modelo apresentou ótima separação entre as classes, visto não houve erro de classificação e apresentou taxas de seletividade, especificidade e eficiência de 100%. Isto significa que o modelo apresenta taxa de 100% de classificação correta de amostras tanto da classe 0 quanto da classe 1, conforme mostra a Tabela 8.

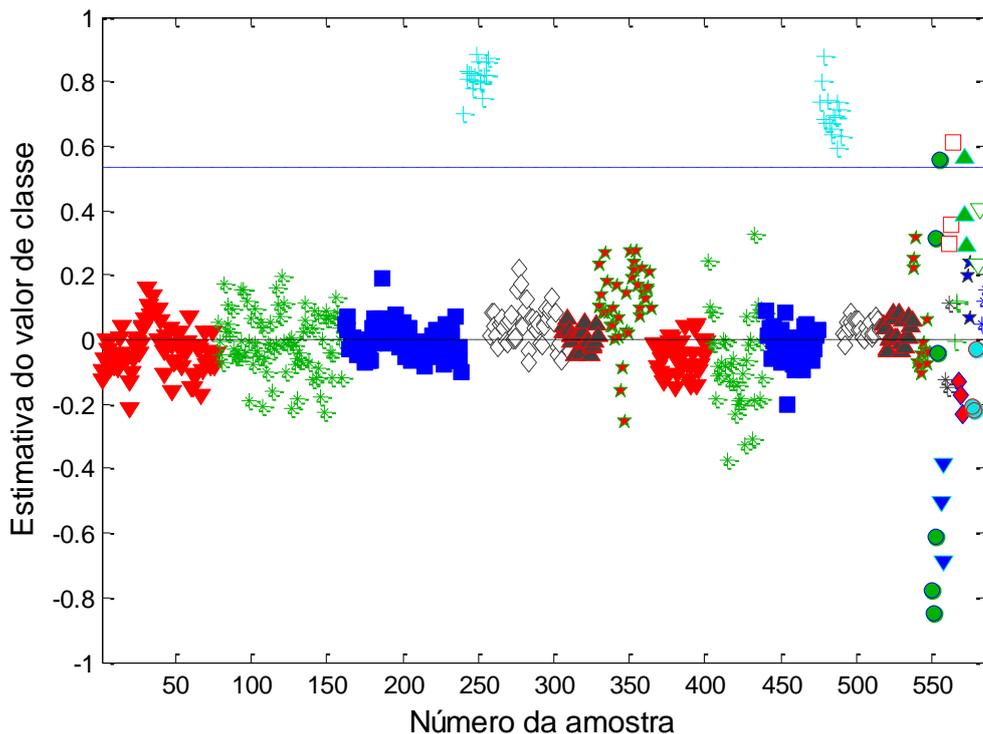


Figura 18. Resultados obtidos para modelo de discriminação entre amostras autênticas de Chivas Regal (+) e as demais marcas.

3.3.1.5 Validação com amostras autênticas de Ballantine's Finest

Para a classe Ballantine's, nenhuma amostra apresentou valores de influência (T^2) e resíduos (Q) acima do limite de 99,9% de confiança. Entretanto, uma amostra de uma marca que não estava incluída no conjunto de calibração apresentou esses parâmetros dentro do limite de confiança, sendo as três replicatas de uma mesma amostra. O erro de classificação foi confirmado com o valor de classe desta amostra compatível com a classe 1 (Ballantine's). Não houve ocorrência de falso negativo.

A Figura 19 apresenta o gráfico de dispersão dos valores de classe estimados do modelo de discriminação entre amostras autênticas de Ballantine's e as demais marcas. Embora tenha ocorrido um erro de classificação, apenas uma dentre 74 amostras foi classificada incorretamente. Portanto, com base na Figura 19 é possível concluir que o modelo apresentou ótima separação, com taxas de seletividade, especificidade e eficiência de 100%, 98,59% e 98,59%, respectivamente. Isto significa que o modelo possui uma taxa de 100% de classificação correta de amostras de Ballantine's como pertencentes à classe 1, possui uma taxa de 98,59% de classificação correta de amostras de outras marcas como pertencentes à classe 0, uma vez que uma amostra foi classificada como pertencente à classe 1, e possui taxa de 98,59% de classificação correta global, ou seja, tanto de classe 1 quanto de 0, devido ao fato de 1 amostra ter sido classificada incorretamente.

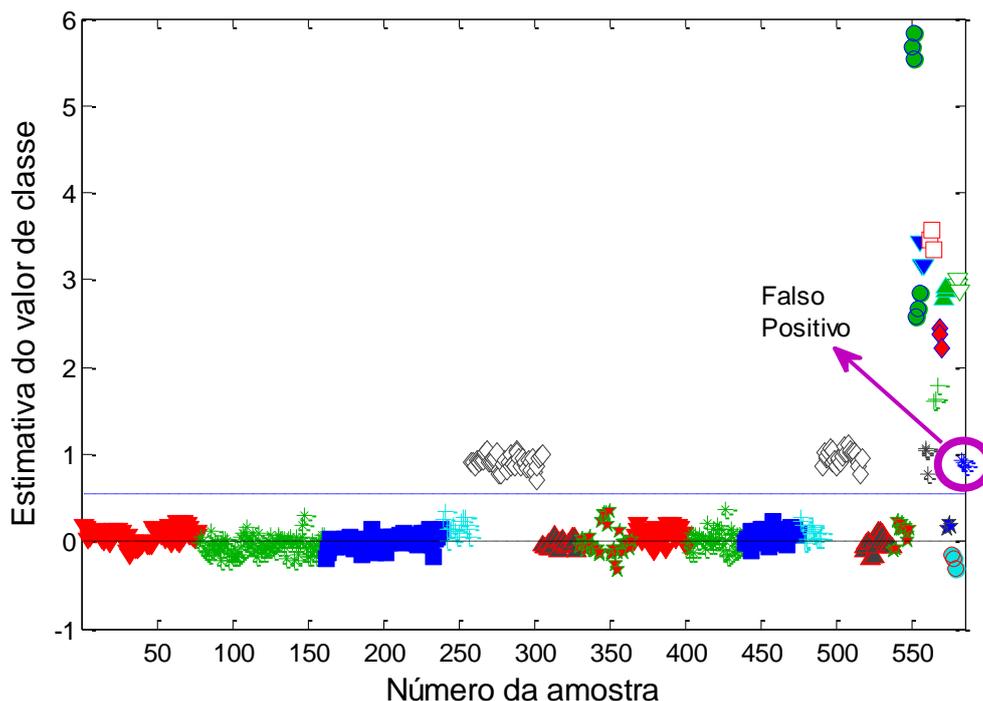


Figura 19. Resultados obtidos para modelo de discriminação entre amostras autênticas de Ballantine's Finest (\diamond) e as demais marcas.

3.3.1.6 Validação com amostras autênticas de Natu Nobilis

Para a classe Natu Nobilis, apenas 2 replicatas apresentaram valores de influência (T^2) e resíduos (Q) acima do limite de 99,9%, sendo cada uma correspondente a uma amostra diferente. Outro fato observado foi a ausência de falsos negativo e positivo. Ambos caracterizam a não ocorrência de erros de classificação.

A Figura 20 apresenta o gráfico de dispersão dos valores de classe estimados do modelo de discriminação entre amostras autênticas de Natu Nobilis e as demais marcas. Com base na Figura 20 é possível concluir que este modelo apresentou ótima separação entre as classes, visto que não houve erro de classificação e que apresentou taxas de seletividade, especificidade e eficiência de 100%. Isto significa que o modelo apresenta taxa de 100% de classificação correta de amostras tanto da classe 0 quanto da classe 1.

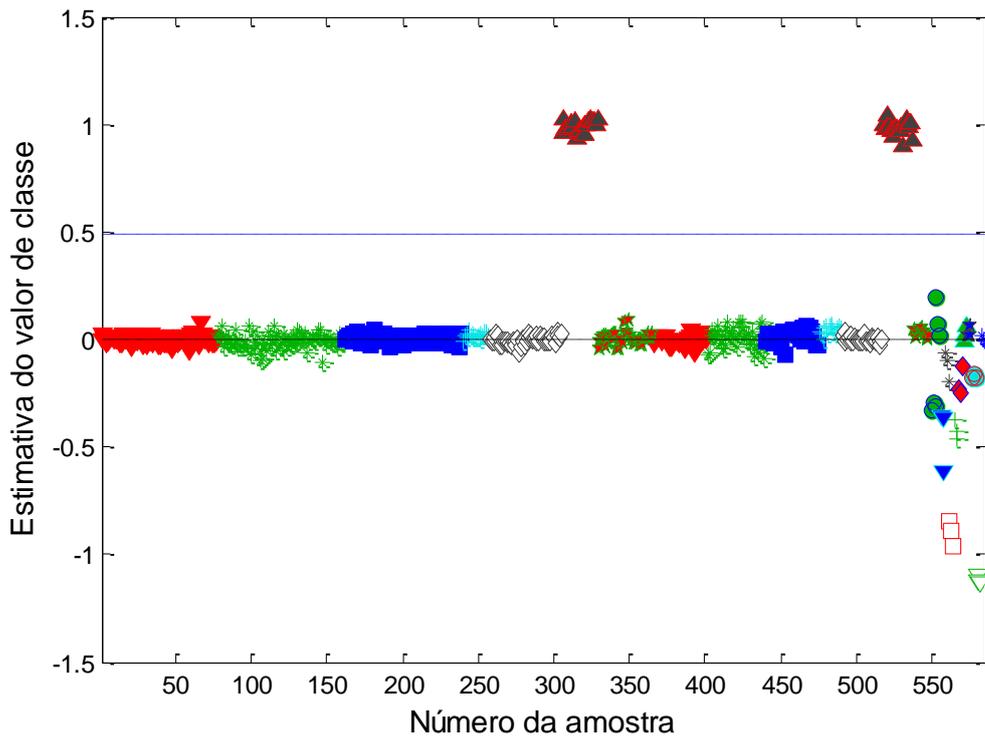


Figura 20. Resultados obtidos para modelo de discriminação entre amostras autênticas de Natu Nobilis (▲) e as demais marcas.

3.3.1.7 Validação com amostras autênticas de Old Parr

A classe Old Parr apresentou apenas 3 replicatas com valores de influência (T^2) e resíduos (Q) acima do limite de 99,9% de confiança, sendo cada uma pertencente a uma amostra diferente. Entretanto, três replicatas de uma amostra de uma marca que não foi incluída no conjunto de calibração apresentou esses parâmetros dentro do limite de confiança. O erro de classificação foi confirmado com o valor de classe desta amostra compatível com a classe 1 (Old Parr). Não houve ocorrência de falso negativo.

A Figura 21 apresenta o gráfico de dispersão dos valores de classe estimados do modelo de discriminação entre amostras autênticas de Old Parr e as demais marcas. Embora tenha ocorrido um erro de classificação, apenas uma dentre 74 amostras foi classificada incorretamente. Portanto, com base na Figura 21 é possível concluir que o modelo apresentou ótima separação, com taxas de seletividade, especificidade e eficiência de 100%, 98,56% e 98,56%, respectivamente. Isto significa que o modelo possui uma taxa de 100% de classificação correta de amostras de Old Parr como pertencentes à classe 1,

possui uma taxa de 98,56% de classificação correta de amostras de outras marcas como pertencentes à classe 0, uma vez que uma amostra foi classificada como pertencente à classe 1, e possui taxa de 98,56% de classificação correta global, ou seja, tanto de classe 1 quanto de 0, devido ao fato de 1 amostra ter sido classificada incorretamente.

