Universidade do Porto
Faculdade de Engenharia
FEUP

Helena Maria Pereira Pinto Dourado e Alvelos

Análise, Desenvolvimento e Teste de Métodos e Técnicas para Controlo Estatístico em Análise Sensorial

Porto, 2002

Helena Maria Pereira Pinto Dourado e Alvelos

Análise, Desenvolvimento e Teste de Métodos e Técnicas para Controlo Estatístico em Análise Sensorial

Tese apresentada à Faculdade de Engenharia da Universidade do Porto
para a obtenção do grau de Doutor em Ciências de Engenharia,
realizada sob a orientação científica do
Professor Doutor José António Sarsfield Pereira Cabral,
Professor Associado da Faculdade de Engenharia da Universidade do Porto

Departamento de Engenharia Mecânica e Gestão Industrial
Faculdade de Engenharia da Universidade do Porto
2002

UNIVERSIDADE DO PORTO
Faculdade de Engenharia
BIBLIOTECA
N.º 65099
CDU
Data 26 / 11 / 2002

Os trabalhos de investigação apresentados nesta tese foram apoiados pelo Programa PRODEP / FSE

A Meus Pais

Ao Ilídio

À Leonor

Agradecimentos

Ao meu orientador, Professor Doutor José António Sarsfield Pereira Cabral, dirijo um agradecimento muito especial pelo seu apoio e incentivo. As suas preciosas sugestões e críticas, a sua disponibilidade e a forma amiga como sempre me acompanhou foram imprescindíveis no desenvolvimento deste trabalho.

Agradeço ainda:

Ao Instituto do Vinho do Porto, à sua Direcção e aos seus Colaboradores, pela oportunidade única que me proporcionaram de contactar com a realidade prática da Análise Sensorial e que permitiu um importante enriquecimento do presente trabalho.

Ao Sr. Dr. Paulo Barros, por ter possibilitado que este trabalho se realizasse em colaboração com o Instituto do Vinho do Porto e pela confiança que demonstrou ao longo da sua progressão.

Ao Sr. Eng. Bento Amaral, pelo interesse e atenção que dedicou a este estudo e pela disponibilidade com que sempre me acolheu no Instituto do Vinho do Porto.

Ao Departamento de Economia, Gestão e Engenharia Industrial da Universidade de Aveiro, na pessoa de todos os Presidentes dos Conselhos Directivos que exerceram funções durante a elaboração desta tese, pelas condições de trabalho e pelas facilidades que me foram concedidas e que permitiram essa realização.

Ao Professor Doutor Pedro Vilarinho, coordenador da área de Gestão das Operações e responsável pela minha Formação e Orientação no Departamento de Economia, Gestão e Engenharia Industrial da Universidade de Aveiro, pela confiança e pelo interesse que sempre manifestou.

Aos meus colegas do DEGEI, pelo apoio e amizade que demonstraram.

Ao Ilídio, à minha Mãe e aos meus Irmãos, que sempre me apoiaram e souberam compreender ausências.

Resumo

A Análise Sensorial é a disciplina que estuda as reacções das pessoas às características dos produtos tal como são percepcionadas pelos sentidos humanos. Os Painéis de Provadores são um dos seus pilares fundamentais, uma vez que é a eles que cabe pronunciarem-se sobre a qualidade dos produtos do ponto de vista sensorial.

Embora o papel dos Painéis de Provadores seja unanimemente reconhecido como sendo determinante em Análise Sensorial, o facto é que os métodos para avaliar e controlar o seu desempenho são escassos e insuficientes. Verifica-se, portanto, a existência de um certo desequilíbrio entre a responsabilidade que é exigida aos provadores quando avaliam os produtos e os meios para verificar se a avaliação foi bem efectuada.

O objectivo principal desta tese é desenvolver e testar métodos que possam ser utilizados, na prática, para avaliar e controlar os julgamentos e as decisões dos Painéis de Provadores.

Após uma análise às técnicas tradicionalmente utilizadas em Controlo Estatístico de Processos, concluiu-se que estas não são fácil nem directamente aplicáveis ao problema em estudo. Assim, e dada a complexidade e a especificidade de toda esta temática, a abordagem proposta recorre a um conjunto de procedimentos que se traduzem nas quatro vertentes seguintes:

(i) caracterização do processo de decisão de cada provador e controlo da sua estabilidade;

(ii) avaliação diária do desempenho de cada provador em termos de reprodutibilidade dos julgamentos;

(iii) avaliação periódica do desempenho global de cada provador;

(iv) aferição periódica de capacidades específicas dos provadores.

A validade e a qualidade das decisões do Painel no seu conjunto são dependentes da validade e da qualidade das decisões de cada um dos seus membros. Desta forma, é finalmente proposto um procedimento para o estabelecimento da decisão final do Painel que integra os resultados demonstrados por cada provador na respectiva avaliação.

Muitos dos métodos propostos foram testados e implementados na Câmara de Provadores do Instituto do Vinho do Porto, tendo revelado resultados prometedores. Salienta-se que no Instituto do Vinho do Porto a situação se reveste de uma complexidade invulgar, pelo que não se prevêem problemas adicionais ao pretender-se, eventualmente, utilizar os procedimentos propostos noutros contextos.

Abstract

Tasting Panels play a crucial role in Sensory Evaluation, since they are responsible for judging the quality of products as perceived through human senses.

Although the organoleptic evaluation of products has a great deal of importance, scientific and practical methods for the control of Tasting Panels decisions are still insufficient. In fact, there is a gap between Tasting Panels responsibility in judging products quality and the means to control their judgments.

The main goal of this thesis is to develop and test methods and procedures for the assessment and control of Tasting Panels decisions, given that the traditional Statistical Process Control techniques are not easily nor directly applied to the Sensory Evaluation processes.

The proposed approach addresses the following four topics:

(i) how to monitor and control the panelists decision process over time;

(ii) how to measure the reproducibility of the panelists perceptions on a daily basis;

(iii) how to assess the global performance of panelists over a given period of time;

(iv) how to evaluate and test periodically the panelists specific skills.

Based on the results obtained for each panelist, a new procedure for establishing the Tasting Panel final decision is proposed.

Most of the methods and procedures developed were tested and implemented within the Tasting Panel of the Port Wine Institute, and revealed promising results.

The case of the Port Wine Institute is a very complex one, essentially due to the great variety of Port types. Therefore, no major problems are expected in using the referred procedures in other contexts.

Résumé

L'Analyse Sensorielle est la discipline où sont étudiées les réactions des individus aux caractéristiques des produits tel qu'elles sont perçues par les sens humains. Les Chambres de Dégustateurs sont l'un de ses piliers fondamentaux, vu que c'est à eux qu'est attribué le jugement sur la qualité sensorielle des produits.

Quoique le rôle des Chambres de Dégustateurs soit unanimement reconnu comme étant déterminant dans le domaine de l'Analyse Sensorielle, les méthodes d'évaluation et de contrôle de leur performance sont rares et insuffisantes. On peut donc vérifier l'existence d'un certain déséquilibre entre la responsabilité que l'on exige aux dégustateurs quand ils évaluent les produits et les moyens pour vérifier si cette évaluation a été bien réalisée.

L'objectif principal de cette thèse est de développer et tester des méthodes qui puissent être utilisées dans la pratique pour évaluer et contrôler les jugements et les décisions des Chambres de Dégustateurs.

Après une analyse des techniques utilisées traditionnellement dans le Contrôle Statistique des Processus, on a conclu que celles-ci ne sont pas facilement ni directement applicables au problème en étude. Donc, et en tenant compte de la complexité e de la spécificité de cette thématique, l'abordage proposé intègre un ensemble de procédures qui se traduisent dans les quatres actions suivantes:

(i) caractérisation du processus de décision de chaque dégustateur e le contrôle de sa stabilité;

(ii) évaluation journalière de la performance de chaque dégustateur du point de vue de la reproductibilité de ses jugements;

(iii) évaluation périodique de la performance globale de chaque dégustateur;

(iv) test périodique des capacités spécifiques des dégustateurs.

La validité et la qualité des décisions de la Chambre de Dégustateurs dans son ensemble, dépend de la validité et de la qualité des décisions de chacun de ses membres. Dans ce contexte, on propose finalement une procédure pour le support à la décision finale de la Chambre de Dégustateurs, qui intègre les résultats démontrés par chaque dégustateur dans l'évaluation respective.

Une partie des méthodes proposées a été testée et implantée dans la Chambre de Dégustateurs de l'Institut du Vin de Porto, avec des résultats prometteurs. Il faut souligner qu'à l' Institut du Vin de Porto la situation est vraiment complexe. On ne s'attend donc pas à des problèmes additionnels lors du transfer des procédures utilisées vers un autre contexte.

Índice

1 Introdução Geral 1

1.1 Objectivo da Tese 2

1.2 Relevância do Problema 2

1.3 Estrutura da Tese 3

2 A Análise Sensorial 6

2.1 Introdução 7

2.2 Definição e Objectivos da Análise Sensorial 9

2.3 As Características Organolépticas 12

2.3.1 Sensação e Percepção 12

2.3.2 Aspecto 16

2.3.3 Odor 17

2.3.4 Consistência e Textura 17

2.3.5 Sabor 18

2.3.6 Factores que Influenciam a Percepção Sensorial 19

2.3.6.1 Factores Fisiológicos 20

2.3.6.2 Factores Psicológicos 20

2.3.6.3 Enviezamento das Respostas 22

2.4 Métodos e Técnicas Utilizados em Avaliação Sensorial 24

2.4.1 Ensaios de Diferenciação Geral 26

2.4.1.1 Teste Triangular 26

2.4.1.2 Teste Dois em Cinco 27

2.4.1.3 Teste Duo-Trio 27

2.4.1.4 Teste da Diferença Simples 28

2.4.1.5 Teste A-Não A 28

2.4.1.6 Teste ABX 29

2.4.1.7 Teste do Padrão Duplo (Dual Standard) 29

2.4.1.8 Teste da Diferença em Relação ao Controlo 29

2.4.1.9 Testes Sequenciais 30

2.4.2 Ensaios de Diferenciação de Atributos 31

2.4.2.1 Comparação por Pares Direccional (2-AFC) 31

2.4.2.2 Comparação Direccional de Três Amostras (3-AFC) 32
2.4.2.3 Teste de Ordenação por Pares (Análise de Friedman) 32
2.4.2.4 Teste de Comparação por Pares de Scheffé 32
2.4.2.5 Teste de Ordenação Simples 33
2.4.2.6 Teste de Cotação de Amostras 33
2.4.2.7 Testes que Utilizam Desenhos de Experiências 34
2.4.3 Ensaios para Determinação de Limiares 34
2.4.4 Ensaios Descritivos 36
2.4.4.1 Perfil de Sabor 37
2.4.4.2 Perfil de Textura 38
2.4.4.3 Análise Descritiva Quantitativa 39
2.4.4.4 Método Spectrum 40
2.4.4.5 Perfil de Escolha Livre 41
2.4.4.6 Análise Descritiva de Tempo-Intensidade 41
2.4.5 Ensaios Afectivos 43
2.4.5.1 Testes de Preferência 43
2.4.5.2 Testes de Aceitação 44

3 A Análise dos Resultados e a Organização dos Ensaios em Análise Sensorial . 45
3.1 Introdução 46
3.2 Análise Estatística dos Resultados dos Ensaios de Avaliação Sensorial 46
3.2.1 Ensaios de Diferenciação e de Preferência 50
3.2.1.1 Testes de Hipóteses à Proporção Binomial 51
3.2.1.2 Abordagem da Teoria da Detecção do Sinal 54
3.2.1.3 Técnicas Estatísticas de Análise de Dados Expressos em Escalas Nominais ou Ordinais 73
3.2.1.4 Comparação dos Vários Ensaios de Diferenciação 73
3.2.2 Ensaios para Determinação de Limiares 80
3.2.3 Ensaios Descritivos e de Aceitação 80
3.2.3.1 Técnicas de Estatística Univariada – Análise de Variância .. 81
3.2.3.2 Técnicas de Estatística Multivariada 84
3.3 Organização de um Programa de Análise Sensorial 90
3.3.1 Condições Físicas 91
3.3.2 Efeitos da Ordenação e do Número de Amostras 92
3.4 Selecção e Treino de um Painel de Provadores Profissional 94

3.4.1 Selecção de Provadores 94
3.4.2 Treino de Provadores 97
3.4.2.1 Exactidão, Repetibilidade, Reprodutibilidade e Calibração . 97
3.4.2.2 Fases do Treino de Provadores 99

4 Métodos de Controlo de Processos 102
4.1 Introdução 103
4.2 Instrumentos de Medida e sua Calibração 105
4.2.1 Calibração e Incertezas 106
4.2.2 Calibração para Grandezas Expressas por Variáveis Discretas 108
4.3 Cartas de Controlo 109
4.3.1 Carta de médias 112
4.3.2 Carta p 114
4.3.3 Carta CUSUM 116
4.3.4 Medidas da Sensibilidade das Cartas de Controlo 117
4.3.5 Outras Cartas de Controlo 119
4.4 Utilização do Controlo Estatístico em Processos Não Industriais 123
4.4.1 Simulação da Calibração de uma Grandeza Expressa numa Escala de Cotação 128
4.4.2 Simulação de um Processo Multinomial e Seu Controlo Estatístico . 132
4.4.2.1 As Cartas de Controlo $\overline{X}_S$, p modificada e χ^2 133
4.4.2.2 Comparação dos Resultados Obtidos com as Cartas χ^2 e $\overline{X}_P$ para Alguns Processos Sob Controlo 138
4.4.2.3 Comparação dos Resultados Obtidos com as Cartas χ^2 e $\overline{X}_P$ para Alguns Processos Fora de Controlo 139
4.4.3 Em Conclusão 145

5 O Problema do Controlo de Câmaras de Provadores 147
5.1 Introdução 148
5.2 O Caso do Instituto do Vinho do Porto 149
5.2.1 O Vinho do Porto 150
5.2.2 O Papel da Câmara de Provadores 155
5.2.3 Descrição e Análise da Situação Inicial 156
5.2.3.1 Ficha de Prova Inicial 157
5.2.3.2 Metodologia Seguida na Análise dos Dados da Ficha de Prova Inicial 159

5.2.3.3 Resultados Obtidos na Análise Inicial 160

5.3 A Complexidade do Problema 174

6 Avaliação e Controlo dos Julgamentos de Câmaras de Provadores 176

6.1 Introdução 177

6.2 Estudo da Ficha de Prova 179

6.3 Caracterização do Comportamento dos Provadores e Controlo da Sua Estabilidade 181

6.3.1 Modelo para o Processo de Decisão Individual dos Provadores 181

6.3.1.1 Regressão Logística 182

6.3.1.2 Modelos Obtidos para o Processo de Decisão dos Provadores do IVP 184

6.3.2 Controlo da Estabilidade do Processo de Decisão dos Provadores 188

6.3.2.1 Utilização da Carta CUSUM no Controlo da Estabilidade do Processo de Decisão dos Provadores do IVP 192

6.4 Avaliação do Desempenho Diário dos Provadores 195

6.4.1 Metodologia Proposta para a Avaliação da Reprodutibilidade dos Provadores Inseridos num Painel 195

6.4.1.1 Índice de Reprodutibilidade (RI) 196

6.4.1.2 Índice de Estratégia Defensiva (DI) 197

6.4.1.3 Índice de Fiabilidade 198

6.4.2 Aplicação da Metodologia Proposta ao Caso do Painel de Provadores do IVP 199

6.5 Avaliação Periódica do Desempenho Global dos Provadores 205

6.5.1 Avaliação Periódica do Desempenho Global do Provadores do IVP 205

6.6 Aferição Periódica de Capacidades Específicas dos Provadores 214

6.6.1 Caracterização e Distinção entre Alguns Tipos de Vinho do Porto ... 216

6.6.1.1 Estudo das Análises Químicas e de Cor 217

6.6.1.2 Estudo dos Resultados Obtidos com o Nariz Electrónico ... 218

6.6.1.3 Estudo das Análises Químicas, de Cor e dos Sensores em Conjunto 220

6.6.1.4 Caracterização da Idade do Vinho do Porto através de Análises Químicas 223

6.6.2 Aferição da Capacidade dos Provadores do IVP para a Detecção de Algumas Características do Vinho do Porto 224

6.7 Decisão Final do Júri de Prova 226

7 Conclusões 229

7.1 Conclusões Gerais 230

7.2 Perspectivas de Desenvolvimento Futuro 232

Referências Bibliográficas 234

Apêndice A - Ficha de Prova Utilizada no IVP no Início do Estudo 250

Apêndice B - Tabelas para os Parâmetros H e F da Versão Tabelada das Cartas CUSUM 252

Índice de Figuras

Figura 2.1 Esquema do mecanismo dos sentidos humanos 13

Figura 2.2 Função de detecção ideal e função psicométrica 15

Figura 2.3 Diagrama que representa o enviezamento das respostas para um exemplo de biscoitos doces e não doces 23

Figura 2.4 Exemplo de um diagrama "teia de aranha" para os resultados obtidos com dois sabonetes 40

Figura 2.5 Exemplo de uma curva tempo-intensidade 42

Figura 3.1 A lógica dos testes de hipóteses aplicada a um caso em Análise Sensorial 49

Figura 3.2 Ilustração do teste de hipóteses à proporção binomial (unilateral à direita), amostras grandes 53

Figura 3.3 Ilustração dos conceitos de sensibilidade e de critério 56

Figura 3.4 Distribuições de frequências que representam as sensações (N e S) criadas por dois estímulos sujeitos a confusão, segundo o modelo de Thurstone 58

Figura 3.5 Distribuições de frequências que representam as sensações N, S e S' e ilustram a diferença de sensibilidades entre dois provadores 59

Figura 3.6 Tabelas de respostas que ilustram o procedimento de escalas de "cotação" na teoria de detecção do sinal 62

Figura 3.7 Curva ROC desenhada a partir dos dados da figura 3.6 63

Figura 3.8 Teste de hipóteses ao valor esperado de uma distribuição normal 65

Figura 3.9 Interpretação do paradoxo *discriminatory nondiscriminators* segundo a teoria de Thurstone para os testes triangular e 3-AFC 74

Figura 4.1 Ilustração do aspecto de uma carta de controlo clássica 111

Figura 4.2 Carta de controlo de médias - relação entre o teste de hipóteses e a carta de controlo 113

Figura 4.3 Curva Característica Operacional (OCC) para uma carta de controlo de médias com $N = 4$ 118

Figura 4.4 Cartas de Controlo para Atributos 120

Figura 4.5 Cartas de Controlo para Variáveis 122

Figura 4.6 Cartas de Controlo para Duas ou Mais Variáveis 123

Figura 4.7 Comparação entre a função densidade de probabilidade χ^2_4 e os histogramas da estatística Q obtidos para quatro distribuições multinomiais diferentes 137

Figura 4.8 Risco β em função da distribuição fora de controlo (1º cenário) 140

Figura 4.9 Risco β em função da distribuição fora de controlo (2º cenário) 142

Figura 4.10 Histograma da estatística de teste ($\overline{X}$) para as distribuições nºs 1 e 2 da tabela 4.7 143

Figura 4.11 Histograma da estatística de teste (Q) para as distribuições nºs 3 e 8 da tabela 4.7 e função densidade de probabilidade χ^2_4 sobreposta 144

Figura 5.1 Processo produtivo do Vinho do Porto 151

Figura 5.2 Histogramas da pontuação atribuída, por cada provador, à característica Qualidade 161

Figura 5.3 Valores médios da pontuação atribuída à característica Qualidade em função do provador, para cada tipo de vinho 162

Figura 5.4 Valores médios da pontuação atribuída à característica Qualidade em função do tipo de vinho, para cada provador 163

Figura 5.5 Valores da variância da pontuação atribuída à característica Qualidade em função do provador, para cada tipo de vinho 164

Figura 5.6 Valores da variância da pontuação atribuída à característica Qualidade em função do tipo de vinho, para cada provador 165

Figura 5.7 Diagramas das classificações atribuídas à variável Qualidade pelos pares de provadores (D,E) e (B,D) 168

Figura 5.8 Diagramas das classificações atribuídas à variável Limpidez pelos pares de provadores (B,E) e (D,F) 169

Figura 5.9 Valores da proporção de vinhos aprovados em função do provador, para cada tipo de vinho 170

Figura 5.10 Valores da proporção de vinhos aprovados em função do tipo de vinho, para cada provador 171

Figura 5.11 Diagramas de concordância entre os provadores (B,E) e (B,D), relativamente à aprovação ou não aprovação dos vinhos 172

Figura 6.1 Tabelas de classificação de casos 187

Figura 6.2 Esquema do procedimento a seguir para construir tabelas CUSUM para a proporção binomial 191

Figura 6.3 Exemplos de tabelas CUSUM para alguns provadores e para vinhos com indicação de idade 193

Figura 6.4 Exemplos de tabelas CUSUM para alguns provadores e para vinhos sem indicação de idade 194

Figura 6.5 Histogramas de frequências relativas do módulo das diferenças das classificações atribuídas às réplicas pelos provadores G, H, I e K 200

Figura 6.6 Histogramas de frequências relativas das classificações atribuídas à Designação pelos provadores G, H, I e K e pela Câmara no seu conjunto 201

Figura 6.7 Evolução do índice *RI* para os provadores G, H, I e K, ao longo de 137 dias 203

Figura 6.8 Evolução do desvio padrão para a Câmara e para os provadores G, H, I e K, ao longo de 137 dias 203

Figura 6.9 Evolução do índice *DI* para os provadores G, H, I e K, ao longo de 137 dias 204

Figura 6.10 Evolução do índice *I* para os provadores G, H, I e K, ao longo de 137 dias 204

Figura 6.11 Classificações atribuídas por cada provador às diversas amostras de vinhos "maus" e "bons" provadas ao longo de um ano 207

Figura 6.12 Coordenadas dos pontos das curvas ROC calculadas a partir das tabelas da figura 6.11 208

Figura 6.13 Curvas ROC para quatro dos provadores do IVP 209

Figura 6.14 Representação gráfica das pontuações obtidas por cada amostra no 1º e 2º componentes principais da ACP_1 218

Figura 6.15 Representação gráfica das pontuações obtidas por cada amostra no 1º e 2º componentes principais da ACP_2 220

Figura 6.16 Representação gráfica das pontuações obtidas por cada amostra no 1º e 2º componentes principais da ACP_3 222

Figura 6.17 Representação gráfica dos pesos das variáveis consideradas na ACP_3 em cada um dos dois componentes principais 223

Índice de Tabelas

Tabela 4.1 Frequências das respostas do provador para os vários padrões 28

Tabela 4.2 Probabilidades estimadas de o produto ser de classe $X = x$ 29

Tabela 4.3 Probabilidades calculadas de o provador atribuir a classificação $D = d$ 29

Tabela 4.4 Tabela de "calibração" do provador 30

Tabela 4.5 Intervalos de confiança a 95% para α, $\hat{\alpha}$ e ARL para as cartas $\overline{X}_P$ e χ^2 37

Tabela 4.6 Valores de $\hat{\beta}$ e ARL correspondentes ao 1º cenário de situações de fora de controlo 39

Tabela 4.7 Valores de $\hat{\beta}$ e ARL correspondentes ao 2º cenário de situações de fora de controlo 41

Tabela 5.1 Distribuição do número de provas pelos provadores e pelos tipos de vinho 13

Tabela 5.2 Correlações entre as classificações atribuídas pelos vários provadores 21

Tabela 6.1 Coeficientes da regressão logística múltipla obtidos para os vários provadores 11

Tabela 6.2 Cálculo dos vários componentes de I_k para quatro dos provadores 28

Tabela 6.3 Estatística de Wilcoxon e respectivo desvio padrão calculados para 4 dos provadores do IVP 38

Tabela 6.4 Valores da estatística de teste (z) para os testes à diferença entre as áreas debaixo das curvas ROC (W) considerando os vários pares de provadores 39

Tabela 6.5 Valores de W e de $\bar{I}$ para os 4 provadores considerados 40

Tabela 6.6 Valores simulados de I_k e de W_k para 5 provadores 53

Tabela 6.7 Valores de FD_A, de FD_R e da *Decisão Final* para as situações nas quais 3 dos 5 provadores aprovam a amostra 54

1

Introdução Geral

Conteúdo:

- **Objectivo da Tese**
- **Relevância do Problema**
- **Estrutura da Tese**

1.1 Objectivo da Tese

Em Análise Sensorial os Painéis de Provadores têm um papel determinante: são eles que se pronunciam sobre os produtos tal como são percepcionados pelos sentidos humanos. A função principal dos provadores profissionais é, portanto, a de avaliar as características sensoriais dos produtos. Neste sentido, os provadores podem ser encarados como "instrumentos de medida" dessas características.

Os instrumentos de medida convencionais são, normalmente, calibrados e controlados de forma a garantir que os resultados que produzem sejam fiáveis. A especificidade dos provadores na qualidade de instrumentos de medida não permite, no entanto, que seja possível aplicar-lhes, de uma forma directa, os mesmos métodos de calibração e controlo que se utilizam nos instrumentos de medida convencionais.

O problema que se estuda nesta tese é, precisamente, o da avaliação e do controlo dos julgamentos de provadores em Análise Sensorial e o seu principal objectivo é o de desenvolver e testar métodos de controlo de qualidade adequados aos processos inerentes à Análise Sensorial e, em particular, aos Painéis de Provadores.

1.2 Relevância do Problema

Em processos industriais, o controlo estatístico é tradicionalmente utilizado como forma de melhorar a qualidade da conformidade dos produtos fabricados. Assim, os parâmetros controlados referem-se, habitualmente, a propriedades físicas ou químicas desses produtos.

Há, no entanto, inúmeros produtos e serviços que são avaliados pelos seus consumidores (actuais e potenciais) com base em critérios que dificilmente se traduzem no tipo de características referido. O facto de não ser possível avaliar as características em causa através de instrumentos de medida convencionais devidamente calibrados, dificulta a abordagem estatística do seu controlo de qualidade.

Uma das áreas em que este problema se coloca com pertinência é a da Análise Sensorial. As decisões dos membros de um júri de prova não são, normalmente, avaliadas nem controladas, embora se reconheça que o papel de cada provador é o de

"instrumento de medida" de características sensoriais, tais como o aroma ou o sabor de um bem alimentar.

Pode mesmo dizer-se que, aparentemente, há uma contradição entre a responsabilidade que é exigida aos provadores quando avaliam os produtos e a falta de meios para verificar se a avaliação foi bem efectuada.

Este facto traduz-se em dificuldades para todas as partes: (i) para as empresas produtoras, que têm problemas com a formulação dos seus produtos; (ii) para as empresas responsáveis pela fiscalização, às quais é por vezes difícil justificar objectivamente algumas decisões menos favoráveis; (iii) para os consumidores, que adquirem produtos cujo controlo sensorial pode ser deficiente.

Neste contexto, a concepção e a utilização de procedimentos que permitam avaliar e controlar as decisões dos painéis de provadores são aspectos centrais para o desenvolvimento da área da Análise Sensorial.

Reconhecendo a relevância desta matéria, o Instituto do Vinho do Porto (IVP) decidiu estudar o problema da avaliação e do controlo dos membros da sua Câmara de Provadores. Nesse sentido, foi efectuado um protocolo com a Faculdade de Engenharia da Universidade do Porto, no qual se integrou este trabalho. O objectivo final do protocolo era o da acreditação, pelo Instituto Português da Qualidade, da Câmara de Provadores do Instituto do Vinho do Porto, o que aconteceu em finais de 1999.

1.3 Estrutura da Tese

A Análise Sensorial é uma área que se pode caracterizar como multidisciplinar, envolvendo conceitos da Psicologia, da Estatística e da Qualidade, entre outras. Por outro lado, o Controlo Estatístico de Processos é uma disciplina que é tradicionalmente utilizada nos processos produtivos convencionais, mas em relação à qual se vem fazendo um esforço de aplicação a outro tipo de processos, nomeadamente os relacionados com serviços.

O tema estudado neste trabalho implicou uma abordagem a três áreas que se podem considerar relativamente estanques. Em primeiro lugar, a área da Análise Sensorial de uma forma geral, a sua definição e objectivos e os ensaios mais utilizados. Em segundo lugar, o processamento dos resultados dos ensaios, que envolve teorias e técnicas essencialmente da área da Estatística. Finalmente, o Controlo de Processos. Esta área foi

incluída no trabalho com o objectivo de verificar até que ponto os métodos utilizados nos processos convencionais se poderiam adaptar aos processos inerentes a outro tipo de actividades, em particular, à avaliação sensorial.

O conteúdo dos capítulos 2, 3 e 4 reflecte, precisamente esta organização. No capítulo 2, é abordado o tema da Análise Sensorial de uma forma geral. Saliente-se que, neste capítulo, houve a preocupação de introduzir conceitos que podem parecer demasiado elementares, mas que são fundamentais para a compreensão do assunto que está a ser tratado. Grande parte do capítulo é, no entanto, dedicado à descrição dos ensaios sensoriais mais comuns.

No capítulo 3, estudam-se os métodos e técnicas utilizados para analisar os resultados dos ensaios apresentados no capítulo 2, após o que se analisam os problemas da organização das experiências sensoriais e da selecção e treino de provadores profissionais.

No capítulo 4, são abordados os métodos de controlo de processos. Numa primeira fase, estudam-se os métodos tal como são normalmente utilizados nos processos industriais ou produtivos. Numa segunda fase, analisa-se a possibilidade de aplicação de alguns dos conceitos e técnicas referidos a processos não industriais, onde se enquadram os processos inerentes à prova em Análise Sensorial. O facto de o controlo estatístico ser uma área quase inexplorada em Análise Sensorial, levou a que se analisassem experiências efectuadas noutras áreas, tais como a Saúde e o Ensino. Por exemplo, os problemas associados ao tratamento de inquéritos de opinião têm muito em comum com as questões envolvidas na análise de fichas de prova. Assim, nesta segunda fase estuda-se a utilização do controlo estatístico nos serviços de uma forma geral. Finalmente, e ainda no contexto dos processos não industriais, apresentam-se dois exemplos de possíveis aplicações práticas dos conceitos e das técnicas em análise a dados expressos em escalas de cotação (este exercício é feito com base em dados simulados).

No capítulo 5, é abordado o tema específico do controlo de câmaras de provadores, com especial ênfase no caso do Instituto do Vinho do Porto. Como se terá oportunidade de verificar ao longo deste capítulo e do seguinte, este caso revela-se particularmente complexo, essencialmente devido à diversidade de tipos de Vinho do Porto existente. Assim, no início do capítulo são descritos os diversos tipos de vinho do Porto e o papel da Câmara de Provadores do Instituto do Vinho do Porto, após o que é descrita e analisada a situação no início do projecto. No fim do capítulo, salienta-se a complexidade do problema do controlo de Câmaras de Provadores.

No capítulo 6, aborda-se a questão da qualidade das apreciações feitas por um júri de prova e desenvolvem-se métodos e técnicas para avaliar o seu desempenho. A avaliação de cada provador é feita em quatro vertentes:

(i) caracterização do processo de decisão do provador e controlo da sua estabilidade;

(ii) avaliação diária do desempenho do provador em termos de reprodutibilidade dos julgamentos;

(iii) avaliação periódica do desempenho global de cada provador;

(iv) aferição periódica de capacidades específicas dos provadores.

A validade e a qualidade das decisões do júri no seu conjunto são, obviamente, dependentes da validade e da qualidade das decisões de cada um dos seus membros. Neste capítulo são também analisadas e propostas metodologias para o estabelecimento da decisão final do painel, com base nos resultados demonstrados por cada provador em cada uma das fases de avaliação referidas.

Muitos dos métodos estudados foram sendo testados e implementados na Câmara de Provadores do Instituto do Vinho do Porto, o que possibilitou a existência de uma forte interacção entre as fases de desenvolvimento e de teste. Ao longo deste capítulo essa interacção é posta em evidência através da descrição da referida implementação e da apresentação de alguns dos resultados obtidos.

No Capítulo 7, são apresentadas as principais conclusões retiradas do trabalho realizado e são analisadas as perspectivas de desenvolvimento futuro dos temas estudados, quer a nível teórico quer a nível da sua implementação no âmbito da Análise Sensorial.

2

A Análise Sensorial

Conteúdo:

- Introdução
- Definição e Objectivos da Análise Sensorial
- As Características Organolépticas
- Métodos e Técnicas Utilizados em Avaliação Sensorial

2.1 Introdução

Os sentidos humanos são usados desde que o Homem existe, para avaliar a qualidade de tudo o que pode ser usado ou consumido. O comércio e a sua expansão, ao permitirem a comparação dentro de um leque muito mais vasto de produtos, contribuíram decisivamente para uma abordagem mais formal dos testes a esses mesmos produtos.

Os bens alimentares ocuparam sempre um lugar privilegiado no que diz respeito à avaliação das qualidades percepcionadas pelos sentidos. A explicação é óbvia: a alimentação é indispensável à sobrevivência da espécie humana e a frequência com que as pessoas consomem os bens alimentares é elevada.

Meilgaard et al. (1987a, p. 1) referem que, ao longo do tempo, se foram criando rituais de atribuição de classificações a alguns bens, tais como o vinho, o chá, o café, a manteiga, o peixe e a carne, alguns dos quais ainda se cumprem actualmente. Os mesmos autores sugerem que, da evolução desses rituais, surgiu o conceito de provador profissional que assumiu o papel de consultor das indústrias emergentes de comidas, bebidas e cosméticos no início do século.

Abordando a mesma questão, Lawless e Heymann (1999, p. 6) salientam que a produção de bens alimentares de boa qualidade sensorial ficava, muitas vezes, dependente de um único provador, que era responsável pela produção ou tomava decisões relativas a alterações do processo, de forma a garantir que o produto final possuía as características desejadas.

Esta foi a origem histórica dos *brewmasters*, dos provadores de vinhos, dos *dairy judges* e de outros inspectores de produtos alimentares.

A Análise Sensorial moderna substituiu as autoridades singulares por painéis de provadores que participam em experiências devidamente planeadas, que se baseiam em testes sensoriais específicos. É nestes métodos que assenta, actualmente, a avaliação dos produtos do ponto de vista sensorial.

A existência de inúmeras oportunidades de estudo na área da Análise Sensorial que não se encontram ainda exploradas é um facto mencionado por muitos dos autores que se dedicam a este tema. Pangborn (1984, p. 76) refere o facto de a avaliação sensorial

ser um campo ainda pouco explorado da área das ciências alimentares, e que a sua abordagem deverá passar pela constituição de grupos de investigação multidisciplinares. Meiselman (1993) e Lawless e Heymann (1999, p. 9) realçam carências no âmbito da comparação da fiabilidade dos vários métodos. Stone e Sidel (1993, p. 20) referem-se à prática industrial, onde as actividades relacionadas com a análise sensorial são, muito frequentemente, organizadas e planeadas por profissionais sem formação na área. Hunter et al. (1995) salientam a necessidade da existência de mais *software*.

Os problemas citados relacionam-se com o objectivo principal da análise sensorial aplicada à indústria: as características sensoriais dos produtos corresponderem àquilo que os clientes esperam. Assim, os métodos são desenvolvidos e utilizados com o objectivo de avaliar os produtos, e os provadores profissionais são encarados como os "instrumentos de medida" das características sensoriais.

Tal como qualquer outro instrumento de medida, os painéis de provadores profissionais devem ser calibrados e controlados. Há, no entanto, que reconhecer que os processos inerentes à apreciação das características organolépticas dos produtos são complexos e que as técnicas utilizadas para calibrar os instrumentos de medida convencionais não lhes podem ser directamente aplicadas.

Stone (1999) chama, inclusivamente, a atenção para o perigo que, segundo ele, existe, do re-surgimento dos *experts* e de alguns conceitos errados, tais como a ideia de que as pessoas podem e devem ser treinadas para serem invariantes, a utilização de escalas universais e o "terrorismo estatístico" (a utilização de algoritmos extremamente complexos como substitutos de uma boa organização e condução de testes).

Neste contexto, em que, por um lado, há a necessidade de calibrar e controlar os provadores e, por outro, é fundamental não reduzir a Análise Sensorial a estudos puramente estatísticos (esquecendo a sua natureza humana), é essencial conhecer, de uma forma profunda, os métodos existentes e testados nesta área. Assim, na secção 2.4, apresentam-se os métodos mais utilizados em Análise Sensorial e, no próximo capítulo, analisam-se as suas aplicações, estudam-se as várias formas de processar os seus resultados e estabelecem-se algumas comparações.

Antes, porém, na secção 2.2, apresenta-se a definição de Análise Sensorial tal como é actualmente reconhecida e descrevem-se os seus objectivos. Na secção 2.3, introduzem-se as características sensoriais e alguns conceitos relativos à forma como estas são percepcionadas pelo Homem.

Questões como a organização de um programa de Análise Sensorial e a selecção e o treino de um painel de provadores profissional serão abordadas no próximo capítulo, depois de estudados os vários métodos de processamento dos resultados dos ensaios.

2.2 Definição e Objectivos da Análise Sensorial

A Análise Sensorial pode ser definida como "a disciplina científica usada para evocar, medir, analisar e interpretar as reacções às características de bens alimentares e outros bens materiais tal como são percepcionadas pelos sentidos da visão, do olfacto, do sabor, do tacto e da audição." Esta definição, adoptada por diversos autores (Stone e Sidel (1993, p. 12) e Lawless e Heymann (1999, p. 2)) e organizações (Institute of Food Tecnologists, American Society for Testing and Materials), revela o carácter experimental da Análise Sensorial e, paralelamente, salienta as suas componentes de percepção e de comportamento reactivo.

A definição apresentada coloca, também, em evidência a natureza multidisciplinar da Análise Sensorial. O envolvimento de áreas do conhecimento tão diversas como a Estatística, a Psicologia, a Química ou o *Marketing* é uma das explicações que Stone e Sidel (1993, p. 12) adiantam para o facto de, tradicionalmente, as funções e as tarefas da Análise Sensorial e dos respectivos responsáveis, tanto em ambientes académicos como industriais, serem difíceis de estabelecer. A definição apresentada, embora suficientemente geral para abarcar a avaliação de todo o tipo de produtos, contribui para a clarificação dessas funções. As quatro actividades nela mencionadas referem-se aos princípios e às práticas que devem estar subjacentes a qualquer estudo nesta área:

(i) **Evocar** – a evocação das sensações faz-se através da apresentação dos produtos aos indivíduos que os vão testar. Os cuidados a ter com a preparação das amostras e com as condições em que os ensaios deverão decorrer são um dos pilares em que deve assentar qualquer projecto de Análise Sensorial.

(ii) **Medir** – as percepções humanas são, pela sua própria natureza, difíceis de medir. Tal como referem Lawless e Heymann (1999, p. 2), "em Análise Sensorial recolhem-se dados numéricos com o objectivo de estabelecer relações válidas e específicas entre as características dos produtos e as percepções humanas". À semelhança do que acontece com a evocação das sensações, a medição das percepções condiciona fortemente os resultados e as conclusões do

estudo, pelo que a selecção dos testes adequados a cada situação é um factor essencial para que um projecto seja bem sucedido.

(iii) **Analisar** – Esta actividade encontra-se intimamente relacionada com a da medição, devendo ser planeadas em conjunto. A análise adequada dos dados é uma parte crítica na Análise Sensorial. O facto de haver muitas fontes de variação nas respostas humanas que não se conseguem controlar completamente em nenhum teste sensorial, complica a análise. O humor, a motivação dos participantes nos testes ou a sua sensibilidade natural às características testadas são exemplos de alguns desses factores. A análise estatística dos dados tem como objectivo investigar as variáveis que se revelem de interesse e, eventualmente, estabelecer relações entre elas, de forma a permitir que se possam retirar conclusões válidas do estudo.

(iv) **Interpretar** – A interpretação dos resultados dos testes sensoriais deverá, tal como acontece com a prática experimental de uma forma genérica, ser feita no contexto das hipóteses colocadas *a priori* e do conhecimento já adquirido. Assim, e tal como referem Lawless e Heymann (1999, p. 3), "as conclusões deverão envolver considerações sobre o método, as limitações da experiência, o conhecimento adquirido e o contexto do estudo".

O'Mahony (1988) refere-se à Análise Sensorial de uma forma mais resumida como "o uso dos sentidos humanos para medir as características sensoriais dos alimentos e de outros produtos" e divide-a em dois tipos, com objectivos claramente distintos, a que chama *Análise Sensorial I* e *Análise Sensorial II*.

A Análise Sensorial I refere-se à Análise Sensorial quando é usada como uma ferramenta para estudar características químicas e/ou físicas dos produtos. Um exemplo será o de determinar se uma dada alteração no processo produtivo de um bem alimentar altera a sua acidez. A ênfase é, portanto, colocada nos produtos, sendo as pessoas encaradas como instrumentos cuja função é a de medir as suas várias propriedades.

Tal como na análise instrumental, o que se pretende é que o *aparelho* seja sensível às características a medir e que os seus resultados sejam fiáveis. Transpondo directamente o que se passa na análise instrumental para a Análise Sensorial tipo I, o ideal seria escolher o melhor provador e utilizá-lo sempre que fosse necessário analisar as características sensoriais de um ou mais produtos.

Os problemas colocados por uma abordagem deste tipo são óbvios e decorrem do facto de as pessoas não poderem ser tratadas como aparelhos em determinadas

situações. Alguns desses problemas são: (i) o melhor provador para certas características pode não ser o melhor para outras; (ii) as pessoas são influenciadas por factores externos que podem alterar o seu desempenho; (iii) por vezes é difícil definir *o melhor* provador devido à natureza subjectiva dos estímulos; (iv) nem sempre as pessoas estão disponíveis, nomeadamente em situação de doença ou em períodos de férias.

A forma de ultrapassar as questões referidas é a utilização de vários provadores simultaneamente. Esses provadores devem ser seleccionados, recorrendo a uma grande variedade de testes que permitam avaliar as suas capacidades para desempenhar a função pretendida. No entanto, e tal como é referido nas normas francesas (AFNOR (1991)), "uma vez seleccionados, os provadores deverão ser controlados regularmente, de forma a assegurar que os critérios que conduziram à sua selecção continuam a verificar-se".

A Análise Sensorial II refere-se à Análise Sensorial quando é usada para prever a reacção dos consumidores aos produtos. Um exemplo será o de verificar se os consumidores detectam a diferença entre duas formulações distintas do mesmo produto. Neste caso, a ênfase é colocada nas pessoas e na sua reacção aos produtos, e os provadores devem ser seleccionados de forma a representar o mais fielmente possível a população dos consumidores actuais ou potenciais.

Neste contexto, os provadores não deverão ser treinados para detectar quaisquer características em particular e não se pretende escolher os *mais sensíveis*, mas sim os *mais representativos* da população em questão. Como em qualquer estudo de mercado, neste tipo de investigação os resultados são tanto mais fiáveis quanto mais provadores responderem ao ensaio. No entanto, ao seleccionar as pessoas, é necessário ter em conta factores que influenciam os seus hábitos como, por exemplo, a região onde vivem, a idade ou a classe económica e/ou social.

Meiselman (1993) aborda a mesma temática, referindo-se à dicotomia sensorial/afectiva. Refere-se ao estudo analítico (que corresponde à Análise Sensorial I) versus o estudo hedónico ou afectivo (correspondente à Análise Sensorial II), e salienta que, enquanto no primeiro se utilizam painéis de provadores treinados e se estudam os produtos, no segundo utilizam-se os próprios consumidores, estudando-se simultaneamente os consumidores e os produtos.

O tema analisado nesta tese refere-se essencialmente à Análise Sensorial I e, portanto, sem perda de generalidade, quando neste trabalho é utilizada a expressão

Análise Sensorial, assume-se que se trata do tipo I. Não obstante, note-se que muitos dos ensaios sensoriais descritos e das considerações feitas são comuns aos dois tipos de análise.

2.3 As Características Organolépticas

As características ou propriedades organolépticas de um produto são aquelas que são perceptíveis pelos órgãos dos sentidos (ver IPQ (1994)). São propriedades físicas ou químicas que se tornam organolépticas no momento em as pessoas tomam consciência delas através dos órgãos dos sentidos. Peynaud e Blouin (1997, p. 16) dão o exemplo do açúcar cristalizado, que tem os seus atributos físicos e químicos próprios (forma, densidade, composição, etc.) e que tem uma propriedade organoléptica que lhe é inerente – o gosto doce. Esta propriedade requer, todavia, a existência de papilas na língua ligadas ao cérebro de uma pessoa.

2.3.1 Sensação e Percepção

Na figura 2.1, é apresentado um esquema onde se pode observar, de um modo simplificado, a forma como os produtos actuam nos sentidos humanos.

Os produtos criam estímulos que podem ser visuais, auditivos, olfactivos, gustativos ou tácteis e que provocam a excitação de receptores específicos dos sentidos respectivos. A sensações são as reacções aos estímulos: a sensação não é mais do que o reflexo que é originado quando a "informação" contida no estímulo é enviada dos nervos sensíveis (receptores neuro-sensoriais) para o cérebro. A percepção refere-se à fase da "descodificação" da informação: com base na experiência e na memória, a pessoa reconhece a sensação, interpretando-a. A percepção é definida no Projecto de Norma Portuguesa prNP 4263 (IPQ (1994)) como "a tomada de conhecimento dos efeitos dos estímulos sensoriais simples ou complexos".

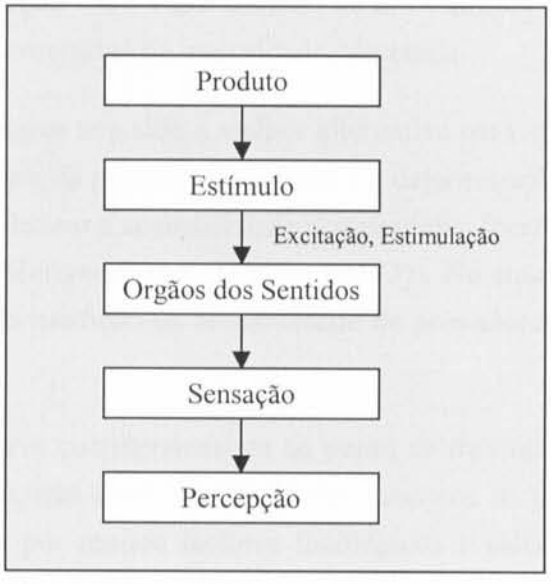

Figura 2.1 – Esquema do mecanismo dos sentidos humanos

A propósito de sensação e percepção, Peynaud e Blouin (1997, p. 15) salientam: "A sensação é inconsciente; a percepção é a tomada de consciência. Uma sensação desconhecida não pode ser interpretada e passa despercebida, ou então é confundida com uma sensação vizinha, situação em que o cérebro funciona por analogia" .

No processo de percepção, as características organolépticas dos produtos sobrepõem-se, ou seja, o sujeito recebe um conjunto de sensações quase simultâneas que lhe causa a percepção global do produto. A maior parte das pessoas não é capaz, sem ser treinada especificamente para tal, de avaliar cada característica independentemente das outras.

Um conceito que se encontra intimamente relacionado com o de percepção é o de limiar. Meilgaard et al. (1987a, p. 113) definem limiar como o limite das capacidades sensoriais de uma pessoa ou de uma população. Bi e Ennis (1998a) referem que os limiares são, muitas vezes, encarados como o limite inferior da sensibilidade do sistema sensorial. Os quatro tipos de limiares mais referenciados em Análise Sensorial são:

- **Limiar de detecção** – é o valor mínimo de um estímulo necessário ao aparecimento de uma sensação (que poderá não ser ainda identificada).
- **Limiar de reconhecimento** – é o valor mínimo de um estímulo que permite reconhecer e identificar a sensação percebida.
- **Limiar diferencial** – é o valor mínimo da alteração de um estímulo necessário para que a diferença (entre o estímulo inicial e o alterado) seja perceptível.