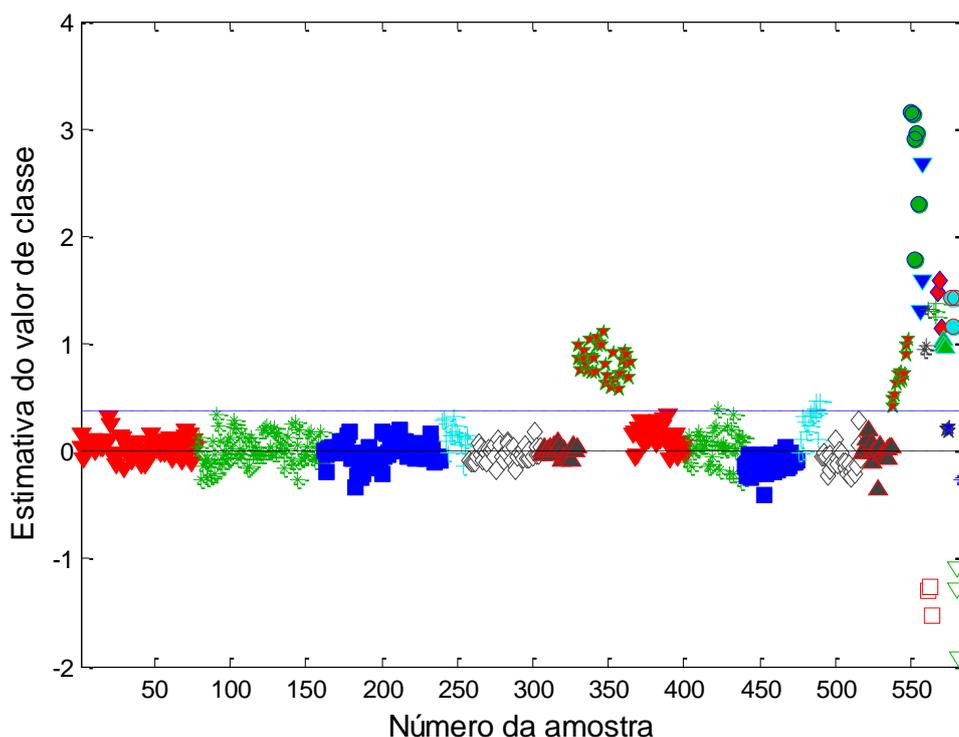


Figura 21. Resultados obtidos para modelo de discriminação entre amostras autênticas de Grand Old Parr (★) e as demais marcas.

3.3.2 Predição de amostras falsas

Após a predição com amostras autênticas para avaliação da capacidade preditiva dos modelos, estes foram aplicados a amostras provenientes de apreensões da Polícia Federal e Polícia Civil do DF, que já haviam passado por análises prévias para elaboração dos laudos, sendo que através dessas análises todas as amostras foram atestadas como falsas.

Como todas as amostras são falsas, espera-se que todas apresentem valores de T^2 , Q e valor de classe maiores que os limites de 99,9%. Amostras que apresentaram este comportamento foram classificadas como não pertencentes às classes modeladas ou não autênticas. Só foram considerados

erros de classificação nos casos em que as amostras apresentaram estes parâmetros dentro dos limites estabelecidos e valor de classe compatível com a classe discriminada.

O cálculo da taxa de classificação correta foi feito a partir das fórmulas que constam na seção 2.3.1.1.5. Para fazer o cálculo desta figura de mérito, foi realizada uma adaptação nos critérios de identificação dos falsos positivo e negativo, assim como os verdadeiros positivo e negativo, com relação aos critérios da Tabela 2. Os novos critérios estão listados na Tabela 9.

A taxa de seletividade indica a taxa de classificação correta de amostras como classe 1 e a taxa de especificidade indica a taxa de classificação correta de amostras como classe 0. Como este conjunto de predição se trata de amostras falsas em que o esperado é não haver classificação como classe 0 ou 1, então estas figuras de mérito não podem ser calculadas, uma vez que não há hipótese em que haja verdadeiro positivo ou negativo. A taxa de classificação correta indica a eficiência do modelo em classificar estas amostras como não autênticas.

Tabela 9. Critérios de identificação de falsos positivo e negativo e verdadeiros positivo e negativo de predições de amostras falsas

Fatores	Abreviação	Significado
Falso Positivo	FP	Amostra falsa classificada como amostra autêntica (classe 1)
Falso Negativo	FN	NÃO SE APLICA
Verdadeiro Positivo	VP	NÃO SE APLICA
Verdadeiro Negativo	VN	NÃO SE APLICA

3.3.2.1 Análise de amostras falsas com o modelo de discriminação de Red Label

O gráfico de influência (T^2) e resíduos (Q) da predição de amostras falsas através do modelo de discriminação de amostras de Red Label é mostrado na Figura 23. Como esperado, a grande maioria das amostras falsas apresentaram valores de T^2 e Q acima do limite de 99,9% de confiança. A Figura 23B mostra uma ampliação da região do gráfico abaixo do limite de confiança, no qual é possível ver que algumas amostras falsas apresentaram valores de T^2 e Q compatíveis com os valores das amostras autênticas. A estas amostras foram aplicados os critérios da Figura 12 a fim de identificar se ocorreu erro de classificação.

Com base na Figura 24 pode-se observar que a maioria das amostras apresenta valores de classe muito superiores a 1 ou muito inferiores a 0, indicando amostras com comportamento diferente daquelas usadas na etapa de calibração. Combinando os resultados das Figuras 23 e 24, é possível concluir que este modelo apresentou boa discriminação entre amostras falsas e autênticas, visto que de 73 amostras falsas verificou-se apenas 1 falso positivo, que levou a uma taxa de classificação correta de 98,6%, conforme mostra a Tabela 10.

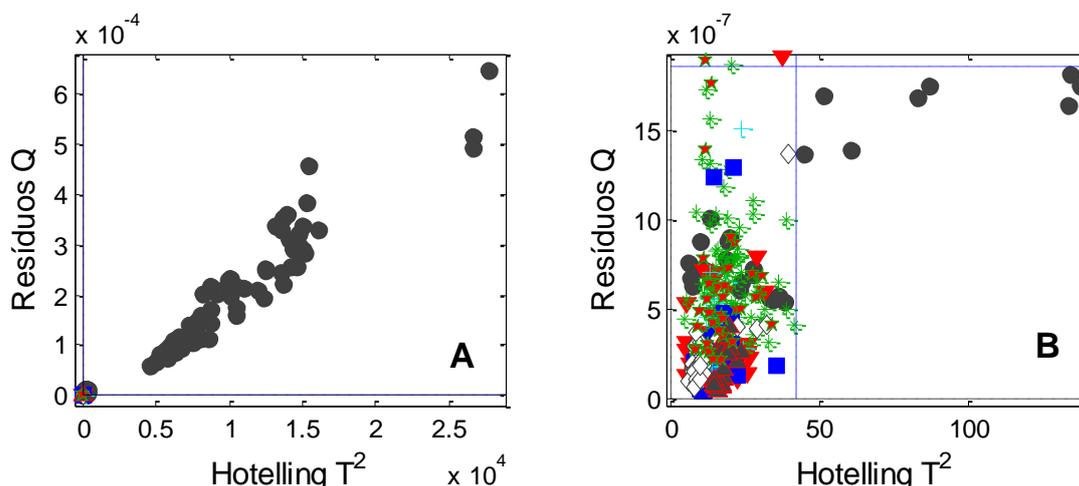


Figura 22. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas Red Label (▼) e as amostras de Black Label (*), White Horse (■), Chivas Regal (+), Ballantine's (◇), Natu Nobilis (▲), Grand Old Parr (★). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.

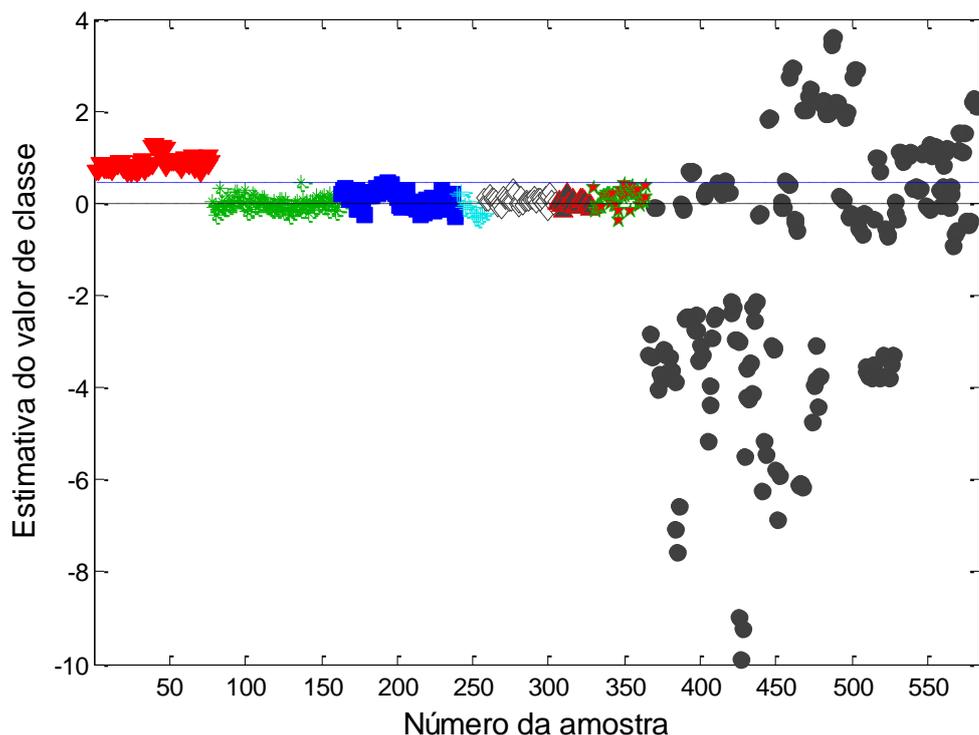


Figura 23. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Red Label (▼). As demais marcas do modelo são: Black Label (*), White Horse (■), Chivas Regal (+), Ballantine's (◇), Natu Nobilis (▲), Grand Old Parr (★).

Tabela 10. Parâmetros e figuras de mérito dos modelos de predição de amostras falsas.

	Red	Black	White	Chivas	Ball	Natu	Old
N	23	15	9	7	8	9	2
NA	67	67	65	65	64	67	67
T²Q	6	6	8	8	9	6	6
FP	1	1	5	0	1	0	0
TCC*	98,6	98,6	93,1	100	98,6	100	100

N = número de amostras; NA = número de amostras classificadas como não autênticas; FP = Falso Positivo; FN = Falso Negativo; TCC = Taxa de Classificação Correta; *Expresso em percentagem; T²Q = valores de T² e Q abaixo do limite de 99,9%

3.3.2.2 Análise de amostras falsas com o modelo de discriminação de Black Label

Para o modelo de discriminação da marca Black Label, os valores de T^2 e Q (Figura 25) evidenciam que a grande maioria das amostras falsas se apresentaram além do limite de 99,9% de confiança. A Figura 25B mostra uma ampliação da região do gráfico abaixo do limite de confiança, no qual é possível ver que algumas amostras falsas apresentaram valores de T^2 e Q compatíveis com os valores das amostras autênticas. A estas amostras foram aplicados os critérios da Figura 12 a fim de identificar se ocorreu erro de classificação, onde verificou-se 1 erro falso positivo, que levou a uma taxa de classificação correta de 98,6%, conforme mostra a Tabela 10. O gráfico de dispersão dos valores de classe é apresentado na Figura 26, que se assemelha muito ao resultado obtido para a marca Red Label.

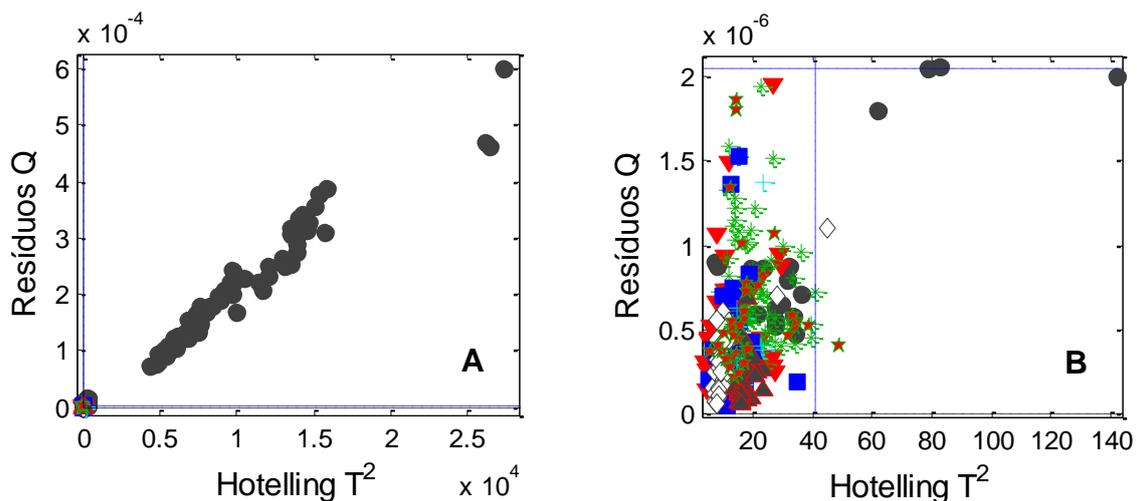


Figura 24. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Black Label (*). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.

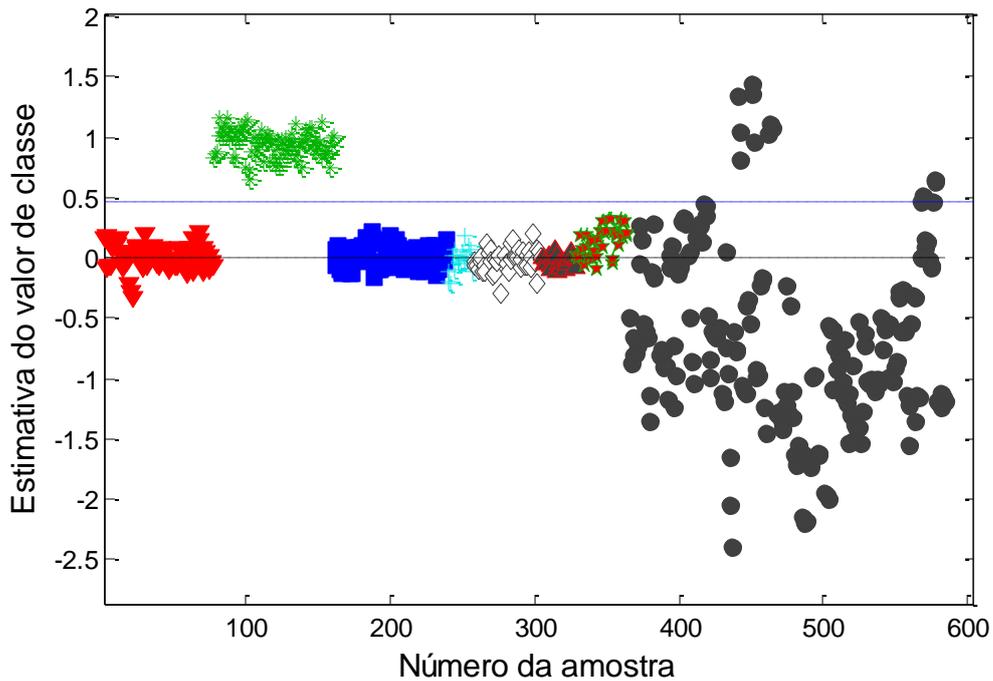


Figura 25. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Black Label (*)

Os gráficos de T^2 e Q para os modelos de White Horse, Chivas, Ballantine's, Old parr e Natu Nobilis (Figuras ímpares de 27 a 35) apresentaram resultados muito semelhantes aos observados nas Figuras 23 e 25 dos modelos das marcas Red Label e Black Label, respectivamente. Com base nesses resultados e nas Figuras pares de 28 a 36, foram obtidas taxas de classificação correta de que variaram de 93,1% a 100% para essas marcas, conforme apresentado na Tabela 10.

3.3.2.3 Análise de amostras falsas com o modelo de discriminação de White Horse

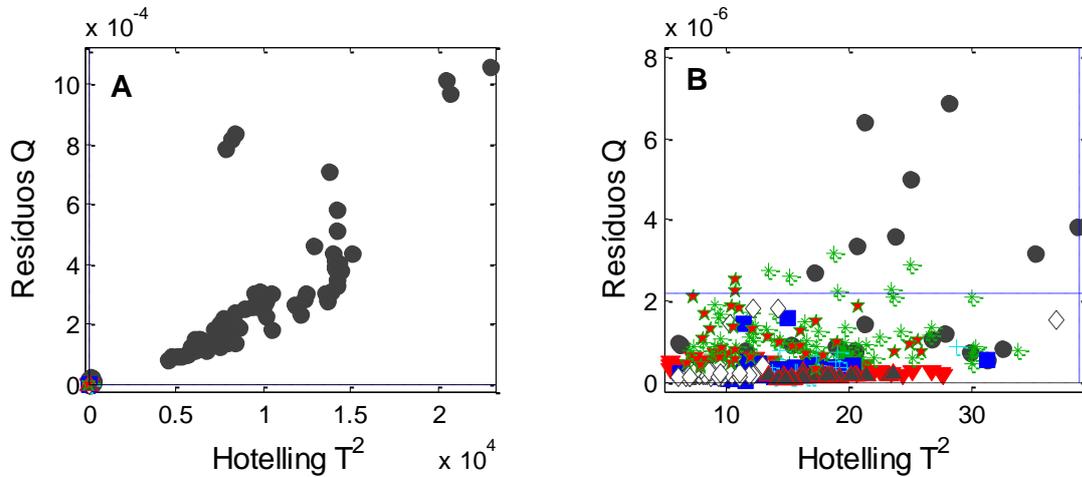


Figura 26. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de White Horse (■). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.

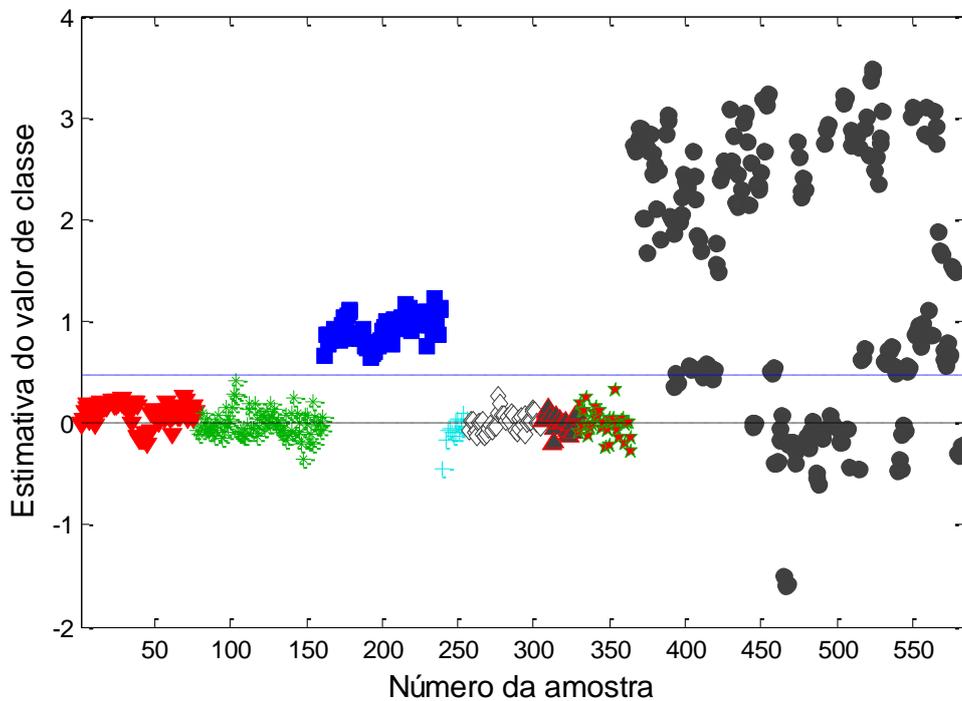


Figura 27. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de White Horse (■).

3.3.2.4 Análise de amostras falsas com o modelo de discriminação de Chivas Regal

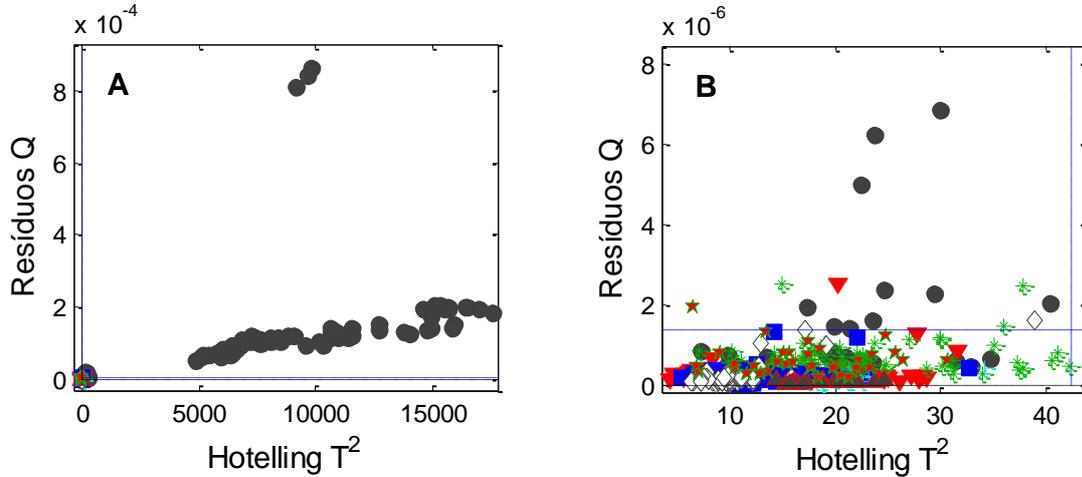


Figura 28. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Chivas Regal (+). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.