- **Limiar de saturação** – é o valor mínimo de um estímulo acima do qual deixa de haver diferença perceptível na intensidade percebida.

A utilização dos limiares tem sido a melhor alternativa para analisar problemas como a deliberação sobre níveis de poluição aceitáveis e a determinação de níveis a partir dos quais certos aditivos reduzem a aceitabilidade dos produtos (perfumes, cosméticos, bens alimentares, etc.) (ver Meilgaard et al. (1987a, p. 113)). No entanto, este conceito pode também ser utilizado na medição da sensibilidade de provadores profissionais e até na sua selecção.

Idealmente, os limiares corresponderiam ao ponto de transição entre os dois estados em causa (por exemplo, não sentir/sentir). O que acontece na prática é que a resposta das pessoas é afectada por muitos factores fisiológicos e psicológicos que alteram o ponto de transição (ver Bi e Ennis (1998a)).

Na figura 2.2, apresenta-se um gráfico que esquematiza a frequência de respostas positivas obtidas por um indivíduo na detecção de um estímulo, em função da intensidade desse mesmo estímulo. A curva representada a tracejado, com a forma de degrau, corresponde à situação ideal, em que a transição entre os estados de "não sentir" e "sentir" é brusca. A curva representada a cheio, com a forma de "S", corresponde à situação real, na qual as respostas do mesmo indivíduo ao mesmo estímulo não são sempre iguais – umas vezes identifica a sensação, outras vezes não identifica. Esta curva é, muitas vezes, designada por função psicométrica do indivíduo.

A probabilidade de o estímulo ser detectado varia, assim, entre 0% e 100%, aumentando gradualmente com a sua intensidade. A definição de limiar sensorial mais utilizada actualmente é, precisamente, a que se baseia nesta interpretação estatística do fenómeno em estudo e que lhe faz corresponder o valor da intensidade do estímulo que corresponde à probabilidade de detecção de 50% (ver ASTM (1997b)). Na figura 2.2, ilustra-se, também, esta definição de limiar.

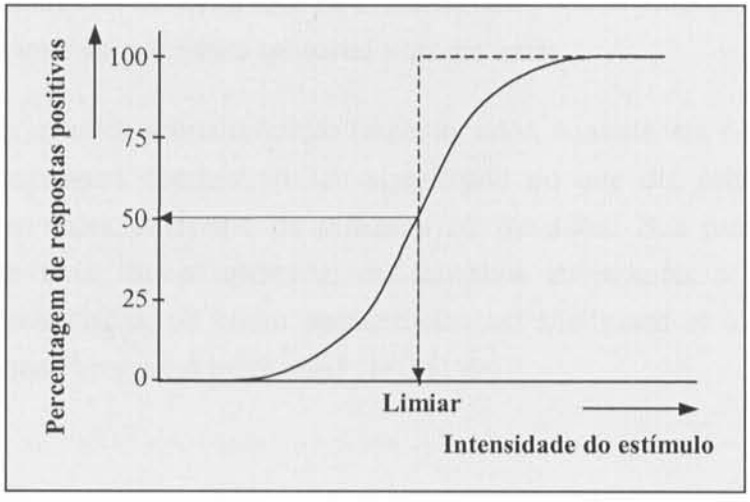

Figura 2.2 – Função de detecção ideal e função psicométrica

No que foi dito até agora, ficou implícito que o conceito de limiar se refere a um indivíduo. Na realidade, existem os dois conceitos, o limiar individual (que corresponde ao que foi referido) e o limiar de grupo. Tradicionalmente, o limiar de grupo é definido como uma medida da tendência central (que pode ser a mediana, o valor esperado, a média geométrica ou outra qualquer que se justifique) dos membros do grupo (ver ASTM (1997b)).

No entanto, Bi e Ennis (1998a) chamam a atenção para o facto de tal definição obrigar à determinação dos limiares individuais dos membros do grupo. Os autores referem que, dado que a variabilidade entre as pessoas é substancialmente maior do que a que se verifica dentro das apreciações efectuadas pela mesma pessoa, não se pode esperar um aumento da precisão no cálculo do limiar pelo facto de perguntar às mesmas pessoas muitas vezes a mesma coisa. Assim, propõem que o limiar de detecção do grupo seja definido como a intensidade do estímulo que produz resposta em metade do grupo, o que, obviamente, facilita a tarefa prática da determinação desta grandeza.

Ao longo deste trabalho, estes conceitos serão referidos frequentemente, dada a relevância que possuem no contexto da Análise Sensorial. Chama-se, porém, a atenção para a secção 2.4.3, onde serão referidas várias formas de determinar os limiares individuais e os de grupos.

Independentemente de haver a tendência de percepcionar os produtos alimentares como um todo, há uma ordem natural pela qual as suas características organolépticas tendem a ser percebidas: inicialmente o seu aspecto, seguindo-se o seu odor, a sua consistência e textura e, finalmente, o seu sabor (ao qual estão associados o aroma, as sensações químicas e o gosto). Os estímulos auditivos não serão aqui referidos, embora,

em casos particulares de determinados bens alimentares, o som produzido durante a sua mastigação seja uma característica sensorial a ter em conta.

Embora as expressões acima referidas (aspecto, odor, consistência e textura e sabor) pertençam à linguagem corrente, o seu significado no que diz respeito à Análise Sensorial é, por vezes, diferente do utilizado no dia-a-dia. Nas próximas secções, explicam-se, de uma forma resumida, os conceitos subjacentes a cada uma das características enunciadas, tal como apresentados em Meilgaard et al. (1987a) e no Projecto de Norma Portuguesa prNP 4263 (IPQ (1994)).

2.3.2 Aspecto

A primeira impressão que um indivíduo retém de um produto é, quase sempre, a do seu aspecto. O aspecto é definido em IPQ (1994, p. 8) como "o conjunto das propriedades visíveis de uma substância ou de um produto".

As características mais relevantes que compõem o aspecto de um produto alimentar são a cor, a forma, a textura exterior, o brilho, a limpidez e a transparência. Em bebidas com gás, o grau de efervescência também contribui para o aspecto do produto.

Vários estudos (DuBose et al. (1980), Christensen (1983) e Maga (1974), por exemplo) indicam que o aspecto de um bem alimentar condiciona fortemente as percepções de características referentes ao seu aroma e ao seu sabor. É, portanto, essencial que os produtos apresentem um aspecto agradável. Por outro lado, é também necessário que a aparência do produto vá de encontro à imagem que as pessoas têm dele: dificilmente comprariam um iogurte de morango que fosse verde, mesmo que reconhecessem que tinha um aspecto atraente.

O aspecto dos bens alimentares tem uma outra vertente que é fundamental considerar e que não assume a mesma relevância no caso de outros bens como os de cosmética ou perfumaria: a de revelar possíveis deteriorações ou defeitos que possam influenciar a saúde humana. Quando, através do aspecto do produto, as pessoas interpretam que o seu consumo pode vir a prejudicar a sua saúde, a reacção é, naturalmente a de não adquirir nem consumir o produto em causa.

Dado que o "bom aspecto" dos produtos alimentares é essencial para a sua aceitação pelo mercado e que esta característica é a primeira a ser apreciada, a sua avaliação deve, também, ser a primeira a ser efectuada durante o processo de prova.

2.3.3 Odor

O odor de um produto é detectado quando os seus componentes voláteis penetram no nariz e são percepcionados pelo sistema olfactivo. Normalmente, a sensação do odor em bens alimentares precede a do sabor, visto que, ainda antes de se levar o alimento à boca, os seus componentes voláteis são libertados, atingindo o sistema olfactivo.

Carola et al. (1992, p. 416) referem que o nosso sentido do olfacto é cerca de 20.000 vezes mais sensível do que o do gosto, pois somos capazes de detectar concentrações muito mais baixas através do olfacto do que as concentrações detectáveis pelo gosto. Os mesmos autores mencionam ainda que os adultos são capazes de detectar cerca de 10.000 odores diferentes, sendo as crianças mais sensíveis e, consequentemente, capazes de se aperceberem de mais.

A intensidade do estímulo depende de vários factores, tais como o tipo de substância (mais ou menos volátil), a maneira como se inspira (a intensidade é maior se se inspirar rapidamente e de uma só vez) e a temperatura a que se encontra o produto.

A saturação das células receptoras (adaptação ao odor) é um problema a ter em consideração quando se trata da prova. Para a evitar, deve cheirar-se o produto durante períodos de tempo curtos (1-2 segundos), fazer intervalos (de, no mínimo, 5 a 20 segundos) e "limpar" as vias respiratórias respirando num ambiente não odorado ou, em alternativa, cheirando substâncias diferentes.

A sensibilidade aos vários odores varia conforme as pessoas e, para cada pessoa, está sujeita a factores como o humor, a concentração ou a saúde (constipações, por exemplo).

2.3.4 Consistência e Textura

A textura é definida no referido Projecto de Norma (IPQ (1994, p. 9)) como "o conjunto das propriedades mecânicas, geométricas e de superfície de um produto, detectáveis pelos receptores mecânicos e tácteis e, eventualmente, pelos receptores visuais e auditivos".

As propriedades mecânicas são as que estão ligadas à reacção do produto a uma pressão e dividem-se em cinco características primárias: dureza, coesão, viscosidade,

elasticidade e aderência. As propriedades geométricas são as que estão ligadas às dimensões, à forma e ao arranjo das partículas. As propriedades de superfície são as que se relacionam com sensações como as provocadas pela água ou pela gordura. Na boca, correspondem ao modo como os constituintes dos produtos são libertados.

Nos alimentos sólidos, as pessoas atribuem muita importância a características como a dureza, a coesão e a elasticidade, que se relacionam com a mastigação, enquanto nos líquidos a viscosidade e as propriedades de superfície são mais valorizadas.

2.3.5 Sabor

O processo de degustação (avaliação sensorial de um produto na boca) é complexo, compreendendo um conjunto de sensações olfactivas, gustativas e químicas.

Em IPQ (1994, p. 7), este conjunto de sensações é denominado de "flavor", com base na terminologia inglesa. No entanto, e dado que o referido documento é apenas um projecto de norma e que o vocábulo "flavor" não faz, ainda, parte da língua portuguesa, neste trabalho adoptar-se-á o termo "sabor" para caracterizar o conjunto de sensações em causa. Este termo é, aliás, adoptado com este sentido pela maioria dos especialistas em Análise Sensorial em Portugal.

As sensações olfactivas da degustação são as que permitem a detecção do aroma do produto. Esta detecção dá-se, não através da entrada dos componentes voláteis do produto pelo nariz (a que, como já foi visto, corresponde o odor), mas através da detecção dos componentes voláteis que se encontram na boca (por vezes, diferentes dos correspondentes ao odor) e que atingem o órgão olfactivo por via retro-nasal.

Consequentemente, em Análise Sensorial, os conceitos de odor e de aroma são distintos.

As sensações gustativas são apenas quatro: doce, salgado, ácido e amargo. São detectadas por células da língua, quando as substâncias que as estimulam se encontram dissolvidas em água, óleo ou saliva.

As células que detectam o doce encontram-se na extremidade da língua, pelo que esta é, normalmente, a primeira sensação a ser transmitida ao cérebro. Seguem-se as sensações de salgado (os receptores encontram-se nos lados da língua, imediatamente a seguir aos do doce) e de ácido (cujos receptores se situam também nos lados da língua, mas mais atrás do que os de salgado). A sensação de amargo é a última a ser detectada e

as células que lhe são sensíveis situam-se no meio da língua, na sua parte de trás. É comum, em alguns alimentos, as pessoas terem uma sensação inicial de alguma doçura, e, no fim, quando engolem, uma sensação de amargo. Isto acontece precisamente porque estas células são as últimas a ser estimuladas.

Dado que o sabor abrange um conjunto de sensações das quais o gosto faz parte não integrante, salienta-se que, em Análise Sensorial, sabor e gosto de um produto não são sinónimos.

As sensações químicas resultam da estimulação de terminais nervosos de algumas membranas das cavidades bucal e nasal. Estas sensações são: a adstringência, o sabor picante, a sensação de frio na boca, a pungência e o sabor metálico.

Segundo Meilgaard et al. (1987a, p. 14), as células receptoras da boca encontram-se ainda mais sujeitas à saturação do que as do nariz, visto que o contacto entre a solução (do alimento com a saliva) e as células é mais regular (cada receptor fica imerso na solução, pelo menos por alguns segundos). Os autores referem que há gostos, como o amargo, que podem durar horas ou mesmo mais do que um dia.

Os provadores são, assim, aconselhados a degustar poucas quantidades de cada vez, retê-las na boca por poucos segundos e esperar entre 15 a 60 segundos (dependendo da intensidade dos estímulos) antes de provarem outra vez.

2.3.6 Factores que Influenciam a Percepção Sensorial

A propósito do facto de os provadores, em Análise Sensorial, serem encarados como instrumentos de medida, Meilgaard et al. (1987a, p. 1) referem: "Os provadores, encarados como instrumentos de medida, são altamente variáveis e muito propensos ao enviezamento, mas são os únicos instrumentos que medem o que nós queremos medir, pelo que devemos minimizar a variabilidade e controlar o enviezamento, utilizando o melhor possível as melhores técnicas existentes ...".

Uma boa utilização das técnicas pressupõe, não só o seu conhecimento, mas também a compreensão de fenómenos que, sendo exteriores às técnicas, fazem parte do processo da medição e influenciam os seus resultados.

Os factores que influenciam a percepção sensorial dividem-se em dois grandes grupos: os factores fisiológicos e os factores psicológicos. Seguidamente, descrevem-se alguns desses factores.

2.3.6.1 Factores Fisiológicos

Os factores fisiológicos mais referidos na literatura são os de adaptação e de potenciação/supressão, que a seguir se abordam.

- **Adaptação** – modificação temporária na sensibilidade de um órgão sensorial a um determinado estímulo, resultante de uma exposição prolongada ao estímulo em causa ou a um estímulo similar.

- **Potenciação / Supressão** – este factor envolve a interacção de estímulos que se encontram presentes simultaneamente no produto a testar. Pode tomar a forma de potenciação, de sinergia ou de supressão. A potenciação ocorre quando a presença de uma substância causa um aumento da intensidade percebida de uma segunda substância. A sinergia ocorre quando a presença de uma substância causa um aumento da intensidade percebida da combinação de duas ou mais substâncias (a intensidade percebida da mistura é superior à soma das intensidades de cada um dos componentes). A supressão ocorre quando a presença de uma substância causa uma diminuição da intensidade percebida de uma ou mais substâncias diferentes.

2.3.6.2 Factores Psicológicos

Os factores psicológicos referidos na literatura são muitos. Neste estudo, optou-se por salientar aqueles que se considerou serem mais relevantes no contexto em análise, os quais se descrevem em seguida.

- **Erro de Expectativa** – a informação relativa às amostras facultada aos provadores antes ou durante a prova pode induzi-los em erro ("as pessoas tendem a encontrar o que estão à espera de encontrar" Meilgaard et al. (1987a, p. 34)). Por exemplo, se um produto foi devolvido porque apresentava um aroma desagradável, os provadores, ao saberem disso, tenderão a encontrar características desagradáveis no aroma. Este tipo de erros pode pôr em causa a validade dos testes, pelo que não deve ser fornecida qualquer informação sobre as amostras antes ou durante o teste.

- **Erro de Habituação** – este tipo de erro sucede pelo facto de as pessoas terem tendência para atribuir sempre a mesma classificação, por exemplo, a uma série de estímulos cuja intensidade cresce lentamente. Como o crescimento é lento, a

pessoa habitua-se a atribuir sempre a mesma cotação. Este erro pode ser minimizado variando o tipo de produtos apresentado.

- **Erro Lógico** – este tipo de erro tende a ocorrer quando duas ou mais características se encontram associadas na mente do provador. Por exemplo, em dois vinhos com a mesma idade, mas com cores diferentes, o provador pode ter a tendência de atribuir mais idade ao que apresentar um tom mais acastanhado.

- **Erro de Associação** – quando são apreciadas mais do que uma característica, a avaliação de umas tende a influenciar a avaliação das outras. Por exemplo, quando um produto está a ser avaliado simultaneamente quanto à intensidade do seu aroma e quanto à sua aceitação, os resultados podem ser diferentes do que seriam caso os dois testes fossem feitos separadamente.

- **Ordem de Apresentação das Amostras** – há, no mínimo, cinco tipos de efeitos que podem decorrer da ordem de apresentação das amostras e que se descrevem seguidamente. O efeito de contraste e o efeito de grupo são fenómenos inversos. O primeiro refere-se à tendência das pessoas para, quase imediatamente após terem julgado uma amostra de boa qualidade, classificarem outra de pior qualidade com uma pontuação inferior à que lhe atribuiriam no caso de ser avaliada isoladamente. Já o efeito de grupo ocorre quando uma amostra de melhor qualidade, por ser apresentada em conjunto com outras de pior qualidade, é pontuada com uma classificação inferior à que lhe seria atribuída isoladamente. O erro de tendência central decorre do facto de as amostras apresentadas aproximadamente no meio de uma série, terem tendência a ser melhor cotadas do que as apresentadas nos extremos. O efeito de padrão acontece quando se repete algum modelo de apresentação das amostras, e os provadores, legitimamente, detectam esse padrão. Finalmente, o erro de tempo/posição ocorre pelo facto de a atitude do provador sofrer mudanças ao longo de uma série de provas, levando a que, frequentemente, a primeira amostra seja sobre ou sub-cotada e as últimas (em testes longos) sejam julgadas já com alguma indiferença.

- **Erro de Sugestão Mútua** – refere-se à situação em que a resposta de um provador é influenciada pela resposta de outro ou outros.

- **Falta de Motivação** – quando há falta de motivação, os graus de concentração e de empenhamento do provador diminuem, provocando uma diminuição da sua reprodutibilidade e da sua capacidade de detectar pequenas diferenças entre estímulos.

- **Capricho vs. Timidez** – este factor relaciona-se com a utilização de escalas que usam vários níveis (por exemplo, de 1 a 7) e ocorre quando o provador tende a abusar do uso dos extremos da escala (capricho) ou se concentra demasiado nos seus valores centrais (timidez).

2.3.6.3 Enviezamento das Respostas

O enviezamento das respostas é ainda um factor psicológico (cognitivo) que se caracteriza por uma variação do critério utilizado pelo provador. No entanto, dada a complexidade e importância deste conceito, decidiu-se incluí-lo numa secção independente dos outros factores psicológicos já abordados.

Este factor é particularmente relevante quando se pretende detectar ou medir pequenas diferenças entre produtos (ver testes de diferenças na secção 2.4). Quando as diferenças são grandes, o problema não se coloca, uma vez que o provador não tem dúvidas em relação ao critério a utilizar para classificar os estímulos.

O conceito de enviezamento das respostas é apresentado com base num exemplo descrito em O'Mahony (1992) e em O'Mahony (1995). Admite-se, então, que se fazem vários biscoitos semelhantes, em que o único factor que varia é a quantidade de açúcar, e que se vai analisar a percepção de um provador relativamente à doçura dos biscoitos.

Na figura 2.3 estão representados três biscoitos: N, X e S. O biscoito N é definitivamente reconhecido pelo provador como um biscoito não doce, e o biscoito S é reconhecido como doce. O biscoito X encontra-se numa região de incerteza correspondente a vários graus de doçura.

À medida que se vai "caminhando" de N para S, há uma região de incerteza onde o provador tem dificuldade em dizer se o biscoito é ou não mais doce do que N. A sua resposta vai depender de onde situa o seu critério de doçura, ou seja, onde é que "desenha a linha" entre "não doce" e "doce". Admita-se que o biscoito X se situa nessa região, tal como está indicado na figura. Se o critério se situar em A (critério restrito), X é classificado como não doce, enquanto, se o critério se situar em D (critério lato), X é classificado como doce.

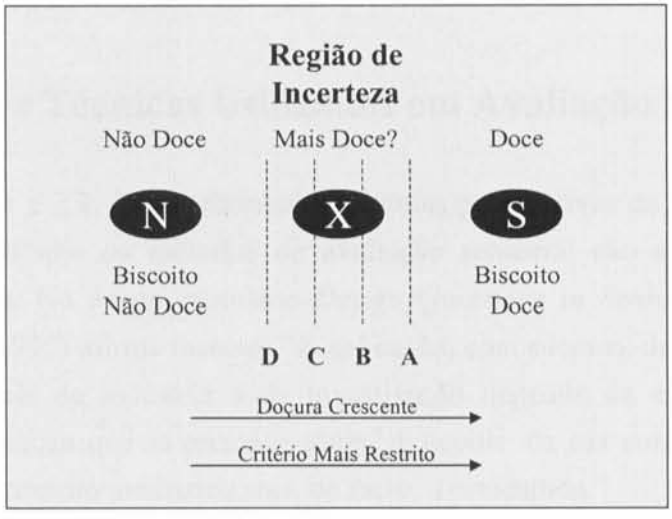

Figura 2.3 – Diagrama que representa o enviezamento das respostas para um exemplo de biscoitos doces e não doces (fonte: O'Mahony (1992))

A tendência para o provador utilizar o critério mais para a direita ou mais para a esquerda é chamada o enviezamento da resposta. Note-se que se trata de um factor puramente cognitivo, não se relacionando com a sensibilidade que o provador tem à característica em causa. O enviezamento da resposta refere-se, sim, à segurança que o provador sente para dizer que o biscoito X é mais doce do que o N.

O'Mahony (1992) refere que, ao elaborar os testes sensoriais, se devem adoptar estratégias para evitar alterações não controladas do critério de decisão dos provadores. Este tema será abordado mais adiante, mas convém referir que a estabilização do critério implica que o provador saiba, antes de começar o teste, quantas amostras são diferentes e/ou iguais.

Para minimizar os efeitos dos factores referidos, é essencial estar atento e ter presentes os tipos de situações que podem surgir. Os métodos de avaliação devem, também, ser escolhidos consoante os riscos que lhes estão associados. Na próxima secção descrevem-se alguns desses métodos, e, em alguns casos, chama-se a atenção para os riscos associados à sua utilização, nomeadamente no que diz respeito aos factores mencionados neste ponto.

Por outro lado, ao planear uma experiência sensorial, devem tomar-se medidas que permitam reduzir a probabilidade de os erros ocorrerem. Na secção 3.3, serão apresentadas algumas dessas medidas.

2.4 Métodos e Técnicas Utilizados em Avaliação Sensorial

Nas secções 2.1 e 2.2, foram abordadas algumas perspectivas da Análise Sensorial, tendo ficado claro que os métodos de avaliação sensorial são a base da Análise Sensorial moderna. No artigo intitulado *Design Questions in Sensory and Consumer Science,* Piggott (1995) afirma mesmo: "A aplicação, com sucesso, da Análise Sensorial aos problemas reais da indústria e da investigação depende da escolha do método adequado à informação que se pretende obter, e, depois, da sua correcta aplicação, de forma a que a informação produzida seja, de facto, a pretendida."

A selecção dos testes adequados a cada situação é, assim, essencial para o sucesso de qualquer programa de Análise Sensorial. Para que essa selecção seja eficaz, é necessário que quem a faz tenha um conhecimento profundo das potencialidades e das limitações dos vários testes.

Meilgaard et al. (1987a) referem que os testes sensoriais podem, legitimamente, tomar variadíssimas formas, mas que, na prática, existem procedimentos que adquiriram um estatuto próprio e que têm uma longa história de utilização. Esses procedimentos encontram-se já extensamente estudados e os problemas que poderão surgir com a sua aplicação prática já se encontram largamente referenciados na literatura especializada. Alguns desses métodos fazem já parte de normas francesas (AFNOR (1991)) e são aconselhados por organizações como a *American Society for Testing and Materials* (ver ASTM (1996)). Assim, torna-se vantajoso utilizar os referidos métodos em alternativa a outros menos conhecidos.

Os métodos em Análise Sensorial são, normalmente, divididos em três classes, consoante o tipo de objectivo do estudo em questão (ver, por exemplo, Meilgaard et al. (1987a), Stone e Sidel (1993) e Lawless e Heymann (1999)): ensaios de diferenciação, ensaios descritivos e ensaios afectivos.

Os ensaios de diferenciação, também chamados testes de diferenças, envolvem, sempre, dois ou mais produtos distintos e têm como objectivo responder à questão: "Será que existe alguma diferença a nível sensorial entre os produtos?".

Quando se pretende detectar pequenas diferenças entre produtos os estímulos são de fraca intensidade. Este facto dá origem a que as escalas que se usam normalmente em procedimentos de classificação (por exemplo as multinomiais) não sejam aconselháveis

nestes casos, dado que se corre o risco de o erro da medição ser superior à diferença que se pretende medir (O'Mahony (1992)). Assim, os testes de diferenças são, de uma forma geral, direccionados para a detecção de amostras iguais e/ou diferentes de uma amostra-testemunho ou para o agrupamento de amostras semelhantes, estando-lhes associados testes de hipóteses, nos quais a hipótese nula se refere à ausência de diferença sensorial entre os produtos.

Neste tipo de testes, é necessário ter em conta que a probabilidade de um provador que responda ao acaso acertar é, normalmente, grande. Este problema resolve-se repetindo o teste várias vezes com o mesmo provador, utilizando vários provadores, ou optando por uma combinação das duas técnicas (vários provadores, cada um com várias provas). Neste último caso, é possível não só determinar se os produtos em causa são diferentes de um ponto de vista sensorial, mas também se existem diferenças significativas na percepção dos vários provadores.

Os testes de diferenças podem, por sua vez, dividir-se em dois grupos: testes de diferenciação geral (*overall difference tests*) e testes de diferenciação de atributos (*attribute difference tests*). Na secções 2.4.1 e 2.4.2, são estudados os métodos pertencentes a cada um desses grupos, respectivamente. Na secção 2.4.3, são abordados os ensaios normalmente utilizados para determinar os limiares, que, no fundo, correspondem a casos particulares dos ensaios de diferenciação.

Os ensaios descritivos (ou análises descritivas) são métodos mais sofisticados do que os testes de diferenças e, na maior parte dos casos, os seus resultados traduzem-se na quantificação da intensidade percebida das características sensoriais. Os testes mais referidos neste tipo de ensaios são descritos em 2.4.4.

Tanto os ensaios de diferenciação como os descritivos são utilizados na Análise Sensorial tipo I ou estudo analítico (ver secção 2.2).

Os ensaios afectivos têm como objectivo quantificar o grau de aceitação (gostar mais ou menos) dos produtos. São, portanto, utilizados na Análise Sensorial tipo II ou estudo hedónico ou afectivo. Embora este trabalho se concentre essencialmente na Análise Sensorial tipo I, optou-se por incluir em 2.4.5 uma síntese dos principais ensaios afectivos, essencialmente devido ao facto de ser possível estabelecer uma correspondência directa entre alguns testes afectivos e de diferenciação e de, portanto, o tratamento estatístico dos respectivos resultados ser semelhante.

Note-se que, em 3.3, são apresentadas, além de outras considerações, as precauções a ter na preparação das amostras para os testes, pelo que este assunto não será abordado de uma forma exaustiva na descrição dos vários ensaios.

2.4.1 Ensaios de Diferenciação Geral

O objectivo dos ensaios de diferenciação geral consiste na detecção da diferença/semelhança sensorial entre as amostras de uma forma não especificada, ou seja, sem precisar em que é que as amostras diferem ou se assemelham.

Este tipo de testes é, pois, particularmente útil, quando se pretende averiguar se existem alterações ou pequenas diferenças nos produtos que não podem ser caracterizadas como afectando apenas um ou dois atributos. São exemplos de aplicação controlar a estabilidade de um produto, testar um novo método de produção (e verificar se o produto se mantém com as mesmas características), estudar o efeito da substituição de ingredientes, etc.

Seguidamente, apresentam-se os testes de diferenciação geral mais referidos e estudados na literatura: triangular, dois em cinco, duo-trio, diferença simples, A-Não A, ABX, padrão duplo, diferença em relação ao controlo e testes sequenciais.

2.4.1.1 Teste Triangular

O teste triangular é, deste tipo de ensaios, o mais conhecido e mais utilizado (ver, por exemplo, Hunter (1996) ou Stone e Sidel (1993, p. 152)).

Neste ensaio são apresentadas ao provador duas amostras de um determinado produto e uma de outro, sendo-lhe solicitado que determine qual a amostra ímpar. As amostras devem ser previamente codificadas e apresentadas todas com o mesmo aspecto (em pratos ou copos iguais e sob as mesmas condições).

O'Mahony (1995) salienta a importância de as instruções dadas aos provadores serem claras, referindo inequivocamente a existência de uma amostra diferente entre as três e forçando o provador a identificá-la, mesmo que lhe pareçam todas iguais – procedimento de escolha forçada.

A probabilidade de o provador acertar respondendo ao acaso é bastante elevada (1/3), pelo que o teste deverá ser repetido pelo mesmo provador ou realizado por vários provadores de forma a poder determinar-se, com um nível de confiança considerado

suficiente, se a diferença entre os produtos é ou não detectável. O número de vezes que se deve realizar o teste depende do nível de confiança pretendido.

2.4.1.2 Teste Dois em Cinco

Neste teste são apresentadas ao provador cinco amostras, sendo duas de um produto e três de outro, pretendendo-se que ele as agrupe convenientemente. A eficiência estatística do teste é bastante mais elevada do que a do teste triangular, dado que a probabilidade de o provador acertar respondendo ao acaso é, agora, de 1/10. Assim, quando se pretender verificar com um determinado grau de confiança se a diferença entre dois produtos é detectável, serão necessárias muito menos repetições do teste dois em cinco do que seriam, caso se utilizasse o teste triangular.

Os cuidados a ter com a realização deste teste são basicamente os mesmos que os descritos para o teste triangular, mas trata-se de um teste mais difícil de realizar e mais sujeito a fadiga sensorial. Assim, deve ser utilizado com provadores já com alguma experiência, sendo frequentemente recomendado para situações nas quais as amostras são comparadas visualmente ou pelo tacto.

2.4.1.3 Teste Duo-Trio

No teste duo-trio é apresentada ao provador uma amostra-testemunho, após o que lhe é solicitado que identifique qual de outras duas amostras é igual à testemunho. Mais uma vez, para evitar o enviezamento das respostas, deve alertar-se o provador para o facto de uma amostra ser igual à testemunho e a outra ser diferente, excluindo a possibilidade de as amostras serem todas iguais.

A eficiência estatística deste teste é muito reduzida, sendo a probabilidade de o provador acertar respondendo ao acaso de 50%. As várias réplicas que se deverão fazer do teste podem ser elaboradas sempre com o mesmo produto como testemunho, caso em que o teste se designa de referência constante. Alternativamente, podem usar-se os dois produtos em teste como testemunho, situação em que o teste se intitula de referência balanceada.

Normalmente é aconselhado o uso da referência constante, sendo que, quando os provadores estão, à partida, familiarizados com um dos produtos em causa, deve ser este o escolhido para referência (Stone e Sidel (1993, p. 150) e Lawless e Heymann (1999, p. 122)).

Uma das situações em que este teste é recomendado é quando se pretende introduzir modificações na composição de produtos que já se encontram à venda no mercado, sem que os consumidores se apercebam da diferença.

2.4.1.4 Teste da Diferença Simples

Meilgaard et al. (1987a) referem que o teste da diferença simples deve ser usado para determinar se dois produtos são ou não diferentes do ponto de vista sensorial e quando não for conveniente utilizar o teste triangular ou o duo-trio, o que acontece, por exemplo, em sabores muito fortes ou em amostras de produtos cosméticos para colocar na pele. Geralmente este método é utilizado colocando a questão a muitos provadores.

O teste da diferença simples consiste em apresentar aos provadores vários pares de amostras (uns em que as duas amostras são do mesmo produto e outros em que as duas amostras são de produtos diferentes) e solicitar-lhes que se pronunciem sobre se as amostras são iguais entre si ou diferentes. Há quatro possibilidades para a ordem de prova das amostras (AA, AB, BA e BB) que devem ser testadas em igual número.

A probabilidade de um provador acertar respondendo ao acaso para este teste é, também, de 50% e, portanto, o teste deverá ser repetido para aumentar a eficiência estatística. A quantidade de testes a realizar por cada provador depende dos provadores em causa e do tipo de estímulo.

Note-se que a natureza deste teste não permite a estabilização do critério, pelo que se trata de um teste propenso ao enviezamento das respostas (não se sabe, à partida, quantos pares de amostras são iguais e quantos são diferentes).

2.4.1.5 Teste A-Não A

No teste A-Não A, é inicialmente apresentada uma amostra-testemunho (A), seguindo-se uma série de outras amostras que o provador deverá classificar como igual a A ou não igual a A. Na realidade, este teste não é mais do que uma sequência de testes de diferença simples, em que a amostra-testemunho só é apresentada uma vez no início do teste (e depois é retirada). Assim, o que foi dito em relação ao teste descrito no ponto anterior aplica-se, também, a este.

Uma variante do teste que, por vezes, se usa é a de apresentar, no início, duas amostras-testemunho: a "A" e a "não A".

Tal como acontece com o teste da diferença simples, este ensaio é propenso ao enviezamento das respostas (não se sabe, à partida, quantas amostras são A e quantas são não A).

2.4.1.6 Teste ABX

O teste ABX pode considerar-se o inverso do teste duo-trio. Enquanto no duo-trio é apresentada ao provador uma amostra-testemunho e é solicitado que identifique qual de outras duas amostras lhe corresponde, no ABX são apresentadas duas amostras-testemunho (A e B) e uma outra amostra (X) que deverá ser identificada como igual a A ou como igual a B.

A vantagem deste teste relativamente à maior parte dos ensaios de diferenciação é que o provador tem a oportunidade de descobrir, por ele próprio, as eventuais diferenças existentes entre os produtos, devido ao procedimento de "aquecimento" (*warm up*). A prova de X é, por conseguinte, mais dirigida. Trata-se, no entanto, de um teste propenso ao enviezamento.

A probabilidade da acertar quando não se detecta diferença é, tal como no duo-trio, de 50%.

2.4.1.7 Teste do Padrão Duplo (*Dual Standard*)

Neste ensaio são inicialmente apresentadas ao provador duas amostras-testemunho, uma de cada produto, sendo, em seguida, apresentadas outras duas amostras devidamente codificadas, também uma de cada produto. O provador deve associar cada uma das amostras codificadas a uma das testemunho, não podendo associar as duas à mesma testemunho.

Este teste tem a mesma vantagem do ABX, visto que também permite o "aquecimento", e tem a vantagem adicional de não ser propenso ao enviezamento das respostas. A probabilidade de responder certo ao acaso é de 50%.

2.4.1.8 Teste da Diferença em Relação ao Controlo

O teste da diferença em relação ao controlo consiste num teste A-Não A, onde é solicitado aos provadores que avaliem a magnitude da diferença entre os produtos, quando a encontram. Os provadores deverão ser alertados para o facto de algumas das amostras serem iguais à de controlo.

Meilgaard et al. (1987a) referem que os resultados deste teste permitem obter uma estimativa do efeito placebo, uma vez que é possível avaliar a amplitude da diferença percebida quando os produtos são iguais (A). No entanto, o facto de implicar uma avaliação da magnitude da diferença, normalmente feita numa escala multinomial, aponta no sentido de ser utilizado em situações onde os produtos em causa sejam suficientemente diferentes para que a escala utilizada seja adequada.

Os mesmos autores sugerem a utilização do teste da diferença em relação ao controlo nos casos em que o teste triangular ou o duo-trio não sejam aconselháveis devido à heterogeneidade natural dos produtos (tal como carnes, saladas e comidas cozinhadas).

2.4.1.9 Testes Sequenciais

Os testes sequenciais aproximam-se mais de um procedimento do que de um teste sensorial, tal como os que foram apresentados anteriormente. No entanto, visto que são, normalmente, analisados em conjunto com os testes de diferenças, optou-se por os incluir nesta secção.

Os testes sequenciais utilizam-se quando se pretende reduzir o número de repetições de um ensaio para retirar uma conclusão válida (com um nível de significância previamente estabelecido). Um exemplo será o de aceitar ou rejeitar um lote de produtos. Se os primeiros n estiverem "bons", não há necessidade de testar mais, pois, mesmo que os restantes m sejam "maus", o lote será aceite. Este tipo de procedimento é, pois, baseado em testes de diferenças para os quais exista uma resposta correcta (como o teste triangular ou o duo-trio).

Tal como o próprio nome indica, os testes sequenciais consistem na condução de uma série de testes iguais em que as respostas que o provador vai dando são imediatamente avaliadas (como correctas ou incorrectas) e processadas estatisticamente, de forma a verificar se a sequência prossegue ou se o teste acaba.

Em Meilgaard et al. (1987a, p. 74-77) são apresentados os procedimentos estatísticos a seguir, para conduzir correctamente este tipo de testes.

2.4.2 Ensaios de Diferenciação de Atributos

O objectivo dos testes de diferenciação de atributos consiste na detecção ou na medição da diferença sensorial existente entre amostras, relativamente a um atributo em particular.

Os resultados destes testes pretendem responder à pergunta: "Como é que o atributo X difere entre as amostras apresentadas?"

Note-se que as amostras podem diferir umas das outras de uma forma geral, mas são comparadas apenas na característica em questão.

Seguidamente, apresentam-se os testes de diferenciação de atributos mais referidos e estudados na literatura: 2-AFC, 3-AFC, ordenação por pares, comparação por pares de Scheffé, ordenação simples, cotação de amostras e testes que utilizam desenhos de experiências.

2.4.2.1 Comparação por Pares Direccional (2-AFC)

O teste de comparação por pares direccional, também denominado 2-AFC (*Two Alternative Forced Choice*), é um teste muito simples, no qual são apresentadas ao provador duas amostras (A e B) e se lhe solicita que seleccione aquela em que a intensidade do atributo em causa é maior. Exemplo: Qual é a mais doce: A ou B?

Como a probabilidade de acertar respondendo ao acaso é muito elevada (50%), o teste deve ser repetido várias vezes.

Este teste é robusto em relação à mudança de critério por parte do provador, pois este sabe exactamente que a uma amostra corresponde uma intensidade do atributo superior à outra. Já a variante do teste que permite ao provador dizer que não há diferença entre as amostras, colocando a hipótese de lhe serem apresentadas amostras iguais, não se deve utilizar, precisamente pelo facto de não estabilizar o critério de decisão.

O'Mahony (1995) refere que este teste é mais sensível do que o teste triangular e que obriga o provador a definir estratégias mais poderosas e mais estáveis, pelo que deve ser utilizado quando o tipo de comparação em causa o permitir.

2.4.2.2 Comparação Direccional de Três Amostras (3-AFC)

Trata-se de um teste idêntico ao teste triangular mas em que a natureza da diferença existente entre as amostras é previamente especificada.

O teste 3-AFC permite ao provador adoptar a estratégia *skimming*, em que procura "a amostra mais" em vez da estratégia de comparação de distâncias, em que procura "a amostra diferente". A primeira estratégia é mais eficiente. Para além disso, o ensaio 3-AFC evita o risco de o provador alterar a sua estratégia (da comparação de distâncias para a *skimming*) ao longo das repetições dos testes. Esta situação ocorre muito frequentemente quando, no teste triangular, o provador identifica a natureza da diferença em causa (ver O´Mahony (1995)).

2.4.2.3 Teste de Ordenação por Pares (Análise de Friedman)

Este teste utiliza-se quando se pretende comparar várias amostras (tipicamente entre 3 e 6) numa só característica, sendo indicado para casos em que os provadores não possuem muito treino, pois trata-se de um teste simples de executar.

São apresentados a cada provador todos os pares de amostras que é possível formar com os produtos em causa, e é-lhes solicitado que, para cada par, indiquem qual a amostra onde a intensidade da característica em apreciação é mais forte.

Os resultados são analisados através de um teste de hipóteses não paramétrico (T de Friedman), que se baseia no número de vezes que cada amostra foi eleita como mais intensa do que as outras (ver Meilgaard et al. (1987a, p. 88)).

Este teste não deve ser usado com produtos que causem muita fadiga sensorial, dado que exige que sejam provados todos os pares de amostras em análise.

2.4.2.4 Teste de Comparação por Pares de Scheffé

Trata-se de um procedimento semelhante ao teste anterior, mas em que, além de indicar a amostra mais intensa, o provador indica a magnitude da diferença numa escala multinomial (tal como acontecia com o teste de "diferença em relação ao controlo").

A utilização deste teste é aconselhada quando o objectivo é o de avaliar a extensão da diferença entre os vários produtos na característica em estudo.

Os resultados são analisados recorrendo à Análise de Variância de Scheffé, descrita em Scheffé (1952), onde se verifica se há diferenças significativas entre as intensidades da característica para as várias amostras, e a testes de Tuckey para separação de médias (onde se conclui quais são as amostras que diferem umas das outras, no caso de haver diferenças). Um exemplo da utilização deste teste pode ver-se em Meilgaard et al. (1987a, p. 92-95).

2.4.2.5 Teste de Ordenação Simples

O teste de ordenação simples é um teste fácil de executar e permite comparar vários produtos simultaneamente. A tarefa solicitada aos provadores é a ordenação das amostras dos produtos relativamente à intensidade da característica em estudo.

Este teste não permite quantificar as diferenças entre as amostras e não deve ser utilizado quando há "degraus" muito pronunciados entre as amostras no que diz respeito à intensidade da característica, ou seja, quando, por exemplo, há duas amostras com pouca diferença entre si e a amostra seguinte tem uma diferença muito maior. Os resultados podem ser analisados, mais uma vez, com recurso à análise de Friedman.

Pangborn (1984) analisa as vantagens e desvantagens deste método e aconselha a sua utilização em situações exploratórias, onde se pretenda adquirir informação para servir de base a estudos posteriores.

2.4.2.6 Teste de Cotação de Amostras

Tal como no caso anterior, este teste permite comparar várias amostras simultaneamente (tipicamente entre 3 e 6, no máximo, 8).

No teste de cotação de amostras pede-se aos provadores para atribuírem uma pontuação a cada amostra para a característica em estudo. As escalas utilizadas podem ser estruturadas como, por exemplo, as escalas tipo Likert, onde são usadas pontuações de 1 a 5 (ver Vavra (1997, p. 155)), ou não estruturadas, como, por exemplo, um segmento de recta onde apenas se identificam as extremidades e a resposta é assinalada graficamente.

Os resultados são avaliados com recurso à ANOVA tradicional, e, quando se rejeita a hipótese da igualdade da intensidade média da característica para as amostras em causa, podem obter-se estimativas da sua diferença através da determinação de intervalos de confiança pelo método de Tuckey.

Kim e O'Mahony (1998) fazem referência a outro tipo de teste, que é um misto de teste de ordenação e de teste de cotação. Neste teste, a numeração da escala é colocada em frente a cada provador, e é-lhe solicitado que ordene as amostras colocando-as à frente dos números que lhes correspondem. Os provadores podem provar cada amostra mais do que uma vez, o que é, aliás, característico de um procedimento de ordenação. Os autores apelidaram a escala referida como de *rank-rating*.

2.4.2.7 Testes que Utilizam Desenhos de Experiências

Estes testes são utilizados quando há uma quantidade grande de amostras para comparar, visto que a capacidade dos provadores é limitada. Normalmente, nos testes de comparação é aconselhado que cada provador avalie, no máximo, 6 produtos diferentes na mesma sessão. Quando o número de produtos é superior a 6 e inferior a 16, Meilgaard et al. (1987a) aconselham o uso de desenhos de experiências, nomeadamente *Balanced Incomplete Block Designs*, tanto para os testes de ordenação como para os de cotação.

O desenho de experiências aplicado à Análise Sensorial tem como objectivo reduzir o número de amostras que os provadores deverão ensaiar. Assim, em vez de serem apresentadas a todos os provadores todas as n amostras simultaneamente, apresentam-se em b blocos mais pequenos, cada um contendo $k < n$ amostras. Cada provador avalia um bloco e, como os blocos diferem uns dos outros, este tipo de abordagem requer um número considerável de provadores para que todos os blocos sejam devidamente testados.

As k amostras deverão ser seleccionadas segundo algumas regras, para garantir que todas as amostras são avaliadas o mesmo número de vezes e que todos os pares de amostras aparecem igual número de vezes no conjunto dos b blocos.

Em Meilgaard et al. (1987b, p. 85-96), são apresentados métodos para desenhar as experiências e para processar os seus resultados.

2.4.3 Ensaios para Determinação de Limiares

Tal como já foi referido em 2.3.1, os limiares sensoriais são uma medida da sensibilidade das pessoas aos estímulos. A sua utilização reveste-se de muita importância para várias áreas ligadas à investigação e à indústria alimentar.

Exemplos de aplicações deste conceitos são muito frequentes na literatura. Hough et al. (1995) utilizam os limiares individuais de detecção dos sabores básicos para seleccionar provadores profissionais a partir de um grupo de candidatos.

González-Viñas et al. (1996) determinam limiares de grupo para dois tipos de vinagre espanhol e para os ácidos cítrico e acético. Mais exemplos podem ser encontrados em Meilgaard et al. (1987a, p. 115) ou em Gatchalian et al. (1990).

A determinação dos limiares passa pela realização de ensaios sensoriais que os permitam detectar. Em 1979, a ASTM adoptou como método *standard* para a determinação de limiares o método ascendente dos limites (ASTM E679-91 (1997a)), que, posteriormente, foi também adoptado por outras organizações como a *American Society of Brewing Chemists* (ASBC) (Meilgaard (1991)).

Este método baseia-se numa sequência de ensaios de escolha forçada (n-AFC). Nessa sequência, as amostras são apresentadas por ordem crescente de concentrações e o limiar individual é calculado com base na última concentração que o provador falhou e na concentração seguinte.

Após o aparecimento de várias críticas, nomeadamente em relação à aproximação matemática utilizada (Meilgaard (1991)), este método passou a ser considerado pela ASTM como método rápido, e foi definido um outro método (ver ASTM (1997b)), onde o limiar é calculado com base na definição estatística apresentada em 2.3.1.

Em ASTM (1997b), são apresentados as definições e os procedimentos a seguir para calcular limiares individuais a partir dos resultados de ensaios 3-AFC (num mínimo de 20 testes por indivíduo). No mesmo texto, é ainda mencionado que, em princípio, os outros testes de diferenças, como o 2-AFC ou o duo-trio, também poderão ser usados na determinação de limiares.

Alguns autores chamam a atenção para o facto de os novos procedimentos envolverem um número muito elevado de ensaios, e sugerem que se utilizem os antigos, em alguns casos com pequenas alterações. Costell et al. (1994) e González-Viñas et al. (1998) apresentam e comparam vários métodos simplificados para avaliar limiares, e concluem que os seus resultados são semelhantes.

Relativamente ao conceito de limiar de grupo, a primeira norma referida define-o como a média geométrica dos limiares individuais, enquanto a segunda assume que deverá ser o valor que melhor representa a tendência central da distribuição dos limiares individuais dos elementos do grupo.

2.4.4 Ensaios Descritivos

Os ensaios descritivos são os ensaios sensoriais mais complexos e envolvem a detecção e a descrição das características dos produtos, sejam estas qualitativas ou quantitativas, por painéis de provadores treinados.