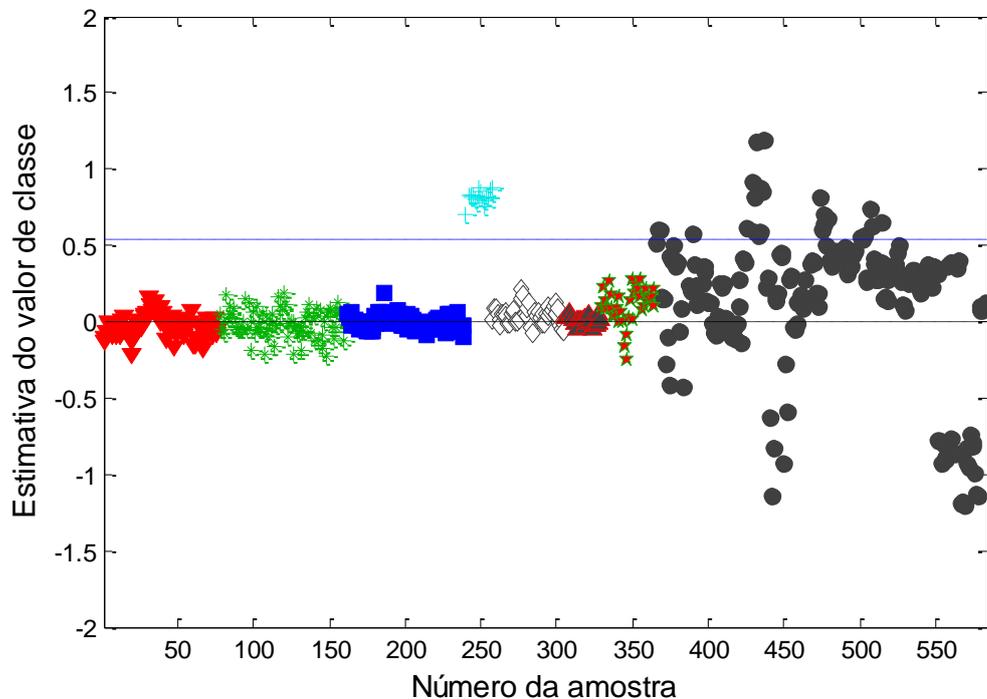


Figura 29. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Chivas Regal (+).

3.3.2.5 Análise de amostras falsas com o modelo de discriminação de Ballantine's Finest

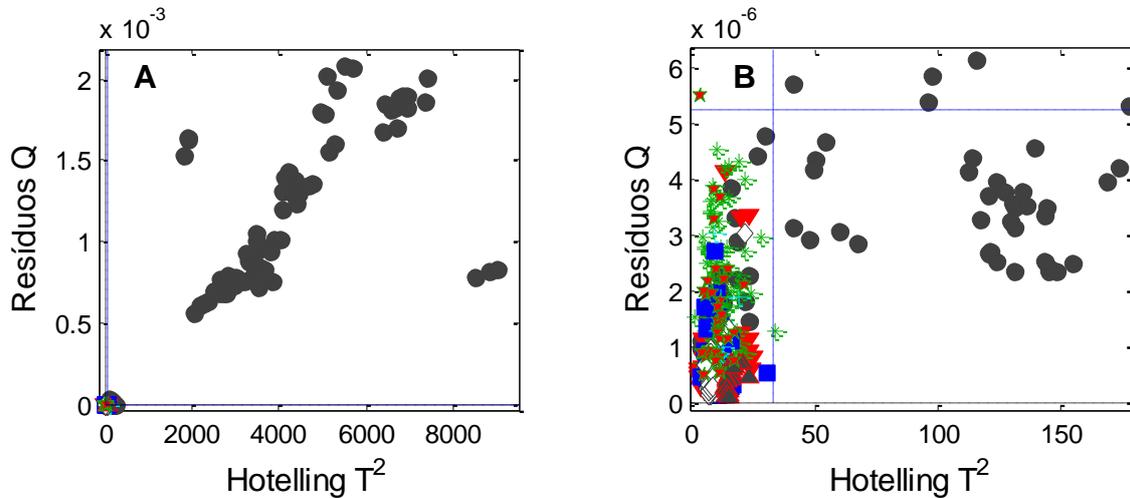


Figura 30. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Ballantine's (◇). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T² e Q abaixo do limite de 99,9% de confiança.

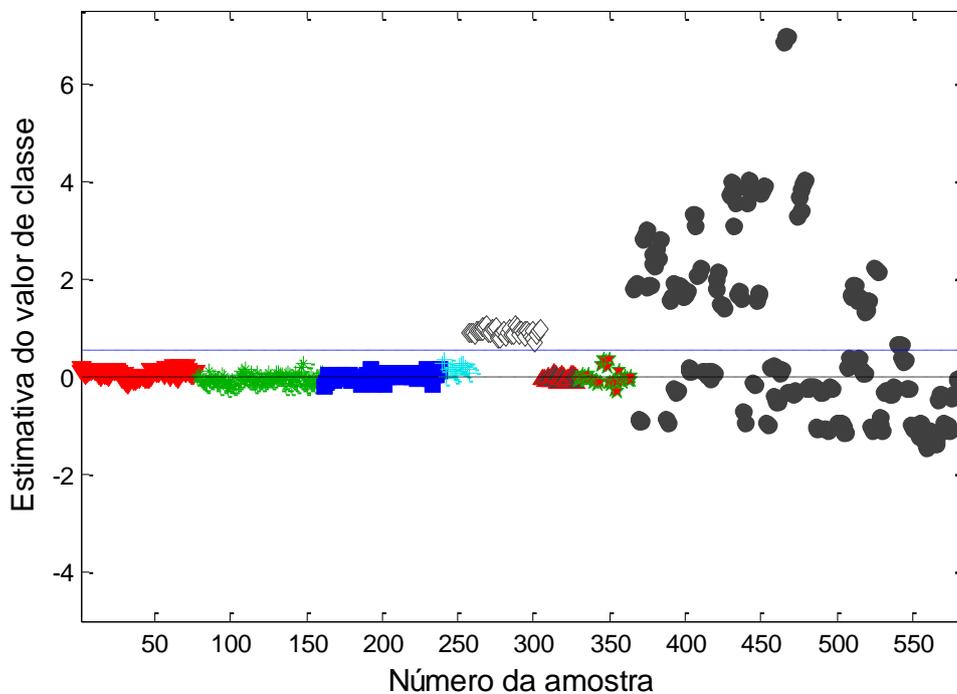


Figura 31. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Ballantine's (◇);

3.3.2.6 Análise de amostras falsas com o modelo de discriminação de Natu Nobilis

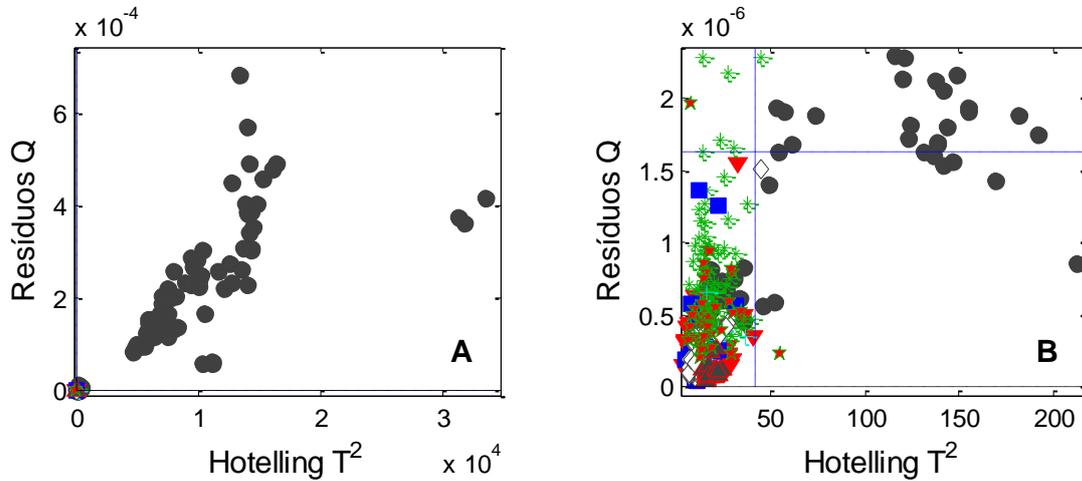


Figura 32. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Natu Nobilis (▲). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.

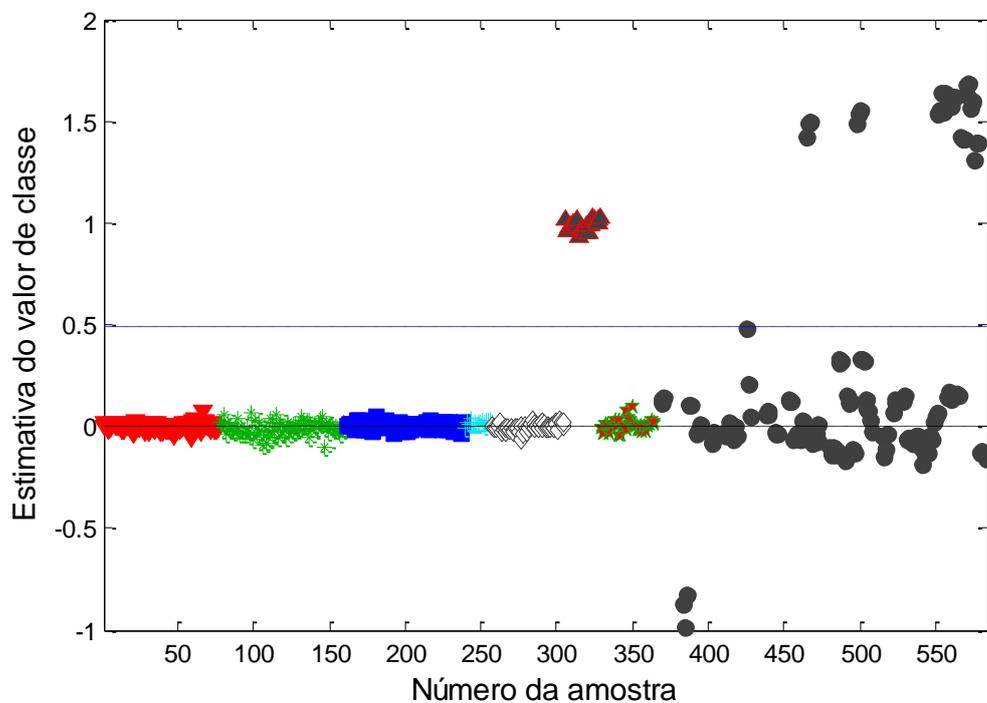


Figura 33. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Natu Nobilis (▲).