Os aspectos qualitativos referem-se à definição do produto em termos sensoriais e à sua descrição, nomeadamente em relação a aspectos que o diferenciam dos outros produtos. A selecção dos termos a usar para as características sensoriais nos estudos descritivos é essencial para o sucesso de um programa de Análise Sensorial. Lawless e Heymann (1999, p. 344) apresentam uma tabela que contém as características que os descritores deverão possuir, por ordem de importância.

A capacidade de discriminação, o facto de não ser redundante e a relação com a aceitação por parte dos consumidores e com as características fisico-químicas dos produtos são os factores mencionados nos primeiros lugares. A precisão do termo e o facto de o seu significado ser consensual são características também referidas como importantes no desenvolvimento do vocabulário a utilizar nos testes descritivos.

Para além deverem ser capazes de avaliar os produtos de uma forma qualitativa, os provadores deverão também conseguir quantificar a intensidade ou o grau de presença da cada característica nas amostras. Meilgaard et al. (1987b, p. 1) referem que, mesmo que dois produtos sejam descritos qualitativamente da mesma maneira, eles podem ser bastante diferentes, possuindo graus de presença de cada característica distintos.

As situações em que se utilizam os testes descritivos são muito diversas. No entanto, o objectivo da sua utilização é comum: a obtenção de descrições detalhadas das características sensoriais dos produtos. Por exemplo, no processo de desenvolvimento de um produto, é essencial concentrar a atenção em variáveis que são identificadas como diferentes em relação a outros produtos concorrentes (essas diferenças podem, eventualmente, ter sido previamente detectadas em testes de diferenças).

A necessidade de utilizar provadores treinados é óbvia: eles têm que entender o objectivo do estudo, saber usar a terminologia adequada, estar familiarizados com os produtos em estudo, conseguir utilizar devidamente as escalas (quando tal se revele necessário) e ser dotados de sensibilidade às características que se pretendem avaliar.

A temática da selecção e treino dos provadores e os cuidados a ter neste domínio quando se pretende utilizar os testes descritivos, são abordados no próximo capítulo (secção 3.4).

Seguidamente, são apresentados alguns dos ensaios descritivos mais utilizados: Perfil de Sabor (*Flavor Profile*), Perfil de Textura (*Texture Profile*), Análise Descritiva Quantitativa, Método Spectrum^{R}, Perfil de Escolha Livre (*Free Choice Profiling*) e Análise Descritiva de Tempo-Intensidade.

2.4.4.1 Perfil de Sabor

O método do perfil de sabor (FP) é a única técnica qualitativa formal e, porventura, a mais conhecida de entre as técnicas de análise descritiva (o termo qualitativo refere-se, aqui, à forma como são apresentados os resultados).

O FP é uma técnica de consenso que utiliza um painel de 4 a 6 provadores seleccionados e treinados. Na fase de treino, os provadores, com a ajuda do líder do painel, desenvolvem e definem a terminologia a ser adoptada. Na fase do teste, os provadores começam por avaliar, individualmente, o produto e registar os julgamentos sobre as características sensoriais ligadas ao sabor, as suas intensidades, a ordem por que surgem e a sensação de "fim-de-gosto". Os resultados são apresentados ao líder, que promove uma discussão geral entre o grupo com o objectivo de chegar a um consenso sobre o perfil do produto em análise e sintetiza os resultados na forma de um relatório.

As intensidades dos atributos são, normalmente, expressas numa escala de "0" (não está presente) a "3" (intensidade forte). O símbolo ")(" (limiar de detecção) faz, também, parte da escala, situando-se logo a seguir ao "0".

Uma extensão deste teste – a análise dos atributos do perfil (*Profile Atribute Analysis* – PAA) utiliza uma escala numérica para avaliar a intensidade percebida das características, o que permite que os resultados sejam analisados com o recurso à estatística.

Uma das críticas feitas FP é o facto de se tratar de uma técnica cujos resultados dependem muito do líder do grupo e da personalidade dos seus membros, dado que exige o consenso. Os defensores da técnica contra-argumentam que o líder tem que ter formação adequada e saber conduzir a discussão por forma a não permitir que as personalidades dos provadores interfiram no resultado do teste. Lawless e Heymann

(1999, p. 350) referem mesmo que esta técnica tem resistido, apesar de haver outras que caracterizam matematicamente os dados.

Neilson et al. (1988) apresentam uma descrição bastante pormenorizada tanto do FP como da PAA.

2.4.4.2 Perfil de Textura

O método de perfil de textura (TP) foi desenvolvido com o objectivo de permitir avaliar as características de textura de um produto desde o primeiro contacto até à mastigação completa (ver, por exemplo, Stone e Sidel (1993, p. 213)). Tal como acontece com o método anterior, baseia-se no julgamento feito por um painel de provadores, mas, neste caso, sobre as propriedades relacionadas com a textura dos produtos.

Os termos utilizados, embora específicos para cada tipo de produtos, baseiam-se no vocabulário expresso nas primeiras publicações sobre este método, citadas em Meilgaard et al. (1987b, p. 6) e são escolhidos por consenso entre os elementos do painel.

Embora inicialmente o TP tenha sido implementado com recurso a uma escala do tipo da referida para o perfil de sabor, mas com 13 níveis, actualmente é utilizado com vários tipos de escalas, desde escalas meramente categóricas a linhas contínuas.

Lawless e Heymann (1999, p. 356) referem que há escalas-padrão específicas para cada característica associada à textura e que, ao longo de cada escala, estão definidos valores ("âncoras") aos quais estão associados produtos que têm essa característica como uma das principais componentes. Note-se que, para graduar a escala com esses produtos, é feita uma avaliação instrumental da característica.

Os mesmos autores apresentam o exemplo de uma escala-padrão a ser usada em testes de dureza e que varia entre 1 (queijo creme) e 14.5 (rebuçado duro), passando por mais 6 "âncoras" intermédias.

A selecção dos provadores para os testes de perfil de textura é feita de acordo com as suas capacidades discriminatórias no tipo de estímulos em questão, e o seu treino é feito com vista a adquirirem um bom conhecimento sobre a textura do tipo de produtos em estudo e sobre o vocabulário e as escalas utilizadas.

As amostras são avaliadas por cada provador individualmente e, dependendo do tipo de escalas usadas, os resultados podem ser obtidos por consenso ou por tratamento estatístico.

2.4.4.3 Análise Descritiva Quantitativa

A análise descritiva quantitativa (QDA) foi desenvolvida com o objectivo de colmatar algumas das falhas detectadas nos métodos de perfil de sabor e similares, nomeadamente a de falta de tratamento estatístico. De facto, a natureza dos problemas sensoriais é particularmente adequada a análises estatísticas, dada a variabilidade inerente às diferentes sensibilidades.

Nesta técnica, as amostras são avaliadas por cada provador individualmente numa série de atributos previamente definidos pelo painel e com recurso a uma escala representada por um segmento de recta com os extremos "ancorados" por termos também definidos pelo painel. Cada provador deverá repetir a sua prova (de uma forma "cega"), a fim de se conseguir avaliar a sua consistência.

A selecção e o treino dos provadores são, mais uma vez, factores determinantes no sucesso deste método. Os cuidados a ter nas diversas fases não diferem substancialmente dos já referidos para os outros métodos de análise descritiva.

Os resultados são processados estatisticamente, sendo, normalmente, utilizadas as técnicas de análise de variância e de estatística multivariada.

A forma como os resultados são apresentados é, tipicamente, a de uma "teia de aranha", em que cada braço corresponde a um atributo e o seu comprimento corresponde à intensidade percebida. Pode ser feita uma "teia" para cada provador, ou uma "teia" média, ou ainda podem ser feitas "teias" em que os braços correspondam a factores encontrados através de técnicas como a análise factorial. Por vezes, inclui-se na representação gráfica uma estimativa da variabilidade associada a cada atributo/factor.

Na figura 2.4, pode ver-se um exemplo (retirado e adaptado de Lawless e Heymann (1999, p. 355)) de uma teia que corresponde à média das respostas dos provadores e que resulta de um estudo cujo objectivo é a caracterização e comparação de dois sabonetes em barra.

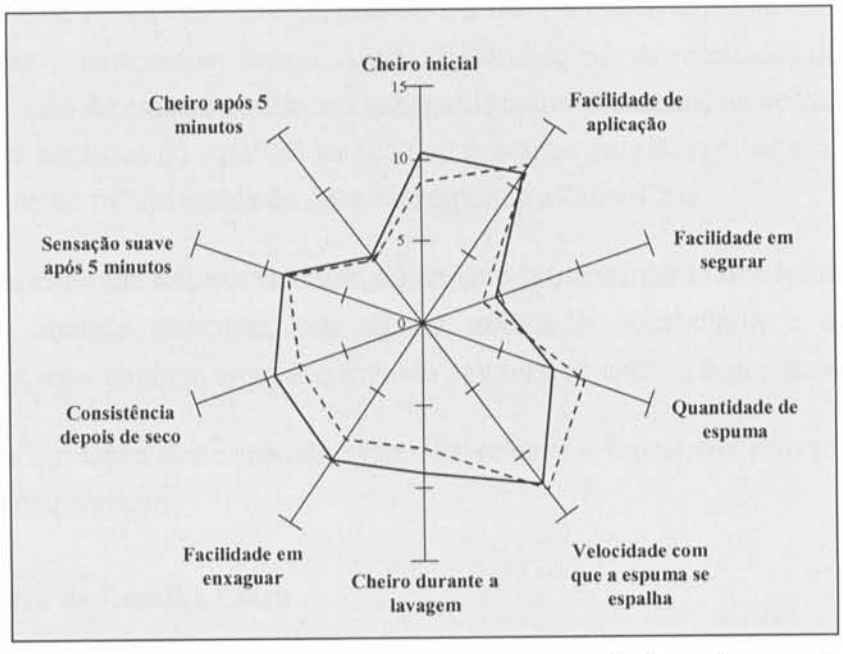

Figura 2.4 – Exemplo de um diagrama "teia de aranha" para os resultados obtidos com dois sabonetes (sabonete 1 – cheio; sabonete 2 - tracejado) (fonte: Lawless e Heymann (1999, p. 355))

Algumas das deficiências apontadas a este método são: (i) os provadores poderem desenvolver e utilizar vocabulário menos correcto; (ii) os provadores poderem utilizar a escala de forma diferenciada; (iii) o facto de a ausência de interacção imediata com os provadores reduzir as oportunidades de aprendizagem.

2.4.4.4 Método Spectrum^{R}

O método Spectrum^{R} foi desenvolvido com o objectivo de providenciar as ferramentas essenciais ao planeamento de uma experiência descritiva para um dado produto. Na prática, o método permite escolher o sistema adequado tendo em vista o produto em estudo, o programa sensorial no seu conjunto e o tipo de resultados pretendido.

A sua característica principal é a de que os participantes no painel não desenvolvem um vocabulário próprio para descrever as características sensoriais dos produtos, mas utilizam um conjunto de termos pré-definidos (ver Meilgaard et al. (1987b, p. 10-16)).

As escalas usadas assumem, tipicamente, a forma de um segmento de recta ao longo do qual existem várias "âncoras", tal como foi descrito em 2.4.4.2. Em Meilgaard et al. (1987b, p. 16-22), são apresentados várias "âncoras" que deverão servir como ponto de referência para construir as escalas quando se utiliza este método. Uma das vantagens associada ao Spectrum^{R} é a de as "âncoras" terem a função de calibrar os provadores,

que, treinados de forma conveniente, passam a utilizar a escala todos da mesma maneira e duma forma consistente no tempo. Assim, assume-se que os resultados dos testes que utilizam este tipo de escalas podem ser encarados como absolutos, no sentido em que se podem tomar decisões do tipo "só se aceitam produtos cuja doçura seja superior a 9". Registe-se que ao "9" na escala de doce corresponde a Coca-Cola.

O planeamento dos testes a elaborar, tal como o processamento dos resultados, ficam a cargo do analista sensorial, que deverá utilizar o vocabulário e as "âncoras" pré-definidos, mas também adaptar o método aos seus objectivos específicos.

A grande limitação deste método relaciona-se com o treino dos provadores, que se torna muito dispendioso.

2.4.4.5 Perfil de Escolha Livre

O método de perfil de escolha livre (*Free Choice Profiling*), segundo Stone e Sidel (1993, p. 236), apareceu, precisamente, para evitar a desvantagem referida no parágrafo anterior: o treino demorado dos provadores.

Neste método, os provadores são confrontados com as amostras e descrevem-nas com o seu próprio vocabulário, atribuindo-lhes, depois, uma pontuação (a escala usada pode ser definida por eles próprios ou pré-definida pela organização).

A análise dos resultados é feita recorrendo a uma técnica estatística – análise Procrustes generalizada (*Generalised Procrustes Analysis*), cujo objectivo é encontrar dimensões de consenso entre os provadores. No fundo, esta técnica permite "compatibilizar" os vários descritores utilizados pelos vários provadores, atribuindo um significado único a cada conceito expresso de forma diferente por cada provador. No ponto 3.2.3.2 faz-se uma descrição mais detalhada desta técnica.

Os resultados propriamente ditos são, então, expressos por um conjunto de factores devidamente classificados que resultam do consenso entre os provadores.

2.4.4.6 Análise Descritiva de Tempo-Intensidade

Este método difere dos que foram apresentados até aqui pelo facto de envolver mais uma variável: o tempo.

A forma como algumas características de alguns produtos evoluem ao longo do tempo é um aspecto essencial da sua apreciação. Um produto muito conhecido – a

pastilha elástica – é um exemplo do que foi dito: o tempo que o seu sabor se mantém é um factor importante na sua avaliação.

A análise descritiva de tempo-intensidade envolve o estudo das características e da evolução da sua intensidade ao longo do tempo. Normalmente, é solicitado aos provadores que se pronunciem (numa escala adequada) sobre a intensidade das características do produto no momento do consumo ou uso e em instantes subsequentes separados por intervalos de tempo iguais.

O número de características avaliadas simultaneamente não deve ser exagerado (entre 3 e 6) e os intervalos de tempo entre as apreciações deve ser estabelecido consoante o tipo de produto em questão (tipicamente entre 15 e 60 segundos).

Os resultados são apresentados sob a forma gráfica – curva tempo-intensidade (TI) – que tem o aspecto do representado na figura 2.5. Os resultados do estudo de tempo-intensidade são, habitualmente, expressos pelas variáveis indicadas no gráfico da referida figura: intensidade máxima (I_{max}), tempo para atingir a intensidade máxima (T_{max}), tempo de reacção, tempo para atingir metade da intensidade máxima na parte decrescente da curva e duração total do estímulo.

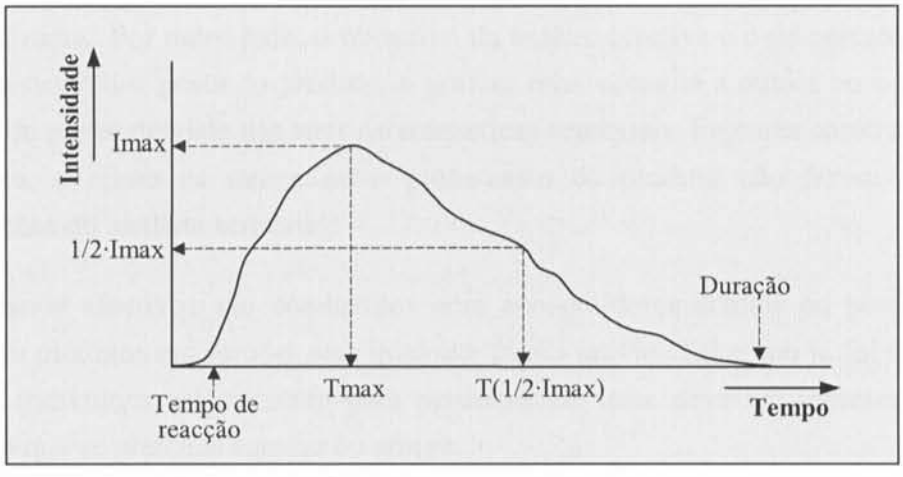

Figura 2.5 – Exemplo de uma curva tempo-intensidade (adaptada de Lawless e Heymann (1999, p. 267))

Lawless e Heymann (1999, p. 266), ao compararem os métodos em que a intensidade dos atributos é medida num só momento (como os referidos de 2.4.4.1 a 2.4.4.5) com o método tempo-intensidade, referem que os valores de intensidade que os provadores atribuem quando se utilizam os primeiros são muito correlacionados com os picos da curva TI, dando a impressão que os provadores retêm o máximo da curva e não fazem a sua integração.

Os mesmos autores, no Capítulo 8 do livro referido, discutem extensamente este método e apresentam várias técnicas que podem ser utilizadas para elaborar as curvas TI.

2.4.5 Ensaios Afectivos

Quando o objectivo do estudo se relaciona com a avaliação dos produtos por parte dos consumidores – o que, no ponto 2.2, foi apresentado como Análise Sensorial tipo I ou estudo hedónico ou afectivo – os testes sensoriais são classificados como testes afectivos.

As razões para conduzir um ensaio afectivo relacionam-se, normalmente, com o desenvolvimento de novos produtos, com a avaliação do potencial do mercado, com a melhoria do produto ou com a sua manutenção (ver Meilgaard et al. (1987b, p. 25)).

A distinção entre os testes afectivos e os efectuados nas pesquisas de mercado é evidenciada por Lawless e Heymann (1999, p. 430): os primeiros são conduzidos de uma forma cega (os inquiridos não sabem de que marca são os produtos que provam), enquanto os segundos são, frequentemente, efectuados com os produtos tal como são comercializados. Por outro lado, o objectivo da análise afectiva é o de perceber até que ponto o consumidor gosta do produto, o prefere relativamente a outros ou o considera aceitável do ponto de vista das suas características sensoriais. Factores como a intenção de compra, o efeito da marca ou o preço/custo do produto não fazem parte das preocupações do analista sensorial.

Os ensaios afectivos são conduzidos com consumidores actuais ou potenciais do produto ou produtos em estudo, sem qualquer treino prévio. Tal como já foi referido, o grupo de indivíduos seleccionado para participar no teste deve ser representativo da população que se pretende estudar ou atingir.

Há dois grandes grupos de testes afectivos: testes de preferência e testes de aceitação. Seguidamente, apresentam-se, de uma forma sucinta, os objectivos e os métodos referentes a cada um destes grupos.

2.4.5.1 Testes de Preferência

Os testes de preferência são utilizados quando o objectivo do estudo é determinar qual o produto preferido dos consumidores.

Quando apenas estão em causa duas amostras, o teste mais utilizado é o da comparação por pares, no qual se pergunta directamente aos consumidores qual das duas amostras preferem. Quando há mais do que duas amostras para avaliar, pode usar-se o método de comparação por pares de mais do que duas amostras, no qual as amostas são avaliadas aos pares, ou o método de ordenação das amostras, no qual os provadores ordenam os produtos pela sua preferência.

Os resultados deste tipo de testes são processados de uma forma semelhante aos dos testes de diferenças cujo princípio se assemelha.

2.4.5.2 Testes de Aceitação

Os testes de aceitação têm como objectivo determinar o grau de aceitação de um produto pelos seus consumidores. Meilgaard et al. (1987b, p. 32) dizem mesmo que este tipo de testes deve ser usado quando se pretende determinar o "estado afectivo" de um produto.

O ensaio habitualmente uilizado nestas situações é o da escala hedónica, em que é solicitado aos consumidores que exprimam o que sentem em relação ao produto numa escala que, tipicamente, tem 9 níveis. A escala é verbal e vai de "não gosto nada" até "gosto muitíssimo".

Frequentemente faz-se o teste a dois produtos simultaneamente: o produto que se pretende testar na realidade e um outro, referência, que se sabe, à partida, ser bem aceite pelo mercado. Desta forma, é possível comparar a aceitação dos dois produtos e analisar o posicionamento relativo do produto em estudo.

Os ensaios que acabaram de ser descritos são os mais frequentemente referidos na literatura da área da Análise Sensorial. O processamento dos seus resultados é, sem dúvida, uma das etapas determinantes para que as conclusões retiradas do estudo sejam válidas e úteis. Assim, no próximo capítulo abordam-se algumas das técnicas que são normalmente utilizadas para processar os resultados da maior parte dos ensaios aqui referidos.

3

A Análise dos Resultados e a Organização dos Ensaios em Análise Sensorial

Conteúdo:

- Introdução
- Análise Estatística dos Resultados dos Ensaios de Avaliação Sensorial
- Organização de um Programa de Análise Sensorial
- Selecção e Treino de um Painel de Provadores Profissional

3.1 Introdução

O facto de os provadores serem encarados na Análise Sensorial moderna como instrumentos de medida implica que haja, da parte de quem organiza os ensaios, um bom conhecimento dos cuidados a ter para que esses "instrumentos" meçam, de facto, o que se pretende medir e o façam da forma mais fiável possível.

No capítulo anterior, descreveram-se muitos dos ensaios sensoriais que os analistas podem utilizar. No entanto, para organizar um programa de avaliação sensorial, não basta, a partir dos seus objectivos, escolher os ensaios adequados. É necessário, também, saber como processar os seus resultados, como organizá-lo em termos físicos e como escolher e treinar os provadores. Neste capítulo abordam-se, precisamente, estes três últimos aspectos.

Na secção 3.2, apresenta-se um conjunto de técnicas que permitem processar os resultados da maior parte dos ensaios referidos no capítulo 2. Na secção 3.3, é abordada a questão da organização de um programa de Análise Sensorial, são apresentados alguns factores que podem influenciar a prova e são sugeridas algumas medidas para minorar o seu efeito negativo. Finalmente, na secção 3.4, discute-se a problemática da selecção e do treino de um painel de provadores profissional e a sua relação com a qualidade das decisões por ele tomadas.

3.2 Análise Estatística dos Resultados dos Ensaios de Avaliação Sensorial

Os métodos apresentados na secção anterior podem, de um ponto de vista estatístico, ser agrupados consoante o tipo de escalas que utilizam. Em Guimarães e Cabral (1998, p. 13-16), pode encontrar-se uma explicação dos vários tipos de escalas e das suas características.

Nas escalas nominais, os dados são apenas identificados pela atribuição de um nome que designa uma classe e as classes não são ordenáveis. As escalas ordinais diferem das nominais apenas pelo facto de se poder estabelecer uma ordem entre as classes. Essa ordem é estabelecida segundo um critério relevante que, em Análise Sensorial, corresponde frequentemente ao grau de presença da característica em estudo. No

método do perfil de sabor, por exemplo, as classes vão do "inexistente", a que corresponde o "0", à "intensidade forte", a que corresponde o "3".

O processamento estatístico de dados expressos em escalas nominais ou ordinais é limitado. A descrição das amostras é, normalmente, feita através de tabelas de frequências e/ou histogramas e a inferência estatística acerca da população é feita com recurso a testes de hipóteses sobre as proporções do número de casos obtidos em cada classe.

Os dados quantitativos podem ser expressos em dois tipos de escalas: de intervalo e absolutas. Nas escalas de intervalo, os dados são diferenciados e ordenados por números expressos numa escala cuja origem é arbitrária, pelo que se pode atribuir um significado à diferença entre os números, mas não à razão entre eles.

A utilização deste tipo de escalas em Análise Sensorial exige muito cuidado: de um ponto de vista psicométrico, é exigido aos provadores que a diferença que encontram entre as amostras se reflicta em graus iguais de diferença na escala sensorial (por exemplo, a diferença entre "3" e "6" é a mesma que entre "6" e "9"). Gay e Mead (1992) chamam a atenção para o facto de que, quando há dúvidas sobre se os provadores estão a utilizar correctamente a escala de intervalo, se deve analisar apenas a informação da ordem, ou seja, assumir que se trata de uma escala ordinal.

As escalas absolutas têm uma origem fixa, que corresponde à ausência completa da característica em estudo e, portanto, é possível atribuir significado à razão entre os números expressos neste tipo de escalas. Por exemplo, numa escala de distâncias, onde o zero corresponde a uma distância nula, pode dizer-se que 200 m corresponde ao dobro da distância de 100 m. Esta propriedade não se aplica ao caso anterior, onde não se pode afirmar que um produto classificado na doçura com "6" é duas vezes mais doce do que um classificado com "3". Note-se que as escalas absolutas raramente são usadas em Análise Sensorial.

O tratamento estatístico dos resultados dos testes depende não só do tipo de escalas utilizadas, mas também do próprio objectivo do estudo. Quando se pretende obter informação e tirar conclusões acerca de cada característica do(s) produto(s) individualmente, recorre-se à análise univariada – estuda-se uma variável de cada vez, ainda que essa variável se possa referir ao produto como um todo. Já quando se pretende estudar várias características em conjunto, estabelecer as relações e graus de semelhança entre elas ou relacionar e comparar as opiniões dos vários provadores, recorre-se à estatística multivariada.

Na maior parte dos testes de diferenças e de preferência são utilizadas escalas nominais ou ordinais e, dada a natureza destes testes – estuda-se um atributo em particular ou o produto no conjunto, sem relacionar as características umas com as outras – o tratamento estatístico dos resultados é univariado.

Já em relação aos ensaios descritivos e aos testes de aceitação, a situação é diferente: podem ser utilizadas escalas dos vários tipos referidos e o tratamento estatístico dos resultados pode ser univariado – no caso em que se pretenda descrever o produto característica por característica – ou multivariado – quando o objectivo é caracterizar o produto através das inter-relações dos seus atributos.

Uma das técnicas estatísticas univariadas mais utilizadas em Análise Sensorial são os testes de hipóteses.

Guimarães e Cabral (1998, p. 333) afirmam: "Os testes de hipóteses têm como objectivo fundamental verificar se dados amostrais (ou estimativas obtidas a partir deles) são, ou não, compatíveis com determinadas populações (ou com valores previamente fixados dos correspondentes parâmetros populacionais)" e, mais adiante, decompõem o procedimento básico da metodologia dos testes de hipóteses nas quatro fases seguintes:

(i) definição das hipóteses;

(ii) identificação da estatística de teste e da sua distribuição;

(iii) definição da regra de decisão, com especificação do nível de significância do teste;

(iv) cálculo da estatística do teste e tomada de decisão.

As hipóteses são definidas com base numa conjectura acerca da população ou populações em estudo (ou ainda dos seus parâmetros). Essa conjectura é normalmente consagrada na hipótese alternativa (H_1). Na hipótese nula (H_0) assume-se que tal conjectura é infundada.

H_0 é considerada verdadeira ao longo de todo o procedimento de teste. Nesse pressuposto, a estatística escolhida para o teste (estatística de teste – *ET*) segue uma determinada distribuição, conhecida à partida. A regra de decisão é estabelecida com base na probabilidade α (nível de significância do teste) de cometer um erro do tipo I: o de rejeitar H_0 quando ela é verdadeira. O valor escolhido para α varia, normalmente, entre 1% e 10%, sendo 5% o número mais utilizado em Análise Sensorial.

Com base na estatística de teste calculada a partir dos dados amostrais e face ao valor fixado na regra de decisão (valor crítico), decide-se pela rejeição, ou não, de H_0.

Na figura 3.1, ilustra-se o conceito de testes de hipóteses considerando uma situação fictícia onde se pretende verificar se dois grupos de biscoitos (A e B), confeccionados segundo o mesmo processo mas com quantidades ligeiramente diferentes de açúcar (A contém mais açúcar do que B), são percepcionados pelos consumidores como diferentes.

Assume-se que a amostra de provadores, constituída por N indivíduos, é representativa da população em questão e definem-se as hipóteses:

H_0: os consumidores não detectam a diferença entre os produtos;

H_1: os consumidores detectam que o produto A é mais doce do que o B;

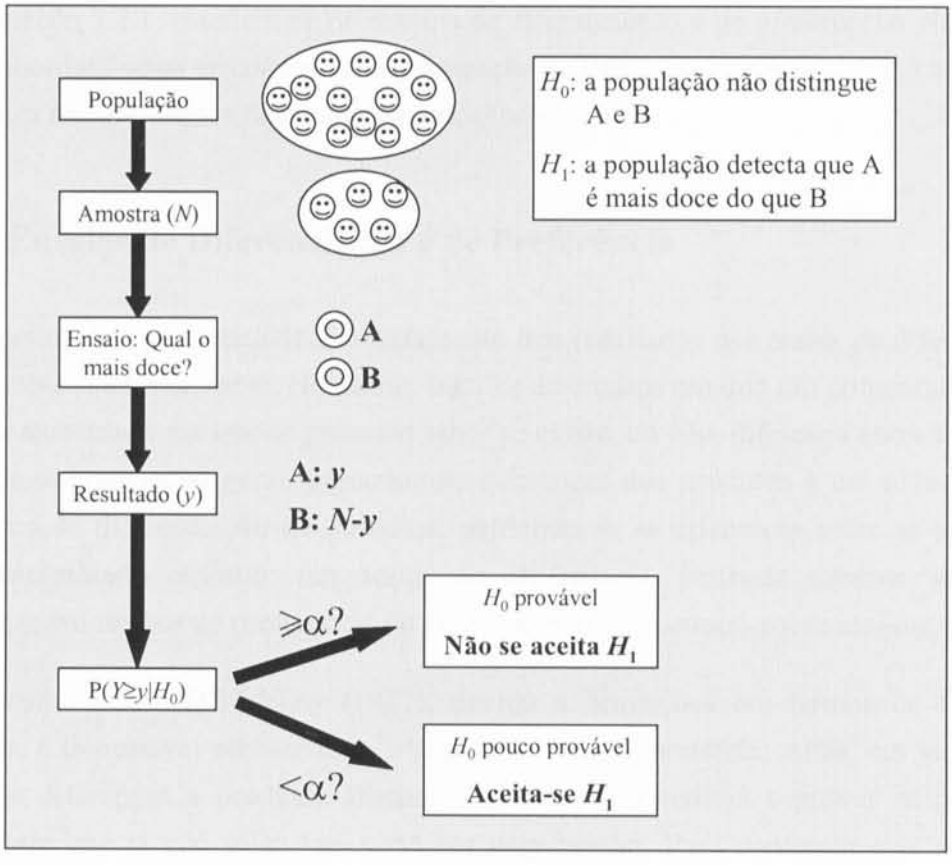

Figura 3.1 – A lógica dos testes de hipóteses aplicada a um caso em Análise Sensorial

Cada provador seleccionado é sujeito a um ensaio 2-AFC. Admita-se que a estatística de teste é Y e corresponde à variável "número de pessoas que consideram A mais doce do que B". Nestas condições, o valor que Y assume num ensaio em concreto denota-se

por y. Escolhido o nível de significância pretendido (por exemplo, $\alpha = 5\%$), determina-se a probabilidade de o resultado obtido (y) ser proveniente de uma população onde a proporção dos que distinguem A de B é de 50% (o que corresponde a H_0 verdadeira). Ou seja, verifica-se se $P(Y \geq y \mid H_0)$ é maior (ou não) do que $\alpha = 5\%$.

Se $P(Y \geq y \mid H_0)$ for inferior a 5%, rejeita-se H_0 e assume-se que os consumidores (a população) distinguem os produtos A e B. No caso de $P(Y \geq y \mid H_0)$ ser superior ou igual a 5%, não se coloca de parte a hipótese de os consumidores não distinguirem os produtos, pelo que não se pode concluir que os distinguem (ou seja, não se rejeita H_0).

Nas secções seguintes, onde se estudam e comparam os métodos estatísticos mais utilizados para processar os resultados dos diversos tipos de ensaios sensoriais, são analisados alguns testes de hipóteses em particular (teste à proporção binomial, teste à diferença entre valores esperados, testes ANOVA e alguns testes não paramétricos).

Na secção 3.2.1, estudam-se os ensaios de diferenciação e de preferência. Na secção 3.2.2 , abordam-se os ensaios para determinação de limiares, e na secção 3.2.3 estuda-se a temática das aplicações dos métodos descritivos.

3.2.1 Ensaios de Diferenciação e de Preferência

Do ponto de vista estatístico, o tratamento dos resultados dos testes de diferenças e de preferência é semelhante. No fundo, trata-se de ensaios em que são comparadas duas ou mais amostras e em que se pretende saber se existe, ou não, diferença entre elas. Nos testes de diferenciação geral, procuram-se diferenças nos produtos a um nível global; nos testes de diferenciação de atributos, exploram-se as diferenças entre as amostras num determinado atributo; nos testes de preferência, pretende saber-se quais as diferenças em termos de preferência de uma (ou mais) amostra(s) sobre a(s) outra(s).

Tal como refere O'Mahony (1982), devido a limitações em termos de tempo e dinheiro, é impossível estudar toda a população que se pretende. Aliás, em termos de testes de diferenças a produtos alimentares, tal corresponderia a provar os produtos todos, pelo que já não sobrariam produtos para vender. Para contornar esta situação, fazem-se as medições pretendidas numa amostra representativa e, com base nos resultados obtidos, infere-se acerca da população.

No caso dos testes de diferenças, a população são os produtos, os provadores são os instrumentos de medida e a amostra estatística são as amostras dos produtos em prova. Já no caso dos testes de preferência, a população são os consumidores – pretende

concluir-se sobre a sua reacção aos produtos – e a amostra estatística são os provadores que, portanto, devem ser representativos da população em causa.

Em muitos dos ensaios de diferenças e de preferência, os provadores limitam-se a dizer se encontram diferenças entre as amostras em prova ou a indicar qual ou quais as amostras diferentes (os dados encontram-se, portanto, em escalas nominais constituídas por duas classes – produtos iguais e produtos diferentes). Tal como foi salientado na apresentação destes ensaios, a sua eficiência estatística é bastante reduzida, dado que a probabilidade de uma pessoa acertar na resposta a um único ensaio respondendo ao acaso é, na maior parte das situações, bastante elevada.

A solução para o problema referido é, então, a repetição dos ensaios o número suficiente de vezes para que possam retirar-se conclusões válidas. A estratégia a adoptar para determinar o número de ensaios a que cada provador deve responder depende, além de outros factores, tais como a referida probabilidade e o nível de confiança que se pretende obter, do objectivo do estudo sensorial. No caso dos testes afectivos, cada provador é, normalmente, submetido apenas a um ensaio. No caso dos testes de diferenças, é frequente cada provador ser submetido a mais do que um ensaio, não só para aumentar a eficiência estatística da análise, mas também para permitir o estudo do comportamento individual dos provadores.

Na secção que se segue, descreve-se a técnica estatística mais utilizada no processamento deste tipo de resultados: o teste de hipóteses à proporção binomial.

3.2.1.1 Testes de Hipóteses à Proporção Binomial

Para que possa utilizar-se este teste, as respostas aos ensaios devem pertencer apenas a um de dois tipos (bom, mau ou doce, não doce, etc.) e ser independentes entre si. Desta forma, o processo em causa pode considerar-se uma experiência de Bernoulli (ver, por exemplo, Guimarães e Cabral (1998, p. 160)) e o número de vezes que o provador acerta (Y) em N experiências é uma variável aleatória que segue uma distribuição binomial com parâmetros N e p, sendo p a probabilidade de "sucesso".

Para ilustrar a aplicação do teste à proporção binomial a este tipo de ensaios, tome-se o exemplo do teste triangular (descrito em 2.4.1.1), onde o provador é forçado a responder qual a amostra diferente (em três amostras), mesmo que não encontre diferenças.

Admita-se que o provador não consegue distinguir as amostras e que, portanto, responde ao acaso. Num só ensaio, a probabilidade de acertar na amostra ímpar é de $p = 1/3$. Em $N = 10$ ensaios, a probabilidade de o provador acertar em todos é de aproximadamente 0.002% ($(1/3)^{10}$), mas a probabilidade de acertar em 5 ou mais, errando nos restantes, é de aproximadamente 21.31% ($\sum_{y=5}^{10} C_y^{10} \cdot (\frac{1}{3})^y \cdot (\frac{2}{3})^{10-y}$).

O valor de 0.002% (bastante menor do que 5%, e mesmo do que 1%), que representa a probabilidade de o provador identificar as 10 amostras ímpares sem as distinguir, é um valor muito baixo. No caso de o provador identificar, de facto, as 10 amostras ímpares, pode, então, concluir-se, com um grau de certeza grande, que o provador não se encontra a responder ao acaso. Já em relação ao valor de 21.31%, não se poderá afirmar o mesmo. O valor é suficientemente elevado para que não se possa excluir a possibilidade de o provador estar, de facto, a responder sem distinguir as amostras.

Formalizando o raciocínio exemplificado nos dois últimos parágrafos, e que explicita o funcionamento do teste de hipóteses à proporção binomial, tem-se:

H_0: $p = p_0 = 1/3$ (os provadores não distinguem as amostras)

H_1: $p > p_0$ ou $p > 1/3$ (os provadores distinguem as amostras, sendo capazes de identificar correctamente a amostra ímpar)

em que p é a proporção binomial da população e p_0 é valor de p correspondente à hipótese nula.

Quando a amostra é de pequena dimensão (normalmente considera-se $N < 20$), a estatística de teste (ET) é representada pelo número de vezes que ocorre a escolha do produto ímpar (Y). No caso de H_0 ser verdadeira, Y segue uma distribuição binomial com parâmetros (N, p_0). Uma vez obtido o valor da ET para a amostra (y), determina-se a probabilidade de ocorrer um valor igual ou superior ao que realmente ocorreu, continuando a assumir H_0 como verdadeira – $P(Y \geq y \mid H_0)$. O valor desta probabilidade, também chamado valor de prova do teste, compara-se directamente com o valor de α previamente estabelecido, retirando-se a conclusão devida – Se $P(Y \geq y \mid H_0) < \alpha$, rejeita-se H_0, caso contrário, não se rejeita H_0.

No caso de a amostra ser suficientemente grande ($N \geq 20$), utiliza-se o resultado do teorema do limite central (ver, por exemplo, Guimarães e Cabral (1998, p. 239-242)), e assume-se que, quando H_0 é verdadeira, a proporção binomial amostral segue uma distribuição aproximadamente normal cujo valor esperado é p_0 e cuja variância é

$\frac{p_0 \cdot (1-p_0)}{N}$. Nestas condições, a estatística de teste é dada por $ET = \frac{\frac{Y}{N} - p_0}{\sqrt{\frac{p_0 \cdot (1-p_0)}{N}}}$

que segue uma distribuição aproximadamente normal com valor esperado 0 e variância 1.

Na figura 3.2, representa-se a distribuição da ET referida, as regiões de rejeição e de não rejeição de H_0, a área correspondente a α e a localização de $ET(\alpha)$.

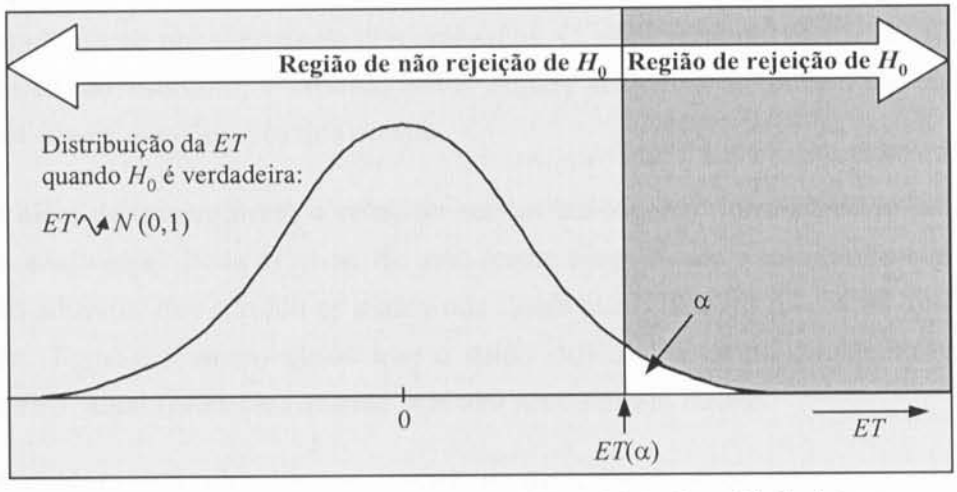

Figura 3.2 – Ilustração do teste de hipóteses à proporção binomial (unilateral à direita), amostras grandes (retirada de Guimarães e Cabral (1998, p. 338))

O facto de o teste efectuado ser unilateral tem a ver com a natureza do ensaio triangular: as pessoas ou não distinguem as amostras, e então a proporção binomial será 1/3, ou distinguem e a referida proporção é maior. A hipótese de se considerarem valores para p inferiores a 1/3 corresponderia a admitir que os provadores se encontravam de má fé e que, identificando a amostra diferente, escolheriam, sistematicamente, uma das outras. Dado que este comportamento não é, felizmente, nada habitual, considera-se que a natureza do teste de hipóteses apropriado para este ensaio é, efectivamente, unilateral.

Embora o processamento estatístico dos testes de diferenças e preferência aqui analisado tenha sido referente ao ensaio triangular, pode verificar-se facilmente que o teste de hipóteses à proporção binomial é directamente aplicável a outros ensaios sensoriais. Seguidamente, especificam-se os ensaios apresentados em 2.4, cujos resultados são susceptíveis de serem analisados através deste teste e, para cada grupo de ensaios, indicam-se as hipóteses a ser testadas:

- Teste dois em cinco – H_0: $p = 1/10$; H_1: $p > 1/10$;
- Testes duo-trio, diferença simples, A-Não A, ABX, 2-AFC – H_0: $p = 1/2$; H_1: $p > 1/2$;
- Teste 3-AFC – H_0: $p = 1/3$; H_1: $p > 1/3$;

3.2.1.2 Abordagem da Teoria da Detecção do Sinal

Na aplicação da teoria de detecção de sinal à Análise Sensorial, o corpo humano é considerado como um sistema de comunicações. O sinal sensorial (*sensory input*), que corresponde ao estímulo, é emitido pelos órgãos sensoriais e transmitido às células cerebrais (receptores) através dos nervos.

Para além de transmitirem o sinal, os nervos introduzem ruído aleatório no sistema (*background noise*). Pode dizer-se, de uma forma simplificada e assumindo que ruído e sinal são aditivos, que o ruído se traduz nos sinais que o cérebro recebe na ausência de estímulo. Torna-se, assim, óbvio que o ruído dificulta a tarefa de identificação do "verdadeiro" sinal.(estímulo) quando este tem níveis muito baixos.

Cliff et al. (1997) afirmam que a teoria de detecção do sinal aplicada à psicometria foi desenvolvida por Green e Swets com o objectivo de estabelecer uma medida do sinal sensorial necessário para que seja possível detectar a diferença entre produtos. O'Mahony (1992) salienta que a base teórica desta abordagem surgiu da necessidade de compreender a medição da sensibilidade sensorial humana. Embora, à primeira vista, possam parecer distintas, estas duas perspectivas referem-se ao mesmo fenómeno: quanto maior é a sensibilidade sensorial de uma pessoa, mais facilmente ela detecta "sinais" de fraca intensidade (ou pequenas diferenças entre produtos).

Quando, no capítulo anterior, foram apresentados os testes de diferenças, afirmou-se que o seu objectivo era o de responder à questão: "Será que existe alguma diferença a nível sensorial entre os produtos?". Considerando outra perspectiva da mesma questão, pode afirmar-se que, sendo os produtos diferentes do ponto de vista físico ou químico, os testes de diferenças têm como objectivo verificar até que ponto as pessoas (provadores ou consumidores) são sensíveis a essas diferenças. A ausência total de sensibilidade traduz-se na inexistência de diferença sensorial entre os produtos, e a presença de uma sensibilidade forte revela a existência de uma diferença relevante entre os produtos.

A medida da sensibilidade dos provadores às diferenças entre os produtos é, assim, simultaneamente, uma medida da própria diferença entre os produtos.

Neste contexto, a "sensibilidade" de uma pessoa é definida como a sua capacidade para discriminar entre dois ou mais estímulos (note-se que um dos "estímulos" pode corresponder à ausência de estímulo) (ver Irwin e Stillman (1996)).

Ao aplicar a teoria de detecção do sinal aos testes de diferenças, o sinal refere-se à diferença sensorial existente entre os dois produtos e o ruído corresponde à sensação provocada pela amostra proveniente do produto tomado como referência.

Embora os conceitos de sensibilidade do sistema sensorial e de critério (este último definido em 2.3.6.3) sejam, pela sua natureza, independentes, o critério utilizado pelo provador pode influenciar a medição da sua sensibilidade. A teoria da detecção do sinal possibilita uma interpretação simples destes dois conceitos e, consequentemente, ajuda a clarificar as causas da referida influência e as formas de a evitar.

Um sistema sensorial é tanto mais sensível quanto maior for o sinal recebido pelas células cerebrais, para um dado estímulo e um dado nível de ruído. O critério caracteriza-se pelo nível a partir do qual o provador interpreta o *input* como sinal e não como mera variação no ruído.

Na figura 3.3, ilustram-se os conceitos em análise com um exemplo simplificado e fictício, em que se consideram dois sistemas sensoriais distintos (X e Y) que são sujeitos aos mesmos estímulos. Os dois sistemas podem ser identificados com dois provadores distintos ou com o mesmo provador em momentos diferentes.

No eixo dos *XX* representa-se o tempo e no eixo dos *YY* a intensidade da sensação do provador. O ruído é representado na figura pela variação da intensidade da sensação ao longo do tempo, na ausência do sinal. Admite-se, duma forma irrealista, que os dois sistemas se encontram sujeitos ao mesmo ruído.

Os estímulos produzem os sinais 1 e 2 em X e Y e os níveis de critério utilizados são quatro (*A*, *B*, *C* e *D*), assumindo-se (mais uma vez de uma forma simplificada) que são os mesmos para os dois sistemas sensoriais. Note-se que estes níveis podem ser identificados com os representados na figura 2.3.

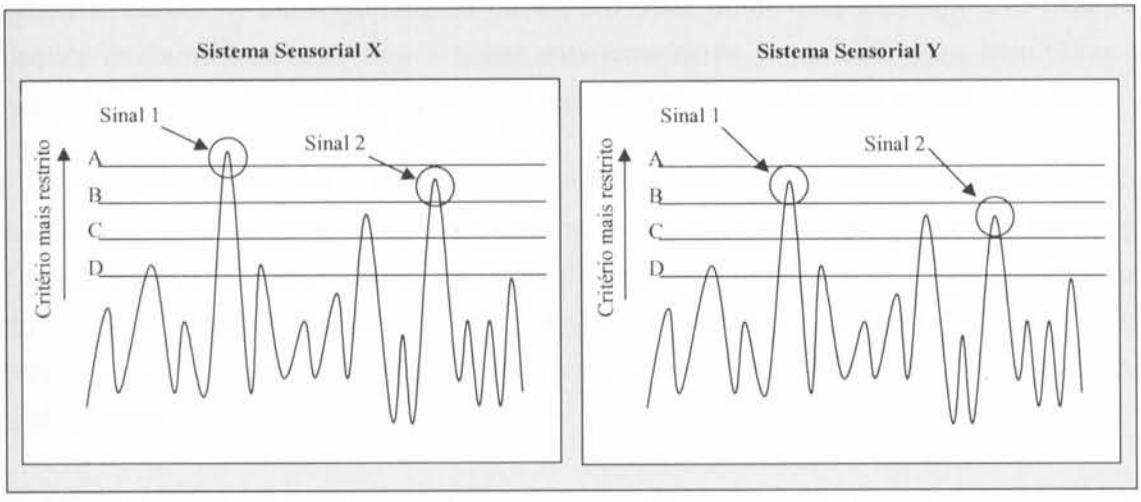

Figura 3.3 – Ilustração dos conceitos de sensibilidade e de critério

Analisando a figura, pode verificar-se que o sistema X é mais sensível do que o Y, dado que, para um critério restrito como o critério *A*, X detecta o sinal 1 (S_1), enquanto Y não, e para o critério *B*, X já detecta os dois sinais (S_1 e S_2), enquanto Y apenas detecta S_1. Quando o critério é mais lato (C ou D), tanto X como Y detectam S_1 e S_2, mas simultaneamente interpretam algum do ruído como sinal ("falso alarme").

O critério tende a variar de sessão para sessão e, por vezes, de ensaio para ensaio. Esta tendência, que foi já identificada em 2.3.6.3 como o enviezamento das respostas, não coloca problema nenhum quando a intensidade dos sinais é elevada (quando a diferença entre as amostras é grande). No entanto, em testes de diferenças e, muitas vezes, em testes de preferências, as diferenças que se pretendem medir são pequenas (o que corresponde a "sinais" de fraca intensidade) e a alteração do critério pode causar problemas, não permitindo que a sensibilidade seja medida com rigor.

O'Mahony (1992) aconselha duas alternativas para evitar alterações não controladas do critério de decisão dos provadores: (i) a utilização de procedimentos de escolha forçada, que permitem estabilizar o critério, ou (ii) a colocação das questões propositadamente enviezadas, fazendo variar o critério de uma forma controlada.

A primeira corresponde a utilizar, no estudo sensorial, apenas testes onde o provador saiba, à partida, quantas amostras correspondem a cada resposta que ele terá de dar (quantas são diferentes, quantas são mais doces, quantas são iguais a N, etc.). Este procedimento permite estabilizar o critério, porque evita indecisões. Considere-se o exemplo do teste 2-AFC, em que é solicitado ao provador que indique qual a amostra mais doce. Ao saber que uma amostra é mais doce do que a outra, o provador escolhe a que lhe parece mais correcta. Se lhe for colocada a questão sem excluir a hipótese de as

amostras serem as duas igualmente doces, ele pode notar uma pequena diferença na doçura das amostras, mas não a achar suficientemente importante para identificar a amostra como "mais doce" (confundindo, assim, o sinal + ruído com o ruído simples).

A segunda alternativa – utilização de critérios múltiplos – baseia-se numa técnica analítica, a análise ROC (*Receiver Operating Characteristic* ou *Relative Operating Characteristic*), que teve origem na teoria de decisão estatística e na teoria de detecção do sinal. O significado da expressão ROC, incluindo o da letra "R" que pode corresponder tanto à palavra "*Receiver*" como à palavra "*Relative*", serão explicados mais adiante.

Swets (1996a) faz uma retrospectiva histórica da evolução do estudo destes assuntos até à análise ROC, e apresenta os conceitos básicos que lhe são inerentes. Seguidamente, explicam-se, de uma forma simplificada e resumida, alguns destes conceitos e analisam-se as medidas da sensibilidade com eles relacionadas.

- **Modelo de Thurstone**

Thurstone, tal como refere Swets (1996a), deu um grande impulso ao estudo de indicadores para medir as diferenças sensoriais. A ideia por detrás do modelo criado por Thurstone é a de que, cada vez que um produto é provado, a intensidade do seu sabor varia. Os factores que influenciam essa variação são muitos, alguns dos quais foram apresentados em 2.3.6, mas é impossível eliminá-los e/ou controlá-los a todos.

Thurstone assume que a variação da intensidade do sabor de um produto ao longo de várias provas idênticas segue uma distribuição normal. Umas vezes, a intensidade é mais forte, outras vezes, mais fraca, e há uma intensidade média, em torno da qual se concentra o maior número de ocorrências.

A figura 3.4 traduz a interpretação do fenómeno ilustrado pela figura 2.3 à luz do modelo de Thurstone e representa a distribuição de frequências para a intensidade do sabor de dois produtos ligeiramente diferentes (P_1 e P_2, por exemplo), avaliados em testes de diferenciação. N e S correspondem à intensidade da sensação do provador, respectivamente para P_1 e P_2. A sensação N pode ser considerada como ruído, e a sensação S, como o mesmo ruído, ao qual foi adicionado o sinal.

Pode verificar-se que as distribuições N e S se sobrepõem parcialmente, havendo a possibilidade de o provador confundir os dois estímulos (S estará, portanto, na região de incerteza representada na figura 2.3).

Note-se que este modelo é generalizável aos outros tipos de estímulos sensoriais diferentes dos relativos ao sabor.

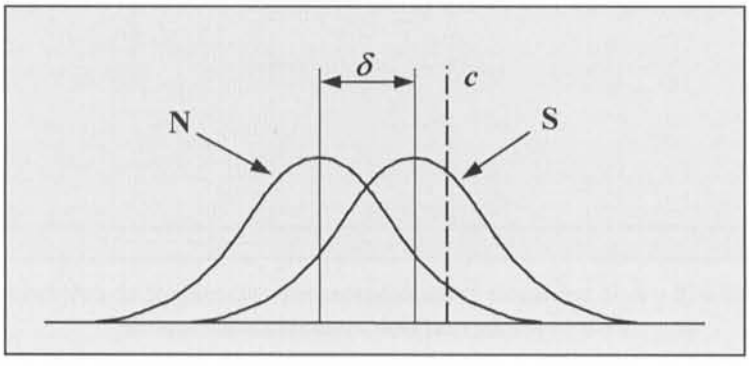

Figura 3.4 – Distribuições de frequências que representam as sensações (N e S) criadas por dois estímulos sujeitos a confusão, segundo o modelo de Thurstone (adaptada de O'Mahony et al. (1994))

Na figura 3.4, também se encontram representados o indicador δ e a linha de critério c. O indicador δ é a distância entre as intensidades médias das duas sensações, medida em unidades de desvio padrão (que se assume igual para as duas distribuições). $\delta = 1$ corresponde, assim, a uma diferença entre as intensidades igual a um desvio padrão, e $\delta = 2$, a uma diferença duas vezes superior ao desvio padrão. Note-se que, neste último caso, as duas distribuições se sobrepõem menos do que no primeiro, o que significa que os estímulos se distinguem mais facilmente.

A linha de critério representa apenas a localização do critério do provador, e pode variar, tal como já ficou explícito, de prova para prova. Uma alteração do critério corresponde a uma deslocação de c ao longo do eixo dos XX – quanto mais para a direita, mais restrito é o critério.

Do que foi dito e da observação da figura 3.4, pode inferir-se que δ é uma medida da sensibilidade à diferença entre dois estímulos que não é afectada por alterações arbitrárias do critério, podendo confirmar-se que, seja qual for o critério utilizado pelo provador (imaginando a linha c a movimentar-se ao longo do eixo dos XX), o valor de δ não se modifica.

Na figura 3.5, apresenta-se o exemplo (mais uma vez irrealista) de dois provadores (X e Y) que respondem ao produto P_1 com a sensação N. No entanto, X responde a P_2 com S, mas Y responde a P_2 com S', que tem a mesma forma de S, mas cujo valor esperado é superior.

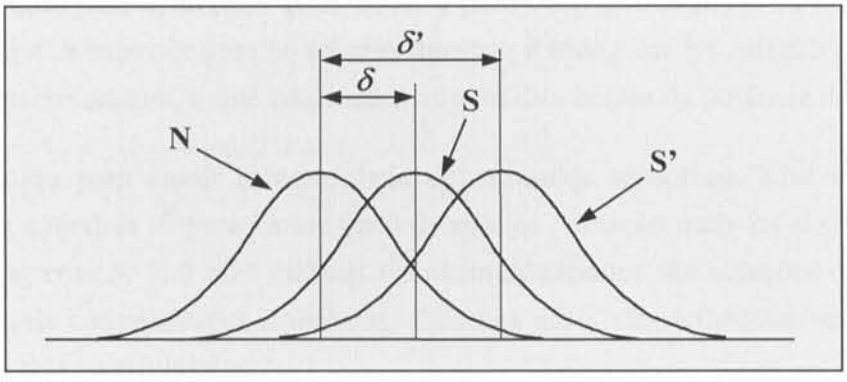

Figura 3.5 – Distribuições de frequências que representam as sensações N, S e S' e ilustram a diferença de sensibilidades entre dois provadores (δ e δ')

É óbvio que Y é mais sensível do que X à diferença entre os estímulos proporcionados por P_1 e P_2 pois, para um dado critério, a probabilidade de Y confundir P_1 e P_2 é inferior à de X os confundir. A diferença entre as duas sensibilidades traduz-se na diferença entre δ' e δ, como se pode verificar na figura 3.5.

O parâmetro δ pode ser estimado directamente a partir dos resultados dos ensaios de diferenciação, assumindo que tanto a distribuição do ruído como a do sinal + ruído são normais e têm a mesma variância. A estimativa de δ é representada por d'.

Em O'Mahony (1992), encontra-se uma explicação pormenorizada da maneira de calcular o valor de d' e Ennis (1993) apresenta tabelas que fornecem o valor de d' directamente a partir da proporção de respostas correctas obtidas para os ensaios triangular, 3-AFC, duo-trio e 2-AFC. As tabelas são diferentes consoante o tipo de ensaio em causa, facto que será justificado mais adiante.

No mesmo artigo, Ennis (1993) apresenta, também, tabelas que fornecem o número de testes que se devem realizar, em função de δ, para obter uma dada potência. Pode verificar-se que, para os ensaios duo-trio e triangular, o referido número é superior ao correspondente aos ensaios 2-AFC e 3-AFC, respectivamente, diminuindo a diferença entre eles à medida que δ aumenta. Conclui-se, assim, que os testes 2-AFC e 3-AFC são bastante mais potentes do que os duo-trio e triangular, respectivamente.

O facto de d' ser uma estimativa do parâmetro δ calculada com base em proporções amostrais, implica que lhe esteja associada uma variância. Bi et al. (1997) apresentam várias tabelas que permitem calcular a variância de d' para os ensaios referidos no parágrafo anterior e afirmam que o conhecimento de d' e da sua variância permite comparar vários d's, assim como estabelecer intervalos de confiança e elaborar testes de hipóteses sobre o parâmetro original (δ).

Note-se que, para o mesmo valor de *d'* e para o mesmo número de ensaios (*N*), a variância de *d'* é superior para os ensaios duo-trio e triangular em relação aos 2-AFC e 3-AFC, respectivamente, o que confirma o que foi dito acerca da potência destes testes.

Num estudo para medir a intensidade de estímulos sensoriais, Kim et al. (1998) determinam a medida *d'* para vários tipos de escalas - cotação tradicional com 9 classes e *rank-rating* com 5, 7, 9 e 15 classes. Os estímulos usados são soluções de cloreto de sódio com seis concentrações diferentes, duas das quais são suficientemente próximas para poderem ser confundidas.

As conclusões do estudo apontam para que ao procedimento *rank-rating* estejam associados valores de *d'* mais elevados do que à escala tradicional, o que será devido à redução do número de erros por falta de memória (note-se que as amostras podem ser provadas mais do que uma vez). Por outro lado, nas escalas *rank-rating* com menos classes, verificou-se um efeito de *bunching*, que consiste na tendência para os provadores atribuírem sempre a classificação mínima aos estímulos mais fracos e a máxima aos mais fortes. Esta convergência prejudica a avaliação de *d'*, que pressupõe que haja variabilidade nas classificações e que estas sejam normalmente distribuídas (e, portanto, não de forma assimétrica ou bimodal). Este problema coloca-se, igualmente, quando se pretende processar os dados com recurso à análise de variância.

O parâmetro δ (ou melhor, a sua estimativa *d'*) foi a primeira medida da sensibilidade do provador associada à teoria de detecção do sinal. Tem, no entanto, algumas desvantagens, das quais se salientam: (i) a sua validade como medida da sensibilidade está condicionada à normalidade das distribuições de frequências de N e de S; (ii) pressupõe a aditividade de sinal e ruído; (iii) o seu valor é expresso em múltiplos do desvio padrão do ruído, o que dificulta o seu entendimento e a comparação entre vários *d'*.

Finalmente, deve ser referido que há outra forma de calcular *d'* diferente da aqui descrita envolvendo o desenho da curva ROC (ver secção seguinte), e que é possível generalizar o conceito de *d'* a situações em que não há normalidade de N e de S. O'Mahony (1992) descreve algumas dessas situações.

- **Curva ROC**

A análise ROC é aplicada em vários domínios científicos, dos quais se salientam as Ciências da Saúde e a Análise Sensorial. Nas Ciências da Saúde é utilizada frequentemente para comparar o desempenho de meios de diagnóstico ou para decidir quais os critérios a adoptar em casos em que o diagnóstico é duvidoso. Metz (1978)

explica os princípios básicos desta técnica e a sua aplicação a algumas destas situações. Swets et al. (1979) apresentam um caso prático de comparação entre duas técnicas radiológicas, Rifkin et al. (1990) comparam as técnicas de Ressonância Magnética Nuclear e de Ecografia na detecção precoce de cancros e Braga et al. (1997) comparam o desempenho de unidades de cuidados intensivos neonatais de quatro hospitais.

Por vezes, a terminologia utilizada nas Ciências da Saúde é diferente da utilizada em Análise Sensorial, nomeadamente no que diz respeito ao conceito de sensibilidade. No entanto, uma vez identificados os sinais que se pretendem comparar e o ruído, a análise é semelhante nos dois casos, encontrando-se mais explorada na literatura científica médica.

Tradicionalmente, em Análise Sensorial, o ensaio em que assenta a construção da curva ROC é o teste A-Não A. A forma de o provador responder a cada pergunta é, todavia, ligeiramente diferente da descrita em 2.4.1.5. O provador associa a cada julgamento "igual ou diferente de A" um grau de certeza. Admitindo quatro níveis de certeza, as respostas possíveis para cada teste são, então, as seguintes: (i) tenho a certeza de que a amostra é diferente da A (*different-sure*); (ii) a amostra parece-me diferente de A, mas não tenho a certeza (*different-not sure*); (iii) a amostra parece-me igual a A, mas não tenho a certeza (*same-not sure*); (iv) tenho a certeza de que a amostra é igual a A (*same-sure*). O'Mahony (1992) refere-se a este procedimento como a utilização de uma escala de "cotação"[1] (*rating*), mas sublinha a diferença entre esta e as escalas convencionais de cotação.

Considere-se o exemplo já referido (ver figuras 2.3 e 3.3), em que N e S correspondem aos biscoitos "não doce" e doce", respectivamente, e admita-se que são suficientemente próximos para suscitar alguma confusão no provador. Ao utilizar o teste A-Não A nesta situação, o estímulo A corresponde a N e o estímulo Não A corresponderá a *S*.

Na figura 3.6 (i), ilustra-se a forma como se controla a variação do critério através de um exemplo possível de respostas a um destes ensaios. Por uma questão de simplificação da notação, admita-se que à resposta "diferente de A" corresponde a letra S (identificação do sinal), e à resposta "igual a A" corresponde N (não identificação do sinal) e que os julgamentos em que o provador manifesta incerteza são representados pela respectiva letra seguida de um ponto de interrogação (S? e N?, respectivamente).

---

[1] O termo cotação entre aspas refere-se, daqui em diante, a este tipo de escalas (julgamento com grau de certeza)

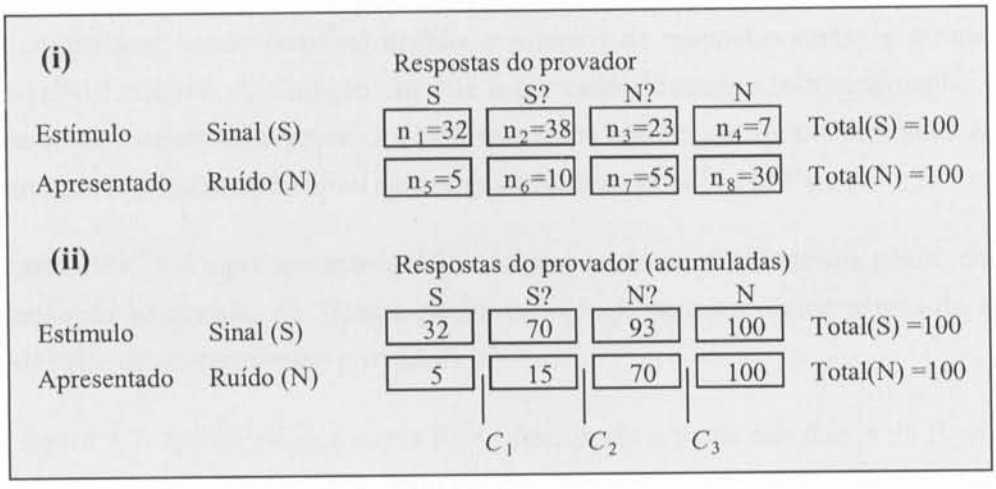

Figura 3.6 – Tabelas de respostas que ilustram o procedimento de escalas de "cotação" na teoria de detecção do sinal (adapatado de O'Mahony (1992))

Analisando a figura 3.6 (i), pode verificar-se que, em 200 testes, dos quais 100 correspondem a ruído e os outros 100 a sinal, o provador acertou com certeza em 32 sinais e em 30 ruídos, tendo duvidado em (38 + 23) sinais e em (10 + 55) ruídos, e ainda tendo errado com certeza em 7 sinais (que identificou como ruídos) e 5 ruídos (que identificou como sinais).

Se o provador utilizar um critério restrito para responder à questão "será que a amostra é diferente de A?", considera que apenas são realmente diferentes aquelas amostras para as quais tem a certeza. Na dúvida, assume que as amostras são iguais a A. Este critério corresponde a C_1 na figura 3.6 (ii). Pode verificar-se que, no caso de C_1 ser aplicado, o provador identifica o sinal em 37 testes, 5 dos quais se referem, na realidade, a ruído. Os resultados de todos os outros testes (163) são considerados como ruído, embora 68 deles sejam, na realidade, sinal.

Se o critério utilizado for lato, o provador assume que as amostras são diferentes de A, desde que não tenha a certeza de que são iguais. Na figura, este critério corresponde a C_3. No caso de C_3 ser utilizado nos 200 testes, 163 deles (93 + 70) serão classificados como sinal, sendo apenas 37 ($n_4 + n_8$) classificados como ruído. Note-se que destes, 7 são sinal.

O critério moderado (C_2) equivale a assumir que se identificam os resultados S e S? com o sinal e os resultados N e N? com o ruído. Assim, aos 200 testes correspondem 85 respostas sinal (70 + 15) e 115 respostas ruído ($n_3 + n_4 + n_7 + n_8$). Das 85 respostas sinal, 15 são, de facto, ruído e das 115 respostas ruído, 30 são, efectivamente, sinal.

A utilização da escala de "cotação" permite ao analista fazer variar o critério de uma forma controlada, sendo possível definir o número de respostas certas e erradas para cada nível do critério. A situação em que o provador identifica, correctamente, o sinal designa-se por verdadeiro positivo (*TP*), enquanto aquela em que o provador assume, erradamente, a existência do sinal se designa por falso positivo (*FP*).

A curva ROC é a representação gráfica da proporção de "verdadeiros positivos" (p_{TP}) em função da proporção de "falsos positivos" (p_{FP}), para os vários níveis do critério utilizados por um determinado provador.

Na figura 3.7, apresenta-se a curva ROC desenhada a partir dos dados da figura 3.6 e na qual os pontos são unidos através de segmentos de recta. Trata-se de uma aproximação, visto que, na realidade, a ROC é uma linha curva (este assunto será retomado mais adiante nesta secção).

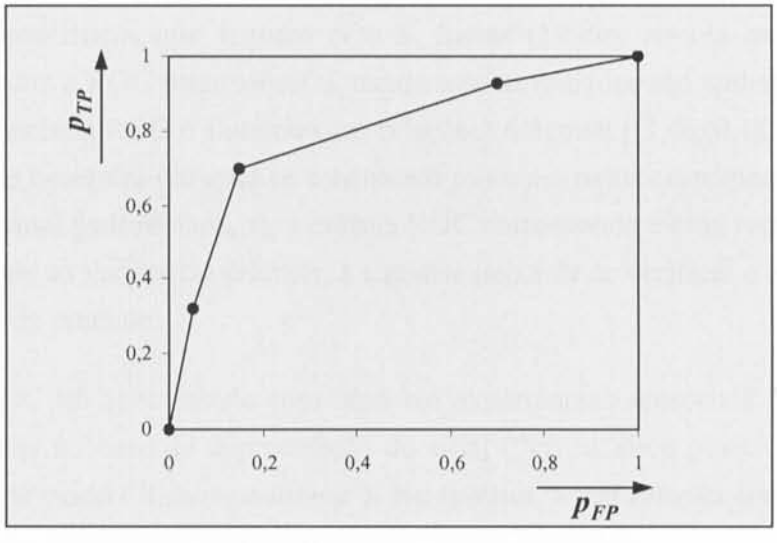

Figura 3.7 – Curva ROC desenhada a partir dos dados da figura 3.6

O ponto (0,0) corresponde à situação em que o critério é o mais restrito possível: o provador identifica todas as amostras com o ruído e, portanto, não acerta em nenhum sinal, mas também não erra em nenhum ruído. O ponto (1,1) corresponde à situação contrária: o provador reconhece o sinal em todos os testes (mesmo naqueles que não são), acertando em todos os sinais, mas errando em todos os ruídos.

O ponto (0.05, 0.32) corresponde ao critério C_1 da figura 3.6, onde a proporção de "falsos positivos" é de apenas 5%, e a proporção de "verdadeiros positivos" é de 32%. Ao considerar a evolução para critérios mais latos (pontos sucessivamente mais para a direita), pode verificar-se que tanto p_{FP} como p_{TP} tomam valores cada vez maiores. A explicação para este facto torna-se mais clara ao analisar a figura 3.4. Quando se

considera a evolução do critério no sentido de o relaxar, o que corresponde a mover a linha c para a esquerda, a identificação do sinal é cada vez mais provável, correspondendo-lhe um aumento natural de p_{TP}; no entanto, verifica-se que p_{FP} também aumenta, pois também se torna mais provável que o ruído seja detectado como sinal.

A curva ROC tem um aspecto côncavo e traduz a sensibilidade do provador. Ao examinar de novo a figura 3.7, pode concluir-se que o provador é tanto mais sensível quanto mais arqueada for a curva: para uma dada proporção de "falsos positivos", a sensibilidade do provador é tanto maior quanto maior for a proporção de "verdadeiros positivos". A um provador que não tenha capacidade para detectar o sinal em estudo, corresponde uma recta a 45º, pois as duas proporções representadas tendem a ser iguais, seja qual for o critério utilizado. Ao provador com o máximo de sensibilidade e que, portanto, detecta sempre o sinal, corresponde a recta horizontal $p_{TP} = 1$.

Relativamente à forma da curva ROC, deve dizer-se que ela depende do tipo de distribuições estatísticas que seguem N e S. Swets (1996b) aborda extensamente as várias formas que a ROC pode tomar. Quando as distribuições são ambas normais com a mesma variância, a ROC é simétrica em relação à diagonal ((1,0),(0,1)). Se se utilizar a representação binormal (na qual se substituem p_{FP} e p_{TP} pelos correspondentes valores da variável normal padronizada, z), a mesma ROC corresponde a uma recta com declive unitário. Quando as variâncias diferem, a simetria deixa de se verificar e o declive passa a ser diferente da unidade.

A curva ROC foi apresentada com base em experiências sensoriais e contagens de resultados certos no caso de apresentação do sinal ("verdadeiros positivos") e errados nas situações de ruído ("falsos positivos"). No entanto, a sua relação com o modelo de Thurstone e com os testes de hipóteses não foi, ainda, discutida.

Analise-se, de novo, a figura 3.4, comparando-a com a figura 3.8, onde se representa um teste de hipóteses ao valor esperado de uma distribuição normal (μ), considerando uma hipótese alternativa particular (H_1: $\mu = \mu_1$). As semelhanças são óbvias. À hipótese nula corresponde a ausência de sinal (N) e à hipótese alternativa, uma determinada intensidade do estímulo (S). Considerando uma dada observação (x) que corresponde a uma prova, e dependendo da localização do critério do provador, μ é considerado igual a μ_0 ou a μ_1.

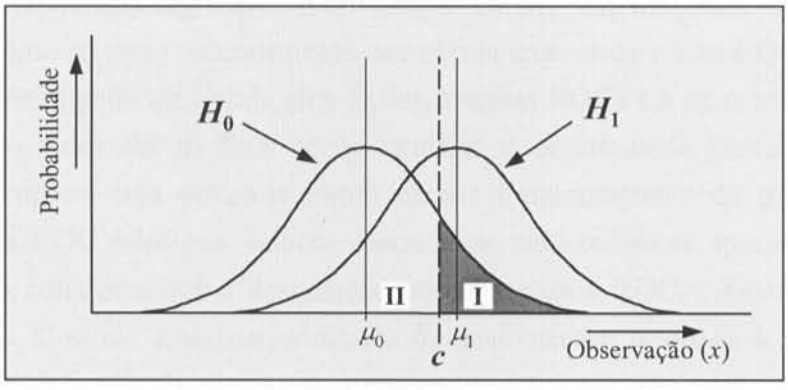

Figura 3.8 - Teste de hipóteses ao valor esperado de uma distribuição normal (adaptado de Swets (1996a))

O critério de decisão do provador é equivalente, na linguagem dos testes de hipóteses, à linha de divisão entre as regiões de rejeição e de não rejeição de H_0 (linha c na figura 3.8). Assim, o erro do tipo I (α), que é a probabilidade de rejeitar H_0 indevidamente, corresponde à probabilidade de ocorrer um resultado falso positivo (p_{FP}), representada no eixo dos *XX* da curva ROC. O erro do tipo II (β) é a probabilidade de não rejeitar H_0 quando ela é falsa e, portanto, corresponde à ocorrência de um resultado falso negativo. O resultado verdadeiro positivo, cuja probabilidade é representada no eixo dos *YY* da curva ROC, é complementar do falso negativo, pelo que $p_{TP} = (1 - \beta)$.

A diferença $(1 - \beta)$ designa-se por potência do teste (note-se que representa a probabilidade de rejeitar H_0 quando ela é, efectivamente, falsa) e, em Estatística, é muito frequente representá-la graficamente em função da distância entre as duas hipóteses (δ, na figura 3.4), para um determinado valor de α. O gráfico resultante denomina-se curva da potência do teste.

Ao encarar o problema da discriminação sensorial como um teste de hipóteses, pode interpretar-se a curva ROC como sendo a representação gráfica da potência do teste em função de α, para um determinado valor de H_1. Assim, a curva ROC pode ser obtida através de várias curvas da potência do teste (para vários valores de α) e vice-versa (a curva da potência obtida através de várias ROCs).

Neste ponto, é importante realçar que quanto mais arqueada for a curva ROC maior será a capacidade de discriminação do provador. É neste sentido que se emprega o termo "sensibilidade" em Análise Sensorial.

Finalmente, salienta-se o significado da sigla ROC. As letras OC referem-se a característica operacional (*operating characteristic*) que, na linguagem dos testes de

hipóteses, é a curva que representa β em função de H_1, para um dado valor de α. Esta curva pode, como se disse anteriormente, ser obtida através de várias ROCs. Uma outra curva que também pode ser obtida através das mesmas ROCs é a de α em função de H_1 para um dado valor de β. Esta curva também é denominada curva característica operacional, embora seja utilizada muito menos frequentemente do que a primeira. Como a curva ROC relaciona valores destas duas características operacionais, Swets (1996a, p. 19) considera que a designação correcta para a ROC é *Relative Operating Characteristic*. O termo *Receiver*, utilizado frequentemente, refere-se à já mencionada perspectiva de detecção de sinais em sistemas de comunicação, onde a detecção faz parte do sistema receptor.

- **Medidas da Capacidade de Discriminação Baseadas nas Curvas ROC**

Swets (1996c) descreve e compara muitas das medidas da capacidade de discriminação que se baseiam nas curvas ROC. Algumas delas são criticadas pelo facto de implicarem, à partida, um critério em particular, o que, segundo o autor, não se coaduna com as situações encontradas na prática, estando, portanto, estas medidas sujeitas a instabilidade (ou ao enviezamento, tal como foi definido em 2.3.6.3). Tal é o caso da proporção de respostas correctas (P_c) ou da probabilidade, corrigida, de ocorrerem verdadeiros positivos (H_c).

Outras das medidas referidas por Swets (1996c) e que não implicam a adopção, à partida, de um dado critério, são d', Δm, η, LOR (*log odds ratio*) e Q. A medida d' é a mesma que foi referida anteriormente na apresentação do modelo de Thurstone e pode ser calculada através do grau de curvatura da ROC. A medida Δm é semelhante a d', sendo apenas introduzida uma correcção para situações em que os desvios padrão de N e S são diferentes. As medidas η e LOR são, também, semelhantes a d', sendo que se baseiam em distribuições logísticas em vez de distribuições normais. A medida Q é, igualmente, do mesmo tipo das anteriores e, segundo o autor, é também consistente com as distribuições logísticas, embora não pressuponha tanto rigor nas formas das distribuições.

O índice que Swets (1996c) recomenda e que refere ser o mais utilizado actualmente para medir a capacidade de discriminação em casos onde N e S seguem distribuições normais (podendo, no entanto, as variâncias diferir), é A_z. A_z corresponde à área debaixo da curva ROC que, obviamente, é tanto maior quanto maior a concavidade e, portanto, quanto maior a capacidade de discriminação.

Uma outra medida da capacidade de discriminação muito utilizada, particularmente em Análise Sensorial, é *P(A)*. Tal como A_z, esta medida é definida como a área que se encontra debaixo da curva ROC, mas quando os pontos são unidos por segmentos de recta. O índice *P(A)* tem a grande vantagem de ser independente da distribuição do sinal e do ruído.

Os índices A_z e *P(A)* têm um significado estatístico importante, que permite uma interpretação clara dos seus valores: são estimativas da probabilidade de o provador escolher o sinal num teste de comparação por pares (2-AFC). Note-se que *P(A)* é uma estimativa conservadora: dada a forma côncava da curva, a união dos pontos através segmentos de recta reflecte-se numa área menor do que a obtida através da curva propriamente dita.

Em Swets (1996d), são referidos vários programas informáticos para desenhar curvas ROC e para calcular os diversos índices de capacidade de discriminação com elas relacionados. Recentemente, Braga (2001) faz uma análise dos programas informáticos disponíveis para a elaboração dos cálculos associados à análise ROC. No mesmo trabalho, a autora apresenta um programa que desenvolveu com o objectivo de colmatar algumas das deficiências que detectou na referida análise.

É interessante verificar que um outro índice – o índice *R* – desenvolvido por Brown em 1974 (ver O'Mahony (1992)) com base em estudos sobre a memória conduzidos sem recorrer à teoria da detecção do sinal, coincide com *P(A)*. O índice *R* é apresentado como uma probabilidade e não como uma área e, segundo O'Mahony (1988), é mais fácil de conceptualizar e foi, por isso, melhor aceite na prática da Análise Sensorial.

A fórmula de cálculo do índice *R* a partir dos dados em escalas do tipo "cotação" e considerando a notação da figura 3.6 (i), é:

$$R = \frac{n_1 \cdot (n_6 + n_7 + n_8) + n_2 \cdot (n_7 + n_8) + n_3 \cdot n_8 + 1/2 \cdot (n_1 \cdot n_5 + n_2 \cdot n_6 + n_3 \cdot n_7 + n_4 \cdot n_8)}{(n_1 + n_2 + n_3 + n_4) \cdot (n_5 + n_6 + n_7 + n_8)}$$

A explicação do cálculo de *R* implica a consideração de uma situação imaginária em que são apresentados ao provador todos os pares possíveis de amostras N e S, devendo este identificar a amostra S em cada par. O número total de pares avaliados aparece em denominador e é, no caso da figura 3.6, de 10.000 (100 · 100).

O numerador consiste no somatório do número de vezes que o provador acertou na identificação do sinal no ensaio imaginário (número de casos favoráveis). A primeira parcela corresponde ao número de vezes que o provador acertaria na comparação entre

as amostras a que respondeu "S" e todas as outras (a que respondeu "S?", "N?" e "N"). No caso da figura 3.6 (i), este número é de $32 \cdot (15 + 55 + 30)$. A segunda parcela corresponde ao número de vezes que o provador acertaria na comparação entre as amostras a que respondeu "S?" e aquelas a que respondeu "N?" e "N", o que, no caso da referida figura, equivale a $38 \cdot (55 + 30)$. A terceira parcela envolve o número de vezes que o provador acertaria nas comparações entre "N?" e "N", o que, no caso em análise, resulta em $23 \cdot 30$.

Os números apresentados referem-se às três primeiras parcelas do numerador. A quarta parcela representa metade do valor dos empates, ou seja, dos casos em que o provador atribuiu a mesma designação (seja ela S, S?, N ou N?) aos dois produtos que lhe foram apresentados. Note-se que esta situação não acontece quando os testes são, efectivamente, conduzidos, sendo consequência do facto de se tratar de uma situação imaginada através dos resultados de um outro ensaio – o A-Não A. O número de empates é, assim, distribuído igualmente entre respostas correctas e incorrectas. O valor 1/2 que precede a soma relativa à quarta parcela refere-se, portanto, à metade correcta das identificações empatadas. No caso da figura 3.6 (i), o número de identificações empatadas é de $32 \cdot 5 + 38 \cdot 10 + 23 \cdot 55 + 7 \cdot 30$.

Uma vez apresentado desta forma, é simples entender o índice R como a probabilidade de identificar correctamente o sinal quando comparado com o ruído, em testes de comparação por pares. Esta é, também, a interpretação já apresentada para $P(A)$. Pode, aliás, verificar-se facilmente que, sendo P(A) a área debaixo da curva ROC (ver figura 3.7), a sua fórmula de cálculo é exactamente a apresentada para R.

O'Mahony (1992) descreve uma outra forma de calcular o índice R, ligeiramente diferente da apresentada, que se baseia em ensaios de ordenação, em vez de ensaios de "cotação". Quando se comparam apenas dois produtos (caso da maior parte dos testes de discriminação), a ordenação faz-se com quatro ou mais amostras em número par (sendo metade de cada produto) e conta-se o número de ensaios em que os produtos S e N foram ordenados em cada lugar (1º - mais intenso; último - menos intenso). A partir desses resultados, constrói-se uma tabela do tipo da da figura 3.6, onde o primeiro lugar corresponde a S, o segundo a S?, o terceiro a N? e o quarto a N (no caso de serem quatro amostras), e, com base nela, calcula-se R da maneira já descrita.

Quando se comparam simultaneamente mais do que dois produtos diferentes em ensaios de ordenação, determina-se um índice R para cada um dos estímulos apresentados (sinais). O cálculo de R para cada sinal pode ser feito de dois modos

distintos que não são equivalentes, pelo que os dois índices nem sempre assumem o mesmo valor.

Em O'Mahony (1992) pode encontrar-se uma descrição pormenorizada da maneira de calcular cada um destes índices e da relação existente entre eles. Por sua vez, Ishii et al. (1992) fazem uma comparação entre os índices *R* obtidos para 3 soluções aquosas de NaCl com concentrações distintas e concluem que os índices baseados nos ensaios de ordenação revelam melhor discriminação do que os que se baseiam nos ensaios de "cotação". As diferenças encontradas são atribuídas ao facto de, no ensaio de ordenação, não se admitir que duas amostras tenham a mesma ordem, o que reduz o número de empates.

Cliff et al. (1997) apresentam um estudo em que comparam os índices *R* com escalas de cotação e com escalas de ordenação para testes de preferência com painéis de consumidores e de provadores treinados. Neste caso, a escala de cotação foi utilizada da forma convencional com 4 níveis. No caso do painel de consumidores, os índices *R* resultantes da cotação e da ordenação foram muito semelhantes e verificou-se que a correlação entre estes e a média das cotações atribuídas a cada produto foi muito elevada (0.98 no caso da cotação e 1 no caso da ordenação). Já quanto aos painéis treinados, os índices *R* revelaram uma dispersão muito maior, que os autores atribuem ao facto de a dimensão do painel ser bastante mais reduzida ($n = 18$ vs. $n = 89$). Uma outra conclusão deste estudo é que, pelo menos neste caso, os resultados obtidos com painéis de provadores treinados não devem ser utilizados como indicadores preliminares da preferência dos consumidores.

Tal como já foi afirmado, o índice *R* representa uma probabilidade. É fundamental, para o analista, verificar se essa probabilidade é significativamente diferente da que se obteria se os provadores estivessem a responder ao acaso (que será, obviamente, de 50%). Bi e O'Mahony (1995) apresentam uma tabela para testar a significância do valor de *R*, dada a dimensão da amostra considerada. A construção da tabela baseia-se num teste de hipóteses à proporção binomial (no pressuposto de que a distribuição de *R*, quando as respostas são dadas ao acaso, é normal).

Um exemplo de utilização da tabela permite verificar que, se ao testar 30 estímulos de sinal e 30 de ruído se obtiver $R = 65\%$, pode considerar-se que os provadores distinguem o sinal do ruído ($\alpha = 5\%$), visto que o valor de 65% é significativamente diferente de 50%. No entanto, se o mesmo valor de *R* for obtido para 15 estímulos de sinal e 15 de ruído, já não é possível retirar tal conclusão.

Uma outra medida da capacidade de discriminação que coincide com $P(A)$ e, portanto, com o índice R tal como foi aqui apresentado, é a estatística de Wilcoxon (W), referida em Hanley e McNeil (1982). No capítulo 6, onde se aborda o caso prático da Câmara de Provadores do Instituto do Vinho do Porto, define-se e utiliza-se esta estatística, que permite, além de estimar a área debaixo da curva ROC, estimar a sua variância, o que possibilita o estabelecimento de intervalos de confiança para a referida área e a comparação entre várias áreas.

- **Modelo Beta-Binomial**

Nos métodos para processar os resultados dos ensaios de diferenciação considerados até aqui, assumiu-se, implícita ou explicitamente, que o número de respostas correctas seguia uma distribuição binomial com parâmetros N e p. Na abordagem dos testes de hipóteses à proporção binomial tal pressuposto é óbvio, sendo o parâmetro p estimado directamente a partir dos resultados dos ensaios.

Quanto às curvas ROC e ao modelo de Thurstone, pode verificar-se que se admite, igualmente, que o número de respostas correctas segue uma distribuição binomial, cujo parâmetro p depende do critério considerado pelo provador. É, aliás, a variação de p com o critério que permite desenhar as curvas ROC e que implica que as tabelas de d' sejam diferentes consoante os ensaios sensoriais utilizados (ver Ennis (1993)). Note-se que a cada um dos ensaios – triangular, 3-AFC, duo-trio e 2-AFC – corresponde um critério estabilizado (dado que os testes não são propensos ao enviezamento das respostas) e, portanto, um determinado parâmetro p.

Um dos pressupostos da distribuição binomial é, precisamente, que a probabilidade de sucesso (p) não se altera ao longo do ensaio. No entanto, e tal como já foi referido anteriormente, em testes de diferenciação ou de preferência, é quase sempre utilizado mais do que um provador. Frequentemente, os resultados são processados englobando as várias respostas (réplicas) dos vários provadores e tratando-as como sendo provenientes da mesma distribuição estatística.

Ennis e Bi (1998) referem que nas situações semelhantes à descrita se viola o pressuposto de que p se mantém constante ao longo da experiência, dada a existência de diferenças de sensibilidade entre os provadores. Este fenómeno, conhecido por *overdispersion*, pode levar a conclusões erradas, nomeadamente a de que a diferença entre os produtos é perceptível, quando, na realidade, não o é.

A solução proposta pelos autores é a utilização da distribuição beta-binomial (BB) que é uma generalização da distribuição binomial desenvolvida para acolher dados excessivamente dispersos e onde o parâmetro p segue uma distribuição beta.

Bi et al. (2000) apresentam um gráfico que representa a distribuição de p para várias distribuições beta e em Ennis e Bi (1998) são propostos procedimentos para: (i) testar a hipótese de os dados provirem de uma distribuição binomial; (ii) calcular os parâmetros da distribuição BB; (iii) testar a hipótese de os produtos ensaiados serem iguais; (iv) determinar a dimensão amostral.

A dimensão amostral (N) é função de duas variáveis: o número de provadores (k) e o número de réplicas que cada provador efectua (n), sendo que $N = n \cdot k$.

Os resultados das várias simulações feitas por Ennis e Bi (1998) confirmam que a distribuição binomial não é adequada para dados com mais do que uma fonte de variação e que, no caso simulado de vários provadores com valores de p diferentes, o modelo BB se mostrou adequado.

Numa das simulações sucede mesmo que, para um α teórico de 5%, se verifica que o erro do tipo I para a distribuição binomial chega a atingir o valor de 44%, contra 7% para a beta-binomial.

O facto de a distribuição binomial não ser adequada à modelização do tipo de dados referido, origina que, também na abordagem da teoria de detecção do sinal, não seja legítimo misturar os resultados das respostas dos vários provadores, nomeadamente na elaboração das curvas ROC. Nestas condições, aconselha-se a elaboração de uma ROC para cada provador.

Quanto à medida d', verifica-se que, quando os dados são provenientes de vários provadores (dados com *overdispersion*), a sua variância é superior à fornecida nas tabelas. Bi e Ennis (1998b) apresentam uma forma de calcular a variância de d' para estas situações, que se baseia nas tabelas referidas e nos parâmetros da distribuição BB. Desta forma, torna-se possível continuar a analisar o problema dos ensaios de diferenciação em termos de d'.

Bi et al. (2000), que aplicam o modelo BB ao caso prático do estudo de dois desodorizantes diferentes, consideram que ele deve ser utilizado não só nas situações em que há vários provadores com sensibilidades diferentes, mas também naquelas em que haja outro tipo de variações entre as experiências. Como exemplo, os autores

referem a existência de pequenas diferenças nos produtos ou os casos em que o tempo decorrido entre experiências influencia a probabilidade *p*.

Dorfman e Berbaum (1986) afirmam que a elaboração de curvas ROC a partir dos dados de vários provadores agrupados (como se se tratasse apenas de um provador) é aconselhada apenas nas situações em que os resultados de cada provador individualmente são dispersos e quando a sensibilidade e os critérios dos vários provadores não diferem muito. Identificam, no entanto, um problema: a escassez de métodos para testar hipóteses ou definir intervalos de confiança para os parâmetros estimados a partir das referidas ROCs.

Para ultrapassar o problema, Dorfman e Berbaum (1986) apresentam um programa informático que, através da conjugação do método da máxima verosimilhança e do método *jack-knife*, permite estimar vários parâmetros associados às curvas ROC e as suas variâncias.

Priso et al. (1994) abordam a questão do tratamento dos resultados das réplicas de ensaios e afirmam que é incorrecto tratar os resultados das réplicas dos provadores todos em conjunto, pelo que sugerem um teste de hipóteses cuja estatística entra em consideração com os dois efeitos existentes: o de cada provador e o da diferença entre provadores. Este último é um efeito variável, dado que os autores se referem a testes de preferência e, portanto, os provadores escolhidos são considerados como uma amostra aleatória dos potenciais consumidores dos produtos em causa.

Curiosamente, nos exemplos apontados (ensaios triangulares), o teste proposto revela-se mais potente do que o teste à proporção binomial do conjunto das respostas, verificando-se mesmo, num dos casos, que a hipótese nula (inexistência de diferenças sensoriais entre os produtos) é rejeitada pelo primeiro e não o é pelo segundo. Os autores atribuem esta diferença ao facto de, no teste proposto, se entrar em consideração com o efeito provador, ou seja, com o facto de haver pelo menos um provador que consegue detectar a diferença entre os produtos.

Note-se que este resultado vai no sentido contrário ao do modelo BB, onde a rejeição da hipótese nula é mais difícil do que no teste à proporção binomial.

Dados os problemas associados ao tratamento de julgamentos de provadores tomados em conjunto, mais adiante, no capítulo 6, a abordagem à avaliação do desempenho dos provadores será efectuada individualmente.

3.2.1.3 Técnicas Estatísticas de Análise de Dados Expressos em Escalas Nominais ou Ordinais

Quando as escalas em que são expressos os resultados dos ensaios sensoriais são do tipo nominal ou ordinal, o leque de opções para o seu processamento estatístico é limitado. Tipicamente são utilizados testes não paramétricos, que, como o próprio nome indica, não pressupõem qualquer distribuição para a população da qual são retiradas as amostras.

Em O'Mahony (1982) são referidos alguns testes e as situações em que podem ser utilizados. Destacam-se, aqui, os seguintes casos: (i) o teste à proporção binomial, já referido na secção 3.2.1.1, que é um teste paramétrico mas aplicado a situações em que as escalas são nominais e quando há apenas duas categorias; (ii) o teste qui-quadrado, referido no próximo capítulo, e que é um teste à qualidade de ajuste entre a distribuição da população e uma distribuição teórica (ou entre as distribuições de duas populações distintas); (iii) o teste de Kruskal-Wallis, que será usado no capítulo 5 e que se utiliza na comparação entre medianas (como medidas da tendência central) de várias populações. Este último teste é, também, muitas vezes, referido como correspondendo à ANOVA não paramétrica.

Uma aplicação interessante da estatística não paramétrica à avaliação do desempenho de provadores encontra-se em Cardinal et al. (1994). Com o objectivo de medir a capacidade de discriminação dos candidatos a provadores, os autores elaboraram testes de ordenação de amostras com diferentes concentrações e calculam o coeficiente de correlação ordinal de Spearman para cada provador. Neste caso concreto, o coeficiente, que pode tomar valores entre –1 e 1, mede o grau de concordância entre a ordem das concentrações reais das amostras e a ordem atribuída pelos provadores. No capítulo 6, faz-se uma referência mais pormenorizada a este coeficiente.

3.2.1.4 Comparação dos Vários Ensaios de Diferenciação

O'Mahony et al. (1994) comparam os resultados obtidos nos testes triangular, 3-AFC, duo-trio e 2-AFC, para vários valores de d'. Verifica-se que, para o mesmo valor de d', os provadores exibem um melhor desempenho (medido através da percentagem de respostas correctas) nos testes 3-AFC e 2-AFC do que nos triangular e duo-trio, respectivamente.

Este fenómeno foi detectado por Byer e Adams (ver O'Mahony et al. (1994), que verificaram que uma pequena alteração nas instruções do teste triangular (convertendo-o

no 3-AFC) originava melhores desempenhos da parte dos provadores. Mais tarde, este fenómeno ficou conhecido por paradoxo *discriminatory nondiscriminators*.

Na figura 3.9, ilustra-se a interpretação do paradoxo à luz da teoria de Thurstone. Admita-se que as distribuições de N e de S são as representadas na figura e que se realizam quatro ensaios - (a) a (d), cujos resultados são simbolizados pelos pequenos círculos (círculo branco - amostra ímpar; círculos pretos - amostras iguais).

Ao analisar os casos ilustrados na figura, deve ter-se presente que as estratégias usadas pelo provador são diferentes no caso dos dois ensaios. No ensaio triangular, o provador escolhe a amostra que lhe parece "mais diferente" – estratégia de comparação das distâncias, ao passo que no 3-AFC escolhe a que lhe parece "mais intensa" (mais à direita, na figura) – estratégia *skimming*.