3.3.2.7 Análise de amostras falsas com o modelo de discriminação de Old Parr

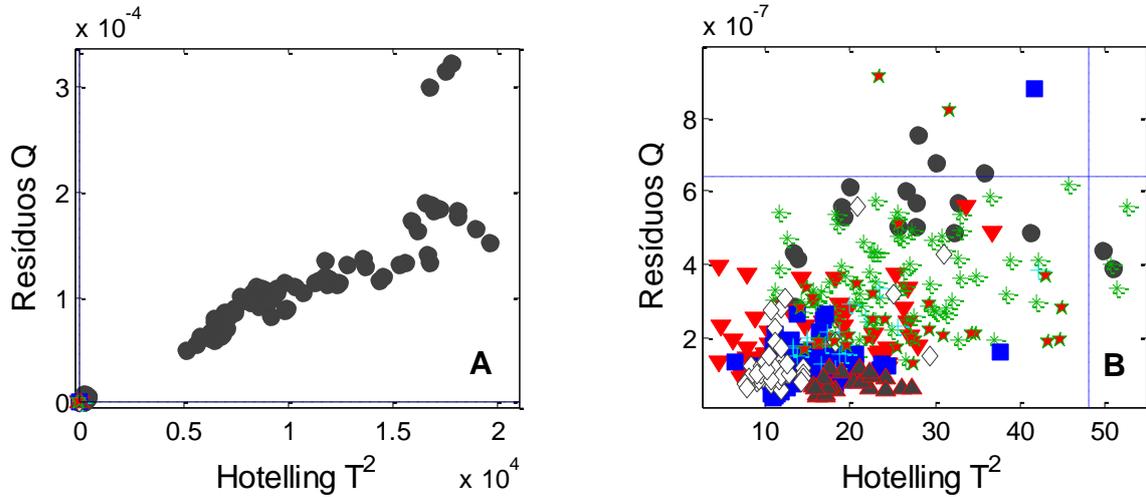


Figura 34. Gráficos da estatística de Hotelling para predição de amostras falsas (●) através do modelo de discriminação de amostras autênticas de Old Parr (★). A) Gráfico sem ampliação; B) Gráfico com ampliação da região com T^2 e Q abaixo do limite de 99,9% de confiança.

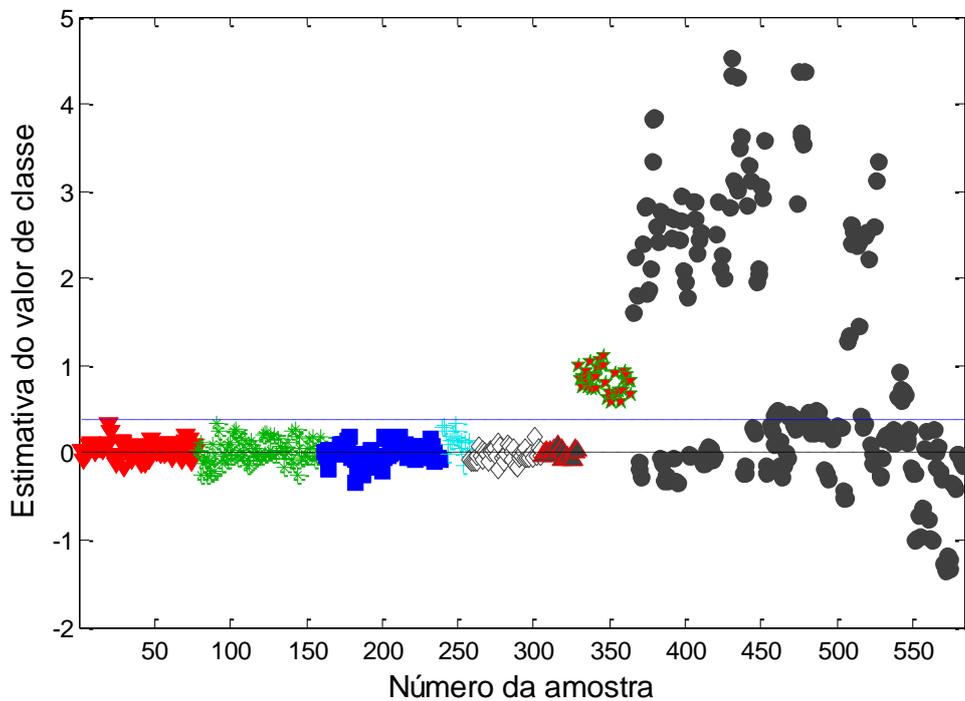


Figura 35. Resultados obtidos para modelo de discriminação entre amostras autênticas e amostras falsas (●), através do modelo de discriminação de Old Parr (★).

Com relação aos erros observados na análise das amostras falsas, foi observado que 5 dessas amostras (todas supostamente de Red Label) se repetiram em todos os modelos e foram classificadas como falso positivo pelo modelo de White Horse. Desta forma, estas amostras foram classificadas como autênticas da marca White Horse pelo modelo desta marca. Contudo, a percentagem de erros observados foi relativamente baixa, de forma que podemos considerar que o método proposto apresentou bons resultados na análise de amostras falsas. Cabe ressaltar que as amostras falsas também puderam ser identificadas pelas análises de rotina feitas nos laboratórios da Polícia Federal, dentre elas análise dos selos de importação, do dosador e presença de corpo de fundo no líquido. Logo, é indicado que, em um caso real, estas análises sejam associadas às de referência, a fim de aumentar a segurança do resultado.

Observando a tabela 10, pode-se constatar que apenas na marca White Horse o modelo apresentou um número maior de erros falso positivos (5 amostras), sendo que para as demais marcas apenas 1 ou nenhuma amostra apresentou esse tipo de erro. Portanto, pode-se considerar que o método foi eficiente para a análise de amostras falsas, tendo obtido taxas de classificação correta maiores ou iguais a 93,1 %.

3.3.3 Aplicação dos modelos em caso real

Durante a realização do trabalho, a Polícia Federal fez uma apreensão de bebidas em uma operação. O perito responsável pelas análises cedeu 16 amostras para aplicação destes modelos de calibração. Com base nas análises do perito, apenas duas amostras de Red Label foram classificadas como falsas, as demais se mostraram autênticas. Juntamente com as amostras da apreensão, o perito também forneceu amostras de referência, com o intuito de comparar o comportamento de duas amostras de origens diferentes. Dentre as amostras de referência fornecidas, havia uma amostra reconhecidamente falsa de Red Label, fornecida para se comparar o comportamento dela com o das duas amostras falsas da apreensão.

Os resultados das análises destas amostras pelos modelos quimiométricos não foram muito satisfatórios, pois não foi possível fazer uma

boa classificação das amostras. Várias investigações foram feitas para a adequada interpretação destes resultados. A primeira delas foi sobrepor os espectros destas amostras com os espectros das amostras de calibração das mesmas 4 marcas das amostras de apreensão, com o objetivo de observar se havia visivelmente alguma característica em seu padrão que justificasse tal comportamento. Porém, como pode ser visto na Figura 37, o padrão dos espectros das amostras de apreensão é, ao menos visivelmente, compatível com as amostras de calibração, com exceção das amostras falsas, que apresentam menores valores de absorbância.

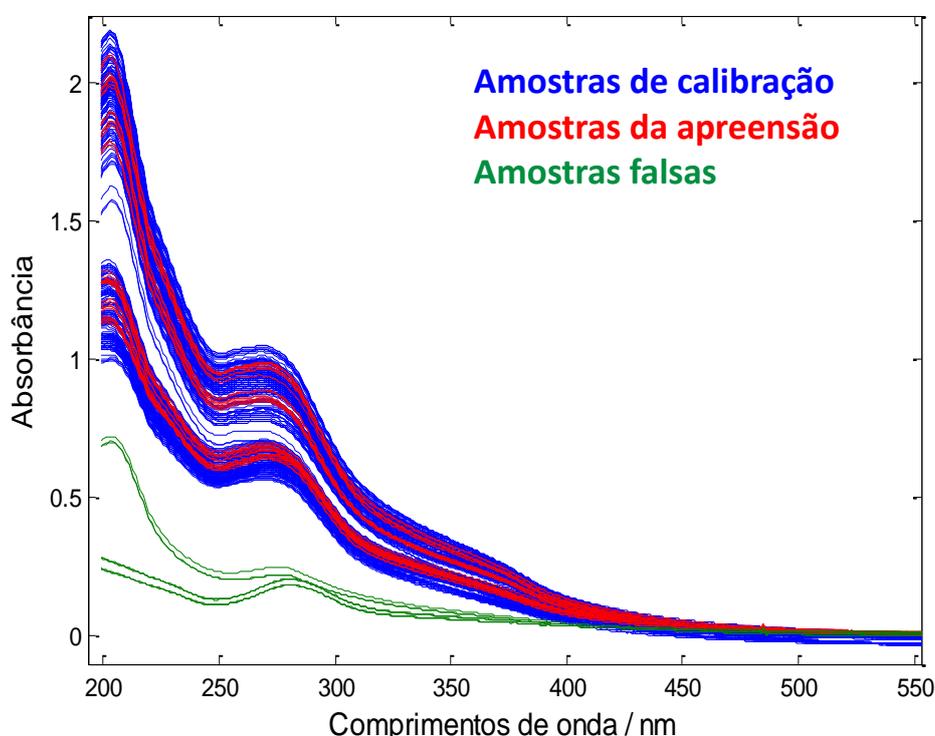


Figura 36. Espectros de absorção de UV-Vis de amostras autênticas da apreensão (vermelho), amostras falsas (verde) e amostras do conjunto de calibração (azul).

Tendo em vista que os espectros das amostras autênticas da apreensão não apresentam diferença visual perceptível com relação às amostras de calibração, levantou-se a possibilidade de ter ocorrido variações no espectrofotômetro UV-Vis durante os 7 meses de intervalo entre as medidas instrumentais das amostras de calibração e as amostras do laudo. Para tentar solucionar o problema, foi feita a recalibração do modelo através da inserção de novos espectros no conjunto de calibração, com o intuito de incluir possíveis

alterações que o instrumento tenha sofrido nesse intervalo de tempo. Entretanto, esta tentativa não surtiu efeitos.

Os resultados dos valores estimados de classe para os modelos são mostrados a seguir e nas tabelas abaixo, sendo que a Tabela 11 mostra os resultados de identificação das amostras e a Tabela 12 detalha a quantidade de amostras, os erros falsos positivos e negativos e mostra as figuras de mérito para os modelos. A Tabela 11 mostra valores de classificação correta e incorreta para cada marca de uísque em seu respectivo modelo. Sobre a Tabela 12 é importante destacar que a taxa de classificação correta foi calculada como:

$$TCC = \frac{N_{total} - FP - FN}{N_{total}} \quad (22)$$

onde N_{total} é o número total de amostras (16), FP e FN são os erros falsos positivo e negativo, respectivamente. Desta forma, a taxa de classificação correta é calculada através da razão entre o número de classificações corretas e o número total de amostras.