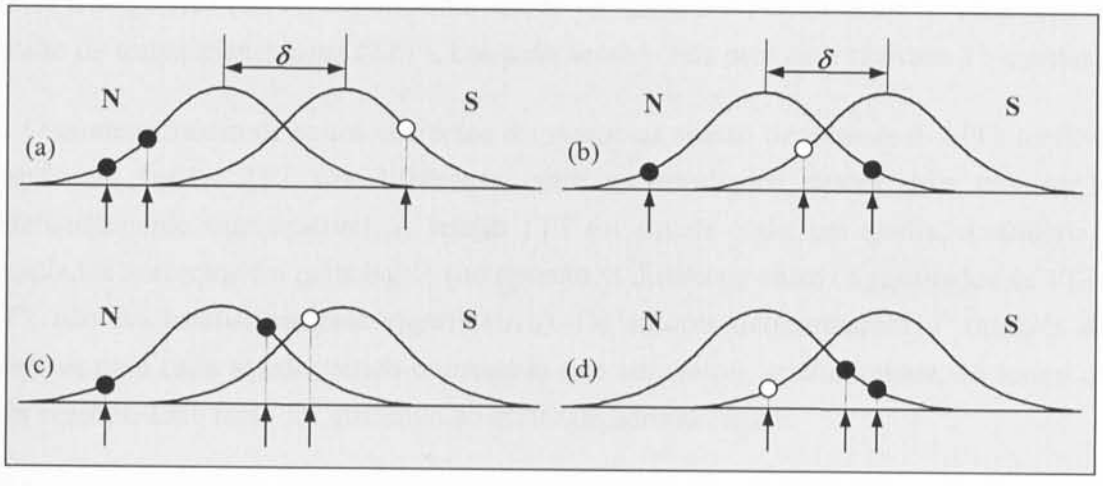

Figura 3.9 – Interpretação do paradoxo *discriminatory nondiscriminators* segundo a teoria de Thurstone para os testes triangular e 3-AFC (adaptado de O'Mahony et al. (1994))

No caso de (a), o provador responde correctamente em ambos os testes. Em (b), o provador responde incorrectamente em ambos os testes, mas por razões diferentes: a amostra ímpar no teste triangular é a correspondente ao círculo preto do lado esquerdo, enquanto no teste 3-AFC é a correspondente ao círculo preto do lado direito. Em (c), a resposta ao teste triangular é errada e ao 3-AFC é correcta e, em (d), a resposta ao 3-AFC está errada e ao triangular está correcta.

Como é fácil verificar pela figura, a situação (c) ocorre mais frequentemente do que a (d) – é mais provável – pelo que, para a mesma sensibilidade do provador (δ), o número de respostas correctas ao teste 3-AFC é maior do que ao teste triangular.

O mesmo raciocínio pode ser aplicado aos testes duo-trio e 2-AFC, verificando-se que este último origina mais respostas correctas do que o primeiro para o mesmo δ.

Quando foi apresentado o modelo de Thurstone, fez-se referência a tabelas que fornecem o valor de *d'* em função da proporção de respostas correctas num determinado ensaio (P_c). Essas tabelas foram construídas, segundo O'Mahony et al. (1994), com base em processos de simulação de situações semelhantes às descritas. Assim, compreende-se que, para testes com probabilidades idênticas de acertar respondendo ao acaso (como, por exemplo, o triangular e o 3-AFC), e para um mesmo *d'*, os valores das referidas proporções sejam diferentes (ao 3-AFC corresponde uma proporção de respostas correctas superior).

Rousseau e O'Mahony (1997) confirmam, na prática, numa experiência em que são avaliados iogurtes com sabor a baunilha, as diferenças previstas pelo modelo que acabou de se expor. Os iogurtes avaliados diferiam ligeiramente na quantidade de açúcar. A experiência consistiu em três sessões de ensaios: primeiro, uma sessão de testes triangulares (1TT); depois, uma sessão de testes 3-AFC; finalmente, uma segunda sessão de testes triangulares (2TT). Em cada sessão cada provador realizou 12 ensaios.

O número médio de testes correctos foi maior na sessão de ensaios 3-AFC, tendo-se seguido a sessão 2TT (as diferenças entre os resultados destes dois ensaios foi estatisticamente significativa). A sessão 1TT foi aquela onde, em média, o número de resultados correctos foi mais baixo (no entanto, a diferença entre os resultados de 1TT e 2TT não foi estatisticamente significativa). Os autores determinaram *d'* (através das tabelas) para cada sessão, tendo constatado que aumentou, gradualmente, ao longo das três sessões. Este facto foi atribuído ao efeito de aprendizagem.

No mesmo artigo, os autores estudam o efeito da apresentação sequencial das amostras (N – não doce; S – doce), considerando as quatro possibilidades de ordem (N-N, N-S, S-N e S-S) e os dois tipos de teste: N ímpar ou S ímpar. Em relação ao ensaio 3-AFC, não foi detectada nenhuma diferença significativa no valor de *d'* entre os testes N ímpar e S ímpar. Já para o teste triangular, as diferenças foram significativas, tendo *d'* sido superior para o teste S ímpar.

Relativamente à ordem de apresentação das amostras, os autores recorrem à Análise de Sensibilidade Sequencial (SSA), onde se reconhece que a ordem pela qual os estímulos são apresentados influencia os resultados dos testes. Segundo a SSA, e admitindo que o estímulo correspondente a S é o mais intenso, a sequência para a qual a detecção dos atributos sensoriais de S e de N é mais fácil é N-S, seguindo-se S-N, N-N, e, finalmente, a mais difícil, S-S. Referindo-se a este fenómeno, Rousseau et al. (1998) explicam que a causa deste fenómeno é o facto de o efeito de adaptação sensorial ser

mais forte com a mesma amostra (*self-adaptation*) do que com amostras diferentes (*cross-adaptation*).

Finalmente, no trabalho em análise, são comparados os resultados obtidos com os testes 2-AFC e 3-AFC e conclui-se que os valores de *d'* são superiores no caso do teste 2-AFC. Este resultado está de acordo com a SSA, na medida em que nos testes 2-AFC só são apresentadas ao provador as sequências mais favoráveis, enquanto no teste 3-AFC, como são 3 amostras e se devem alternar as sequências (para reduzir o efeito de habituação), acabam por aparecer outras sequências menos favoráveis (como a N-N e a S-S). Assim, é aconselhada, sempre que possível, a utilização do teste 2-AFC em vez dos testes que incluem três amostras.

Esta recomendação é, aliás, feita por outros autores, na sequência de vários outros estudos. De facto, o efeito fadiga, que é normalmente evocado para não se utilizarem testes apenas com duas amostras, não foi considerado relevante num estudo onde Stillman e Irwin (1995) comparam os testes triangular, da diferença simples e da diferença simples com escala de "cotação", nem num outro, onde Hautus e Irwin (1995) comparam os testes A-Não A e diferença simples, ambos com escalas de cotação.

São muitos os artigos em que são feitas comparações entre os vários ensaios de diferenciação. Em seguida, descrevem-se, resumidamente, as conclusões de alguns deles.

No que diz respeito ao ensaio mais referido, que é, sem dúvida, o triangular, constata-se que acaba por ser pouco recomendado, quando comparado com ensaios tais como o 2-AFC, 3-AFC ou o da diferença simples. Além dos estudos acima referidos, O'Mahony (1995) aconselha o uso do teste 2-AFC em alternativa ao triangular, advogando que, quando a diferença entre as amostras é difícil de especificar, se deve adoptar um procedimento de *warm-up*, possibilitando que os provadores percebam qual a diferença e utilizem uma estratégia *skimming*. O mesmo autor argumenta, também, que o teste triangular é frágil e não estável e que, se for utilizado por pessoas sem formação adequada, pode produzir resultados confusos.

Rousseau et al. (1998) comparam os testes de diferença simples, triangular e duo-trio, em termos de valores de *d'* no caso do iogurte com aroma de baunilha, concluindo que o teste de diferença simples modificado (com escalas de "cotação" ou ordenação) se revela superior aos outros.

Stillman e Irwin (1995), no estudo já referido, concluem que, em termos de *d'*, os testes triangular, da diferença simples e da diferença simples com cotação se revelaram

equivalentes. Em termos da reacção dos provadores aos testes, o último tem uma clara vantagem, já que os provadores se sentiram bastante mais confortáveis quando lhes foi dada a oportunidade de exprimirem alguma incerteza. Ainda quanto à reacção dos provadores, é sugerido que seja dito a cada um, logo após cada ensaio, qual o resultado correcto, a fim de motivar e ajudar a melhorar a sensibilidade às diferenças entre as amostras. Neste estudo foram analisadas duas concentrações distintas de sumo à base de concentrado de framboesa.

Bi e Ennis (1999) comparam a potência dos testes 2-AFC (uni-direccional e bi-direccional), 3-AFC, duo-trio e triangular com base nos modelo BB e de Thurstone e concluem que, nas condições em que se aplica o modelo BB, os testes 2-AFC e 3-AFC têm potências bastante superiores aos duo-trio e triangular.

Huang e Lawless (1998) estabelecem comparações entre vários ensaios sensoriais, numa primeira fase para sumos de uva com três concentrações diferentes de açúcar (1%, 2% e 3%) e, numa segunda fase, para quatro tipos de chá divididos em dois grupos. As medidas utilizadas são a proporção do número de respostas correctas (P_c) e d'.

Na primeira fase são comparados os testes 2-AFC, triangular, ABX, 3-AFC, duo-trio e padrão duplo. Os resultados revelam que a hipótese nula é rejeitada em todos os testes de hipóteses. Em termos de d', apenas há diferença significativa entre os valores obtidos para os ensaios triangular e duo-trio (d' é superior para o duo-trio) na discriminação entre os sumos com concentrações 1% e 2%.

Na segunda fase, em que são comparados os testes triangular, duo-trio e ABX (o 2-AFC e 3-AFC não são adequados, dado que a diferença entre os vários tipos de chá não é especificada), as diferenças foram significativas no caso de um dos grupos e não foram no caso do outro.

Os autores confirmam o que já foi dito em relação a d': trata-se de uma medida bastante independente do ensaio sensorial utilizado (o que não acontece com P_c) e concluem que, contrariamente às suas expectativas, o teste ABX não se revelou mais sensível do que qualquer um dos outros.

Um resultado semelhante foi obtido por Hautus e Irwin (1995) ao compararem os testes A-Não A e diferença simples, ambos com escalas de cotação com 6 níveis: em termos de d', não foram encontradas diferenças significativas entre estes dois métodos (note-se que em termos de proporção de julgamentos correctos, P_c, os resultados são diferentes para os dois ensaios).

Num outro estudo que, aliás, será referido mais adiante a propósito das escalas de cotação, Kim et al. (1998) confirmam, na prática, o que já foi dito a propósito da utilização deste tipo de escalas quando o objectivo é medir pequenas diferenças entre produtos. Foi feita a comparação entre os valores de *d'* obtidos para um ensaio 2-AFC e para uma escala de cotação com 9 níveis para duas soluções de NaCl com concentrações próximas. Verificou-se que, enquanto no caso do teste 2-AFC o valor de *d'* foi significativamente diferente de zero e, portanto, se pôde concluir que as amostras foram percebidas pelos provadores como diferentes, no caso da escala de cotação, a hipótese de *d'* ser igual a zero não foi rejeitada. No mesmo estudo foi utilizada uma escala de *rank-rating*, que originou um valor de *d'* significativamente diferente de zero, mas inferior ao do teste 2-AFC.

Bi (1995) desenvolve modelos não paramétricos para determinar a probabilidade P_c para ensaios de diferenciação[2] e compara os resultados teóricos obtidos com os ensaios 3-AFC e triangular nas várias combinações possíveis das amostras. Ao todo, são feitas 14 comparações e o autor constata que, para o mesmo ensaio, P_c é sempre superior quando o estímulo ímpar é o de maior intensidade e que, em geral, P_c é maior para o teste 3-AFC do que para o triangular. No entanto, quando se compara o teste 3-AFC com o estímulo de intensidade fraca ímpar, com o triangular com estímulo de intensidade forte ímpar, há situações nas quais P_c é superior para este último ensaio.

Os modelos referidos são definidos com base na separação dos factores de percepção e cognitivo. Ao primeiro, corresponde a probabilidade de o provador percepcionar um dado estímulo de uma forma correcta ou incorrecta (julgamento correcto ou falso alarme) e, ao segundo, corresponde a probabilidade de o provador responder correctamente a cada sensação possível.

O efeito do *feed-back* e o da comparação dos produtos a posteriori no desempenho dos provadores são estudados por Dacremont et al. (1998) em testes triangulares com amostras de *Ice tea* e de dois tipos de sumo de laranja (misturas com diferentes concentrações). Numa primeira fase, são estudadas quatro situações: (i) a de controlo, na qual é administrado o teste triangular convencional; (ii) a de *feed-back*, para a qual, após cada resposta ao teste triangular, o provador é informado se respondeu certo ou errado; (iii) a de *feed-back* com comparação posterior, que consiste na de *feed-back* seguida de uma comparação entre as duas amostras em prova; (iv) a de comparação

[2] Os ensaios para os quais foram desenvolvidos os modelos foram o 2-AFC, o duo-trio, o 3-AFC e o triangular. No caso destes dois últimos, foram desenvolvidos modelos diferentes consoante a amostra ímpar correspondia ao estímulo de intensidade mais fraca ou mais forte.

posterior, que consiste apenas na comparação após o teste triangular, mas sem *feed-back*. Cada experiência envolve 15 ou 30 testes e a comparação do desempenho dos provadores é feita com base nos resultados dos primeiros e dos últimos 5 testes.

Embora o desempenho dos provadores tenha evoluído favoravelmente em quase todos os casos (excepto naquele em que foram feitos 30 testes, onde se notou algum efeito de fadiga), o aumento verificado apenas foi estatisticamente significativo no caso do procedimento de *feed-back* com comparação posterior.

Numa segunda fase, são comparados dois tipos de *Ice tea* e duas misturas diferentes de sumos concentrados de laranja e de frutos silvestres. Nesta experiência apenas se comparam os resultados do teste triangular convencional e do teste com *feed-back*. Verifica-se que os provadores com menos sensibilidade têm uma evolução negativa no desempenho com o procedimento de *feed-back* e nula ou positiva no procedimento convencional. Já os provadores com maior sensibilidade apresentam evoluções positivas nos dois casos, sendo mais pronunciadas no caso da experiência com *feed-back*.

Com base nos resultados das duas experiências, os autores não aconselham a utilização do *feed-back*, a não ser em situações particulares para motivar ou transmitir segurança aos provadores. Já quanto à alternativa *feed-back* + comparação, referem que se trata de um procedimento com resultados ligeiramente melhores, mas que, simultaneamente, consome mais tempo, mais recursos e é mais sujeito à fadiga sensorial.

Para finalizar, é interessante comparar o processamento dos resultados dos ensaios de diferenças através dos testes de hipóteses à proporção binomial (estudado em 3.2.1.1) com a abordagem da teoria da detecção do sinal. Enquanto os testes de hipóteses apenas fornecem a informação sobre se os produtos podem, ou não, ser considerados diferentes, a segunda abordagem proporciona medidas para a diferença entre os produtos (como d' ou o índice R). Por outro lado, verifica-se que a abordagem dos testes de hipóteses trata como equivalentes os ensaios a que correspondem probabilidades iguais de acertar ao acaso, o que, conforme foi visto, não corresponde à realidade.

A comparação que acaba de se efectuar neste ponto sublinha o cuidado e a delicadeza associados à escolha dos ensaios sensoriais a utilizar e revela, também, que não é fácil estabelecer de forma clara quais as vantagens de uns testes relativamente aos outros. No entanto, aparentemente, e desde que não haja nenhuma indicação em contrário, pode aconselhar-se a utilização do ensaio de diferença simples com "cotação" no caso da diferenciação geral e do 2-AFC no caso da diferenciação de atributos.

3.2.2 Ensaios para Determinação de Limiares

A determinação dos limiares sensoriais passa pelo processamento dos resultados de ensaios sensoriais que os permitem detectar, e que já foram referidos de uma forma geral em 2.4.3.

Relativamente ao método ascendente dos limites (ver ASTM (1997a)), o limiar individual é determinado através do cálculo da média geométrica dos valores correspondentes à última concentração falhada e à seguinte. Quanto ao método adoptado actualmente pela ASTM (ver ASTM (1997b)), o limiar individual é calculado com base na definição estatística apresentada em 2.3.1 (ver figura 2.2). Assim, torna-se necessário estimar a curva da função psicométrica de cada indivíduo, o que é feito através do ajuste de uma curva aos valores das proporções de respostas correctas. O ajuste pode ser efectuado através de um método gráfico ou através de um método iterativo utilizando *software* apropriado. As proporções referidas são corrigidas do efeito sorte (acertar ao acaso).

Bi e Ennis (1998a) abordam detalhadamente os vários métodos estatísticos que podem ser utilizados para estimar a curva psicométrica.

3.2.3 Ensaios Descritivos e de Aceitação

Os ensaios descritivos e os testes de aceitação são utilizados em estudos de natureza mais abrangente do que os testes de diferenciação ou de preferência (ver secção 2.4). Uma consequência deste facto é que os ensaios pertencentes ao primeiro grupo não se encontram, à partida, tão bem especificados como os pertencentes ao segundo.

Particularmente no que diz respeito às escalas utilizadas e às análises estatísticas dos resultados, estão disponíveis muitas alternativas. Um exemplo é o do método do perfil de escolha livre, onde a escala a utilizar pode tomar a forma de uma escala de cotação ou de uma escala com âncoras, e ser pré-definida pela organização do ensaio ou definida pelos provadores. A escolha entre as várias opções, tanto a nível de escalas como de tratamento dos resultados, deve basear-se no objectivo específico do estudo.

Neste contexto, optou-se por não descrever exaustivamente a forma de processar os resultados de cada ensaio, abordando-se o processamento estatístico dos resultados de

uma forma geral e, em alguns casos, através de exemplos que podem ser facilmente adaptados e/ou extrapolados para outras situações.

As escalas utilizadas na maior parte dos ensaios descritivos e de aceitação podem, em certas circunstâncias que já foram referidas, ser consideradas escalas de intervalo. Estão nestas condições as escalas de cotação (como as de Likert), as escalas não estruturadas (segmento de recta) e as escalas com âncoras.

Normalmente, a primeira abordagem estatística a este tipo de resultados consiste numa análise descritiva dos dados, onde são calculados e descritos os respectivos parâmetros amostrais (tais como as médias e as variâncias obtidas para cada característica de cada produto e/ou para cada provador).

Quando se pretende extrapolar, a partir dos resultados obtidos nos ensaios, para retirar conclusões sobre a população (seja ela referente aos consumidores ou aos produtos), utilizam-se técnicas de inferência estatística que podem ser divididas em dois grandes grupos: univariadas e multivariadas.

Das técnicas univariadas a análise de variância é, sem dúvida, aquela que é mais utilizada em Análise Sensorial. Assim, na secção 3.2.3.1, aborda-se, sucintamente, este tema.

As técnicas de estatística multivariada utilizam-se quando se pretende analisar as inter-relações de dois ou mais atributos do(s) produto(s) em estudo.

Ennis et al. (1982) chamam a atenção para o facto de a natureza dos problemas em Análise Sensorial ser multivariada. De facto, a quantidade de variáveis envolvida num estudo de Análise Sensorial é, frequentemente, elevada e as relações entre elas são, normalmente, complexas. Trata-se, portanto, de um campo muito propício à aplicação das técnicas de estatística multivariada. Este assunto será abordado na secção 3.2.3.2, onde se analisam algumas das aplicações mais referidas na literatura.

3.2.3.1 Técnicas de Estatística Univariada – Análise de Variância

As técnicas de estatística univariada utilizam-se quando o objectivo é o de estudar uma característica em particular do(s) produto(s), ou várias características independentemente umas das outras.

Muitos autores (ver, por exemplo Gay e Mead (1992) ou Lundahl e McDaniel (1988)) referem que, quando se trata de escalas de intervalo em Análise Sensorial, a

técnica base usada quase invariavelmente para analisar os dados é a análise de variância (ANOVA). Lawless e Heymann (1999) salientam, também, que esta técnica é a mais utilizada na comparação de mais do que dois produtos com base em escalas de intervalo.

A análise de variância possibilita a comparação dos valores esperados de mais do que duas populações. A partir da análise da dispersão total presente num conjunto de dados, esta técnica permite identificar os factores que deram origem a essa dispersão e avaliar a contribuição de cada um deles (ver Guimarães e Cabral (1998, p. 427)).

A aplicação da análise de variância clássica às classificações atribuídas em Análise Sensorial, baseia-se nos seguintes pressupostos:

(i) as classificações são independentes umas das outras;

(ii) as distribuições das classificações têm variâncias iguais;

(iii) as classificações distribuem-se normalmente;

O'Mahony (1982) sugere que se verifique a normalidade dos dados através da sua representação gráfica. Quanto à independência dos resultados e à sua variância, verifica-se que alguns dos problemas podem ser minorados através do planeamento da experiência e da formação dos provadores.

Quando, apesar de todos os cuidados, se verifica que os dados não cumprem minimamente os requisitos referidos, deve, então, assumir-se apenas a informação da ordem, como se os dados se encontrassem expressos numa escala ordinal, e utilizar as análises estatísticas descritas em 3.2.1.3.

Na aplicação da ANOVA à Análise Sensorial, decompõe-se a variação total que existe nos resultados dos ensaios em variações devidas a factores que se supõe que afectam as respostas. A partir da comparação entre a variação devida a cada factor e a variação residual (que é considerada aleatória, e, portanto, sem uma causa facilmente detectável e/ou controlável), testa-se a hipótese de o efeito de cada um desses factores ser nulo.

No caso de se conseguir rejeitar a hipótese nula, conclui-se que o factor em questão explica, de uma forma estatisticamente significativa, uma parte das diferenças existentes entre os resultados dos vários ensaios.

Quando o objectivo é o da comparação apenas entre produtos, utiliza-se a ANOVA só com um factor – o factor produto. A hipótese nula é a da igualdade dos valores esperados das classificações atribuídas pelos provadores aos vários produtos na característica em estudo. No caso da rejeição desta hipótese, conclui-se que há pelo menos um produto que, na característica em questão, é, em média, classificado de uma forma diferente dos outros. Para detectar qual ou quais os produtos percebidos como diferentes e para definir intervalos de confiança para o valor esperado do parâmetro em estudo para cada produto, recorre-se ao método de Tukey ou de Scheffé (ver Guimarães e Cabral (1998, p. 436-440)).

Em Kim e O'Mahony (1998), são comparados 6 estímulos diferentes (soluções de NaCl) em duas situações distintas: (i) utilização do procedimento de cotação convencional com 9 pontos e (ii) escala de *rank-rating*, também com 9 pontos. Os produtos foram avaliados por 30 provadores não treinados. Os resultados de cada procedimento foram processados através de uma análise de variância (factor produto) e as conclusões apontam para a existência de diferenças sensoriais entre as amostras, confirmadas pelos dois métodos.

Todavia, os autores afirmam que, enquanto no caso da escala *rank-rating*, a existência das referidas diferenças é confirmada após serem analisados os dados dos primeiros 6 provadores, no caso da escala de cotação convencional, tal só sucede ao fim do 12º provador.

A justificação encontrada para esta diferença reside no facto de o procedimento *rank-rating* evitar dois tipos de erros: os devidos à falta de memória da intensidade dos estímulos e os devidos ao esquecimento da pontuação atribuída aos estímulos, ao passo que o procedimento convencional apenas evita os erros do segundo tipo.

A análise de variância mais utilizada no contexto da Análise Sensorial é, no entanto, a ANOVA com 2 factores, sendo um deles o correspondente ao produto e o outro o correspondente ao provador. Esta técnica permite analisar a variação existente entre produtos e entre provadores e verificar se algum destes efeitos ou a sua interacção são estatisticamente significativos.

Lundahl e McDaniel (1988) discutem a problemática de considerar o efeito provador como fixo ou variável, concluindo que, na maior parte das situações, este efeito deve ser considerado como variável. A excepção é feita para os painéis de provadores profissionais, que não é suposto terem qualquer relação com qualquer população estatística. Neste caso, o efeito provador deve ser considerado fixo.

Se, por um lado, a análise de variância é a técnica mais utilizada para processar os resultados dos ensaios na área da Análise Sensorial, por outro, verifica-se que, muitas vezes, os estudos são complementados com a aplicação de outras técnicas estatísticas, nomeadamente, técnicas multivariadas. Exemplos deste tipo de estudos encontram-se em Antoniou et al. (2000), Branger et al. (1999) ou Calvino et al. (1996).

3.2.3.2 Técnicas de Estatística Multivariada

Ennis et al. (1982) sugerem que há, basicamente, quatro tipos de questões em Análise Sensorial que as técnicas de estatística multivariada podem ajudar a resolver. Os autores ilustram as referidas questões sob a forma de pergunta, do seguinte modo:

- Até que ponto é que as variáveis medidas são independentes umas das outras? Será possível desenvolver um conjunto de variáveis mais elementar?
- Será que todos os provadores estão a interpretar o significado das variáveis da mesma maneira?
- Quando se comparam amostras ou grupos de amostras, quais são as variáveis responsáveis pela discriminação entre as amostras ou os grupos comparados?
- Quais as amostras que se podem considerar semelhantes e/ou diferentes e em que variáveis?

Como se pode verificar, os problemas abordados relacionam-se, não só com os produtos, mas também com os seus instrumentos de medida – os provadores. As diferenças na interpretação das variáveis e na utilização das escalas por parte dos vários provadores são apenas exemplos de questões que podem ser detectadas através da utilização deste tipo de técnicas, que são cada vez mais referidas na literatura da área da Análise Sensorial.

A abordagem exaustiva das técnicas de estatística multivariada e da sua aplicação à Análise Sensorial encontra-se fora do âmbito deste trabalho. No entanto, e dado que algumas delas se revelam importantes para a melhor compreensão de fenómenos ligados à percepção sensorial, apresentam-se, resumidamente, quatro das técnicas consideradas mais relevantes e alguns exemplos da sua utilização.

Uma análise profunda da utilização deste tipo de técnicas em Análise Sensorial pode encontrar-se no livro *Multivariate Analysis of Data in Sensory Science*, editado por Naes e Risvik (1996). Existem, também, numerosas publicações onde são apresentadas

aplicações das mais diversas técnicas multivariadas. Além das referidas adiante, salientam-se Tepper e Kuang (1996), onde é utilizada a técnica de *multidimensional scaling* e McEwan e Schlich (1992), onde os autores explicam os princípios subjacentes à análise de correspondência e comparam esta técnica com a análise factorial e com a análise Procrustes generalizada, exemplificando com um caso de avaliação de compota de morango.

- **Análise de Variância Multivariada**

O principal objectivo da análise de variância multivariada (MANOVA) é determinar se existem diferenças significativas entre grupos, quando comparados relativamente a todas as variáveis dependentes que se revelem de interesse. Tal como na ANOVA, na MANOVA são testadas as diferenças entre os grupos considerados. No entanto, enquanto no primeiro caso os testes são feitos característica a característica, no segundo são-no simultaneamente a todas as características, o que previne os problemas causados pela correlação entre as diversas variáveis avaliadas.

Lawless e Heymann (1999) aconselham a que se utilize, neste tipo de estudos e numa primeira fase, a MANOVA, com o objectivo de testar a influência das variáveis no seu conjunto. No caso de tal influência se revelar estatisticamente significativa, os autores referem que se devem conduzir várias ANOVAs, uma para cada variável dependente, com o objectivo de interpretar quais dessas variáveis são responsáveis pelas diferenças existentes entre os grupos.

A grande desvantagem desta técnica é, como salientam Dijksterhuis e Heiser (1995), o facto de o incremento do número de variáveis dependentes originar um aumento do grau de complexidade do método, e consequentes dificuldades na interpretação dos efeitos principais e das interacções presentes nos resultados.

- **Análise Factorial**

Esta técnica encontra-se particularmente direccionada para responder à primeira questão colocada no início desta secção, sendo o seu principal objectivo representar um conjunto de variáveis a partir do menor número possível de dimensões (factores) que lhe estão subjacentes.

A sua aplicação à Análise Sensorial permite que as variáveis medidas na experiência sejam substituídas por um conjunto de factores em menor número. Seguidamente, salientam-se algumas das vantagens que mais frequentemente decorrem da utilização da análise factorial para tratar os resultados de experiências na área de Análise Sensorial:

- Redução do grau de redundância que eventualmente se encontre nas variáveis originais.
- Detecção de eventuais diferenças entre os provadores na interpretação do vocabulário ou das amostras referência utilizadas.
- Maior clareza na definição das diferenças existentes entre as várias amostras.
- Melhor interpretação e compreensão dos resultados.
- Utilização, em estudos futuros, dos factores em vez das variáveis originais, e consequente simplificação desses estudos.

As vantagens acima descritas relacionam-se, directa ou indirectamente, com o facto de a análise factorial permitir a identificação de fenómenos latentes.

Risvik (1996) descreve pormenorizadamente a diferença entre estruturas de variáveis latentes e variáveis directamente observáveis, salientando que o nosso cérebro sintetiza a informação que recebe dos órgãos sensoriais, transformando as variáveis directamente observadas numa estrutura de variáveis latentes. O autor compara, assim, o processo estatístico de extracção dos factores a partir das variáveis de entrada com o processo de interpretação que as pessoas elaboram a partir da informação que recebem dos sentidos.

Através da análise factorial, as características sensoriais medidas em escalas de intervalo são reduzidas a um número inferior de factores. Cada factor é fortemente correlacionado com um sub-conjunto das variáveis originais que, por sua vez, se interrelacionam umas com as outras. Os factores extraídos são, normalmente, independentes entre si.

Na expressão seguinte, ilustra-se a relação entre cada variável e os factores:

$$Z_j = a_{j1} \cdot F_1 + ... + a_{jm} \cdot F_m + U_j$$

sendo:

Z_j: j^{ésima} variável de entrada;

F_1: Factor 1;

F_m: Factor m;

U_j: factor único para a variável Z_j;

M: número de factores comuns a todas as variáveis;

a_{ij}: *factor loadings*.

O modelo de análise factorial é tanto melhor quanto menores forem os valores de m e de U_j, devendo os valores de a_{ij} aproximar-se da unidade ou de zero. Para cada matriz de dados de entrada é possível obter vários modelos, seja porque se utilizam diferentes métodos para a extracção dos factores ou para a sua eventual rotação, seja porque se opta por um determinado critério na decisão do número de factores a considerar. Em Hair et al. (1998), pode encontrar-se uma explicação pormenorizada desta técnica e das condições em que se devem utilizar os diferentes métodos.

Uma vez encontrado um modelo satisfatório para um determinado conjunto de variáveis de entrada, podem calcular-se os *factor scores*, que correspondem às classificações que cada caso obtém em cada factor. Na maior parte dos estudos de Análise Sensorial, um caso corresponde ao conjunto das classificações atribuídas, por um dado provador, às várias características medidas (variáveis de entrada) de uma amostra.

Cada factor representa uma variável da estrutura latente e, portanto, cada *factor score* corresponde à pontuação que cada provador atribui a cada amostra no que diz respeito a essa variável. Estes valores são, muito frequentemente, utilizados em estudos posteriores que empregam outras técnicas de estatística multivariada.

Na literatura especializada são relatados inúmeros casos de aplicação da análise factorial à Análise Sensorial. A título de exemplo, refiram-se Calvino et al. (1996), onde é descrita a avaliação de café instantâneo, Antoniou et al. (2000), onde são analisadas e comparadas as características físicas, químicas e sensoriais relacionadas com a textura de queijos franceses, Dijksterhuis (1995), onde se comparam os vários provadores no que respeita à utilização dos vários atributos num estudo referente às características de queijos, ervilhas e bifes, e Dijksterhuis et al. (1994), onde a análise factorial é utilizada para analisar e comparar curvas tempo-intensidade.

- **Análise de *Clusters***

Análise de *clusters* é a denominação atribuída a uma categoria de técnicas cujo objectivo é agrupar objectos (que, neste contexto, se podem identificar com provadores ou com produtos), com base numa medida de semelhança.

A semelhança entre os objectos é calculada a partir das pontuações atribuídas pelos provadores às características dos produtos em estudo. Esse cálculo pode ser feito de várias formas, dando origem a diferentes medidas para a semelhança. Medidas frequentemente utilizadas são, por exemplo, o coeficiente de correlação ou a distância euclidiana entre os objectos (ver Hair et al. (1998)).

Os grupos são constituídos de forma a que a semelhança entre elementos do mesmo grupo seja superior à verificada entre elementos de grupos diferentes.

As aplicações mais frequentes da análise de *clusters* à Análise Sensorial têm como objectivo formar grupos de produtos semelhantes de um ponto de vista de características organolépticas. Por exemplo, Ennis et al. (1982) apresentam um caso de definição de grupos para um conjunto de vários odores e Antoniou et al. (2000) dividem um conjunto de 15 queijos franceses em três grupos que diferem nas características de textura. Ainda acerca desta técnica, salienta-se o estudo de Calvino et al. (1996), já citado, onde 18 cafés instantâneos são agrupados em 4 *clusters* distintos, com base nas suas características sensoriais. Os autores chamam a atenção para a semelhança dos resultados obtidos com as técnicas de análise factorial e análise de *clusters*.

- **Análise Procrustes Generalizada**

O objectivo da Análise Procrustes Generalizada (GPA) é determinar uma configuração de consenso a partir de dois ou mais conjuntos de dados, num espaço multidimensional.

A configuração de consenso é obtida através de um processo iterativo, onde os conjuntos de dados originais são sujeitos a operações de translacção, mudança de escala e rotação/reflexão. A solução final é aquela que minimiza a distância residual entre os conjuntos originais e a própria configuração de consenso (a esta distância dá-se o nome de estatística de Procrustes).

A aplicação mais comum da GPA em Análise Sensorial é a do processamento dos resultados de testes de perfil de escolha livre, nos quais cada provador descreve o produto com o seu próprio vocabulário e atribui, a cada característica identificada, uma pontuação. Nesta situação, cada conjunto de dados corresponde às respostas de um provador (uma dimensão por cada característica identificada e classificada).

Nas fases de translacção e de mudança de escala, as pontuações atribuídas por cada provador são padronizadas e centradas na origem. Estas operações são responsáveis pela minimização do efeito causado pelo facto de os vários provadores utilizarem, originalmente, as escalas de formas diferentes. Na fase de rotação/reflexão, a GPA minimiza as inconsistências reveladas pelos provadores na utilização dos atributos. É nesta última fase que a análise toma em consideração a possibilidade de os provadores utilizarem expressões diferentes para exprimirem as mesmas sensações.

Exemplos do processamento de resultados dos referidos ensaios são apresentados por Williams e Langron (1984), em que 8 amostras de diferentes vinhos do Porto são avaliadas por 10 provadores e em Williams e Arnold (1985), onde são estudadas as características sensoriais de 5 tipos de café.

Além da situação descrita, a GPA pode, também, ser utilizada no processamento dos resultados dos ensaios descritivos de uma forma geral. Esta técnica pode também ser aplicada, tal como referem Lawless e Heymann (1999, p. 593), na comparação ou integração de dados obtidos através de diferentes metodologias. Como exemplos podem referir-se a comparação entre resultados de ensaios descritivos e de aceitação ou entre os resultados obtidos com dois ou mais painéis de provadores, ou ainda entre resultados de técnicas instrumentais e de painéis de provadores. Em Dijksterhuis (1994), são estudadas as relações entre conjuntos de dados sensoriais e instrumentais no caso particular da avaliação de 72 maçãs.

Apesar das vantagens referidas, as opiniões acerca da utilização da GPA em Análise Sensorial encontram-se divididas. Huitson (1989) apresenta um estudo em que a GPA é aplicada a dados gerados aleatoriamente e é encontrada uma configuração de consenso que, aparentemente, faz sentido. Stone e Sidel (1993) criticam, também, a GPA, salientando que se trata de uma técnica muito empírica e que força demasiadamente os resultados até chegar a uma solução satisfatória. Arnold (1990) e Dijksterhuis e Heiser (1995), por outro lado, argumentam que a experiência feita com os dados aleatórios foi limitada e sugerem que as conclusões teriam sido outras se fossem feitas mais simulações. Os últimos autores salientam, ainda, que o procedimento de analisar os dados até se chegar a uma solução satisfatória é comum a muitas análises estatísticas e que nem por isso as pessoas duvidam delas.

King e Arents (1991) e Wakeling et al. (1992) propõem a avaliação do grau de consenso obtido para a configuração final a partir de testes estatísticos de qualidade de ajuste e, após simularem várias situações com base na técnica de Monte-Carlo, aplicam os testes a dados provenientes de ensaios de perfil de escolha livre de sabores de frutos e de produtos lácteos, respectivamente.

A existência da referida divergência de opiniões acerca da utilização da GPA em Análise Sensorial aponta para que essa utilização seja feita com alguma precaução. Em Dijksterhuis e Gower (1992), pode encontrar-se uma síntese dos cuidados a ter na utilização da GPA e na interpretação dos seus resultados.

3.3 Organização de um Programa de Análise Sensorial

O sucesso de um programa de Análise Sensorial encontra-se, obviamente, dependente da forma como ele é organizado. As etapas fundamentais de um estudo deste tipo são enunciadas em Meilgaard et al. (1987a, p. 3) e resumem-se em seguida:

(i) **Determinação do objectivo do projecto em que se insere o programa** – trata-se de definir o objectivo global do projecto. A substituição de ingredientes de um produto, a comparação entre produtos concorrentes ou o melhoramento das características sensoriais do produto, são exemplos de objectivos em situações reais.

(ii) **Determinação do objectivo específico do programa sensorial** – o objectivo específico do programa relaciona-se com o tipo de ensaios a realizar: por exemplo, avaliar diferenças globais ou de atributos entre produtos, descobrir quais as preferências dos consumidores, ou tentar prever a aceitação do produto no mercado.

(iii) **Estudo das amostras** – antes de definir quais os ensaios específicos a realizar, devem estudar-se as características sensoriais dos produtos a analisar, de forma a evitar enviezamentos nas respostas dos provadores. Por exemplo, já foi afirmado anteriormente que a cor, por vezes, influencia o julgamento de outras características. Numa situação dessas, devem ser tomadas precauções no sentido de garantir independência entre as características em avaliação. Nesta fase, e quando se justifique, deve também ser definido o vocabulário a usar na ficha de prova.

(iv) **Planeamento do teste a efectuar** – nesta fase estão envolvidos a selecção dos ensaios a realizar, a selecção e o treino dos provadores neles envolvidos, a definição da folha de prova, a especificação dos critérios para a preparação das amostras, a especificação das condições de apresentação das amostras e a determinação da forma como serão analisados os resultados dos ensaios.

(v) **Condução do teste** – é necessário verificar que todas as especificações definidas na fase anterior são cumpridas na realização efectiva dos ensaios.

(vi) **Análise dos resultados** – os procedimentos a seguir nesta fase já devem estar definidos desde a fase do planeamento do teste. Não devem esquecer-se, no entanto, questões como a da ordem das amostras, a(s) hora(s) do dia a que foram efectuados os ensaios, a eventual realização dos ensaios em dias diferentes, e variáveis demográficas tais como a idade, o sexo, a área geográfica, etc...

(vii) **Interpretação dos resultados e elaboração do relatório** – A interpretação dos resultados deve ser feita com a máxima prudência, na perspectiva de englobar os efeitos das variáveis relevantes para o estudo. O relatório deve ser sucinto e especificar, com clareza, o que foi feito em cada uma das fases descritas.

Ao longo deste capítulo foram sendo abordadas questões que se relacionam com algumas das etapas referidas, nomeadamente com o tipo de ensaios a realizar e com a análise dos resultados desses mesmos ensaios. A discussão pormenorizada das várias fases da organização de um programa sensorial encontra-se fora do âmbito deste trabalho. No entanto, recomenda-se o livro *Sensory Evaluation Pratices*, de Stone e Sidel (1993), no qual se aborda o problema da organização de um programa de Análise Sensorial de uma forma simples e prática, sem, no entanto, deixar de referir os tópicos de maior relevância.

Em seguida, referem-se, de modo sucinto, algumas questões que podem interferir directamente com o desempenho dos provadores, em particular as condições físicas em que devem decorrer as provas e os efeitos eventualmente significativos da ordem das amostras, do número de réplicas dos ensaios em cada sessão e do tempo entre a prova de diferentes amostras.

A temática da selecção e do treino dos provadores profissionais será objecto de análise na secção 3.4.

3.3.1 Condições Físicas

A existência de condições físicas adequadas à prática da prova é essencial para o sucesso do programa. Os factores que devem ser tidos em conta são: (i) a localização do laboratório de prova, que deve ser afastado de locais ruidosos e permitir que as pessoas não sejam interrompidas durante a sessão de prova; (ii) a disposição física dos móveis e dos instrumentos, que deve ser concebida de forma a evitar que os provadores se distraiam (por exemplo, a olhar para a sala de preparação de amostras) e que se influenciem mutuamente; (iii) os materiais de construção, que deverão ser inodoros,

facilmente laváveis e de cores neutras; (iv) as condições ambientais, que deverão ser controladas relativamente à temperatura, à pressão, à luz e à ausência de cheiros.

Para além destas considerações, salienta-se que os provadores deverão sentir-se o mais confortáveis possível, pelo que pormenores como o tipo de cadeiras ou de cabinas de prova não são de desprezar.

O cumprimento destes requisitos exige um investimento grande em aparelhagem e em materiais. Nos livros da especialidade, nomeadamente em Meilgaard et al. (1987a, p. 21-28), ASTM (1996, p. 3-5), Peynaud e Blouin (1997, p. 113-116) e Lawless e Heymann (1999, p. 85-91), podem encontrar-se exemplos e conselhos técnicos para a implementação de um novo laboratório de prova, ou para o melhoramento de laboratórios já existentes.

3.3.2 Efeitos da Ordenação e do Número de Amostras

A ordem pela qual as amostras são provadas pode revestir-se de grande importância. Para além disso, o número de amostras que as pessoas conseguem provar mantendo as suas capacidades num nível aceitável é limitado. Seguidamente, faz-se uma breve análise a estas duas questões, chamando a atenção para alguns aspectos que se julgam relevantes na organização de um conjunto de ensaios sensoriais, referindo-se alguma da literatura da especialidade onde estes assuntos são abordados com profundidade.

- **Ordenação das Amostras**

Os efeitos mais comuns que podem ser originados pela ordem de apresentação das amostras foram já descritos em 2.3.6.2. Há um outro efeito – efeito de *carry-over* – que não tem directamente a ver com a posição da amostra na sequência, mas com a amostra que a precede. Por exemplo, a amostra A pode ser sistematicamente avaliada com pontuações mais elevadas quando é precedida pela prova da amostra B do que quando é precedida pela prova da C.

A forma de minimizar estes efeitos é, quando possível, a apresentação das amostras de uma forma balanceada. O método mais utilizado para balancear as amostras relativamente ao efeito da ordem é o *latin square*. Em Schlich (1993), é apresentado este método e uma variante que, para além de considerar o efeito da ordem, considera também o efeito *carry-over*. A utilização destes métodos é exemplificada com a avaliação de quatro bifes diferentes por um painel de provadores profissional. Em

Durier et al. (1997), é analisado o efeito *carry-over* com base num caso prático de prova de vinhos.

- **Número de Amostras**

O número máximo de amostras a apresentar numa sessão de prova deve ser estabelecido em função do tipo de produto em estudo e das características dos provadores que participam na sessão. Mais concretamente, deve depender do quão rapidamente os provadores começam a ser influenciados pelo efeito fadiga ou pelo efeito adaptação. Por sua vez, o número mínimo de amostras depende do ensaio sensorial escolhido.

De uma forma geral, não se encontram na literatura regras que permitam estabelecer, a priori, qual o número de amostras (ou de ensaios) que se devem realizar numa só sessão. Geralmente é admitido que se utilizem mais amostras numa sessão com provadores profissionais do que com consumidores ou provadores não treinados.

No Manual publicado pela ASTM (1996, p. 20) – *Sensory Testing Methods* – são feitas algumas recomendações que, segundo os autores, deverão ser encaradas como linhas de orientação na ausência de informação específica sobre a situação concreta do teste sensorial. Destacam-se: (i) no teste de comparação por pares podem testar-se até 3 pares na mesma sessão; (ii) em testes de ordenação devem ordenar-se entre 4 e 6 amostras (excepto para estímulos unicamente visuais, onde este número pode ser largamente excedido); (iii) quando os provadores são profissionais, não devem apresentar-se mais amostras do que aquelas que o painel se sinta capaz de avaliar.

Os mesmos autores salientam, ainda, que há casos de sucesso em que consumidores testaram produtos durante 4h, mas advertem para a necessidade de mentalizar previamente as pessoas, espaçar convenientemente a prova das amostras no tempo e fazer intervalos durante as sessões.

- **Tempo que Medeia entre a Prova das Amostras**

Cubero et al. (1995) apresentam um estudo muito interessante sobre a alteração no nível de desempenho de provadores em ensaios de diferenciação quando varia o intervalo de tempo entre a prova das amostras a discriminar (no caso, refresco com sabor de limão). Concluem que, para provadores não familiarizados com o estímulo, o desempenho se deteriora à medida que o intervalo aumenta (0 – 30s – 60s), enquanto, para os provadores familiarizados com o estímulo, o desempenho é superior no caso dos 0s, mas não apresenta diferenças significativas entre os intervalos de 30s e de 60s. O

ensaio utilizado foi o da diferença simples e o intervalo "0s" refere-se à prova da amostra a analisar imediatamente após a prova da amostra testemunho.

A referida pesquisa foi estendida aumentando o intervalo de tempo até 1 dia (ver Almeida et al. (1999)), e as conclusões foram compatíveis com as anteriores, embora se tivesse constatado uma degradação do nível de desempenho, também nos provadores treinados (mas substancialmente inferior à dos não treinados).

Os autores atribuem estas diferenças à utilização de mecanismos de memória diferentes. Este ponto revela-se extremamente importante quando se torna necessário fazer comparações com produtos ou conceitos que não podem estar presentes fisicamente no teste.

3.4 Selecção e Treino de um Painel de Provadores Profissional

Nesta secção, aborda-se o tema da selecção e do treino de provadores profissionais, que são encarados como instrumentos de medida de características sensoriais (ver secção 2.2). Embora as actividades de selecção e de treino sejam distintas, encontram-se relacionadas, dado que, frequentemente, os processos de selecção incluem algum treino dos candidatos.

Em 1981, a ASTM editou uma publicação técnica (ASTM (1981)) cujo objectivo é providenciar linhas de orientação para seleccionar e treinar membros de painéis de avaliação sensorial para ensaios de diferenciação e/ou descritivos. Seguidamente, apresenta-se uma síntese dos tópicos mais relevantes desse guia, complementada com algumas propostas e exemplos referidos por outros autores.

3.4.1 Selecção de Provadores

A selecção dos provadores consiste numa avaliação do potencial dos candidatos e na preparação para a sua formação futura. A selecção deve basear-se tanto nas características pessoais dos candidatos como na sua capacidade potencial para desempenhar tarefas específicas na área da Análise Sensorial.

As características pessoais devem ser avaliadas com base em questionários e/ou entrevistas onde se deve tentar obter, com o maior rigor possível, informação sobre cada

um dos candidatos, nomeadamente: (i) interesse pela actividade que se propõe realizar; (ii) disponibilidade para se dedicar ao trabalho e para ser pontual; (iii) condições de saúde apropriadas ao exercício da função; (iv) capacidade de expressão (particularmente se a metodologia a adoptar incluir ensaios descritivos); (v) atitude perante a classe de produtos que se pretende avaliar. Além dos factores referidos, devem ter-se em conta outras informações como a idade, o sexo, os facto de ser ou não fumador e a experiência de trabalho na área.

Após uma primeira selecção baseada nas suas características pessoais, os candidatos "passam" à fase mais técnica, onde participam em ensaios sensoriais. Esses ensaios devem ser focalizados na(s) classe(s) de produtos e/ou nas características sensoriais que se pretendem testar no futuro.