Tabela 11. Figuras de mérito de modelos de discriminação de amostras provenientes da apreensão da Polícia Federal

	Nº de amostras	Nº de classificações corretas	Nº de classificações erradas
Old	3	1	2
Red aut	5	1	4
Red F	3	3	0
Black	3	1	2
White	2	2	0

Old = Old Parr; Red aut = amostras autênticas de Red Label; Red F = amostras falsas de Red Label; Black = Black Label; White = White Horse

Tabela 12. Figuras de mérito de modelos de discriminação de amostras provenientes da apreensão da Polícia Federal

Figuras de mérito	Red	Black	Old	White
N	5	3	3	2
NA/NP	9	6	6	8
FP	0	2	0	2
FN	2	0	1	0
VP	1	1	1	2
VN	4	7	8	4
TFN*	66,7	0	50	0
TFP*	0	22,2	0	33,3
TST*	33,3	100	50	100
TSP*	100	77,8	100	66,7
TEF*	33,3	77,8	50	66,7
TCC*	80	84,6	92,3	81,8

*Expresso em percentagem; N = número de amostras; NA/NP = número de amostras classificadas como não autênticas ou não pertencentes às classes modeladas; FP = Falso Positivo; FN = Falso Negativo; VP = Verdadeiro Positivo; VN = Verdadeiro Negativo; TFN = Taxa de Falso Negativo; TFP = Taxa de Falso Positivo; TST = Taxa de seletividade; TSP = Taxa de especificidade; TEF = Taxa de Eficiência; TCC = Taxa de classificação Correta

3.3.3.1 Análise de amostras reais a partir de modelo de discriminação de Red Label

As amostras falsas de Red Label (com numeração de 405 a 413) apresentaram valores bastante altos de T^2 e Q em todos os modelos. Este resultado era esperado pois são as únicas amostras falsas deste conjunto.

A Figura 38 mostra os resultados para o modelo de discriminação entre amostras de Red Label e as demais marcas provenientes da apreensão da Polícia Federal. Na Figura 38 o valor de classe estimado para estas amostras foi em torno de 0 e de 3, bastante discrepante do valor 1 esperado para amostras desta classe. Neste ponto, nota-se que o modelo identificou facilmente as amostras falsas. As amostras falsas da apreensão apresentaram valores de

classe próximos de 0, enquanto que a amostra falsa de referência apresentou o valor de classe por volta de 3. Este resultado sugere que o modelo possa ter conseguido distinguir o tipo de falsificação, pois as amostras da apreensão são de origem diferente da amostra falsa de referência. Entretanto, o modelo não conseguiu classificar corretamente nenhuma das amostras de Red Label autênticas, apenas o padrão de referência selecionado pelo perito que analisou as amostras. Os 2 erros falsos negativos contribuíram com a ocorrência de taxas de seletividade, eficiência e classificação correta, de 33,3%, 13,3% e 80%, respectivamente, sendo estas as taxas mais baixas observadas na aplicação do método proposto.

Das amostras autênticas, 9 foram classificadas como não pertencentes às classes modeladas ou não autênticas pelo modelo e dentre elas estão, além da própria Red Label, amostras das marcas Black Label e Old Parr, o que significa que o modelo não foi capaz de reconhecer estas amostras como pertencentes à classe 0. As 9 amostras com a referida classificação correspondem a 56,2% das amostras totais, o que indica que o modelo não apresentou boa capacidade de identifica-las como semelhantes às do conjunto de calibração.

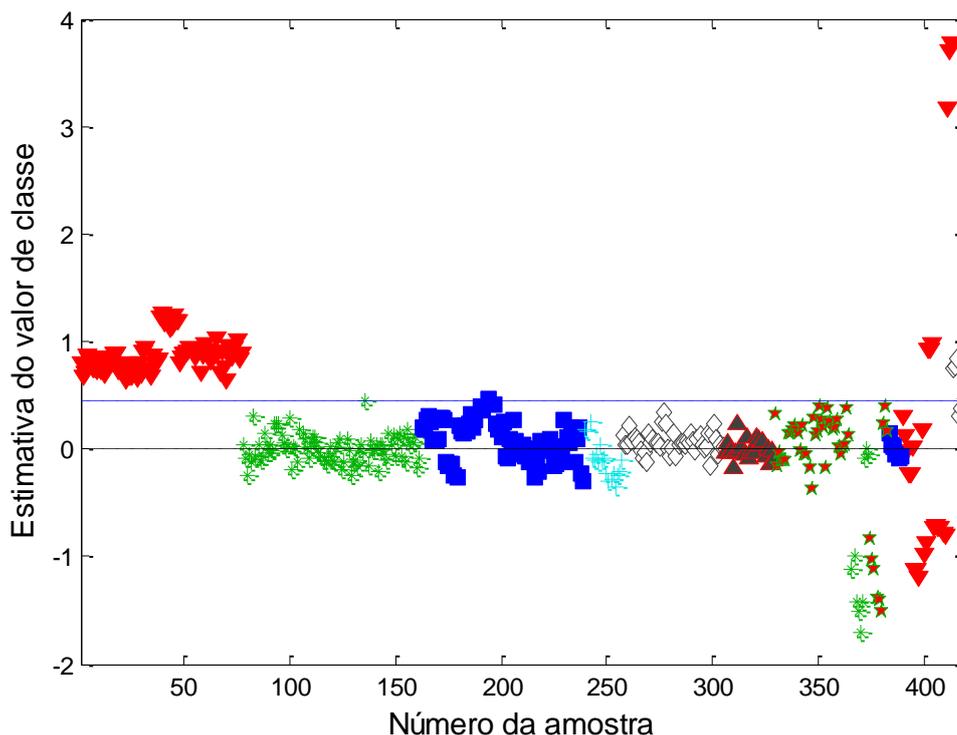


Figura 37. Resultados obtidos para modelo de discriminação entre amostras de Red Label (▼) e as demais marcas provenientes da apreensão da Polícia Federal. As demais marcas do modelo são: Black Label (*), White Horse (■) e Grand Old Parr (★).

3.3.3.2 Análise de amostras reais a partir de modelo de discriminação de Black Label

A Figura 39 mostra o resultado do modelo de discriminação entre amostras de Black Label e as demais marcas de amostras provenientes da apreensão da Polícia Federal. Esta classe é composta por 2 amostras da apreensão e 1 padrão escolhido pelo perito. As duas amostras provenientes da apreensão apresentaram elevados valores T^2 e Q , acima do limite de 99,9% de confiança, sendo assim classificadas como não pertencentes às classes modeladas ou não autênticas. Apenas a amostra de referência foi classificado corretamente. Entretanto, embora o modelo não tenha identificado corretamente as duas amostras de Black Label, como não houve nenhuma ocorrência de erro falso positivo, a taxa de seletividade calculada foi de 100%. A taxa de seletividade indica a taxa de classificação correta como classe 1, porém neste caso, a interpretação correta desta figura de mérito seria a taxa de não ocorrência de falsos negativos, uma vez que que o cálculo independe da

quantidade de amostras que foram classificadas corretamente como classe 1, apenas de quantidade de falsos negativos.

A taxa de eficiência indica a taxa de classificações corretas baseada apenas na ocorrência de falsos positivos e negativos, não considerando a quantidade total de amostras. Para a correta avaliação da capacidade do modelo em classificar corretamente as amostras, seja como classe 0 ou classe 1, a Taxa de Classificação Correta é mais adequada, pois leva em consideração, em seu cálculo, a quantidade de erros em relação à quantidade total de amostras. Para este modelo, a taxa de eficiência foi de 77,8% e a Taxa de Classificação Correta (TCC) foi de 84%. Dentre as 16 amostras totais, o modelo apresentou 6 amostras classificadas como não pertencentes às classes modeladas ou não autênticas, um valor correspondente a 37,5% das amostras. Isto indica que o modelo não apresentou boa capacidade de identificar essas amostras como semelhantes às do conjunto de calibração.

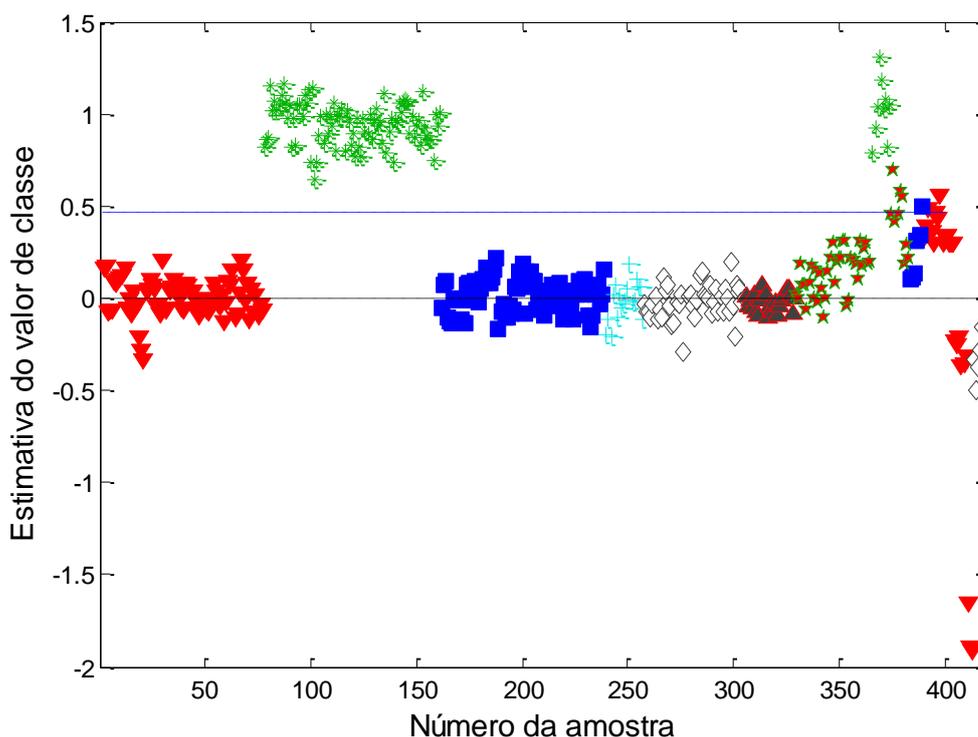


Figura 38. Resultados obtidos para modelo de discriminação entre amostras de Black Label (*) provenientes da apreensão da Polícia Federal.

3.3.3.3 Análise de amostras reais a partir de modelo de discriminação de White Horse

A Figura 40 mostra o resultado do modelo de discriminação entre amostras de White Horse e as demais marcas provenientes da apreensão da Polícia Federal. Analisando a figura é possível notar que as amostras de Black Label, Old Parr e Red Label apresentaram grande dispersão, ultrapassando o limite de discriminação. Por outro lado, as amostras de White Horse apresentaram valor de classe compatível com o valor esperado. Como os valores de influência e resíduo destas amostras apresentaram valor dentro do limite de 99,9% e o valor de classe foi compatível com a classe 1, elas foram classificadas corretamente, não havendo, portanto, nenhum falso negativo no modelo. Dentre os modelos aplicados às amostras provenientes de apreensão, este foi o único que classificou todas as amostras corretamente. Embora seja um modelo com características únicas, a quantidade de amostras consideradas como não pertencentes às classes modeladas ou não autênticas foi elevado, com 8 dentre 16 amostras totais, o que correspondeu a 50% das amostras. Não havia amostra de referência para essa marca, então todas as amostras eram provenientes da apreensão. A taxa de classificação correta foi de 81,8%.