A selecção dos candidatos nesta segunda fase deve garantir que estes possuam as seguintes qualificações básicas: (i) acuidade sensorial normal; (ii) interesse pela actividade de avaliação sensorial; (iii) capacidade para discriminar e reproduzir resultados; (iv) comportamento apropriado em termos de cooperação, motivação e rapidez de resposta.

Os ensaios escolhidos para esta fase devem ser semelhantes aos dos objectivos específicos do ensaio sensorial para o qual os provadores estão a ser seleccionados. No guia referido podem encontrar-se procedimentos para os ensaios de diferenciação (baseados no teste triangular) e os ensaios de análise descritiva (perfil de sabor, perfil de textura, QDA e testes de cotação de atributos). Note-se que, para alguns ensaios, os autores aconselham o uso de produtos químicos e/ou comerciais específicos, e recomendam concentrações e quantidades, sugerindo que se eliminem os provadores com desempenhos de nível inferior ao definido para cada caso.

Hough et al. (1995) referem que, antes de serem elaborados os testes de diferenciação específicos, se devem medir as capacidades dos candidatos para identificarem os sabores básicos, e apresentam um estudo em que, de 324 candidatos, 259 foram seleccionados para passar à segunda fase e, destes, só 226 foram bem sucedidos nos testes dos sabores básicos.

Nos testes dos sabores básicos é solicitado aos candidatos que identifiquem o sabor da solução – foram utilizadas soluções de ácido cítrico para o sabor ácido, de sacarose para o doce, de cafeína para o amargo e de cloreto de sódio para o salgado. As concentrações utilizadas foram escolhidas de entre várias hipóteses (Hough et al. (1995) apresentam uma síntese elaborada a partir de vários estudos de outros autores).

No mesmo artigo é, também, discutida a problemática dos critérios para seleccionar os provadores, tendo sido utilizado o critério de aceitar os candidatos que tenham 65% ou mais respostas certas e rejeitar os que tenham menos de 50% de respostas certas (considerando as concentrações utilizadas). No caso do número de candidatos aprovados não ser suficiente, os candidatos que tenham entre 50% e 65% deverão repetir o teste, sendo-lhes exigido, desta vez, 70% de respostas correctas.

Zamora e Calvino (1996) apresentam um estudo muito completo em termos de número e tipo de ensaios, que foi conduzido com o objectivo de seleccionar um painel sensorial para avaliação de café instantâneo. O número de candidatos que participou nos ensaios sensoriais foi de 20.

A série de ensaios específicos consistiu em: (i) testes aos sabores fundamentais, semelhantes aos já descritos; (ii) testes de correspondência e identificação de odores; (iii) avaliação da intensidade do sabor com recurso a testes de diferença em relação ao controlo; (iv) percepção de pequenas diferenças no sabor, com recurso a testes A-Não A com escalas de cotação; (v) determinação da sensibilidade a um composto químico com sabor amargo.

Os resultados foram processados com recurso a três técnicas diferentes: cálculo da percentagem média de respostas correctas, análise de variância com um factor e análise de componentes principais. Estas técnicas deram origem a diferentes números de candidatos seleccionados (10, 9 e 12, respectivamente), mas constata-se que os 12 seleccionados pela ACP incluem os 10 seleccionados pela percentagem, que, por sua vez, incluem os 9 seleccionados pela ANOVA.

Os autores concluem que, embora cada um dos métodos de processamento dos resultados possa ser utilizado individualmente, é preferível usá-los simultaneamente, de forma a poder obter o máximo de informação e a aumentar o grau de certeza com que se aceitam ou rejeitam candidatos. Deve salientar-se, ainda, que as três técnicas utilizam os mesmos resultados, não sendo, portanto, necessário aumentar o volume de trabalho durante as sessões apenas para se poder utilizar mais uma técnica de processamento.

Em Gatchalian et al. (1990), é apresentado um método para seleccionar provadores, recorrendo a um caso envolvendo fruta Filipina e seus derivados. Depois de completadas duas fases de selecção relacionadas com as características pessoais dos candidatos, foram determinados os vários tipos de limiares (diferença, reconhecimento e absoluto) para 30 candidatos com recurso ao teste da série de diluições (cada candidato

realizou 3 ensaios de cada tipo). A substância utilizada foi o sulfato de quinino, visto que o sabor amargo era a sensação mais pertinente a ser testada.

A capacidade de reproduzir os resultados é outra das características que, de uma forma geral, é referida na literatura como sendo desejável e importante num provador profissional. No entanto, não é comum encontrar muitas referências a casos práticos de selecção de provadores em que esta capacidade tenha sido testada. Gatchalian et al. (1990), por exemplo, comparam os resultados obtidos pelos provadores nas três réplicas de cada ensaio, mas não valorizam muito este critério, visto que consideram que três é um número demasiado baixo para poderem retirar conclusões definitivas. Os mesmos autores referem, também, que este é um aspecto que pode e deve ser objecto de treino.

3.4.2 Treino de Provadores

No já referido guia publicado pela ASTM (1981, p. 3) pode ler-se: "Os instrumentos devem ser calibrados para fornecer resultados padronizados que possam ser universalmente reproduzidos e interpretados. Os provadores também devem ser treinados através de experiências directas e especificamente planeadas para os familiarizar com os procedimentos dos ensaios, aperfeiçoar a sua capacidade de reconhecimento e de identificação de características sensoriais em sistemas complexos e melhorar a sua sensibilidade e memória, possibilitando, assim, julgamentos sensoriais mais precisos e consistentes."

Referindo-se à mesma problemática, Piggott (1995) destaca o facto de a função principal da Análise Sensorial ser a de providenciar resultados de medições que, para terem algum valor, deverão ser exactos, precisos e válidos. O objectivo fundamental do treino é, pois, desenvolver um painel que produza resultados com essas características, tal como um bom instrumento de medida.

3.4.2.1 Exactidão, Repetibilidade, Reprodutibilidade e Calibração

No contexto em análise, torna-se necessário clarificar o significado de cada um dos termos referidos (precisão, exactidão e validade) que, em linguagem corrente, se utilizam muitas vezes como sinónimos. A exactidão refere-se à proximidade do resultado com o verdadeiro valor da grandeza, ou melhor, à inexistência de erros sistemáticos. A precisão refere-se à variabilidade das medições: quanto menos variação apresentarem os resultados do mesmo teste efectuado nas mesmas condições, mais

precisa é a medição. Finalmente, a validade relaciona-se com o facto de os resultados expressarem realmente o que se pretende medir.

Em metrologia, o termo precisão é evitado e utiliza-se o conceito de incerteza. A incerteza da medição é definida em IPQ (1996b) como "o parâmetro associado ao resultado de uma medição que caracteriza a dispersão dos valores que podem ser razoavelmente atribuídos à mensuranda". No capítulo 4, este assunto será aprofundado, estudando-se vários tipos de incerteza.

Ainda relativamente à variabilidade, distinguem-se, normalmente, dois tipos diferentes: a variabilidade dentro de cada experiência e a variabilidade entre experiências. À variabilidade dentro de cada experiência associa-se a expressão repetibilidade dos resultados, que é definida no Vocabulário Internacional de Metrologia (IPQ(1996a)) como "Aproximação entre os resultados de medições sucessivas da mesma mensuranda efectuadas nas mesmas condições de medição." As referidas condições são chamadas condições de repetibilidade. Note-se que esta definição de repetibilidade coincide com a de precisão referida em Piggott (1995) e citada acima.

A variabilidade entre experiências é associada à expressão reprodutibilidade dos resultados, que, em IPQ (1996a), é definida como "Aproximação entre os resultados de medições da mesma mensuranda efectuadas com alterações das condições de medição." As condições que são alteradas devem ser especificadas.

A quantificação da repetibilidade faz-se com o recurso a medidas de dispersão dos dados, tal como o desvio padrão, enquanto a quantificação da reprodutibilidade se faz através da análise de variância e dos testes de diferenças entre valores esperados que lhe estão associados.

Na análise instrumental, a repetibilidade, a reprodutibilidade e a exactidão dos instrumentos de medida são determinadas em calibrações que se devem efectuar periodicamente.

Em IPQ (1996a), a calibração é definida como "O conjunto de operações que estabelecem, em condições especificadas, a relação entre valores de grandeza indicados por um instrumento de medição ou sistema de medição, ou valores representados por uma medida materializada ou um material de referência e os correspondentes valores realizados por padrões."

A leitura desta definição à luz da Análise Sensorial pressupõe que os provadores sejam calibrados e que, para tal, se utilizem padrões. Esta temática será desenvolvida de uma forma específica no Capítulo 6, para o caso prático estudado – o da Câmara de Provadores do Instituto do Vinho do Porto. Mas, de uma forma geral, pode dizer-se que apenas é possível uma calibração convencional para a detecção de sensações simples, onde os padrões são produtos referência (normalmente soluções) estabilizados e em condições previamente definidas (no caso das soluções, as respectivas concentrações).

A necessidade de treinar os provadores para reconhecer as sensações é evidente. Piggott (1995) refere, mesmo, que este é um factor-chave no sucesso das experiências sensoriais. Ainda a propósito da formação dos provadores para reconhecer as sensações, Stone e Sidel (1993, p. 210) referem estudos onde se podem encontrar listas de substâncias referência para determinadas sensações e respectivas instruções de utilização. No entanto, os autores chamam a atenção para o facto de que, ao utilizar como padrões substâncias com um grau de complexidade elevado, se introduzem outros atributos que não se relacionam com o objectivo da calibração e que aumentam as interacções sensoriais, podendo perturbar os julgamentos dos provadores.

3.4.2.2 Fases do Treino de Provadores

A primeira fase do treino deve consistir numa formação, na qual os provadores se familiarizam com os produtos a ser testados e com os procedimentos específicos dos ensaios. Normalmente, realizam-se sessões individuais ou em grupo, nas quais se esclarecem dúvidas, se trocam informações e se discutem problemas relacionados com as características sensoriais relevantes (por exemplo, o vocabulário a utilizar). O uso de amostras de referência é aconselhado, quando possível.

O número de sessões e a sua duração depende do número de produtos diferentes, da sua complexidade, das características a estudar, do procedimento de teste escolhido e da velocidade demonstrada por cada provador. Deve, no entanto, assegurar-se que os provadores, ao longo das várias sessões, adquirem níveis satisfatórios de consistência nos seus resultados.

O sucesso desta fase é essencial para o sucesso das fases seguintes e para assegurar a validade dos resultados. Note-se que a validade está intimamente relacionada com o facto de os provadores utilizarem a terminologia correcta para se referirem às características e/ou sensações que se pretendem medir, ou, inversamente, identificarem correctamente as características e/ou sensações a partir da terminologia apresentada.

Quando os ensaios a realizar têm uma natureza simples, como acontece, por exemplo, com o teste triangular, não são, em princípio, necessárias mais fases no treino dos provadores. Já quando a natureza dos ensaios é mais complexa e envolve, por exemplo, escalas, fichas de prova ou um número elevado de produtos, torna-se necessária uma segunda fase de treino, na qual os provadores participam em ensaios reais.

A segunda fase é necessária para assegurar que os critérios, as escalas e a terminologia são compreendidos e utilizados de uma forma semelhante por todos os provadores. Os ensaios devem, portanto, ser repetidos até que o desempenho dos provadores seja considerado fiável e consistente.

No guia da ASTM para a selecção e o treino de provadores (ASTM (1981)), são apresentadas instruções específicas para o treino de provadores para alguns dos ensaios descritivos mais utilizados.

Quando os ensaios envolvem escalas de intervalo ou absolutas, torna-se necessário treinar os provadores para a sua utilização. A questão das escalas é extensamente abordada na literatura especializada, dado que, naturalmente, as pessoas têm tendência para as utilizar de forma diferente. Algumas das técnicas estatísticas referidas em 3.2 permitem quantificar e harmonizar os efeitos decorrentes dessas diferenças. Riskey (1988) e Moskowitz e Jacobs (1988) analisam os vários tipos de escalas e apresentam técnicas para formar os provadores na sua utilização. Em alguns casos, sugerem mesmo que os provadores sejam devidamente calibrados para o uso das escalas com produtos-padrão.

A complexidade deste problema reside no facto de ser, por vezes, difícil distinguir se a diferença entre as avaliações feitas pelos provadores é devida ao facto de os provadores utilizarem a escala de formas diversas ou ao facto de as sensações que experimentam serem, na realidade, diferentes.

Nesta secção (3.4) foram sendo referidos alguns temas que, pela sua importância e relevância para o trabalho, serão retomados nos capítulos seguintes. De facto, o controlo de Câmaras de Provadores passa por avaliar, ao longo do tempo, as capacidades dos seus elementos para desempenharem as funções de instrumentos de medida. Assim, questões como a definição e utilização de padrões, a calibração periódica de provadores e a capacidade de os provadores reproduzirem os seus julgamentos serão analisadas com mais profundidade nos próximos capítulos.

No capítulo seguinte, será abordada a temática do controlo de processos não industriais de uma forma geral, sem a preocupação de aplicar esses métodos directamente à Análise Sensorial. O objectivo é, antes, o de analisar algumas ferramentas que poderão ser úteis e servir de ponto de partida para o desenvolvimento do sistema de controlo pretendido.

BIBLIOTECA F.E. PORTO

4

Métodos de Controlo de Processos

Conteúdo:

- **Introdução**
- **Instrumentos de Medida e Sua Calibração**
- **Cartas de Controlo**
- **Utilização do Controlo Estatístico em Processos Não Industriais**

4.1 Introdução

O principal objectivo deste capítulo é verificar até que ponto os métodos tradicionais de controlo de processos se podem aplicar ao problema estudado nesta tese: o controlo de processos em Análise Sensorial.

Desta forma, numa primeira fase foi necessário abordar os métodos utilizados no enquadramento tradicional da calibração de instrumentos de medida e de controlo estatístico de processos, para, numa segunda fase, avaliar a exequibilidade da implementação desses métodos em áreas menos convencionais, como é o caso da Análise Sensorial.

O controlo de processos é uma matéria que, tradicionalmente, é analisada no contexto da Qualidade. Particularmente, o controlo estatístico de processos é tido como uma ferramenta indispensável na implementação da Gestão pela Qualidade Total em processos produtivos (ver, por exemplo, Roes e Dorr (1997)).

Acerca do significado da palavra controlo, em Cabral (2001) pode ler-se:

> "Ao termo **controlo** associam-se, em geral, duas ideias aparentemente contraditórias: a ideia de estabilidade e a de regulação. De facto, quando se diz que a velocidade de um automóvel está controlada isso significa que os seus valores se mantêm estáveis dentro de determinados limites (pretendidos ou aceitáveis). Mas, por outro lado, para controlar a velocidade do automóvel (isto é, para o manter à velocidade pretendida, dentro dos limites da faixa de rodagem) o condutor tem que actuar sobre o acelerador, o volante, a caixa de velocidades, o travão, etc. Isto é, quando o fenómeno em causa (a velocidade do automóvel) está sujeito a factores de variação suficientemente fortes para provocarem instabilidade, é necessário actuar para eliminar, reduzir, ou prevenir os efeitos de tais factores e manter o fenómeno estabilizado. Controlar significa assim intervir no sistema por forma a que ele se comporte da forma pretendida e frequentemente, na medida em que se efectuam correcções, significa também melhorar. O conceito de controlo ultrapassa largamente o simples acto de inspecção ou de acompanhamento."

Ainda no contexto da Qualidade, **processo** é definido como "uma sequência de actividades que adicionam valor através da produção de *outputs* a partir de uma

variedade de *inputs*" (EFQM (1999)), ou como "uma actividade que usa recursos e que é gerida com o objectivo de permitir a transformação dos *inputs* em *outputs*" (ISO (2000)). Note-se que a concepção actual dos sistemas de gestão da qualidade tende a basear-se numa abordagem por processos, como se pode ver nas normas internacionais ISO 9001:2000 (ISO (2000)).

Os métodos estatísticos têm-se revelado muito úteis e eficazes no melhoramento dos processos produtivos e na redução do número de defeitos e de produtos defeituosos nos ambientes industriais. De facto, sendo a estatística uma disciplina que estuda os dados cujos valores variam de experiência para experiência, de momento para momento ou de pessoa para pessoa, a aplicação dos métodos estatísticos ao controlo deste tipo de processos surge de uma forma natural.

As ferramentas mais utilizadas no âmbito do controlo estatístico de processos industriais são as cartas de controlo. Há inúmeros livros e artigos publicados sobre esta temática. Em poucas palavras (ver, por exemplo, Montgomery (2000, p. 101)), um processo diz-se sob controlo estatístico quando a variação do seu *output* tem natureza aleatória e uma distribuição estável. Este tipo de variação designa-se, frequentemente, por ruído de fundo.

Há, no entanto, outro tipo de variações que afectam significativamente a distribuição do *output* do processo e que ocorrem ocasionalmente. As suas causas relacionam-se, em geral, com um (ou mais) de três factores: máquinas desajustadas, erros dos operadores ou problemas com a matéria prima. A ordem de grandeza desta variabilidade é, na maior parte dos casos, superior à do ruído de fundo e apresenta níveis incompatíveis com o desempenho pretendido para o processo. Às causas deste tipo de variabilidade dá-se o nome de causas especiais (ou causas assinaláveis), e, na sua presença, o processo diz-se fora de controlo.

O principal objectivo da utilização das cartas de controlo é detectar rapidamente a ocorrência de causas especiais, de forma a que a investigação sobre o processo e as devidas acções correctivas possam ser implementados antes de serem produzidos demasiados produtos defeituosos. Paralelamente, as cartas de controlo também podem ser utilizadas para estimar parâmetros do processo e para providenciar informação que permita melhorar o processo e/ou reduzir a sua variabilidade.

Tipicamente, uma carta de controlo é uma representação gráfica dos valores de uma estatística amostral correspondente a uma característica do produto, medida ou calculada para amostras sucessivas, em função do número da amostra ou do tempo (ver

Montgomery (2000, p. 102)). Essa característica é, normalmente, denominada característica de qualidade.

A correcta medição da característica em causa constitui um requisito essencial para que o controlo do processo seja efectivo. De facto, um erro na medição da característica de qualidade poderá levar a que a carta de controlo emita um sinal de alarme (processo fora de controlo) quando não houve alteração significativa na característica, ou, inversamente, que não emita o sinal quando, na realidade, o deveria emitir.

Como facilmente se compreende, os instrumentos de medida convencionais, particularmente quando utilizados na medição de grandezas importantes, são sujeitos a calibrações periódicas com a finalidade de verificar a sua adequabilidade para medir a grandeza em causa.

Não se pretende, com o presente capítulo, tratar exaustivamente os assuntos da calibração dos instrumentos nem do controlo estatístico de processos. Pretende-se, tal como foi salientado logo no início desta secção, apresentar uma panorâmica geral desta problemática tal como é encarada no ambiente industrial e analisar as possibilidades da sua utilização em processos não industriais. Destes últimos, salientam-se aqueles cujas características são medidas em escalas de cotação ou multinomiais, como é o caso de inquéritos à satisfação dos clientes ou de alguns ensaios sensoriais.

Desta forma, na secção 4.2, aborda-se o tema da calibração de instrumentos de medida e, na secção 4.3, apresenta-se uma síntese dos conceitos associados às cartas de controlo mais utilizadas, ilustrando-se o funcionamento de algumas delas. Finalmente, na secção 4.4, analisa-se o tema da utilização do controlo estatístico em processos não industriais. Nesta última secção, salientam-se duas aplicações (recorrendo a dados simulados) referentes a processos cujo *output* é medido numa escala multinomial com cinco níveis: uma calibração e o controlo estatístico de um processo.

4.2 Instrumentos de Medida e sua Calibração

Os termos calibração, exactidão, repetibilidade e reprodutibilidade foram já definidos no capítulo anterior, a propósito da necessidade de calibrar os painéis de provadores. As definições apresentadas, no entanto, não se aplicam unicamente ao caso da Análise Sensorial, tendo sido, aliás, obtidas a partir de fontes relacionadas com a calibração de instrumentos de medida convencionais. Nesta secção, descreve-se, sucintamente, a forma como são feitas as calibrações dos instrumentos de medida convencionais, na

perspectiva de que sirva como um ponto de partida para a abordagem à calibração de outros tipos de "instrumentos", como é o caso dos provadores.

Por uma questão de simplificação na ilustração dos conceitos, na próxima secção assume-se que as grandezas consideradas se podem representar por variáveis aleatórias contínuas. Na secção 4.2.2, referir-se-ão as condições em que estes conceitos poderão ser generalizados a variáveis de natureza discreta, e apresenta-se uma sugestão para realizar a calibração quando as variáveis em causa não respeitem essas condições e sejam representadas por uma escala do tipo multinomial.

4.2.1 Calibração e Incertezas

No Guia para a Expressão da Incerteza de Medição nos Laboratórios de Calibração (IPQ (1996b)) pode ler-se: "A expressão de um resultado de medição só está completa quando contém o valor atribuído à mensuranda e a incerteza de medição associada a esse valor".

A incerteza de medição (ver definição na secção 3.4.2.1) resulta do facto de haver factores, uns aleatórios e outros mais ou menos sistemáticos, que influenciam a medição e que estão sujeitos a variações impossíveis de controlar totalmente. O conjunto de operações através das quais é determinada a incerteza (que se deve associar ao resultado da medição feita por um determinado instrumento) é a calibração, que já foi definida em 3.4.2.1.

Os processos de calibração diferem muito, dependendo do tipo de grandezas a medir. Quando a grandeza se pode medir directamente e existe um padrão ou material de referência, a calibração pode ser feita por comparação e consiste na realização de N observações da grandeza a medir, independentes e feitas nas mesmas condições de medição.

Por exemplo, quando se pretende calibrar um termómetro para uma determinada temperatura, deve utilizar-se um termómetro padrão e um meio (por exemplo, um banho termostático de água destilada) que se encontre à temperatura pretendida (obviamente medida no termómetro padrão).

A incerteza de medição correspondente ao termómetro a calibrar (u_y) será função das incertezas devidas às diversas causas de variabilidade introduzidas no processo de calibração. No caso em análise, podem considerar-se quatro causas de variabilidade, que dão origem a outras tantas incertezas: a incerteza devida à calibração do termómetro

padrão (u_{ct}), a incerteza devida à estabilidade do meio (u_{em}), a incerteza devida à resolução da escala do termómetro (u_{et}) e, finalmente, a incerteza devida à dispersão das leituras (u_{dl}).

Uma das medidas passíveis de exprimir a incerteza de um dado resultado numérico é o desvio padrão que, neste contexto, é designado por incerteza padrão. Assumindo que os valores correspondentes às três primeiras incertezas referidas no parágrafo anterior se encontram já determinados, resta, no processo de calibração, repetir as leituras do termómetro a calibrar um número N pré-definido de vezes (normalmente usa-se $N \geq 10$) e calcular o desvio padrão correspondente, que será uma estimativa da incerteza padrão devida à dispersão das leituras.

A variância da medição feita pelo termómetro sujeito a calibração será igual à soma das variâncias correspondentes aos diversos factores referidos, dado que estes se podem considerar independentes e aditivos. Deste modo, a incerteza padrão combinada (é assim que se denomina a incerteza padrão quando é obtida através de várias outras incertezas) será dada pela seguinte expressão:

$$u_y = \sqrt{u_{ct}^{\,2} + u_{em}^{\,2} + u_{et}^{\,2} + u_{dl}^{\,2}}$$

A avaliação da incerteza devida à dispersão das leituras, que foi brevemente exemplificada, pertence à categoria Tipo A, na qual o método de avaliação da incerteza se baseia na análise estatística de séries de medições repetidas. No entanto, e quando tal não é possível, seja por razões técnicas ou outras, recorre-se à avaliação Tipo B da incerteza, onde a incerteza padrão é avaliada com base em toda a informação técnico-científica disponível sobre a variabilidade da grandeza em causa.

Uma vez calibrado um instrumento de medida, a informação relativa à incerteza determinada no processo deve estar sempre associada ao resultado das medições feitas com esse instrumento.

No entanto, o valor que habitualmente aparece associado aos resultados das medições efectuadas com os instrumentos é a incerteza expandida (U), que não é mais do que um múltiplo da incerteza padrão ($U = K \cdot u_y$). O valor da mensuranda deverá encontrar-se, com uma determinada confiança p, no intervalo caracterizado pelo resultado da medição mais ou menos a incerteza expandida.

Na prática, é usual admitir-se que a distribuição da estimativa da incerteza padrão da grandeza de saída (no exemplo referido, a leitura do termómetro a calibrar) se aproxima

de uma distribuição normal. Esta aproximação é considerada válida em IPQ (1996b), desde que as componentes da incerteza sejam não correlacionadas, que provenham de distribuições bem conhecidas e que contribuam para a incerteza combinada em montantes comparáveis. No caso de as incertezas serem calculadas por avaliação do tipo A, estas devem ser baseadas em, pelo menos, 10 observações.

Nas condições descritas que, segundo os mesmos autores, são preenchidas na maioria dos casos encontrados em trabalhos de calibração, o valor que se utiliza mais frequentemente para a incerteza expandida é o de $U = 2 \cdot u_y$ (onde $K = 2$), ao qual corresponde um valor de p de aproximadamente 95%.

O exercício de calibração é, em princípio, repetido periodicamente, em intervalos de tempo pré-estabelecidos e as incertezas dos instrumentos calibrados são consideradas válidas até à calibração seguinte (na qual são calculadas novas incertezas). A periodicidade da calibração depende do tipo de instrumentos e processos em que estão envolvidos e faz parte da rotina de funcionamento do instrumento.

Note-se que, quando o funcionamento normal do instrumento de medida é afectado por algum factor externo (uma avaria, por exemplo) deve haver uma calibração extraordinária, no sentido de repor a normalidade da situação.

4.2.2 Calibração para Grandezas Expressas por Variáveis Discretas

Os conceitos relacionados com variabilidade e calibrações introduzidos até agora, são mais directamente aplicáveis aos casos em que as grandezas mensurandas são passíveis de ser expressas por variáveis de natureza contínua do que àqueles em que estas são expressas por variáveis discretas.

No entanto, os problemas que se colocam quando as variáveis são de natureza discreta são semelhantes. A questão é que, em muitos casos, não faz sentido associar um valor da incerteza expandida a resultados expressos através de uma variável discreta, pelo simples facto de os valores contidos no intervalo $[y - U,\ y + U]$ serem, eles próprios, de natureza contínua. Este facto torna-se mais evidente se se considerar o caso extremo da variável binária que apenas toma os valores "0" e "1". Supondo que, numa determinada medição, se obtém o valor "1" para a variável, não faz sentido associar-lhe qualquer intervalo, uma vez que o único valor alternativo que a variável poderá tomar é o "0".

Deverão, portanto, ser estudadas outras formas de caracterizar a incerteza associada a tais variáveis, para que se possa ter a noção de qual o grau de confiança a atribuir à medição em causa. De notar que a incerteza padrão pode ser calculada tanto para variáveis contínuas como discretas, e o seu significado mantém-se o mesmo – quanto maior a incerteza padrão, maior a dispersão dos resultados em torno do seu valor esperado.

Quando a variável discreta é expressa numa escala de cotação com um número de níveis suficientemente elevado, e se a sua distribuição tiver uma forma que se possa considerar próxima de uma distribuição contínua, poderá, teoricamente, fazer-se uma calibração semelhante à descrita, tomando sempre em consideração que se trata de uma aproximação.

Se, pelo contrário, não for possível aproximar a variável multinomial por uma variável contínua, não será possível calcular a incerteza expandida recorrendo ao desvio padrão. Uma possibilidade de ultrapassar este problema é associar, a cada nível da escala (correspondente a um resultado), a probabilidade de ele ser correctamente atribuído.

Na secção 4.4.1, apresenta-se um caso em que este procedimento é descrito e simulado.

4.3 Cartas de Controlo

Nesta secção, descreve-se, sucintamente, a utilização das cartas de controlo em processos industriais, apresentam-se exemplos simples que ilustram o funcionamento de três delas (carta de médias, carta p e carta CUSUM) e referem-se brevemente algumas outras cartas que poderão ter interesse na aplicação a processos não industriais.

Um dos objectivos fundamentais das cartas de controlo é, tal como foi enunciado em 4.1, identificar causas especiais de variação, delimitando os intervalos de tempo em que tais fenómenos ocorrem e, portanto, facilitando a sua posterior eliminação.

Na grande maioria dos processos de produção em grande escala, a inspecção de todos os itens produzidos (inspecção 100%) é impraticável, seja pelos elevados custos envolvidos, seja pela natureza destrutiva das actividades ligadas à própria inspecção. No entanto, a necessidade de conhecer as características dos produtos que saem das linhas

de produção é, cada vez mais, de uma importância vital para a sobrevivência das empresas.

Uma das soluções possíveis para a questão apontada passa pela utilização dos métodos de inferência estatística, que se caracterizam pelo facto de, a partir de uma amostra, se inferir acerca da população.

No que diz respeito à análise dos dados amostrais, Cabral (2001) salienta:

> "Aquilo que mais complica a análise de dados amostrais e a respectiva tomada de decisão é o fenómeno da variabilidade. Perante a diversidade dos valores observados numa amostra poder-se-á concluir que o processo subjacente se manteve estável durante o período de recolha dos dados? Ou, pelo contrário, será que as diferenças entre os valores amostrais permitem afirmar que o processo terá sofrido oscilações significativas originadas por causas que escaparam ao nosso controlo? Como se compreende facilmente, o teste de hipóteses tem, nesta matéria, um enorme interesse prático, uma vez que, a partir de amostras reduzidas, permite efectuar inferências acerca dos parâmetros ou da forma das populações subjacentes."

Mais adiante, o mesmo autor (Cabral (2001)) refere:

> "Neste contexto, estabilidade e controlo são palavras equivalentes. Ou, de outra forma, se um processo está em controlo a característica em avaliação seguirá uma determinada distribuição de probabilidade estável (conhecida ou desconhecida), com parâmetros constantes. (...) Note-se que pelo facto de um processo estar em controlo nada se pode inferir quanto à magnitude da variação que lhe está associada. Por exemplo, de um processo em controlo tanto podem resultar, em média, 50% como 5% de defeituosas: o processo estará em controlo enquanto a distribuição da variável "número de defeituosas" for estável."

As primeiras cartas de controlo que surgiram foram desenvolvidas por Shewhart e são, actualmente, denominadas cartas de controlo clássicas ou cartas Shewhart. O seu aspecto é semelhante ao indicado na figura 4.1.

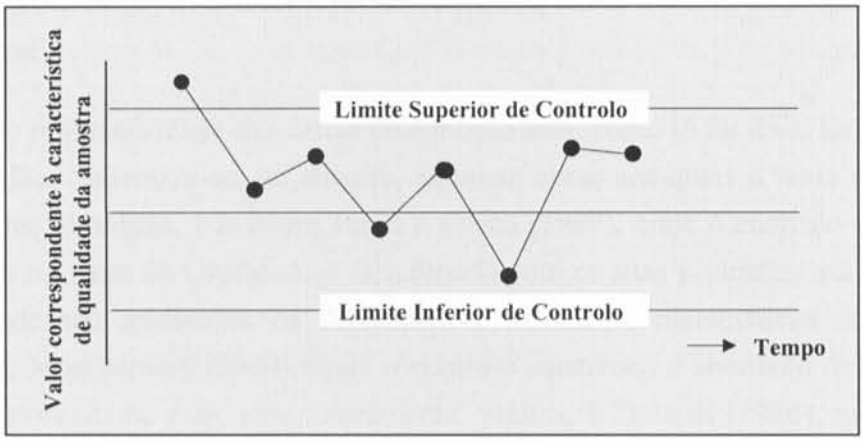

Figura 4.1 – Ilustração do aspecto de uma carta de controlo clássica

Os pontos são assinalados na carta tomando em consideração que o eixo dos *XX* corresponde ao instante temporal da recolha da amostra ou, em alternativa, ao próprio número da amostra, e o eixo dos *YY* corresponde ao valor registado para a estatística amostral referente à característica que se pretende controlar. Quando um ponto sai fora dos limites da carta (seja o limite inferior, seja o superior), conclui-se que, com grande probabilidade, os valores correspondentes à característica da qualidade já não seguem a distribuição que era suposto seguirem, a partir da qual foram calculados os referidos limites.

Fundamentalmente, estas cartas de controlo não são mais do que uma aplicação prática do método de inferência estatística de Teste de Hipóteses, já descrito no capítulo anterior a propósito da análise dos resultados dos ensaios sensoriais. O teste de hipóteses que é elaborado depende do tipo da característica a medir e da distribuição que os seus valores seguem no estado de controlo.

As cartas de controlo clássicas dividem-se em dois grupos: as cartas de controlo de variáveis e as de atributos. Na linguagem relativa à área da Qualidade, quando uma característica se pode exprimir de uma forma quantitativa, como é o caso de uma dimensão de uma peça, do seu peso ou do seu volume, essa característica é denominada variável. As cartas de controlo cujo objectivo é o de acompanhar directamente os valores desse tipo de características são, então, chamadas cartas de controlo de variáveis.

Há, no entanto, muitas características que não é possível exprimir de uma forma quantitativa, tais como o funcionamento de um transistor (funciona ou não) ou de um fio condutor (conduz a corrente ou não). Nestes casos, as medidas designam-se por atributos e resultam da contagem de dados originalmente expressos em escalas nominais

e ordinais. Note-se que os atributos podem, também, tomar a forma de proporções ou de percentagens.

O estudo pormenorizado das cartas de controlo está, como já foi dito, fora do âmbito deste trabalho. Salientam-se, no entanto, algumas obras nas quais o tema é rigorosa e extensamente abordado, tais como Juran e Gryna (1988), onde o controlo estatístico é enquadrado no tema da Qualidade e relacionado com os suas principais vertentes, Burr (1976), onde são analisados os fundamentos teóricos e matemáticos dos métodos estatísticos, Montgomery (2000), onde o controlo estatístico é abordado de uma forma bastante aprofundada, mas, simultaneamente, prática, e Duncan (1986), que se refere muito especialmente aos ambientes industriais e onde se podem encontrar vários exemplos de aplicação.

Em seguida descrevem-se, sucintamente, três tipos de cartas de controlo: a carta de médias, que é uma carta de controlo de variáveis, a carta *p*, que é uma carta de atributos e a carta CUSUM, que foi desenvolvida mais tarde e se enquadra no grupo das cartas de controlo especiais.

4.3.1 Carta de médias

A carta de controlo de médias ($\bar{x}$) baseia-se no teste de hipóteses ao valor esperado da população e é construída com base no pressuposto de que a média amostral ($\overline{X}$) segue uma distribuição aproximadamente Normal com valor esperado μ_0. As hipóteses em análise são, portanto, expressas por:

H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$.

A hipótese nula traduz a situação de controlo estatístico. Nestas circunstâncias, e assumindo que se conhece o desvio padrão da distribuição da população (σ), a probabilidade de se observarem valores no intervalo $\mu_0 \pm k\frac{\sigma}{\sqrt{N}}$ é constante (note-se que N é a dimensão da amostra, que se considera constante, e que $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}$). Normalmente utiliza-se o valor de $k = 3$, para o qual a referida probabilidade é de 99.73%, sendo os limites de controlo os extremos do referido intervalo.

Na figura 4.2, ilustra-se uma carta de controlo de médias com 5 pontos representados. Cada ponto representa a média dos valores obtidos para a característica,

para a amostra correspondente. Na mesma figura pode verificar-se que a cada valor de $\bar{x}$ corresponde um teste de hipóteses (a distribuição da estatística de teste quando H_0 é verdadeira encontra-se representada graficamente do lado esquerdo, invertida e rodada de 90º em relação aos gráficos normalmente utilizados na explicação dos testes de hipóteses). Os limites inferior (LIC) e superior (LSC) de controlo são, precisamente, os valores críticos da estatística de teste, que separam as zonas de rejeição e de não rejeição de H_0.

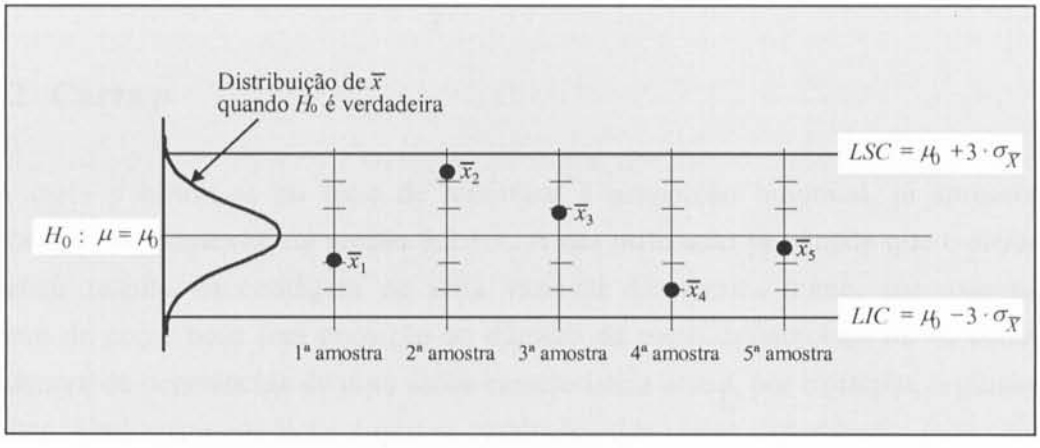

Figura 4.2 – Carta de controlo de médias - relação entre o teste de hipóteses e a carta de controlo

Assumiu-se, até aqui, que tanto o valor esperado como o desvio padrão da distribuição original são conhecidos ou fixados à partida, o que, na maior parte dos casos, não se verifica. Quando não se conhece μ_0 nem σ, os limites de controlo são calculados com base em estimativas dos referidos parâmetros, que, por sua vez, são determinadas recorrendo a resultados obtidos com amostras piloto - m amostras de dimensão N. A recolha dessas amostras deve ser feita obedecendo a regras que evitem que os dados estejam contaminados por efeitos provocados por causas assinaláveis.

Este problema é comum a todos os tipos de cartas de controlo e a sua importância é evidente: se os dados não correspondem ao estado de controlo, os limites da carta reflectem não só a variação devida às causas aleatórias, mas também a que é devida a algumas das causas assinaláveis, e a carta será, no futuro, insensível à presença destas últimas. A maneira de estimar o valor esperado e o desvio padrão da população sob controlo é, portanto, um assunto extensamente abordado na literatura especializada (ver, por exemplo, Juran e Gryna (1988) ou Montgomery (2000)), que deve ser analisado com toda a prudência quando se trata de implementar, na prática, qualquer tipo de controlo estatístico.

A carta de médias é, em geral, utilizada em conjunto com uma carta de amplitudes ou de desvios padrões, consoante $N \leq 8$ ou $N > 8$, respectivamente. Esta diferença resulta do facto de, para dimensões amostrais reduzidas, a estatística "amplitude amostral" (A) ser um estimador de σ mais eficiente do que o desvio padrão amostral (s).

Através da utilização conjunta das duas cartas, é possível detectar alterações na distribuição subjacente tanto no que respeita à tendência central (carta de médias) como no respeita à dispersão (carta A ou carta s).

4.3.2 Carta *p*

A carta p baseia-se no teste de hipóteses à proporção binomial, já apresentado, embora noutro contexto, na secção 3.2.1.1. A sua utilização pressupõe que o atributo a controlar resulte da contagem de uma variável dicotómica como, por exemplo, o número de peças boas (em oposição ao número de peças defeituosas) ou da contagem do número de ocorrências de uma única característica como, por exemplo, o número de defeitos. Uma outra condição é que os resultados das várias experiências (sob controlo estatístico) sejam independentes, o que, no exemplo das peças, equivale a afirmar que o facto de uma peça ser ou não defeituosa não altera a probabilidade de ocorrência do resultado relativo à peça que se retira logo a seguir (ou relativo a qualquer outra).

Continuando a analisar o mesmo exemplo, o teste de hipóteses é feito considerando as hipóteses seguintes:

$H_0: p = p_0$

$H_1: p \neq p_0$

onde p é a proporção de peças defeituosas na população e p_0 é o valor dessa proporção que corresponde ao estado de controlo estatístico.

Admita-se que se retiram periodicamente amostras de N peças da linha da produção e, dessas N, se contam as defeituosas (Y). Quando o processo produtivo se encontra sob controlo estatístico e se a dimensão da amostra é suficientemente grande ($N > 10$, por exemplo), a proporção amostral de peças defeituosas produzidas ($\frac{Y}{N}$) segue uma distribuição aproximadamente normal cujo valor esperado é igual à proporção de defeituosas na população (p_0) e cuja variância é igual a $\sigma^2 = \frac{p_0 \cdot (1 - p_0)}{N}$. Nestas condições, é possível elaborar uma carta de controlo semelhante à carta de médias (ver

figura 4.2), onde a hipótese nula, que corresponde à linha central da carta, é substituída por H_0: $p = p_0$, e em vez dos valores de $\bar{x}$ se registam os de $\frac{Y}{N}$. Os limites de controlo da carta p são expressos por:

$$LIC = p_0 - k\sqrt{\frac{p_0(1-p_0)}{N}}$$

$$LSC = p_0 + k\sqrt{\frac{p_0(1-p_0)}{N}},$$

onde k toma normalmente o valor 3, tal como acontece na carta de médias (sendo a correspondente probabilidade de não rejeitar H_0 quando ela é verdadeira também de 99.73%).

Na maior parte dos casos em que se utiliza esta carta, o valor de p_0 é desconhecido, pelo que se usa uma sua estimativa baseada em dados históricos ou em amostras piloto. As razões pelas quais se devem tomar especiais precauções na definição destas estimativas foram já expostas relativamente à carta de médias e a literatura onde se podem encontrar exemplos e conselhos para a sua utilização é, também, a mesma.

Tanto a carta p como a de médias fazem parte do grupo de cartas de controlo que se designa por cartas de controlo Shewhart. A maior desvantagem deste tipo de cartas é que apenas utilizam a informação sobre o processo contida em cada ponto registado, ignorando qualquer outra informação que seja expressa pela sequência de pontos.

Para evitar este inconveniente, são, muitas vezes, adicionados outros critérios à definição de processo fora de controlo (que não apenas um ponto fora dos limites), critérios esses relacionados com as sequências de pontos. Uma das regras utilizadas é, por exemplo, "o processo encontra-se fora de controlo se houver 7 pontos seguidos do mesmo lado da linha central". Note-se que, embora esta seja uma forma de minorar o inconveniente referido, é óbvio que, simultaneamente, irá aumentar o número de "falsos alarmes", ou seja, o número de vezes que a carta indica que o processo está fora de controlo quando ele, na realidade, se encontra sob controlo.

4.3.3 Carta CUSUM

A carta CUSUM foi proposta como alternativa às cartas Shewhart precisamente com o objectivo de incorporar directamente, em cada ponto registado, a informação contida na sequência dos valores amostrais.

O valor que é registado em cada momento é a soma acumulada dos desvios dos valores amostrais relativamente a um valor de referência. A sigla CUSUM provém, precisamente, da expressão *Cumulative Sum*. Por exemplo, quando se pretende controlar o valor esperado de uma característica resultante de um dado processo, a estatística CUSUM registada na carta para cada amostra (ou instante de tempo) m é dada por:

$$S_m = \sum_{i=1}^{m} (\bar{x}_i - k)$$

onde:

$\bar{x}_i$: média da $i^{\text{ésima}}$ amostra;

k: valor referência.

Se o valor esperado da variável X_i se mantiver estável, S_m tende a evoluir linearmente, sofrendo, em cada amostra, um acréscimo de valor esperado igual a $E(X_i) - k$. Normalmente utilizam-se valores de k próximos de $E(X_i)$. Se, à partida, há um determinado desvio (Δ) que se pretende detectar com uma determinada probabilidade, é normal escolher k aproximadamente entre os valores de $E(X_i)$ e de $E(X_i) + \Delta$.

Os "limites de controlo" desta carta são expressos de uma forma muito mais complexa do que nas cartas Shewhart. Há duas alternativas mais usuais para aplicar os critérios de decisão, que indicam quando é que o parâmetro em avaliação sofre alterações significativas. Uma é a utilização da "Máscara V", que se sobrepõe no último ponto do gráfico onde são representados os valores de S_m em função de m, e que assinala a existência de uma situação de fora de controlo quando algum valor de S_m se situa para cima ou para baixo das linhas da máscara (ver, por exemplo, Harrison et al. (1986)). A outra é a construção de uma tabela onde a estatística CUSUM é registada, adoptando-se a regra do intervalo de decisão, enunciada resumidamente em Cabral (1990, p. 204):

> "Sempre que a observação corrente do CUSUM subir (ou descer) em relação à menor (ou maior) das observações anteriores mais do que uma quantidade H

previamente fixada, considera-se que a média do processo se alterou significativamente."

O valor de H é determinado com base no compromisso definido entre os valores da probabilidade de ocorrer um falso alarme e da probabilidade de detectar correctamente uma alteração do parâmetro e no desvio padrão da população em causa.

Em Cabral (1990), pode encontrar-se um exemplo de construção da tabela CUSUM para o caso do valor esperado de uma característica distribuída normalmente.

As vantagens da utilização da tabela CUSUM em relação à "Máscara V" consistem no facto de não ser necessária a implementação gráfica e na possibilidade de conduzir testes unilaterais (só a subidas ou só a descidas dos valores do parâmetro a controlar), ou ainda testes com especificações diferentes para as subidas e para as descidas do parâmetro. Note-se que também é possível representar graficamente os valores da tabela e os dos limites do intervalo de decisão, sendo que esse gráfico apresenta uma forma semelhante à das cartas de controlo de Shewhart.

Até agora foi assumido que se tratava de controlar o valor esperado de uma característica resultante de um processo. No entanto, a carta CUSUM pode ser utilizada para controlar características expressas por outro tipo de variáveis, tais como proporções binomiais, desvios-padrão amostrais, ou amplitudes amostrais.

Mais adiante, neste trabalho, será descrita com algum pormenor a forma de construir uma tabela CUSUM para controlar características expressas por uma proporção binomial.

4.3.4 Medidas da Sensibilidade das Cartas de Controlo

Também o termo sensibilidade já foi abordado neste trabalho num outro contexto: os provadores mais sensíveis a um determinado estímulo detectam-no mais facilmente. Foram, inclusivamente, definidas medidas de sensibilidade dos provadores aos estímulos.

A sensibilidade de uma carta de controlo corresponde à sua capacidade para detectar situações de descontrolo.

Há várias medidas que podem ser utilizadas para avaliar essa capacidade. Retomando a linguagem dos testes de hipóteses e assumindo que a rejeição da hipótese nula

significa que o parâmetro em análise sofreu uma alteração significativa em consequência da acção de causas assinaláveis de variação, a capacidade para uma carta de controlo detectar uma situação de descontrolo pode ser medida pela potência do teste efectuado, $P_d = (1-\beta)$, face à hipótese alternativa que é colocada.

No contexto do controlo estatístico da qualidade e como alternativa à potência do teste, é comum utilizar-se a probabilidade de não rejeição de H_0 quando ela é falsa (ou de não detecção de descontrolo), que se denota por $P_{\bar{d}} = \beta$. Em geral, $P_{\bar{d}} = \beta$ é representada graficamente em função de hipóteses alternativas cujos valores se expressam como desvios a H_0. Tais desvios são frequentemente definidos como múltiplos do desvio padrão de X ($\Delta = \pm k \cdot \sigma$). Esta representação gráfica designa-se por Curva Característica Operacional (em inglês, *Operational Characteristic Curve*, cuja sigla é *OCC*) e traduz a capacidade (ou sensibilidade) da carta de controlo para detectar situações que se afastam de H_0. Ou melhor, no caso das cartas de controlo, a Curva Característica Operacional traduz a probabilidade de um ponto se situar dentro dos limites de controlo em função de cada valor possível do parâmetro em análise da distribuição subjacente.