As amostras de Black Label e Old Parr apresentaram valores de influência e resíduo acima do limite de 99,9% de confiança, com exceção apenas de suas amostras de referência. As amostras de Red Label apresentaram comportamento semelhante, apenas 2 amostras foram classificadas corretamente como pertencente à classe 0, dentre elas o padrão, 2 foram classificadas erroneamente como pertencentes à classe 1. As falsas e uma autêntica apresentaram os valores de T^2 e Q elevados. Este resultado indica que o modelo classificou corretamente apenas as amostras de White Horse e as de referência das outras marcas, seja como classe 0 ou classe 1.

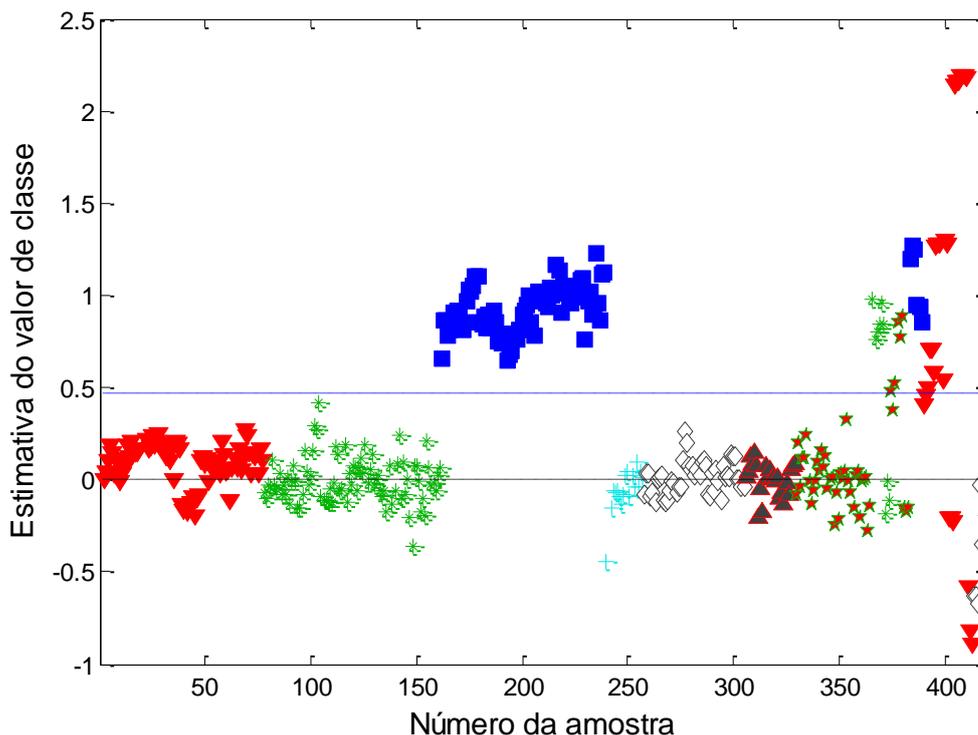


Figura 39. Resultados obtidos para modelo de discriminação entre amostras de White Horse (■) provenientes da apreensão da Polícia Federal.

3.3.3.4 Análise de amostras reais a partir de modelo de discriminação de Old Parr

A Figura 41 mostra o resultado do modelo de discriminação entre as amostras de Old Parr e as demais marcas provenientes da apreensão da Polícia Federal. Analisando a figura, é possível notar que apenas 1 replicata de uma amostra de Black Label, 3 replicatas de amostras de Old Parr e 3 replicatas de Red Label ultrapassaram o limite de discriminação. Destas replicatas, as 2 de Old Parr são da mesma amostra, o que levou a um erro de classificação como falso negativo, as 3 replicatas de Red Label são da mesma amostra, porém não foi classificada como falso positivo porque foram classificadas como não pertencentes às classes modeladas ou não autênticas pelos critérios de T^2 e Q , assim como as amostras falsas de Red Label, 1 amostra de Old Parr e 2 amostras de Black Label. Destas duas últimas marcas, este modelo classificou corretamente apenas as amostras de referência. Das duas amostras de Old Parr da apreensão, uma foi classificada como não pertencente às classes modeladas

ou não autêntica e a outra como falso negativo. As amostras autênticas de Red Label foram classificadas corretamente como classe 0.

Como não houve ocorrência de falso positivo, a taxa de especificidade foi de 100%. Em contrapartida, como das 3 amostras 1 foi classificada como falso negativo e uma como não pertencentes às classes modeladas ou não autêntica, a taxa de seletividade foi de 50%, gerando assim uma taxa de eficiência de 50% e uma taxa de classificação correta de 92,3%.

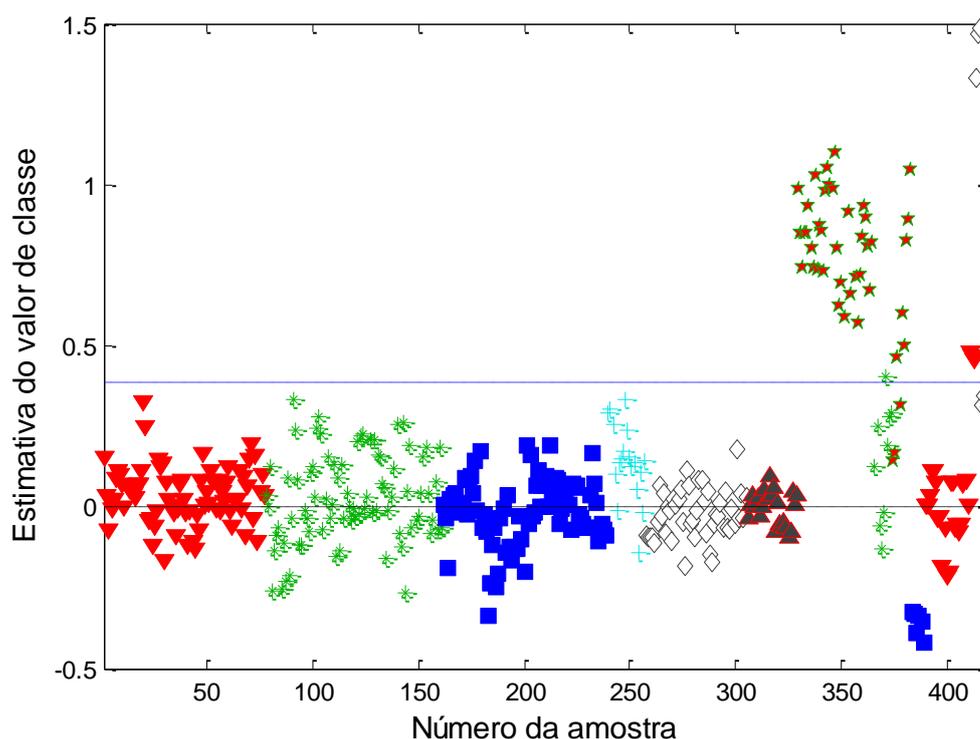


Figura 40. Resultados obtidos para modelo de discriminação entre amostras de Old Parr (★) provenientes da apreensão da Polícia Federal.

Ao se comparar os resultados destas predições com as da etapa de validação, notou-se uma grande discrepância entre a capacidade do modelo em identificar as amostras destes dois conjuntos como pertencentes às marcas da calibração. Não havia dúvidas de que as amostras da apreensão eram autênticas, com exceção das amostras falsas de Red Label, e não foram encontrados indícios de que alguma variação no equipamento causou diferenças nos espectros com relação aos do conjunto de validação que justificasse este comportamento dos modelos de discriminação. A provável causa da menor eficiência do método observada para o caso real foi que o lote das amostras da

apreensão apresenta diferenças sutis na composição química dos uísques em relação aos lotes usados na calibração, a ponto de o PLS conseguir identificar. Essa hipótese se fundamenta no fato de a composição química do uísque ser bastante complexa e depender de inúmeros fatores. A etapa de envelhecimento é de grande relevância para a composição final da bebida, porém é propícia a gerar variações entre lotes. Outro fato que pode contribuir com a alteração na composição da bebida é a reutilização dos barris. Ainda de acordo com Dragone et al¹⁰, nos Estados Unidos é exigido por lei que os barris sejam utilizados apenas uma vez para maturação dos uísques, exceto os de milho. Assim, após a primeira maturação, os barris são desmontados e enviados para outros países produtores, como Escócia e Japão, para serem reutilizados. Nestes casos ainda ocorrem a substituição de tábuas danificadas, o que também pode contribuir para alteração na composição final.

3.4 CONCLUSÕES DO CAPÍTULO

Os modelos apresentaram resultados satisfatórios, mostrando grande eficiência na discriminação de marcas de uísque através de espectros de UV-Vis das bebidas, obtidos sem nenhum preparo de amostra. Desde a calibração os modelos se mostraram promissores, pois apresentaram pequenas quantidades de *outliers*, com um número máximo de 6 *outliers* (espectros, independentes do número de replicatas de cada amostra) excluídos no modelo do Old Parr. Outro fato importante é a não ocorrência de sobreajuste, uma vez que os valores de RMSEC e RMSECV foram próximos.

A predição de amostras autênticas apresentou ótimos resultados. Houve a ocorrência de apenas 2 erros de classificação de falsos positivos, um para a marca Ballantine's e outro para a marca Old Parr. Os modelos apresentaram elevadas taxas de eficiência e classificação correta, sendo a menor delas de 98,6%. Os resultados para as marcas que não foram incluídas na calibração também foram muito satisfatórios, visto que não houve nenhum erro de classificação destas amostras.

A predição de amostras falsas também apresentou resultados satisfatórios, mostrando-se eficiente na identificação de amostras falsas, com taxa de

classificação correta de 93% e superior. Apenas 8 erros de classificação falsos positivos ocorreram dentre todos os modelos.

A aplicação dos modelos em um caso real não apresentou resultados tão bons quanto na predição de amostras autênticas. Além dos erros de classificação falsos positivos e negativos terem sido altos em relação a pequena quantidade de amostras, a ocorrência de classificação como não pertencentes às classes modeladas ou não autênticas foi muito alta, de 37,5% para Black Label e Old Parr e cerca de 50% para Red Label e White Horse. Logo, acredita-se que este resultado seja devido a variações de composição entre diferentes lotes de bebidas de uma mesma marca.

CAPÍTULO 4
Conclusões Gerais

4. CONCLUSÕES GERAIS

A espectroscopia UV-Vis combinada à análise discriminante por mínimos quadrados parciais (PLS-DA) mostrou-se eficiente para a discriminação de sete marcas de uísque, com espectros adquiridos utilizando-se reduzido volume da bebida e nenhum preparo de amostra.

A análise de amostras autênticas apresentou ótimos resultados, com altas taxas de classificação correta. As marcas da bebida que não foram incluídas na calibração apresentaram comportamento anômalo das demais marcas, ocorrendo apenas 1 erro de classificação destas amostras. Da mesma forma, a análise de amostras falsas foi satisfatória, mostrando que o modelo é eficiente neste tipo de discriminação.