Na Figura 4.3, apresenta-se a *OCC* de uma carta de controlo de médias com $N = 4$. Note-se que, mesmo para desvio moderados, a probabilidade de, na primeira amostra, não se observar um valor fora dos limites de controlo é elevada. Por exemplo, se o valor esperado do processo se deslocar de $\pm 1.5 \cdot \sigma$ em relação a μ_0, a probabilidade de esse facto ser detectado na próxima amostra é exactamente igual à de não o ser, ou seja, 50%.

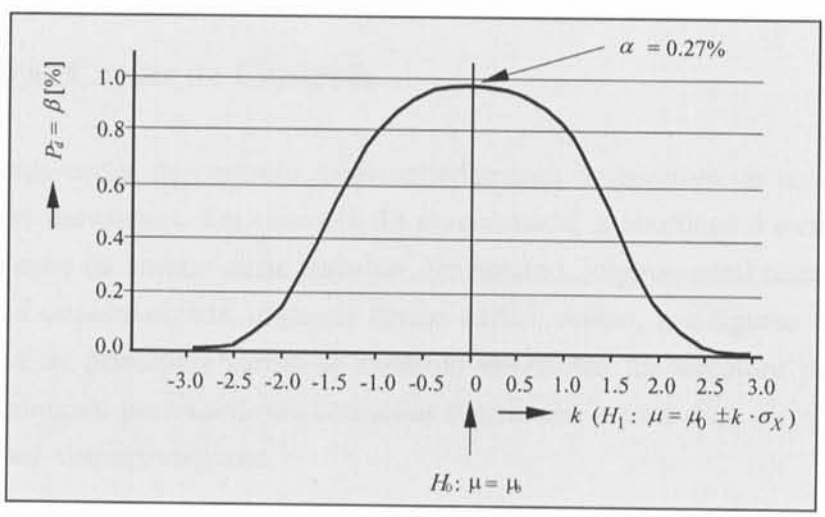

Figura 4.3 – Curva Característica Operacional (*OCC*) para uma carta de controlo de médias com $N = 4$ (retirado de Cabral (2001))

O significado de α é o que já foi referido anteriormente, ou seja traduz a probabilidade de ocorrer um erro tipo I, que corresponde à emissão de um falso alarme. Recorde-se que, neste contexto, a expressão "falso alarme" se refere à situação em que a carta assinala um ponto fora de controlo quando, na realidade, o processo se encontra sob controlo.

Uma outra medida da sensibilidade das cartas de controlo muito utilizada é a "sequência média", normalmente designada por *ARL* (do inglês, *Average Run Lenght*), sendo igual ao número de pontos (ou de amostras) que, em média, são assinalados na carta até ocorrer um ponto fora de controlo.

O valor de *ARL* na situação de controlo é o inverso do valor médio de α, e traduz o número de amostras seguidas que, em média, separam dois "falsos alarmes". É, pois, desejável que este valor seja elevado.

Na situação de fora de controlo, a probabilidade de detectar a mudança logo na primeira amostra é igual a $(1-\beta)$. Nestas circunstâncias, o valor de *ARL* é igual ao inverso de $(1-\beta)$ e traduz o número de amostras que ocorre, em média, antes de se detectar a mudança verificada. Note-se que este valor depende da amplitude do desvio ocorrido. Contrariamente ao que foi dito para a situação de controlo, nas situações de fora de controlo é desejável que o valor de *ARL* seja baixo (quanto mais depressa for detectado o desvio, mais sensível é a carta).

Na secção 4.4.2, podem encontrar-se exemplos (simulados) do cálculo dos valores de *ARL* para várias cartas de controlo na situação de controlo e em várias situações fora de controlo (vários desvios).

4.3.5 Outras Cartas de Controlo

Há inúmeras cartas de controlo desenvolvidas com o objectivo de serem utilizadas nos ambientes industriais. Tal como já foi mencionado, a abordagem exaustiva destas técnicas não cabe no âmbito deste trabalho. No entanto, julgou-se útil referir, ainda que de uma forma esquematizada, algumas dessas cartas. Assim, nas figuras 4.4, 4.5 e 4.6 apresentam-se as principais cartas de controlo abordadas na literatura para variáveis discretas (atributos), para variáveis contínuas (variáveis) e para duas ou mais variáveis (multivariadas), respectivamente.

Na literatura especializada podem encontrar-se diversas aplicações práticas da utilização das cartas referenciadas na figura 4.4 e pormenores sobre aquelas menos

conhecidas. Shore (2000b) trata das cartas de atributos de uma forma geral, Benneyan e Kaminsky (1994) referem-se às cartas g e h, Steiner et al. (1996) e Geyer et al. (1996) abordam a "Sequential Probability Ratio Tests" e a CUSUM, Chan et al. (2000) descrevem a carta *CQC* e, finalmente, Duncan (1950), Wardell e Candia (1996) e Marcucci (1985) abordam as cartas χ^2 e p modificada.

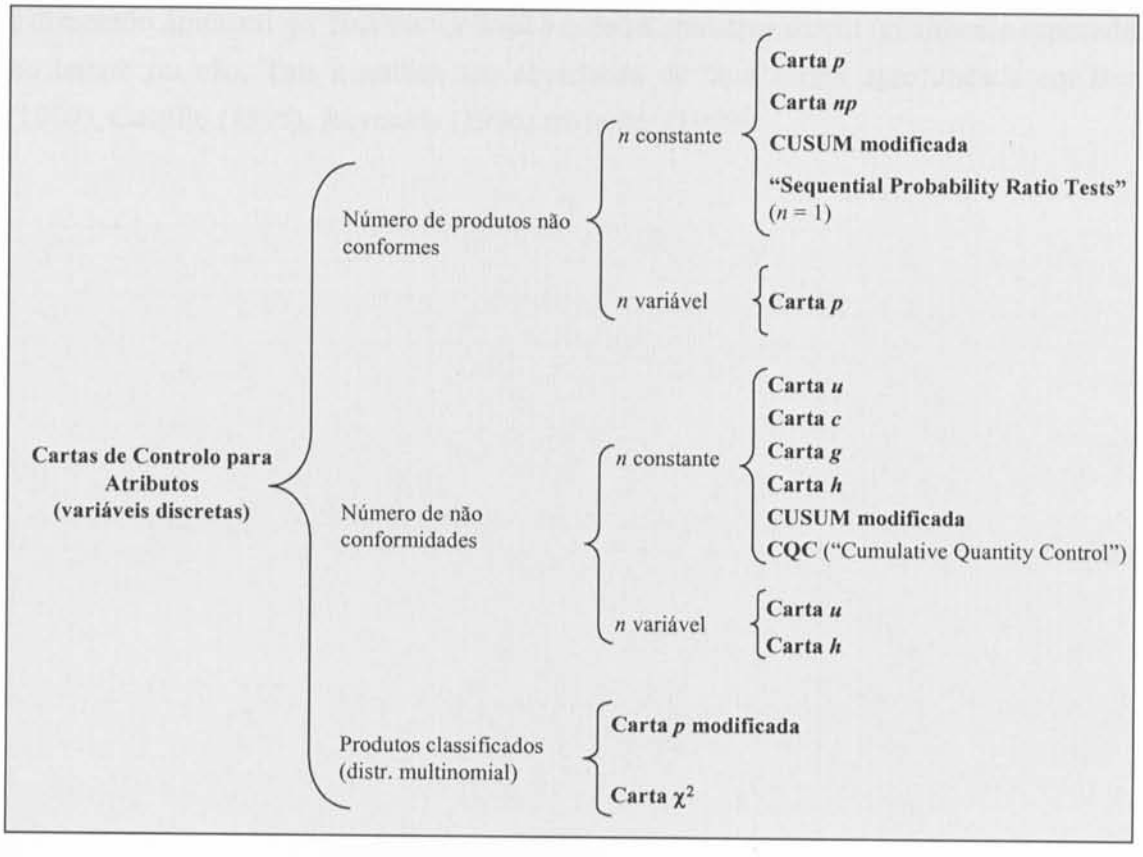

Figura 4.4 – Cartas de Controlo para Atributos

No que diz respeito às cartas de controlo de variáveis, particularmente as referidas na figura 4.5, a literatura é muito mais vasta. Das cartas menos comuns destacam-se a Carta não Linear (Sastri et al. (1996)), a Neuronal (Chang e Aw (1996)), a Semicircular (Chao e Cheng (1996)), a "Bach-Means" (Runger e Willemain (1996)), a SSR (Montgomery (2000)), a Carta Sintética (Wu e Spedding (2000)), a Carta ARMA (Jiang et al. (2000)) e as cartas para detecção de ciclos (Tatum (1996) e Beneke et al. (1988)).

Existem, também, inúmeras publicações que descrevem aplicações práticas e que comparam a utilização das cartas de variáveis utilizadas mais habitualmente. Destas, salientam-se Burr (1967), Castillo (1995), Bothe (1986), Bai e Choi (1995), Woodall e Thomas (1995), Serel et al. (2000), Chinna et al. (2000), Cassady et al (2000) e Shore (2000a) onde são abordadas pormenorizadamente temáticas relacionadas com as cartas $\overline{X}$, R, S, $\overline{MR}$, e ANOVA. Dos artigos que abordam e comparam as cartas CUSUM,

Média Móvel e EWMA, destacam-se Roberts (2000), Bissel (1969), Lucas e Crosier (2000), Geyer et al. (1996), Steiner et al. (1996), Wheeler (1995), Hunter (1996), Borror et al. (1999), Lu e Reynolds (1999), Sun e Zhang (2000), Luceño e Puig-Pey (2000), Koning e Does (2000) e Sparks (2000).

Considerando ainda as cartas de variáveis, há problemas relacionados com o facto de a dimensão amostral ser fixa ou variável e o de as amostras serem igualmente espaçadas no tempo ou não. Tais questões são abordadas de uma forma aprofundada em Burr (1969), Castillo (1995), Reynolds (1996) e Steiner (1999).

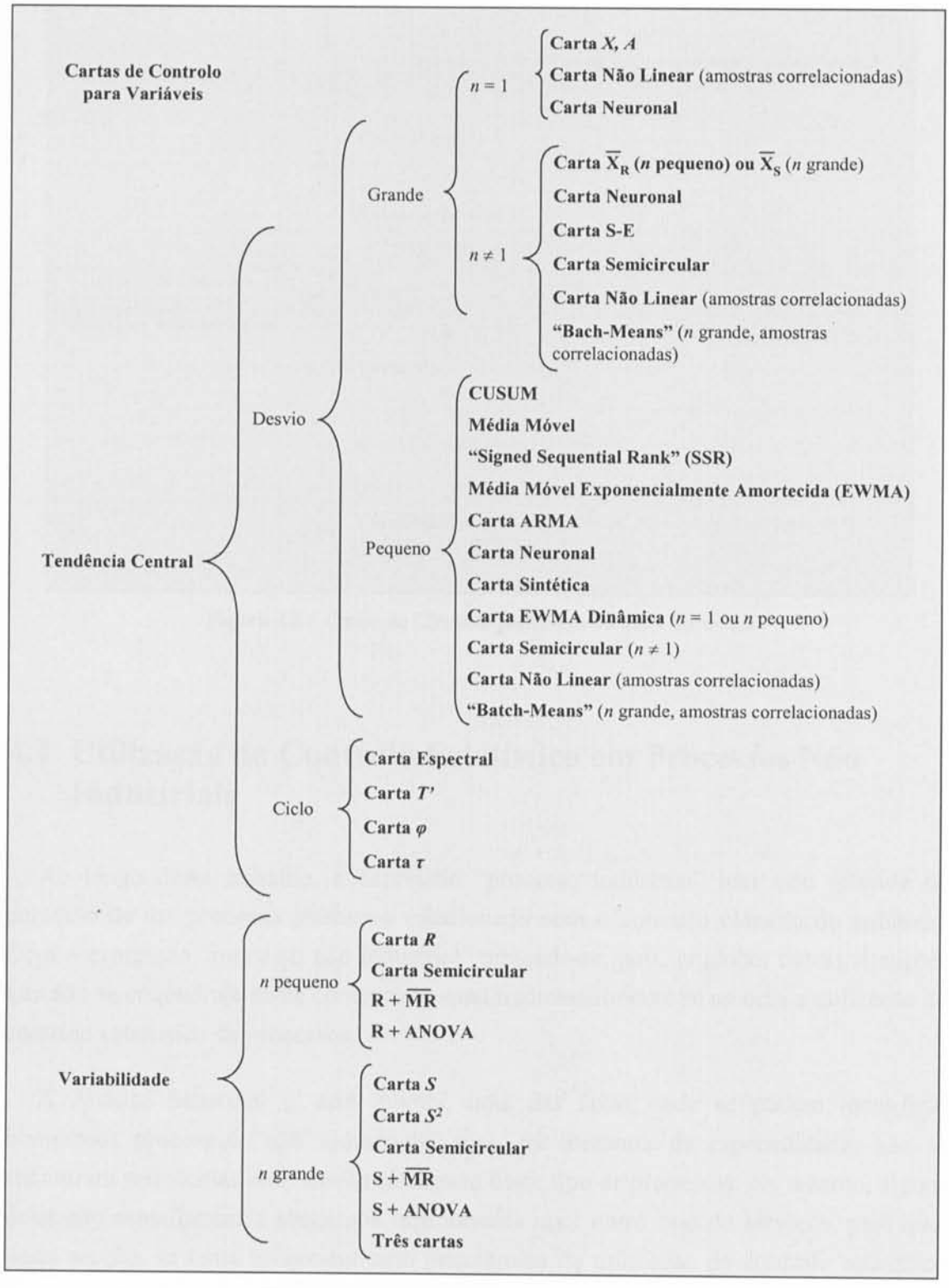

Figura 4.5 – Cartas de Controlo para Variáveis

Finalmente, em relação às cartas de controlo multivariadas (ver figura 4.6), destacam-se as seguintes referências: Runger (1996), Lowry e Montgomery (1995), Liu (1995), Mason et al. (1997) e Bodden e Rigdon (1999).

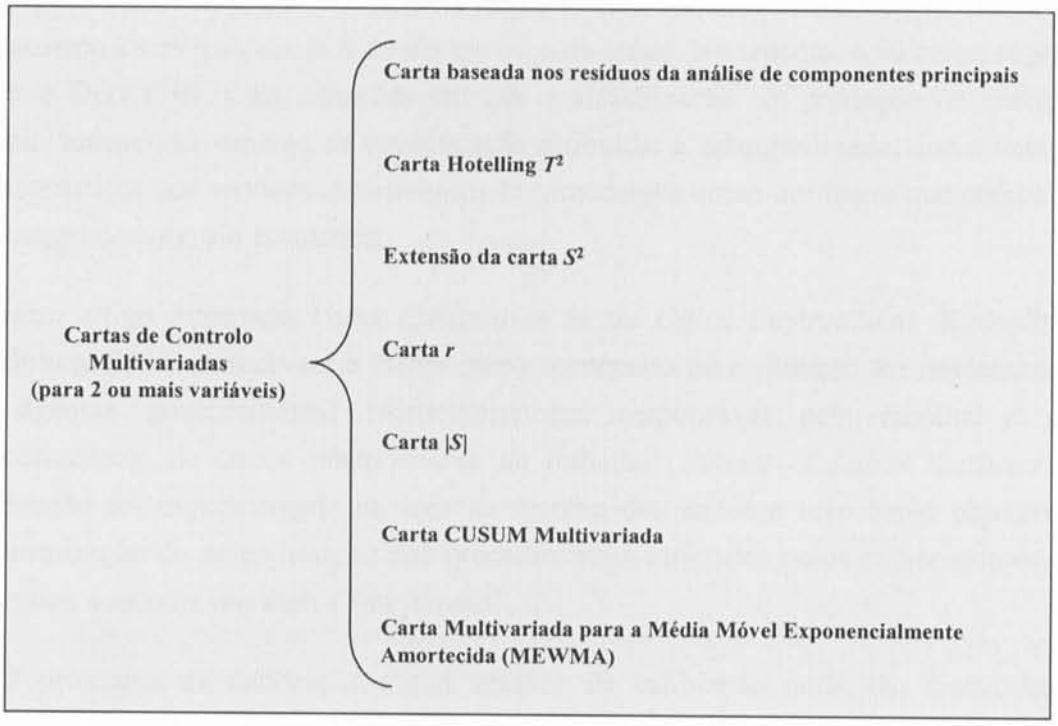

Figura 4.6 – Cartas de Controlo para Duas ou Mais Variáveis

4.4 Utilização do Controlo Estatístico em Processos Não Industriais

Ao longo deste trabalho, a expressão "processo industrial" tem sido referida no contexto de um processo produtivo relacionado com o conceito clássico de indústria. Com a expressão "processo não industrial" pretende-se, pois, englobar outras situações que não se enquadram neste contexto ao qual tradicionalmente se associa a utilização do controlo estatístico de processos.

A Análise Sensorial é, sem dúvida, uma das áreas onde se podem identificar numerosos processos "não industriais" mas, na literatura da especialidade, não se encontram referências ao controlo estatístico deste tipo de processos. No entanto, alguns deles são semelhantes a processos relacionados com outro tipo de serviços, pelo que, nesta secção, se tenta apresentar uma panorâmica da utilização do controlo estatístico nos processos "não industriais" de uma forma geral, enquanto nos capítulos seguintes se aborda a situação particular da Análise Sensorial, apresentando propostas de solução e resultados obtidos no caso estudado.

Alguns dos casos que serão mencionados ou estudados nesta secção referem-se claramente a serviços, como o do ensino ou o da saúde. No entanto, e tal como sugerem Roes e Dorr (1997), há situações em que a classificação em produção ou serviço é difícil. Independentemente da classificação atribuída, a intangibilidade, que é uma das características dos serviços, é normalmente considerada como um factor que complica a utilização do controlo estatístico.

Num artigo intitulado *Using Calibration In an Office Environment*, Kydoniefs e Lindblom (2000) descrevem a forma como o conceito de calibração foi implementado na agência governamental Norte-Americana responsável pela recolha e pelo processamento de dados relativamente ao trabalho (*Bureau of Labor Statistics*). A calibração foi implementada na área da recolha dos dados e teve como objectivo a uniformização do entendimento dos procedimentos utilizados pelos colaboradores que efectuam a recolha dos dados "no terreno".

O programa de calibração inclui sessões de calibração onde são discutidas as respostas dos colaboradores a um conjunto de exercícios desenvolvidos especificamente para cada sessão. Tanto a preparação dos exercícios como o papel do moderador são aspectos críticos para o sucesso do programa e são descritos no artigo em causa.

É interessante verificar que uma das principais preocupações da equipa responsável pela calibração é a de desenvolver medidas para avaliar e melhorar a consistência da recolha de dados. As duas razões fundamentais para essa preocupação são assegurar um produto final de melhor qualidade e responder de uma forma robusta a críticas externas ou auditorias.

A definição de medidas que traduzam os requisitos de qualidade que se pretendem avaliar e/ou controlar é, aliás, uma das grandes dificuldades que se coloca à implementação do controlo estatístico de processos não industriais. Roes e Dorr (1997) consideram mesmo que a investigação futura nesta área se deve concentrar no desenvolvimento e no refinamento de instrumentos de medida que permitam obter *feedback* dos clientes dos serviços.

As medidas actualmente mais utilizadas são o número de erros (por exemplo, num terminal de pagamento automático), o tempo de espera ou de atraso e as estatísticas amostrais correspondentes às respostas a inquéritos em escalas de cotação.

A natureza das variáveis relacionadas com o número de erros e com o tempo de espera ou de atraso não é muito diferente da das variáveis dos processos industriais, dado que se trata de medidas objectivas. Será de prever que, para o seu controlo, no

primeiro caso se utilize uma carta p (ou np) e no segundo uma carta de médias. A este propósito, Bothe (1986) aconselha, em casos particulares de controlo de processos industriais, a classificação dos defeitos numa escala de cotação, de forma a poder utilizar a carta de médias em vez da carta p. O autor refere que, no caso estudado, foi possível, desta forma, readaptar o processo antes de produzir alguma peça defeituosa, o que não acontecia com a utilização da carta p.

Na literatura há diversos exemplos do controlo deste tipo de processos. Em Greenwood e Gaunt (1994), são referidas diferentes variáveis que se relacionam com as mais diversas actividades realizadas numa escola que podem ser sujeitas a controlo estatístico: o número de absentistas, a classificação final dos alunos de determinado curso (ao longo de vários anos) ou a classificação de cada aluno numa determinada disciplina (ao longo de um ano). Em Hoop e Sturgis (2000), são utilizadas cartas de controlo para acompanhar os tempos de atraso em sistemas de produção por encomenda, e em Cheng et al. (2000) é utilizada uma carta de controlo multivariada para controlar um vasto conjunto de parâmetros relacionados com a segurança no espaço aéreo.

No que diz respeito às respostas a inquéritos, verifica-se que a natureza das variáveis em causa é muito distinta das referidas nos parágrafos anteriores, pelo facto de envolver conceitos mais abstractos.

Os inquéritos utilizam-se com o objectivo de medir as expectativas, as percepções ou o estado de satisfação das pessoas relativamente às características de um produto ou de um serviço e, habitualmente, são realizados num só momento no tempo ou em dois ou três momentos bastante espaçados. Neste último caso, o objectivo do primeiro inquérito é a identificação dos problemas e o dos seguintes é verificar se, após a implementação das medidas consideradas adequadas, os problemas foram resolvidos.

Este é o tipo de abordagem que se encontra, de uma forma geral, na literatura que se refere à qualidade dos serviços. Em Lim e Tang (2000), por exemplo, são medidas as expectativas e as percepções de 252 doentes de hospitais de Singapura, utilizando a técnica SERVQUAL (onde são analisadas seis dimensões da qualidade dos serviços). Estes autores concluem que há diferenças grandes entre as expectativas e as percepções dos doentes, pelo que é necessário melhorar o serviço ao nível das seis dimensões estudadas. Zanella (1998), por sua vez, desenvolve um modelo estatístico em que a satisfação global dos clientes é função da avaliação subjectiva dos níveis de qualidade das características dos produtos ou serviços.

O objectivo dos estudos que utilizam cartas de controlo no processamento de resultados de inquéritos é diferente dos atrás enunciados e surge como uma extensão do controlo em processos produtivos. Trata-se de conhecer a evolução do perfil das respostas das pessoas aos inquéritos e de controlar a sua estabilidade.

Normalmente, as alterações dos valores das variáveis traduzem mudanças cujo sentido é fácil de detectar, mas cuja interpretação nem sempre é evidente. Uma alteração nos resultados de um inquérito pode ser ocasional ou sistemática, pode ter um valor relativamente baixo ou elevado e pode ou não ser estatisticamente significativa. Como já foi evidenciado relativamente aos processos industriais, a finalidade do controlo estatístico é, precisamente, a de ajudar a distinguir os diferentes tipos de variação.

Entre as várias limitações que se colocam a esta abordagem, salientam-se as seguintes:

- Os inquéritos devem ser repetidos numa base periódica.
- Em geral, é necessário um grande volume de dados (respostas) tanto para estabelecer os limites de controlo como para assinalar cada ponto na carta.
- Exige-se que o processo a controlar seja caracterizado por um perfil de respostas estável (ou seja, fora da influência de causas assinaláveis as respostas aos inquéritos deverão exibir distribuições estatísticas estáveis).

A última questão merece, pela sua importância e complexidade, alguma reflexão. Ao admitir a existência de um perfil estável para as respostas das pessoas, está a admitir-se que, pelo menos ao longo do período em que foram definidos os limites de controlo, não houve alteração sistemática da variável a controlar. No entanto, a expectativa, a percepção ou a satisfação das pessoas, são sentimentos que podem variar radicalmente de um momento para o outro. Se se tratar de satisfação relativamente ao produto A, por exemplo, basta que apareça no mercado um outro produto concorrente – B – que seja percepcionado como melhor e mais barato, para a satisfação dos consumidores habituais de A diminuir de uma forma sistemática.

Por vezes, as causas das alterações sistemáticas são difíceis de detectar. É de prever, por exemplo, que, quando um produto de consumo (por exemplo, um iogurte) aparece no mercado, a satisfação que proporciona nessa altura seja superior àquela que proporciona após 3 meses. Isto porque é suposto que o novo produto contenha características do tipo de "qualidade inesperada", que se vão transformando ao longo do

tempo em características do tipo de "qualidade requerida", ou mesmo em características de "qualidade básica". Esta classificação das características da qualidade baseia-se no modelo de Kano (ver, por exemplo, Alvelos (1995)).

Note-se que, tal como Roes e Dorr (1997) salientam, o controlo estatístico só deve ser utilizado quando os inquéritos contemplam características de "qualidade requerida" ou de "qualidade básica", que são as características que os inquiridos conhecem bem e em relação às quais existem opiniões estáveis ao longo do tempo e, portanto, é possível definir uma distribuição base para a variável medida. A não verificação desta condição é equivalente à impossibilidade de definir uma hipótese nula que, como se viu, está na base do procedimento adoptado na elaboração das cartas de controlo.

Os problemas referidos são, na realidade, uma limitação importante à utilização do controlo estatístico em processos não industriais, nomeadamente no processamento de resultados de inquéritos. No entanto, na literatura encontram-se algumas referências à sua utilização em áreas como a da saúde, da educação ou do *marketing*, onde é controlada a satisfação de doentes, de estudantes ou de clientes com determinados serviços ou produtos.

Na área do ensino, por exemplo, para além de alguns parâmetros mais objectivamente mensuráveis e já referidos atrás, Bowrring-Carr e West-Burnham (1994) aconselham o controlo estatístico de variáveis como a satisfação dos alunos e dos pais ou a qualidade de atendimento. Na área do *Marketing*, Hayes (1992) sugere esta abordagem aos resultados dos inquéritos feitos periodicamente aos consumidores.

Na área da saúde destacam-se o livro intitulado *How to Use Control Charts for Healthcare* (Kelley (1999)) e o artigo *Statistical Process Monitoring of Customer Satisfaction Survey Data* (Wardell e Candia (1996)). No primeiro, são explicados, de uma forma simples, os conceitos relacionados com o controlo estatístico e a sua aplicação ao funcionamento de um hospital. Realça-se, aqui, a monitorização do grau de satisfação dos doentes e do funcionamento dos serviços de urgência (que se baseia nas causas que levam alguns doentes a voltar a recorrer a este serviço até 72h depois de terem sido atendidos pela primeira vez).

No segundo caso, os autores utilizam e comparam o desempenho de cartas de controlo para o processamento de resultados de inquéritos feitos regularmente aos doentes que estiveram internados num determinado hospital. Partindo dos resultados obtidos neste artigo, e com o objectivo de avaliar o desempenho destas ferramentas,

foram feitas simulações de várias situações e retiradas algumas conclusões que são apresentadas no ponto 4.4.2.

De seguida, na secção 4.4.1, apresenta-se uma proposta de um método para calibrar grandezas expressas em escalas do tipo Likert, e simulam-se alguns resultados.

4.4.1 Simulação da Calibração de uma Grandeza Expressa numa Escala de Cotação

O procedimento proposto pode ser aplicado em várias áreas, tais como a avaliação de desempenho ou a análise de inquéritos. No entanto, e dado que neste trabalho se abordam essencialmente os métodos passíveis de serem utilizados em Análise Sensorial, o exemplo refere-se à calibração de um provador na sua qualidade de "instrumento de medida". Assume-se que se trata de um produto alimentar de um determinado tipo (rebuçados de morango, por exemplo) e que a característica medida é a qualidade global do produto. Refere-se, assim, a um caso de calibração de uma grandeza expressa numa escala de cotação com 5 níveis.

Admite-se, igualmente, que existem produtos padrão para cada um dos níveis da escala e que, portanto, ao valor "1" corresponde um padrão muito mau (Padrão 1) e que ao valor "5" corresponde um padrão muito bom (Padrão 5). Aos valores "2", "3" e "4" correspondem os produtos padrão medíocre, suficiente e bom, respectivamente. Note-se que se está a supor que se trata de uma escala de intervalo, e que, portanto, a distância sensorial entre a qualidade dos vários padrões adjacentes é semelhante.

O ensaio consiste, então, em o provador avaliar a qualidade global dos vários padrões utilizando a escala mencionada. Admite-se que as provas decorrem ao longo de 10 sessões e que em cada uma são provadas e avaliadas 25 amostras (5 de cada padrão), apresentadas por ordem aleatória. O número de vezes que o ensaio é repetido para cada nível da escala é, portanto, de 50 ($N = 50$).

Na tabela 4.1, representam-se, para cada produto padrão, as frequências com que o provador atribuiu cada nível da escala. X representa a classe da escala a que pertence o produto padrão e D a classificação atribuída pelo provador.

Assume-se que, ao longo das provas elaboradas, o provador se encontrava nas condições normais de prova sem que tivesse ocorrido nenhum factor sistemático a influenciá-lo.

Tabela 4.1 – Frequências das respostas do provador para os vários padrões

Decisão	Padrão 1 $X=1$	Padrão 2 $X=2$	Padrão 3 $X=3$	Padrão 4 $X=4$	Padrão 5 $X=5$
$D=1$	46	4	1	0	0
$D=2$	3	43	2	1	0
$D=3$	1	3	44	3	2
$D=4$	0	0	3	44	3
$D=5$	0	0	0	2	45

Através dos resultados apresentados na tabela 4.1 pode já ter-se uma ideia da variação das opiniões do provador para um mesmo produto, podendo, mesmo, estimar-se a probabilidade de o provador se enganar em relação aos produtos padrão, ou seja, de a decisão ser $D=d$, dado que o padrão é $X=x$. Essa probabilidade será a probabilidade condicionada de d dado x e representa-se por P $(D=d \mid X=x)$. O seu valor pode ser estimado a partir da frequência relativa obtida para cada padrão e para cada decisão, que se obtém dividindo os valores da tabela 4.1 por 50 (número de provas de cada padrão).

O que se pretende, no entanto, com a calibração, é saber qual a incerteza a atribuir ao resultado das provas realizadas pelo provador na sua situação de trabalho normal, relativamente a qualquer produto do mesmo tipo (que não só os produtos padrão). Admita-se que o provador atribuiu a classificação de "2" à qualidade global de uma dada amostra. A sua tabela de calibração deverá informar de quais as probabilidades de a amostra em causa pertencer à classe "1", "2", "3", "4" ou "5". Esta probabilidade é a probabilidade condicionada de x dado d e representa-se por P $(X=x \mid D=d)$.

A relação entre as duas probabilidades condicionadas referidas pode encontrar-se em qualquer livro de Estatística (por exemplo, Guimarães e Cabral (1998)) e traduz-se por:

$$P(X=x|D=d)=\frac{P(X=x)\cdot P(D=d|X=x)}{P(D=d)}$$

P $(X=x)$ é a probabilidade de a amostra apresentada ao provador pertencer à classe x. Trata-se de uma probabilidade *a priori*. O seu valor pode ser estimado pela frequência relativa com que ocorreram amostras da classe em causa, nas últimas m provas. Por vezes, a dificuldade situa-se em definir, mesmo *a posteriori*, qual a verdadeira classificação dos produtos. No caso de se dispor de um painel de provadores estável, pode tomar-se a decisão do painel como a decisão "certa". Caso contrário, podem adoptar-se resultados de estudos de aceitação ou de preferência. Outra

abordagem possível é a de assumir probabilidades subjectivas, perguntando a várias pessoas envolvidas no processo da prova quais as probabilidades que atribuiriam a cada uma das classes.

Na tabela 4.2, apresenta-se um exemplo de possíveis valores de P $(X = x)$.

Tabela 4.2 – Probabilidades estimadas de o produto ser de classe $X = x$

	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$
P $(X = x)$	0.03	0.45	0.34	0.11	0.07

P $(D = d)$ é a probabilidade de o provador atribuir a classificação d à amostra que lhe é apresentada. Há duas maneiras de obter este valor. Aquela que, à primeira vista, poderá parecer mais directa é estimar P $(D = d)$ através da frequência relativa das decisões do provador nas últimas m provas, tal como se fez com P $(X = x)$. No entanto, uma vez que os acontecimentos X e D não são independentes e que já se conhecem as estimativas de P $(D = d \mid X = x)$ e de P $(X = x)$, o valor de P $(D = d)$ encontra-se implicitamente estimado. O seu cálculo deverá, então, ser feito com base naqueles dois valores e segundo a expressão:

$$P(D = d) = P(D = d \cap X = 1) + P(D = d \cap X = 2) + \ldots + P(D = d \cap X = 5)$$

onde

$$P(D = d \cap X = x) = P(X = x) \cdot P(D = d \mid X = x).$$

Na tabela 4.3, apresentam-se os valores das probabilidades P $(D = d)$, calculadas segundo o procedimento acima descrito.

Tabela 4.3 – Probabilidades calculadas de o provador atribuir a classificação $D = d$

	$D = 1$	$D = 2$	$D = 3$	$D = 4$	$D = 5$
P $(D = d)$	0.071	0.405	0.336	0.121	0.067

Coloca-se aqui uma questão pertinente: como actuar se o valor de P $(D = d)$ calculado desta forma for muito diferente da frequência relativa das apreciações do provador em causa? Tal significa ou que P $(X = x)$ não foi correctamente estimado ou que houve incongruências da parte do provador quando apreciou os padrões. A diferença entre as duas funções de probabilidade – P $(D = d)$ tal como se apresenta na tabela 4.3 e P $(D = d)$ calculada a partir da frequência relativa, pode ser testada através de um teste de hipóteses do qui-quadrado (qualidade de ajuste).

No caso de a hipótese nula não ser rejeitada, está-se, finalmente, em condições de elaborar a tabela de "calibração" do provador, que se apresenta seguidamente (tabela 4.4).

Tabela 4.4 – Tabela de "calibração" do provador

P ($X = x \mid D = d$)	$D = 1$	$D = 2$	$D = 3$	$D = 4$	$D = 5$
P ($X = 1 \mid D = d$)	0,392	0.004	0.002	0	0
P ($X = 2 \mid D = d$)	0,511	0.957	0.08	0	0
P ($X = 3 \mid D = d$)	0,097	0.034	0.89	0.168	0
P ($X = 4 \mid D = d$)	0	0.005	0.02	0.797	0.065
P ($X = 5 \mid D = d$)	0	0	0.008	0.035	0.935

A tabela 4.4 deve ser interpretada da seguinte forma: admita-se que o provador em causa atribui a classificação de "2" à qualidade global de uma amostra de rebuçados de morango. A partir da tabela conclui-se que, com 95.7% de probabilidade, a qualidade do rebuçado provado é de classe "2", com 3.4% será de classe "3", com 0.5% de classe "4", com 0.4% de classe "1" e não será nada provável que seja de classe "5".

Note-se que este processo de calibração é sensível à política adoptada para definir a tabela de probabilidades *a priori* (semelhante à tabela 4.2). Estas probabilidades, tal como já foi dito, podem ser estimadas com base em critérios distintos, que, naturalmente, originam valores distintos para as probabilidades e que, portanto, influenciam o nível de desempenho dos provadores que se encontram a ser calibrados (ou a sua tabela de calibração). Recomenda-se, assim, a maior precaução na definição da referida tabela.

Tudo o que foi dito nesta secção diz respeito apenas a um provador. A decisão final de um possível Painel de Provadores deverá resultar das várias decisões individuais dos seus membros. Se se assumir que as decisões destes são independentes entre si (o que deve ser garantido no planeamento e na execução de qualquer experiência), é possível calcular as várias probabilidades que compõem a tabela de calibração conjunta. Seguidamente apresenta-se a fórmula a aplicar para calcular as probabilidades correspondentes à tabela de calibração de um painel de dois provadores, A e B, com as respectivas decisões, d_a e d_b. Esta fórmula é generalizável para qualquer número de provadores, embora seja difícil, a partir de 3 provadores, representar os resultados finais sob a forma de tabela.

$$P(X = x|[(D_A = d_A) \cap (D_B = d_B)]) = \frac{P(X = x) \cdot P(D_A = d_A|X = x) \cdot P(D_B = d_B|X = x)}{P(D_A = d_A) \cdot P(D_B = d_B)}$$

A proposta apresentada para a calibração de instrumentos de medida de grandezas expressas por variáveis discretas tem limitações. Salientam-se o elevado volume de dados requerido, a dificuldade em estimar algumas das probabilidades referidas e a possibilidade de haver incongruências, que poderão inviabilizar a calibração. No entanto, acredita-se que esta abordagem, quando possível, é uma contribuição válida para um conhecimento mais profundo da incerteza associada a este tipo de grandezas.

4.4.2 Simulação de um Processo Multinomial e Seu Controlo Estatístico

Tal como se referiu atrás neste capítulo, é cada vez mais frequente o recurso a inquéritos para avaliar o nível das expectativas, das percepções ou da satisfação das pessoas com determinados produtos ou serviços. Neste contexto, a utilização de cartas de controlo no processamento dos resultados dos inquéritos surge como uma forma de distinguir as situações em que a variação no perfil das respostas dos inquiridos é devida a factores aleatórios, daquelas em que há, na realidade, uma alteração no nível de satisfação dos consumidores com o produto ou serviço em causa.

Neste capítulo, foram já enunciadas algumas das dificuldades que surgem quando se pretende controlar a estabilidade das respostas a inquéritos através de cartas de controlo. No entanto, há casos em que essas dificuldades são atenuadas por ser possível obter um volume suficiente de respostas ao inquérito e repeti-lo numa base periódica (por vezes até numa base permanente), e também pelo facto de a variável a controlar reflectir uma característica do tipo "qualidade requerida" ou "qualidade básica", às quais, em princípio, corresponde um perfil de respostas estável ao longo de um período de tempo razoável.

Um dos casos em que esta abordagem parece possível é o descrito em Wardell e Candia (1996), onde são analisadas as respostas dos doentes a uma pergunta que se refere à qualidade da informação que receberam quando saíram do internamento no hospital. A informação em causa diz respeito às instruções recebidas para retomar a vida normal, e a escala utilizada foi uma escala de Likert (5 níveis).

O estudo foi conduzido durante 18 semanas e a dimensão da amostra é variável (rondando $N = 70$), tendo os resultados sido agrupados em períodos de uma semana. As cartas de controlo utilizadas foram a carta $\overline{X}_S$, a p modificada e a χ^2.

Os autores concluem que o desempenho da carta χ^2 é, de uma forma geral, superior ao das outras duas, mas sugerem que, como é mais complexa de implementar e interpretar, se compreende que os decisores optem por utilizar a carta p modificada.

Os ensaios efectuados no âmbito da Análise Sensorial recorrem, muitas vezes a escalas semelhantes às referidas em Wardell e Candia (1996). Justifica-se, assim, que se analisem as potencialidades das referidas cartas de controlo, tendo em vista a sua possível utilização no contexto da Análise Sensorial.

Nesta secção, e recorrendo a duas simulações, avalia-se e compara-se o comportamento das cartas referidas. A situação simulada é a da resposta a uma única pergunta, numa escala semelhante à utilizada por Wardell e Candia (1996), cujo objectivo é o de avaliar o grau de satisfação dos consumidores com o sabor de um determinado produto. O nível "0" corresponde a nada satisfeito e o "4" a muito satisfeito. A dimensão da amostra foi fixada em $N = 70$.

Em cada processo simulado geram-se números entre "0" e "4" pertencentes a uma distribuição multinomial com 5 parâmetros (N, p_0, p_1, p_2, p_3) e que correspondem a uma configuração estável de respostas.

4.4.2.1 As Cartas de Controlo $\overline{X}_S$, p modificada e χ^2

Descrevem-se seguidamente, de uma forma breve, as cartas de controlo referidas em Wardell e Candia (1996).

A carta $\overline{X}_S$, corresponde à carta de médias descrita em 4.3.1 e é utilizada para controlar o valor esperado de uma variável aleatória cuja função densidade de probabilidade é aproximadamente normal. No caso em análise, trata-se de uma distribuição multinomial, e não normal. No entanto, e dado que a dimensão amostral ($N = 70$) é suficientemente elevada para que se possa utilizar o teorema do limite central, assume-se que a média amostral, $\overline{X}$, segue aproximadamente uma distribuição normal. Nestas condições, e no caso de os parâmetros da distribuição multinomial serem desconhecidos, a linha central e os limites de controlo da carta são os seguintes:

$LC = \overline{\overline{X}}$ (linha central)

$LIC = \overline{\overline{X}} - A_3 \cdot \overline{S_x}$ (limite inferior de controlo)

$LSC = \overline{\overline{X}} + A_3 \cdot \overline{S_x}$ (limite superior de controlo)

onde

$\overline{\overline{X}}$: média das médias amostrais;

$\overline{S_x}$: média dos desvios padrão amostrais;

A_3 : constante tabelada, função de N.

A carta p modificada (ou $\overline{X}_p$) é uma extensão da carta p (ver 4.3.2) a mais do que duas categorias. Tal como no caso anterior, a variável a controlar é o valor esperado da média amostral, pelo que a linha central das duas cartas é a mesma. Também neste caso se assume que o teorema do limite central se pode aplicar, de modo a que os limites de controlo possam ser fixados em ± 3 desvios padrão da média. A estes limites corresponde, então, um erro do tipo I (já atrás referido como α) de aproximadamente 0.27%.

A diferença entre as cartas $\overline{X}_S$ e $\overline{X}_p$ é apenas que, na última, o desvio padrão do processo é estimado com base nas estimativas dos parâmetros da distribuição multinomial subjacente.

As distribuições multinomiais aqui consideradas são caracterizadas pelos parâmetros N, p_0, p_1, p_2 e p_3, onde N representa a dimensão amostral e p_i representa o valor esperado da proporção de observações na categoria i.

Os limites de controlo da carta $\overline{X}_p$ são, portanto, dados por:

$$LIC = \overline{\overline{X}} - \frac{3}{\sqrt{N}} \cdot \hat{\sigma}_x \quad \text{e}$$

$$LSC = \overline{\overline{X}} + \frac{3}{\sqrt{N}} \cdot \hat{\sigma}_x .$$

A estimativa do desvio padrão é obtida através da seguinte expressão:

$$\hat{\sigma}_x = \sqrt{\sum_{i=0}^{4} x_i^2 \cdot \hat{p}_i - \left(\sum_{i=0}^{4} x_i \cdot \hat{p}_i \right)^2}$$

onde:

x_i: variável que assume os valores de "0", "1", "2", "3" ou "4" e

$\hat{p}_i$: estimativa de p_i, dada por $\hat{p}_i = \dfrac{\sum_{j=1}^{m} y_{i,j}}{N \cdot m}$.

Nesta última expressão, m representa o número de amostras consideradas (é recomendável que $m \geq 25$) e y_{ij} representa o número de respostas do tipo x_i na amostra j.

A diferença entre os limites de controlo das duas cartas é muito pequena. Em 50 simulações efectuadas com 2000 amostras de 70 elementos cada uma, os pontos fora dos limites das duas cartas foram sempre os mesmos. Conclui-se, assim, que as duas cartas são praticamente equivalentes para dimensões de amostras da grandeza considerada.

Para além disso, quando os parâmetros da distribuição multinomial são conhecidos, os limites das duas cartas coincidem, dado que a estimativa do desvio padrão é substituída pelo seu valor real. Nestas condições, os limites da carta $\overline{X}_S$ são calculados substituindo $\overline{S}_x$ e A_3 por σ_x e $\dfrac{3}{\sqrt{N}}$, respectivamente e na carta $\overline{X}_p$ o desvio padrão estimado, $\hat{\sigma}_x$, deverá ser substituído pelo seu valor teórico, σ_x.

Na prática recomenda-se que, de entre estas duas cartas, se utilize a carta $\overline{X}_p$, uma vez que se baseia nas estimativas dos próprios parâmetros da distribuição multinomial e que se adapta melhor ao caso em que N é variável. As cartas comparadas nas secções seguintes são, portanto, as cartas χ^2 e $\overline{X}_p$.

A carta χ^2 é baseada no teste de hipóteses à qualidade de ajuste. No caso em análise, a hipótese nula (H_0) é a de que a população de que é retirada a amostra segue uma distribuição multinomial para a qual a probabilidade de uma observação pertencer à categoria i é de p_i. Os valores de p_i representam, portanto, a função de probabilidade da distribuição teórica correspondente a H_0.

O teste, que é particularmente adequado ao caso de variáveis discretas tipo multinomial, utiliza a estatística Q, definida pela expressão:

$$Q = \sum_{i=0}^{k-1} \frac{(Y_i - E_i)^2}{E_i},$$

na qual:

k: número de categorias considerado (no caso em estudo $k = 5$);

Y_i: número de observações obtido na categoria i;

E_i: valor esperado do número de observações na categoria i.

Quando H_0 é verdadeira e se verificam determinadas condições que serão referidas adiante nesta secção, a estatística Q segue aproximadamente uma distribuição χ^2 com k - 1 graus de liberdade.

O valor de Q representa a qualidade global do ajuste entre os valores obtidos para a amostra e os valores teóricos. Se H_0 é verdadeira, os valores de Y_i deverão ser próximos dos de E_i, e, portanto, Q deverá assumir um valor baixo. No caso de H_0 ser falsa, Q assumirá valores tanto mais elevados quanto maior for a diferença entre a distribuição real e a assumida em H_0. Assim, o teste χ^2 é unilateral à direita e a carta de controlo correspondente possui apenas um limite superior de controlo.

Como a distribuição χ^2 é assimétrica à direita, o procedimento habitual de estabelecer o limite superior de controlo em $\mu + 3 \cdot \sigma$ (usado para as distribuições aproximadamente normais) não conduz a um valor de α próximo de 0.27%. No entanto, em Duncan (1950) e em Wardell e Candia (1996) encontram-se tabelas onde se estabelece o limite da carta de controlo χ^2 para o referido valor de α em função do número de graus de liberdade.

A aproximação da distribuição da estatística Q a uma χ^2 é válida quando a dimensão amostral é elevada (normalmente assume-se $N > 30$) e quando o valor esperado do número de elementos em cada categoria é razoável. É normalmente aceite (ver, por exemplo, Guimarães e Cabral (1998)) que a aproximação é válida se todas as classes tiverem, pelo menos, 5 elementos ($E_i \geq 5$, para todo o i). Esta regra é uma generalização de uma outra devida a Cochran (1954) e referida em Marcucci (1985), segundo a qual não deverá haver mais do que 20% das frequências esperadas inferiores a 5, não devendo nenhuma delas ser inferior a 1. Note-se que esta regra se torna mais flexível do que a anterior, à medida que aumenta o número de classes.

Uma outra regra apresentada em Yarnold (1970) refere que, quando $k \geq 3$ (em que k representa o número de classes) e r é o número classes com frequência esperada menor do que 5, a frequência esperada mínima deverá ser de, pelo menos, $5 \cdot r / k$.

Para ilustrar o comportamento da estatística Q relativamente à função densidade de probabilidade χ^2, apresentam-se, na figura 4.7, os histogramas dos valores da estatística Q obtidos a partir de quatro diferentes distribuições multinomiais com 5 categorias ("0",

"1", "2", "3" e "4"). Aos histogramas referidos sobrepuseram-se os gráficos da função densidade de probabilidade da distribuição χ^2_4. Os valores representados nos histogramas foram obtidos por simulação de 5000 amostras de dimensão $N = 70$. As distribuições são representadas por um conjunto de cinco números de dois dígitos colocados lado a lado e separados por um ponto. Cada número corresponde ao valor esperado da frequência numa categoria, sendo que o número da esquerda se refere à categoria "0" e o da direita à categoria "4". Por exemplo, no caso B da figura, o conjunto 02.02.10.18.38 identifica a distribuição para a qual $E_0 = 2$, $E_1 = 2$, $E_2 = 10$, $E_3 = 18$ e $E_4 = 38$. Para efeitos de cálculo da estatística Q nos casos em que $E_i = 0$, substituiu-se o valor "0" por um valor muito pequeno.