Ao aplicar os modelos em casos reais, os resultados não foram tão promissores, apresentando altas taxas de erros. Uma recalibração dos modelos foi executada a fim de identificar se o problema constava em variações instrumentais, porém sem efeitos. Portanto, sugere-se que esses resultados foram causados pelas variações na composição da bebida em diferentes lotes de uma mesma marca, devido a diversos fatores das etapas de produção. Desta forma, para se aplicar esses modelos em casos reais, é indicado que o método inclua um número maior de lotes de forma a minimizar a chance de ocorrer classificações como não pertencentes às classes modeladas ou não autênticas. Além disso, devem ser feitas recalibrações dos modelos com inserção de amostras de lotes diferentes das que já contém no conjunto de calibração.

CAPÍTULO 5
Referências Bibliográficas

5. REFERÊNCIAS BIBLIOGRÁFICAS

1. Medeiros, L. A. de *A CPI da pirataria: os segredos do contrabando e da falsificação no Brasil*; 1st ed.; Geração Editorial: São Paulo, 2005.
2. Júnior, E. E. de M. ;*Rev. Jurídica* **2013**, 1, 21.
3. Strehlau, S. O luxo falsificado e suas formas de consumo. 2004. 307 p. Tese (Doutorado em Administração). Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas.São Paulo, 2014. **2014**.
4. Talhavini, M.; da Veiga, M. A. M. S. In *Análise de falsificação de bebidas: o exemplo de uísques*; Bruni, A. T.; Velho, J. A.; de Oliveira, M. F., Eds.; Millenium editora: Campinas SP, 2012; cap. 7.
5. Brasil *Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 15 de 31 de março de 2011.*; Diário Oficial da União, seção I, 1 de abril de 2011, p. 4, 2011.
6. <http://www.abrabe.org.br>. Acesso em maio de 2015.
7. Roche, M. de la *Um brinde à vida: a história das bebidas*. Publicado no site www.abrabe.org.br. Acesso em maio de 2015.
8. Hoffman, M. A. *Whiskey: a fascinating journey through the most famous whiskey and distilleries worldwide*; Parragon: Bath, UK, 2008.
9. *BRASIL. Decreto nº 6.871, de 4 de junho de 2009, do Ministério da Agricultura, Pecuária e Abastecimento.*; Diário Oficial da União, seção I, 4 de junho de 2009, p 20.
10. Dragone, G.; Flórez, M. D. G.; García, M. A. V.; Almeida e Silva, J. B. In *Bebidas alcoólicas: ciência e tecnologia*; Venturini Filho, V. G., Ed.; Edgard Blucher: São Paulo, 2010. cap. 20.
11. Dolan, C. S. In *Whisky: Technology, Production and Marketing*; Russell, I., Ed.; Academic press: London, 2003. cap. 2.
12. Piggott, J. R.; Conner, J. M. In *Fermented beverage production*; Lea, A. G. H. Piggot, J. R., Ed.; Springer Science + Business Nedia: New York, USA, 2003; cap. 11.
13. Conner, J.; Reid, K.; Jack, F. In *Whisky: Technology, Production and Marketing*; Russell, I.; Stewart, G.; Bamforth, C., Eds.; Elsevier, 2003; cap. 7.

14. Aylott, R. In *Whisky: Technology, Production and Marketing*; Russel, I.; Stewart, G.; Bamforth, C., Eds.; Academic Press: London, 2003; cap. 2.
15. Almeida, F. L. A. Caracterização e diferenciação de uísque mediante análise de componentes minoritários. Monografia. (Curso Superior de Polícia e especialização em Gestão de Políticas de Segurança Pública). Academia Nacional de Polícia e Fundação Getúlio Vargas. 53p. Brasília, 2005.
16. Boscolo, M.; Andrade-Sobrinho, L. G.; Lima-Neto, B. S.; Franco, D. W.; Ferreira, M. M. C. ; *J. AOAC Int.* **2002**, *85*, 744.
17. Fitzgerald, G.; James, K. J.; MacNamara, K.; Stack, M. a ; *J. Chromatogr. A* **2000**, *896*, 351.
18. Ng, L.-K.; *Anal. Chim. Acta* **2002**, *465*, 309.
19. Câmara, J. S.; Marques, J. C.; Perestrelo, R. M.; Rodrigues, F.; Oliveira, L.; Andrade, P.; Caldeira, M.; *J. Chromatogr. A* **2007**, *1150*, 198.
20. Lee, K.-Y. M.; Paterson, A.; Birkmyre, L.; Piggott, J. R.; *J. Inst. Brew.* **2001**, *107*, 315.
21. Demyttenaere, J. C. R.; Sánchez Martínez, J. I.; Verhé, R.; Sandra, P.; De Kimpe, N.; *J. Chromatogr. A* **2003**, *985*, 221.
22. Aylott, R.; Mackenzie, W.; *J. Inst. Brew.* **2010**, *116*, 215.
23. Møller, J. K. S.; Catharino, R. R.; Eberlin, M. N.; *Analyst* **2005**, *130*, 890.
24. McIntyre, A. C.; Bilyk, M. L.; Nordon, A.; Colquhoun, G.; Littlejohn, D.; *Anal. Chim. Acta* **2011**, *690*, 228.
25. Mignani, A. G.; Ciaccheri, L.; Gordillo, B.; Mencaglia, A. A.; González-Miret, M. L.; Heredia, F. J.; Culshaw, B.; *Sensors Actuators, B Chem.* **2012**, *171-172*, 458.
26. Garcia, J. S.; Vaz, B. G.; Corilo, Y. E.; Ramires, C. F.; Saraiva, S. a.; Sanvido, G. B.; Schmidt, E. M.; Maia, D. R. J.; Cosso, R. G.; Zacca, J. J.; Eberlin, M. N.; *Food Res. Int.* **2013**, *51*, 98.
27. Cantarelli, M. A.; Azcarate, S. M.; Savio, M.; Al, E.; *Food Anal. Methods* **2014**, agosto.
28. Mackenzie, W. M.; Aylott, R. I.; *Analyst* **2004**, *129*, 607.
29. Lehtonen, P. J.; Keller, L. A.; Ali-Mattila, E. T.; *Z Leb. Unters Forsch* **1999**, *208*, 413.

30. Skoog, D. A.; Holler, F. J.; Nieman, T. A. *Princípios de Análise Instrumental*; Bookman: Porto Alegre, 2002.
31. Pavia, D. L.; Lampman, G. M.; Kriz, G. S.; Vyvyan, J. R. *Introdução à Espectroscopia*; tradução d.; Cengage Learning: São Paulo, 2010.
32. Assis, O. B. G.; *Rev. Bras. Ensino Física* **2013**, 35, 2301.
33. Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Chemometrics: a textbook*, 1st ed.; Elsevier: Amsterdam, 2003.
34. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.; *Quim. Nova* **1999**, 22, 724.
35. Chau, F.-T.; Liang, Y.-Z.; Gao, J.; Shao, X.-G. *Chemometrics: From Basics to Wavelet Transform*; Winefordner, J. D., Ed.; John Wiley & Sons: Hoboken, New Jersey, 2004; Vol. 164.
36. Neto, B. de B.; Scarminio, I. S.; Bruns, R. E.; *Quim. Nova* **2006**, 29, 1401.
37. Brereton, R. G.; *J. Chemom.* **2014**, 28, 749.
38. Scarminio, I. S. *Análises Espectrométricas de Águas Minerais e suas Classificações por Meio de Reconhecimento Padrão*; Universidade Estadual de Campinas: Tese de Mestrado, 1981.
39. Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons: Chichester, UK, 2003.
40. Thomas, E. V.; *Anal. Chem.* **1994**, 66, 795A.
41. Naes, T.; Isaksson, T.; Feran, T.; Davies, T. *A user-friendly guide to Multivariate Calibration and Classification*; NIR Publications: Chichester, UK, 2002.
42. Esbensen, K. H. *Multivariate Data Analysis In Practice*; 5th ed.; CAMO Software: Woodbridge, USA, 2009.
43. Wold, S.; Esbensen, K.; Geladi, P.; *Chemom. Intell. Lab. Syst.* **1987**, 2, 37.
44. Borin, A.; Poppi, R. J.; *J. Braz. Chem. Soc.* **2004**, 15, 570.
45. Trygg, J. ;*Homepage Chemom.* **2003**, setembro.
46. Valderrama, L.; Gonçalves, R. P.; Março, P. H.; Valderrama, P.; *Rev. Bras. Pesqui. em Aliment.* **2014**, 5, 32.
47. Varmuza, K.; Filzmoser, P. *Introduction to multivariate statistical analysis in chemometrics*; CRC Press: Boca Raton USA, 2009.

48. Kalivas, J. H.; Gemperline, P. J. In *Practical guide to chemometrics*; Gemperline, P., Ed.; CRC Press: Boca Raton USA, 2006; pp. 105–160.
49. Wold, S.; Sjöström, M.; Eriksson, L.; *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109.
50. Miller, J. N.; Miller, J. C. *Estadística y quimiometría para química analítica*; 4th ed.; Prentice Hall: Madrid, 2002.
51. Brereton, R. G.; *Analyst* **2000**, *125*, 2125.
52. Gemperline, P. *Practical guide to chemometrics*; Gemperline, P., Ed.; 2nd ed.; Taylor & Francis: Boca Raton, USA, 2006.
53. Braga, J. W. B.; Junior, A. A. D. S.; Martins, I. S. ; *Fuel* **2014**, *120*, 171.
54. Wise, B. M.; Gallagher, N. B.; Bro, R.; Shaver, J. M.; Windig, W.; Koch, R. S. *PLS_Toolbox 4.0: for use with MATLAB*; Eigenvector research: Wenatchee, USA, 2006.
55. Sun, D.-W. *Infrared spectroscopy for food quality analysis and control*; Academic press: New York, USA, 2009.
56. Hobro, A. J.; Kuligowski, J.; Döll, M.; Lendl, B.; *Anal. Bioanal. Chem.* **2010**, *398*, 2713.
57. Savorani, F.; Rasmussen, M. A.; Rinan, A.; Engelsen, B. In *Data handling in science and technology: Chemometrics in Food Chemistry*; Marini, F., Ed.; Elsevier B.V.: Amsterdam, 2013; p. 498.
58. Bergo, M. C. J. *Transferência de calibração na discriminação de mogno e espécies semelhantes utilizando NIRS e PLS-DA*; Universidade de Brasília: Dissertação de Mestrado, 2014.
59. Valderrama, P.; Braga, J. W. B.; Poppi, R. J.; *Quim. Nova* **2009**, *32*, 1278.
60. Trullols, E.; Ruisánchez, I.; Rius, F. X. ; *TrAC - Trends Anal. Chem.* **2004**, *23*, 137.
61. Neto, B. B.; Scarminio, I. S.; Bruns, R. E. *Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria*; 4th ed.; Bookman: Porto Alegre, 2010.
62. Silva, A. A.; Nascimento, E. S. P.; Cardoso, D. R.; Franco, D. W. ; *Quim. Nova* **2012**, *35*, 563.