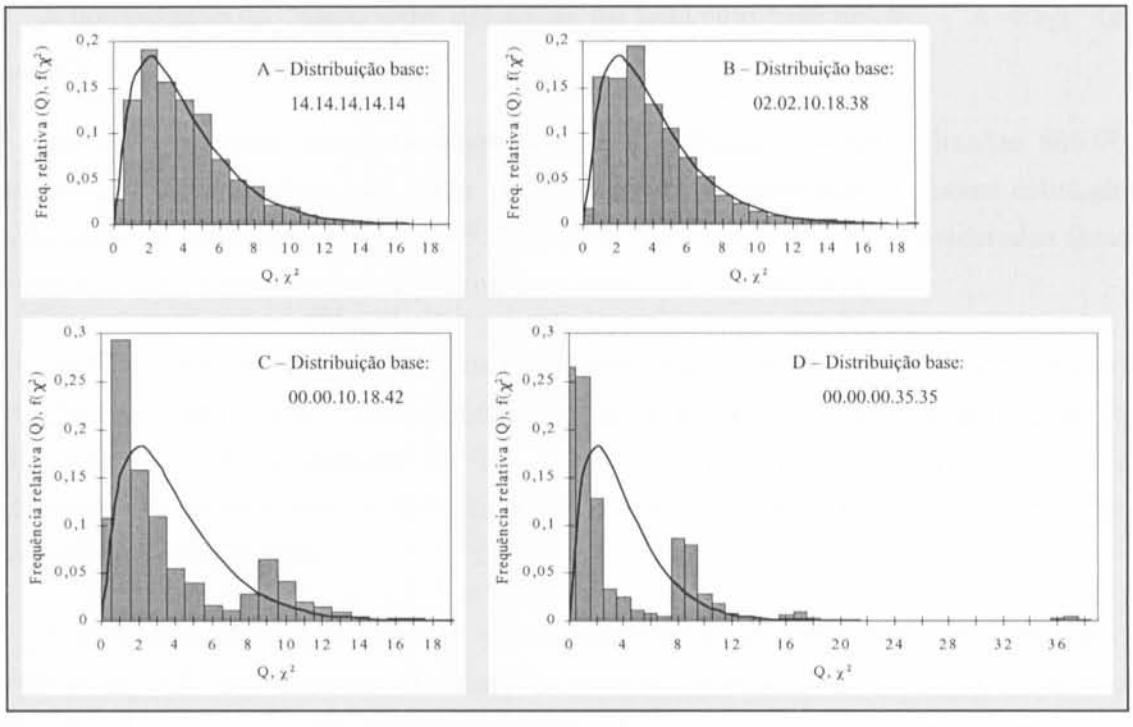

Figura 4.7 – Comparação entre a função densidade de probabilidade χ^2_4 e os histogramas da estatística Q obtidos para quatro distribuições multinomiais diferentes

A análise da figura 4.7 coloca em evidência o facto de não ser legítimo utilizar o teste χ^2 em algumas situações, nomeadamente quando há classes cujo valor esperado é nulo. Este resultado seria de prever, já que as distribuições C e D estão fora das condições atrás mencionadas. Note-se que a distribuição B se encontra nas condições limite da utilização da regra de Yarnold.

4.4.2.2 Comparação dos Resultados Obtidos com as Cartas χ^2 e $\overline{X}_p$ para Alguns Processos Sob Controlo

Nesta secção, compara-se o desempenho das cartas de controlo $\overline{X}_p$ e χ^2 para processos sob controlo, ou seja, para os quais H_0 é verdadeira.

É de notar que, atendendo a que os resultados são simulados a partir de uma distribuição previamente estabelecida, o desvio padrão é conhecido e, consequentemente, os limites de controlo nas cartas $\overline{X}_p$ e $\overline{X}_S$ coincidem.

A comparação do desempenho das cartas foi feita com base no *ARL* ("Average Run Length").

Em cada processo (correspondente a uma distribuição) foram utilizadas 500.000 amostras independentes, cada uma com 70 respostas simuladas e foram estimados intervalos de confiança bilaterais a 95% para α. Os valores de *ARL* considerados foram o inverso dos valores de $\hat{\alpha}$ (valor central do intervalo de confiança).

Na tabela 4.5, são apresentados os resultados obtidos para os referidos intervalos de confiança e para o *ARL*, correspondentes a algumas das distribuições simuladas. As distribuições com os números de 1 a 7 são perfeitamente simétricas, enquanto as restantes são assimétricas. A distribuição nº 10 é aproximadamente igual à referida por Wardell e Candia (1996).

A conclusão geral que se pode retirar dos resultados apresentados é a de que os valores do *ARL* sob controlo são bastante instáveis, revelando-se sensíveis à forma de distribuição da população de onde são retiradas as amostras. No entanto, o risco α associado à carta $\overline{X}_p$ é aparentemente menos sensível à forma da distribuição, mantendo-se mais próximo do valor teórico de 0.27% ($ARL \approx 370$) do que o associado à carta χ^2, que é geralmente mais baixo (exceptuando os casos das distribuições números 7 e 9, que apresentam intervalos de confiança que se sobrepõem).

Note-se que o desempenho da carta χ^2 está claramente associado às condições de validade do respectivo teste: os piores resultados são obtidos para as distribuições onde há categorias com valores esperados mais baixos (entre 2 e 5).

Tabela 4.5 – Intervalos de confiança (I.C.) a 95% para α, $\hat{\alpha}$ e *ARL* para as cartas $\overline{X}_P$ e χ^2

	Distribuição	Carta $\overline{X}_P$			Carta χ^2		
Nº		I.C. para α (%)	$\hat{\alpha}$ (%)	*ARL*	I.C. para α (%)	$\hat{\alpha}$ (%)	*ARL*
1	02.16.34.16.02	[0.267,0.297]	0.282	355	[0.505,0.545]	0.525	190
2	02.17.32.17.02	[0.202,0.227]	0.215	466	[0.485,0.524]	0.505	198
3	03.17.30.17.03	[0.283,0.313]	0.298	336	[0.427,0.463]	0.445	225
4	05.05.50.05.05	[0.232,0.259]	0.245	408	[0.405,0.441]	0.423	236
5	05.12.36.12.05	[0.203,0.229]	0.216	463	[0.349,0.382]	0.365	274
6	11.15.18.15.11	[0.224,0.251]	0.237	422	[0.275,0.305]	0.290	345
7	14.14.14.14.14	[0.238,0.266]	0.252	397	[0.259,0.288]	0.274	365
8	05.08.12.15.30	[0.202,0.227]	0.214	466	[0.311,0.342]	0.326	306
9	20.17.14.11.08	[0.265,0.295]	0.280	357	[0.257,0.285]	0.271	369
10	02.02.10.18.38	[0.300,0.331]	0.316	317	[0.491,0.531]	0.511	196

Este exercício de simulação foi alargado a muitas outras situações, cujos resultados estão de acordo com os que aqui se apresentam. Parece poder concluir-se que, para situações de controlo, a carta $\overline{X}_p$ produz *ARL*s mais elevados e mais estáveis do que a carta χ^2.

4.4.2.3 Comparação dos Resultados Obtidos com as Cartas χ^2 e $\overline{X}_p$ para Alguns Processos Fora de Controlo

Nesta secção, estuda-se o desempenho das cartas $\overline{X}_p$ e χ^2 em situações de fora de controlo. A título de exemplo, seleccionaram-se, de entre as múltiplas simulações efectuadas, dois estados de controlo e, para cada um deles, 12 cenários de fora de controlo.

As comparações entre o desempenho das duas cartas são feitas, mais uma vez, em termos de *ARL*. Quando a situação é de fora de controlo, o desempenho de uma carta de controlo é tanto melhor quanto menor for o *ARL* visto que se espera que a carta detecte rapidamente a transição do estado de controlo para outro qualquer estado.

Na tabela 4.6, apresentam-se os resultados obtidos para $\hat{\beta}$ e para o *ARL* em 12 simulações para as quais o estado de controlo é o correspondente à distribuição número 9 da tabela 4.5 (20.17.14.11.08), representada, aqui, pelo nº 1 (a sombreado). Esta distribuição foi escolhida pelo facto de apresentar valores sob controlo semelhantes para as duas cartas, facilitando, assim, a comparação nos estados de fora de controlo.

Tabela 4.6 – Valores de $\hat{\beta}$ e *ARL* correspondentes ao 1º cenário de situações de fora de controlo

	Distribuição	Carta $\overline{X}_P$		Carta χ^2	
Nº		$\hat{\beta}$ (%)	*ARL*	$\hat{\beta}$ (%)	*ARL*
1	20.17.14.11.08	0.99716	352	0.99718	354
2	21.18.14.10.07	0.99466	187	0.99713	348
3	21.19.14.10.06	0.99003	100	0.99508	203
4	21.20.14.09.06	0.98428	64	0.99121	114
5	22.20.13.09.06	0.97372	38	0.98883	89
6	22.21.13.09.05	0.95601	23	0.97699	43
7	23.21.12.09.05	0.93221	15	0.96993	33
8	24.22.11.08.05	0.86420	7	0.93230	15
9	26.21.13.05.05	0.69969	3	0.86741	8
10	26.23.11.05.05	0.62390	3	0.77523	4
11	26.25.09.05.05	0.55016	2	0.60725	3
12	27.26.07.05.05	0.42763	2	0.41057	2

Na figura 4.8 representam-se graficamente os valores de $\hat{\beta}$ apresentados na tabela 4.6. Note-se que, embora tenha havido a preocupação de colocar nas primeiras linhas os estados mais próximos do controlo e nas últimas os mais afastados, cada linha da tabela representa um estado de fora de controlo independente dos outros. Desta forma, a tabela (assim como o correspondente gráfico) deve ser lida assumindo que o estado de fora de controlo derivou de uma transição abrupta a partir do estado de controlo e não de qualquer evolução gradual.

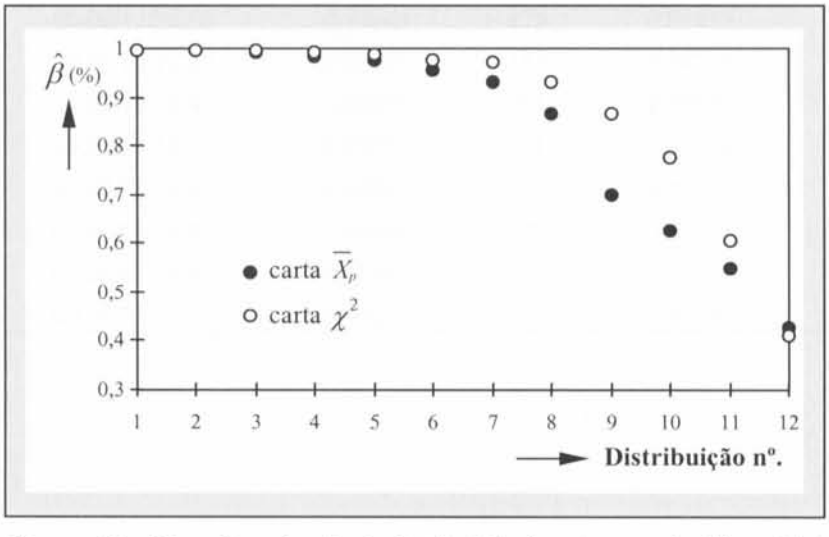

Figura 4.8 – Risco β em função da distribuição fora de controlo (1º cenário)

Ao analisar estes resultados, chega-se à conclusão de que a carta $\overline{X}_p$ detecta mais rapidamente as mudanças ocorridas do que a carta χ^2 (exceptuando uma ligeira vantagem desta última carta para a distribuição nº 12).

Este comportamento deve-se ao facto de, nas situações simuladas, se introduzirem pequenas diferenças no número de elementos de cada classe, transitando, esses elementos, das classes de mais baixa pontuação para as de mais alta. Verifica-se, assim, uma progressiva alteração do valor esperado, facto que é mais facilmente detectado pela carta $\overline{X}_p$.

Na tabela 4.7 e na figura 4.9, representam-se os resultados de simulações efectuadas a partir de uma situação de controlo em que há duas classes com apenas 2 elementos. Esta situação é bastante semelhante à referida por Wardell e Candia (1996).

As situações simuladas correspondem, ao contrário do caso anterior, a uma melhoria do nível médio das respostas, que se vai acentuando à medida que se desce na tabela.

Tabela 4.7 – Valores de $\hat{\beta}$ e *ARL* correspondentes ao 2º cenário de situações de fora de controlo

	Distribuição	**Carta $\overline{X}_P$**		**Carta χ^2**	
Nº		$\hat{\beta}$ (%)	*ARL*	$\hat{\beta}$ (%)	*ARL*
1	02.02.10.18.38	0.99655	290	0.99466	187
2	00.02.12.18.38	0,99870	769	0,99001	100
3	00.00.14.18.38	0,99839	620	0,96053	25
4	00.02.10.20.38	0,99750	401	0,99371	159
5	00.00.12.20.38	0,99641	279	0,98835	86
6	00.00.10.22.38	0,99355	155	0,98716	78
7	00.00.10.20.40	0,98274	58	0,99574	235
8	00.00.10.18.42	0,95907	24	0,99672	305
9	00.00.08.18.44	0,86989	8	0,99199	125
10	00.00.06.18.46	0,66659	3	0,96817	31
11	00.00.06.16.48	0,52359	2	0,92476	13
12	00.00.06.14.50	0,38144	2	0,83680	6

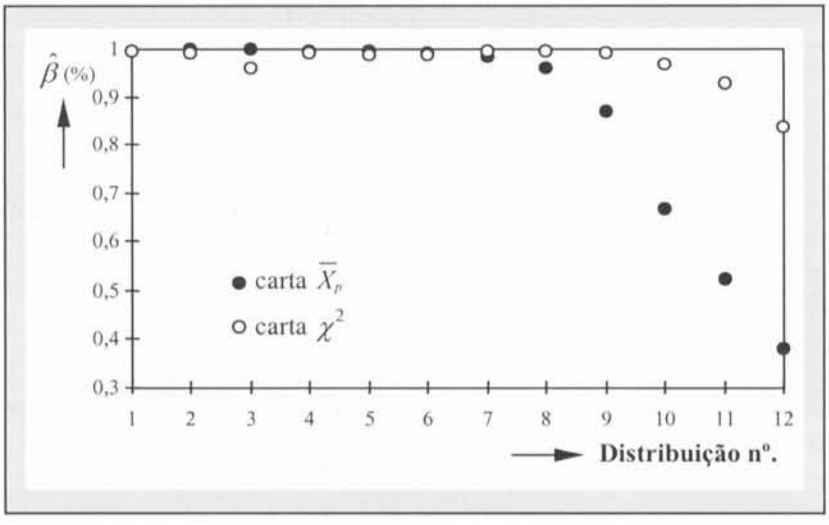

Figura 4.9 – Risco β em função da distribuição fora de controlo (2º cenário)

Os resultados desta simulação são, à primeira vista, surpreendentes.

Analise-se, em primeiro lugar, o que se passa com a carta $\overline{X}_p$, que era suposto detectar desvios no valor esperado, quando, afinal, nas primeiras 4 situações de desvio de controlo (distribuições nos. 2 a 5), não detecta.

À medida que se vai descendo na tabela 4.7, o valor esperado da população vai aumentando. Esse aumento é acompanhado de uma diminuição significativa no desvio padrão, visto que há classes que ficam sem elementos. Assim, nas primeiras distribuições, o efeito de diminuição de σ compensa o do aumento de μ, ficando menos pontos fora dos limites de controlo do que na própria situação de controlo.

Esta situação é ilustrada na figura 4.10, onde se podem observar os histogramas da distribuição da estatística $\overline{X}$ na situação de controlo (à esquerda) e da estatística $\overline{X}$ correspondente à distribuição nº 2 da tabela 4.7 (à direita). Os limites de controlo que estão a ser utilizados são 2.9 e 3.614 ($\mu \pm 3 \cdot \sigma$).

Tal como se pode observar, no gráfico do lado direito verificam-se menos observações fora dos limites de controlo do que no gráfico do lado esquerdo.

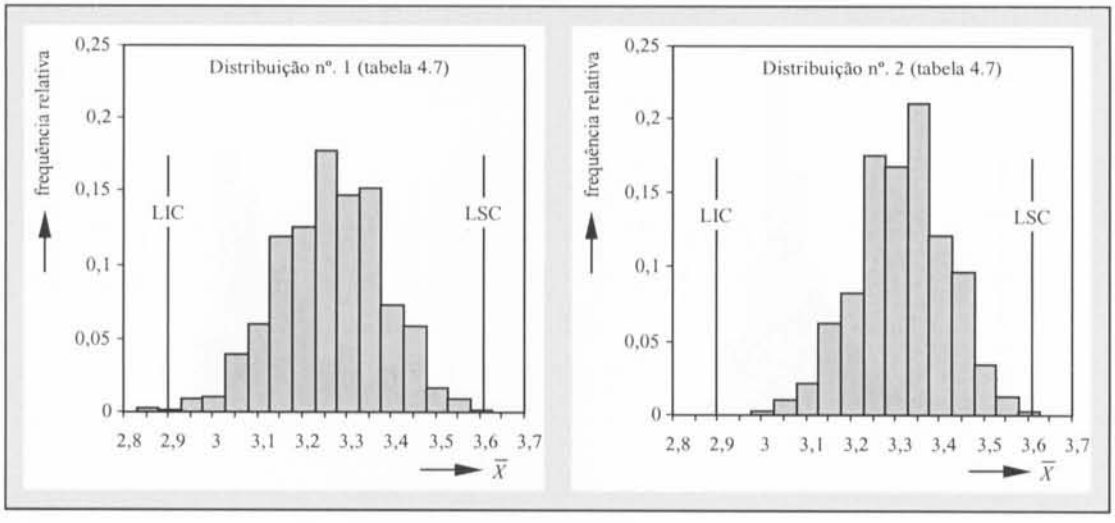

Figura 4.10 – Histograma da estatística de teste ($\overline{X}$) para as distribuições nºs. 1 e 2 da tabela 4.7

No que diz respeito à carta χ^2, sucede um fenómeno semelhante, mas, curiosamente, para distribuições diferentes. As distribuições em que se manifesta maior impossibilidade de detecção do desvio de controlo são as correspondentes aos números 7 a 9 (tabela 4.7).

Neste caso, a distribuição da estatística de teste, embora tenha um valor esperado superior à da χ^2_4 (com a qual se está a comparar), tem uma variância inferior. Este fenómeno é mais acentuado quando o número de transferências entre classes é muito baixo e as transferências ocorrem de classes que ficam sem elementos para classes com frequências esperadas altas.

Saliente-se que a distribuição correspondente ao estado de controlo estatístico se encontra nas condições limite de aplicação do teste χ^2, segundo Yarnold (1970).

Na figura 4.11, estão representados os histogramas da estatística de teste (Q) para as distribuições números 3 e 8 da tabela 4.7, (à esquerda e à direita, respectivamente), a função densidade de probabilidade χ^2_4 e o limite de controlo ($LSC = 16.25$).

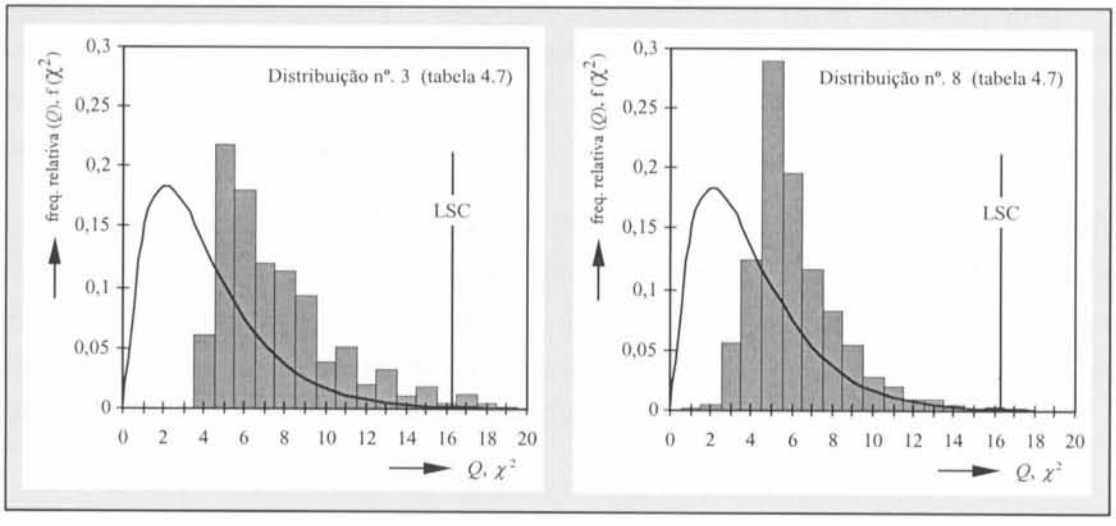

Figura 4.11 – Histograma da estatística de teste (Q) para as distribuições nºs 3 e 8 da tabela 4.7 e função densidade de probabilidade χ^2_4 sobreposta

Comparando os resultados obtidos com as cartas $\overline{X}_p$ e χ^2 (ver tabela 4.7 e figura 4.9), verifica-se que a carta χ^2 revela superioridade nos casos números 2 a 6, e que a carta $\overline{X}_p$ é melhor nos restantes casos (números 7 a 12).

Nas duas situações de fora de controlo aqui analisadas, verificou-se a superioridade da carta $\overline{X}_p$, contrariamente ao que foi concluído por Wardell e Candia (1996).

Uma possível justificação para a discrepância encontrada é que, numa tentativa de resolver os casos em que o valor esperado de algumas categorias é inferior a 5, Wardell e Candia (1996) agruparam categorias adjacentes, de forma a que a cada uma das novas categorias correspondesse um valor esperado superior a 5. Desta forma, os autores satisfazem plenamente as condições de aplicabilidade do teste χ^2 . Este procedimento afigura-se, no mínimo, discutível: perde-se a informação sobre a distribuição das respostas nas categorias agrupadas e sobre as transferências de opinião entre essas mesmas categorias, para além de que as novas categorias podem, mesmo, ser desprovidas de sentido. Por exemplo, Wardell e Candia (1996) agrupam as três primeiras categorias, o que quer dizer que as pessoas que responderam "0" (muito insatisfeitas) foram equiparadas às que responderam "2" (moderadamente satisfeitas). Se, por hipótese, o grau de satisfação das pessoas que responderam na classe "2" evoluísse para a classe "0", esta modificação seria completamente impossível de detectar com o agrupamento efectuado.

Este problema coloca-se de uma forma pertinente quando se discute o controlo de processos que envolvem votação em categorias. Particularmente em áreas como as

pesquisas de mercado, o estudo da satisfação de clientes ou as fichas de prova em Análise Sensorial, há uma tendência natural para as respostas das pessoas se concentrarem em algumas das categorias, ficando as outras com muito poucas ou mesmo nenhumas respostas.

Esta é, sem dúvida, mais uma, e séria, limitação à utilização do controlo estatístico tradicional neste tipo de processos. No entanto, e quando for possível ultrapassar as limitações apontadas, aconselha-se a que sejam utilizadas as duas cartas simultaneamente, já que se não se comprovou a superioridade de nenhuma delas e, por outro lado, se verificou que detectam situações de fora de controlo diferentes.

4.4.3 Em Conclusão

De tudo o que foi dito neste capítulo, parece poder concluir-se que a utilização dos métodos tradicionais de controlo estatístico em processos "não industriais", nomeadamente quando a característica de qualidade a controlar é medida em escalas de cotação, é problemática. As dificuldades são muitas e foram apresentadas ao longo de todo o capítulo, juntamente com algumas respostas para as resolver quando as circunstâncias o permitem.

Em forma de síntese, pode dizer-se que os obstáculos à implementação do controlo estatístico neste tipo de processos se situam em três níveis:

- Estabilidade do perfil de respostas correspondentes à avaliação da característica da qualidade.
- Elevado volume de dados exigido.
- Dificuldade na interpretação dos sinais das cartas de controlo utilizadas.

Quando se considera particularmente a área da Análise Sensorial, na qual em alguns testes são utilizadas escalas de categorias e noutros apenas há a escolha da amostra diferente ou da mais intensa (caso dos testes de diferenças), a abordagem ao problema da qualidade das decisões é, ainda, mais difícil. Como garantir que os instrumentos de medida (provadores) se encontram devidamente calibrados e sob controlo, ou seja, que as amostras estão a ser avaliadas correctamente?

Nos próximos capítulos, propõe-se uma abordagem a este problema que utiliza algumas das ferramentas referidas ao longo destes três capítulos e na qual se procura

integrar os conceitos que lhes estão subjacentes. Pretende-se que a implementação da metodologia proposta, para além de proporcionar um melhor conhecimento dos provadores e da qualidade dos resultados das suas medições, seja exequível nos ambientes práticos do dia-a-dia da Análise Sensorial.

5

O Problema do Controlo de Câmaras de Provadores

Conteúdo:

- Introdução
- O Caso do Instituto do Vinho do Porto
- A Complexidade do Problema

5.1 Introdução

A avaliação sensorial formal de produtos, em particular de produtos alimentares, é cada vez mais utilizada, não só pelas empresas que os fabricam, como por instituições e organismos ligados ao controlo da sua qualidade e/ou aos consumidores. No caso particular dos vinhos, existe já uma longa tradição de avaliação das características organolépticas, que constituem um factor decisivo para a sua aceitação pelo mercado.

Verifica-se, no entanto, que, tal como foi referido nos capítulos em que se apresentou o tema da Análise Sensorial, a avaliação e o controlo da qualidade dos julgamentos dos provadores, na sua condição de instrumentos de medida, são assuntos frequentemente abordados na literatura de uma forma muito geral e pouco quantificada. Esta lacuna deve-se, em muito, ao facto de a actividade de prova se incluir no tipo de processos "não industriais" referidos no capítulo anterior, e que, por isso, requerem um tipo de abordagem diferente da utilizada tradicionalmente.

Com este trabalho pretende-se contribuir para o estudo da qualidade dos julgamentos emitidos pelos elementos de uma câmara (ou painel) de provadores profissionais. Alguns dos métodos e ferramentas desenvolvidos foram testados e implementados na Câmara de Provadores do Instituto do Vinho do Porto, o que constituiu uma excelente oportunidade para verificar, na prática, a exequibilidade dos procedimentos propostos.

Na secção 5.2, descreve-se, sucintamente, a envolvente do estudo do caso da Câmara de Provadores do Instituto do Vinho do Porto. Na secção 5.2.1, é feito um resumo do processo produtivo dos diversos tipos de vinho do Porto, que se espera venha a contribuir para facilitar a compreensão do problema em estudo. Seguidamente, na secção 5.2.2, caracteriza-se o papel do Instituto do Vinho do Porto, particularmente no que diz respeito à sua Câmara de Provadores. Na secção 5.2.3, apresenta-se a descrição e a análise da situação no início do presente estudo, salientando-se, desde já, as óptimas condições que a Câmara de Provadores do Instituto do Vinho do Porto possui.

Na secção 5.3, realça-se a complexidade do problema e a dificuldade de adaptação dos métodos e técnicas utilizadas tradicionalmente à área da Análise Sensorial, nomeadamente à avaliação da qualidade dos julgamentos de membros de painéis de provadores.

5.2 O Caso do Instituto do Vinho do Porto

O Instituto do Vinho do Porto (IVP) é o organismo responsável pela certificação e fiscalização da Denominação de Origem "Porto". Entre outras funções, o IVP verifica a qualidade dos vinhos através de um rigoroso controlo analítico e organoléptico efectuado pelos seus Serviços Técnicos.

Há vários tipos de vinho do Porto, aos quais correspondem diferentes designações (ver secção 5.2.1). Ao submeter um vinho para aprovação no IVP, as empresas especificam qual a designação com que o pretendem comercializar, e o vinho é analisado segundo as características que correspondem a essa designação.

A Câmara de Provadores (CP) é o órgão do IVP responsável pela apreciação organoléptica dos vinhos. Às suas decisões está associada uma grande responsabilidade: se um vinho é rejeitado, a empresa que o submeteu ao IVP vê-se impedida de o comercializar. Neste contexto, o controlo da qualidade das decisões tomadas pela CP assume uma importância especial.

De facto, as decisões de um júri (ou painel) de prova, que têm uma repercussão decisiva na avaliação dos vinhos, assentam em processos dificilmente avaliáveis quanto à sua representatividade, consistência e reprodutibilidade. Por outras palavras, embora a avaliação sensorial seja uma actividade indispensável na produção e na comercialização do vinho, os métodos existentes para o seu controlo são limitados e insuficientes.

Em 1998, o IVP, reconhecendo a necessidade de imprimir um maior rigor científico às decisões decorrentes da prova de vinhos, decidiu acreditar a sua Câmara de Provadores (o laboratório de análises tinha já sido acreditado pelo Instituto Português da Qualidade (IPQ) para vários testes e produtos em 1994). Este projecto obrigou à implementação de um Sistema de Gestão da Qualidade, à estabilização das condições de trabalho dos elementos da CP e ao desenvolvimento de novos métodos para avaliar o desempenho dos provadores. Em Dezembro de 1999, a acreditação da Câmara de Provadores do Instituto do Vinho do Porto estava concluída com sucesso.

A primeira fase da implementação do programa de garantia da qualidade consistiu na definição de uma Política da Qualidade, cujos princípios se encontram explícitos no Manual da Qualidade da Câmara de Provadores do IVP. Esta Política assenta fundamentalmente em procedimentos operacionais normalizados e escritos, na

qualificação e formação dos provadores e demais funcionários afectos à CP e na verificação da conformidade dos processos através de um sistema de auditorias da qualidade.

Paralelamente, o IVP melhorou as condições ambientais e físicas para realização dos ensaios, tendo procurado, tanto quanto possível, neutralizar os factores causadores de variabilidade (ex.: pressurização da sala de prova, estabilização da temperatura e humidade, cabinas individuais, *touch screens* para recolha dos dados, etc.). Os métodos e as rotinas utilizados na prova foram analisados em pormenor e, em alguns casos, redefinidos e testados (ex.: especificação das escalas usadas na avaliação das características dos vinhos). Todos os métodos de ensaio foram incluídos no Manual de Métodos de Ensaio do Serviço de Prova. De igual forma, foram desenvolvidos os procedimentos necessários ao exercício das funções previstas no sistema da qualidade. Um total de 26 procedimentos foram finalmente incluídos no Manual de Procedimentos do Serviço de Prova. O modo de condução dos ensaios, o planeamento da afectação dos provadores aos ensaios, a preparação e colocação das amostras para a análise sensorial, a avaliação da competência dos provadores, são exemplos de actividades consagradas neste Manual de procedimentos escritos.

O estudo deste caso incidiu apenas sobre a avaliação das competências dos provadores e, quando teve início, estavam já implementadas muitas das medidas referidas, nomeadamente no que diz respeito às condições físicas e de recolha de dados da prova. Note-se que o facto de as condições referidas serem excelentes facilitou, em muito, a abordagem ao problema em causa.

5.2.1 O Vinho do Porto

O Vinho do Porto é um produto muito complexo e difícil de definir. Na sua página da internet (http://www.ivp.pt), o IVP refere-se ao Vinho do Porto da seguinte forma:

> "O Vinho do Porto é um vinho licoroso, produzido na Região Demarcada do Douro, sob condições peculiares derivadas de factores naturais e de factores humanos. O processo de fabrico, baseado na tradição, inclui a paragem da fermentação do mosto pela adição de aguardente vínica (benefício ou aguardentação), a lotação de vinhos e o envelhecimento.
>
> O Vinho do Porto distingue-se dos vinhos comuns pelas suas características particulares: uma enorme diversidade de tipos em que surpreende uma riqueza e

intensidade de aroma incomparáveis, uma persistência muito elevada quer de aroma quer de sabor, um teor alcoólico elevado (geralmente compreendido entre os 19 e os 22% vol.), numa vasta gama de doçuras e grande diversidade de cores. Existe um conjunto de designações que possibilitam a identificação dos diferentes tipos de Vinho do Porto."

Na figura 5.1, apresenta-se, sucintamente, o processo produtivo dos tipos de Vinho do Porto mais comuns.

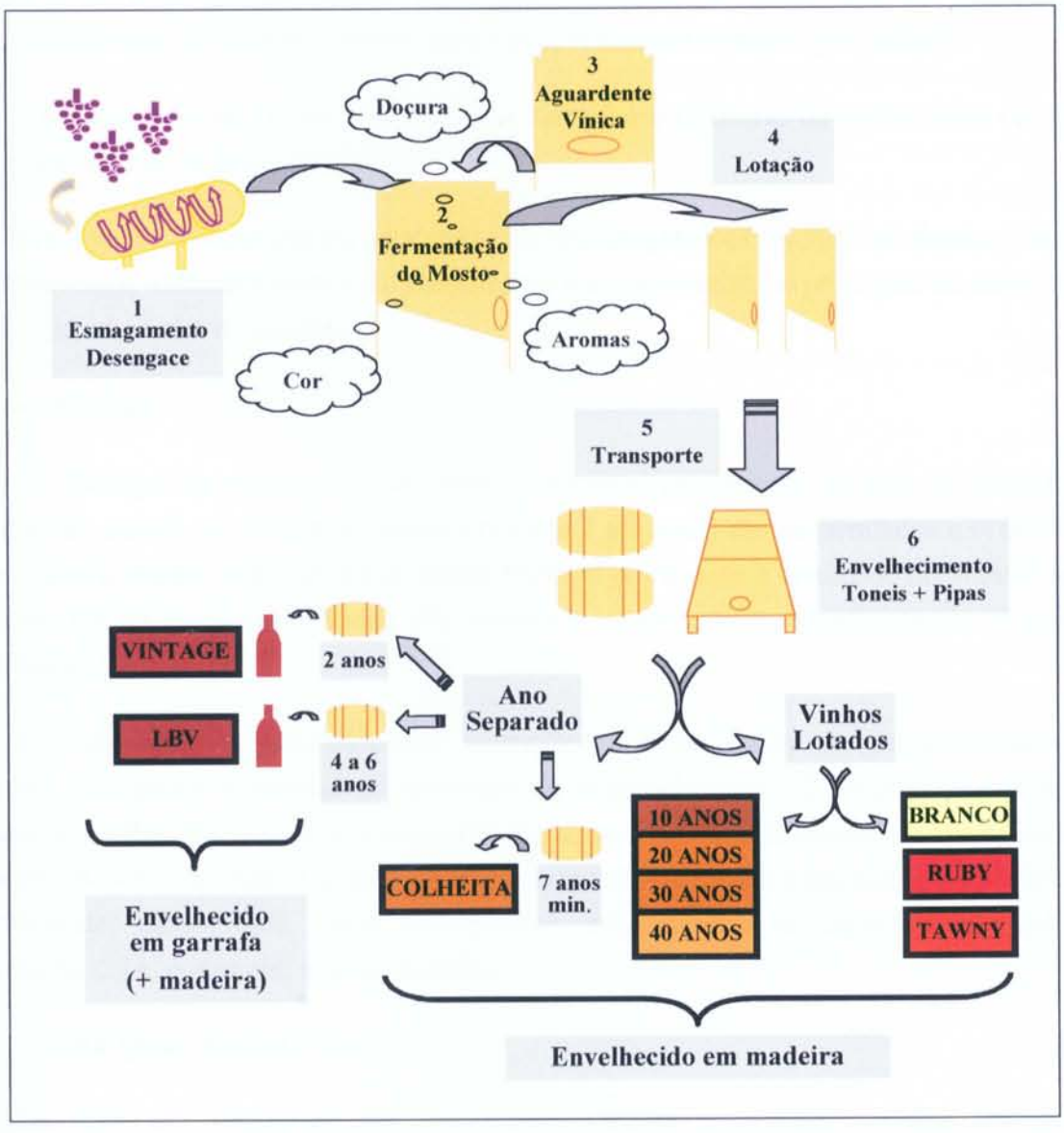

Figura 5.1 – Processo produtivo do Vinho do Porto

Analisando a figura, pode verificar-se que, após o esmagamento e desengace das uvas (fase 1), se inicia o processo de fermentação do mosto (fase 2) que é, em determinado momento, interrompido pela adição de aguardente vínica (fase 3). Depois,

procede-se a uma primeira lotação entre vinhos da mesma colheita (fase 4), seguindo-se o transporte para as caves (fase 5), onde se inicia o envelhecimento (fase 6). Há várias formas de envelhecimento, que dão origem a tipos de vinhos diferentes, cada um com uma dada designação.

O envelhecimento pode dar-se em toneis ou balseiros (vasilhas em madeira de grandes dimensões) ou em cascos (pipas de madeira com a capacidade de 550 litros) e, em alguns casos, prossegue na garrafa. Na madeira, o vinho mantém-se em contacto com o ar e o envelhecimento dá-se por oxidação, ao passo que na garrafa o envelhecimento se dá sem contacto com o ar e, consequentemente, por redução.

Por outro lado, os vinhos podem, ainda, ser lotados (misturas de vários anos) ou ser provenientes de uma única colheita.

Em seguida, descrevem-se, para cada uma das designações referidas na figura, o tipo de processos utilizados para o envelhecimento e as características principais de cada um dos tipos de vinho resultantes.

- ***Vintage***

Os *Vintage* são vinhos de qualidade excepcional proveniente de uma só colheita. Após um estágio em toneis de madeira durante 2 a 3 anos, são engarrafados e evoluem em garrafa durante um período de tempo variável (consoante a qualidade do vinho e as condições em que é conservado), não inferior a 10 anos, mas podendo atingir 50 anos ou mais.

De características organolépticas excepcionais, estes vinhos devem apresentar-se muito encorpados e retintos no momento do engarrafamento. O envelhecimento em garrafa confere-lhes macieza e elegância, desaparecendo gradualmente a adstringência inicial. Aos *Vintage* com alguns anos em garrafa estão associados aromas de torrefacção (chocolate, cacau, café, caixa de charutos, etc), aromas de especiarias (canela, pimenta, ...) e, por vezes, aromas frutados.

- ***LBV* (*Late Bottled Vintage*)**

Os *LBV* são vinhos de um estilo relativamente semelhante ao dos *Vintage*. Permanecem em toneis cerca de 4 a 6 anos, após o que envelhecem em garrafa. O vinho é também de uma só colheita e de qualidade superior, mas não lhe é exigida a excelência dos *Vintage*.

Aquando do engarrafamento, os *LBV* são tintos na cor e revelam-se encorpados e macios. O seu aroma, mais ou menos frutado, pode manifestar alguma evolução, dependente da duração do estágio em madeira. São normalmente menos adstringentes e menos encorpados do que os *Vintage* do mesmo ano.

- **Vinho *Tawny* com Data de Colheita**

Os vinhos deste tipo são vinhos de boa qualidade e provêm de uma só colheita, envelhecendo em cascos durante um período mínimo de 7 anos. No rótulo das garrafas é obrigatória a indicação do ano da colheita e do ano do engarrafamento, sendo assim possível determinar o período de tempo durante o qual o vinho envelheceu em cascos.

Durante este envelhecimento, os aromas jovens, frutados e frescos, evoluem por via oxidativa, dando lugar a um *bouquet* em que sobressaem os aromas de frutos secos, de torrefacção, madeira e especiarias. A cor evolui para o alourado, notando-se mesmo reflexos esverdeados nos vinhos muito velhos.

Um vinho deste tipo não deve permanecer em garrafa por períodos de tempo muito longos, pois corre o risco de adquirir aromas desagradáveis.

- **Vinhos *Tawny* com Indicação de Idade: 10 Anos, 20 Anos, 30 Anos e 40 Anos**

São vinhos de boa qualidade obtidos por lotação de vinhos de colheitas de diversos anos. Estagiam em madeira durante períodos de tempo variáveis e apresentam as características organolépticas próprias de um vinho envelhecido em casco durante o período correspondente à idade indicada. Por exemplo, um vinho 10 Anos revela uma cor, um aroma e um sabor típicos de um vinho que permaneceu durante 10 anos em casco.

Tal como os vinhos com Data de Colheita, estes vinhos apresentam um característico *bouquet* de oxidação que se traduz em aromas de frutos secos, torrefacção, especiarias, mais evidentes nos vinhos com mais idade. Na boca revelam-se vinhos macios e harmoniosos.

- **Vinhos *Tawny* Correntes**

Englobam-se nesta designação vinhos de categorias bastante distintas, que podem ir desde vinhos com aroma e sabor próprios de uma maturação de 3 ou 4 anos em casco, ainda com algumas reminiscências de vinho jovem (nomeadamente o caracter frutado), até vinhos mais velhos com algumas das propriedades anteriormente apontadas para os vinhos com Data de Colheita e com Indicação de Idade.

A cor apresenta evolução, devendo integrar-se nas sub-classes de cor tinto-alourado, alourado ou alourado-claro.

- ***Ruby***

Esta designação engloba vinhos de qualidade aromática e gustativa bastante variável, obtidos por lotação de diversas colheitas em que a maturação é conduzida em recipientes de grande volume. O período médio de envelhecimento situa-se, em geral, entre os 2 e os 3 anos. Os *Ruby* podem ser desde vinhos muito frutados, com corpo e estrutura, até vinhos de aroma discreto e pouco encorpados.

São vinhos em que se procura suster a evolução da cor, devendo ser integrados na sub-classe de cor tinto.

- **Branco**

Os Vinhos do Porto brancos apresentam-se sob vários estilos, sendo, tradicionalmente, classificados em função do grau de doçura. São também obtidos por lotação, mas, contrariamente aos Vinhos do Porto tintos (caso de todos os que foram descritos até aqui), que são em regra meio-doces ou doces, os brancos podem ser extra-secos, secos, meio-secos, doces ou muito doces (ou lágrima).

Trata-se de vinhos jovens, com aromas frutados e florais e, não raramente, com um ténue aroma a mel. A cor varia entre o branco pálido e o branco dourado e é predominantemente função da maceração a que as uvas são sujeitas.

Os tipos de Vinho do Porto descritos são, por vezes, agrupados segundo alguns critérios. O critério clássico e natural é a divisão entre vinhos do Porto brancos e tintos. No entanto, dentro dos tintos, há a definição das categorias especiais que englobam os *Vintage*, os *LBV*, os vinhos com data de Colheita e os vinhos com indicação de Idade. Os vinhos *Ruby* e os vinhos *Tawny* que não se incluem nas categorias especiais são denominados vinhos correntes.

Há um outro critério que se refere ao modo de envelhecimento e que separa os vinhos que são envelhecidos em casco – vinhos com data de Colheita, vinhos com indicação de Idade e *Tawny* correntes – dos que não o são – *Vintage*, *LBV* e *Ruby*. Note-se que este último é envelhecido apenas em madeira, embora por pouco tempo, e em vasilhas grandes (toneis), enquanto os dois primeiros continuam o seu envelhecimento em garrafa.

Ainda acerca do envelhecimento, refira-se que a expressão "o vinho do Porto quanto mais velho melhor", utilizada frequentemente em meios menos profissionais, resulta de uma crença só aplicável aos vinhos tipo *Vintage* e *LBV*, pois os outros tipos não ganham nada em permanecer na garrafa, podendo mesmo, em muitos casos, verificar-se alguma deterioração.

5.2.2 O Papel da Câmara de Provadores

A comercialização de Vinho do Porto engarrafado está dependente da aprovação do vinho pelo IVP e consequente atribuição de um número de registo. Quando um vinho é aprovado, passa, portanto, a poder usar o Selo que garante a autenticidade do Vinho do Porto engarrafado na origem.

Para além dos registos e das suas renovações, há outras situações que exigem a apreciação organoléptica feita pela CP, tais como as fiscalizações ou os registos de cedências.

Os momentos em que os vinhos são submetidos para aprovação pelo IVP variam consoante o tipo de vinho, havendo, no entanto, algumas regras. Por exemplo, os *Vintage* são submetidos para aprovação pelo IVP na altura do engarrafamento (com 2 ou 3 anos de envelhecimento em toneis) e as Colheitas devem renovar os registos de 3 em 3 anos.

A descrição exaustiva das diversas finalidades com que são feitas as provas ou dos momentos em que os vinhos são provados não se encontra no âmbito deste trabalho. Salienta-se apenas que, no que diz respeito às categorias especiais, a idade do vinho é um dos parâmetros que é avaliado na prova e que deverá corresponder à idade indicada na sua designação.

Cabe, então, à Câmara de Provadores do Instituto do Vinho do Porto pronunciar-se relativamente à aprovação, no que diz respeito às características organolépticas, dos vinhos que são submetidos ao IVP. Note-se que, quando uma empresa submete um vinho para aprovação pelo IVP, indica, no processo, qual a designação com que o pretende comercializar.

A CP é composta por 7 provadores, sendo o júri de prova, em cada sessão, formado pelo número ímpar máximo de provadores em serviço no dia da sessão. Em cada semana há um provador responsável pela coordenação das sessões de prova, designado

através de um processo de rotatividade (cada semana um provador diferente), não fazendo parte do júri de prova.

A decisão final da CP de aprovar ou não uma amostra baseia-se na decisão individual de cada um dos membros do júri de prova que, após terem avaliado cada uma das características organolépticas do vinho em questão, emitem um parecer final – o vinho está de acordo com a designação que é proposta (caso em que o provador aprova o vinho) – ou – o vinho não está de acordo com a designação (caso em que o provador não aprova o vinho).

Na secção seguinte, apresentam-se a descrição de alguns aspectos do funcionamento da Câmara de Provadores que se julgam relevantes para o estudo do problema e a análise de alguns dos dados de prova disponíveis à data do início da colaboração com o IVP.

5.2.3 Descrição e Análise da Situação Inicial

A Câmara de Provadores do Instituto do Vinho do Porto possui excelentes condições de trabalho que permitem, tanto quanto possível, neutralizar os factores causadores de variabilidade. A título de exemplo, salientam-se, no que diz respeito às condições ambientais, a pressurização da sala de prova e a estabilização da sua temperatura e humidade. Já no que respeita às cabinas individuais de prova, salienta-se o material com que foram construídas, inodor e facilmente lavável, a iluminação própria com luz normal e com luz vermelha (para ocultar a cor do produto) e a existência de cabinas em número suficiente para que os provadores provem todos em simultâneo.

Finalmente, e no que diz respeito à ficha de prova e ao processamento dos resultados, verifica-se que a ficha se encontra informatizada, sendo os resultados da prova inseridos pelo provador directamente numa base de dados através do *touch screen*, instalado na cabina de prova (cada cabina está equipada com um computador ligado à rede e um *touch screen*). Os resultados de todos os provadores são processados automaticamente depois de terminada a sessão de prova. Desta forma, evitam-se os erros que possam ser cometidos ao "passar" os resultados da prova de cada provador da ficha em papel para a base de dados, no computador.

Em seguida, descreve-se, resumidamente, a ficha de prova utilizada quando se iniciou este estudo.

5.2.3.1 Ficha de Prova Inicial

A ficha utilizada para a prova de vinhos no IVP no início deste estudo encontra-se reproduzida no Apêndice A. Embora as respostas dos provadores sejam, tal como foi referido no parágrafo anterior, inseridas directamente na base de dados do computador através de *touch screens*, a informação obtida é a que resulta do preenchimento dos campos indicados na ficha apresentada.

- **Ficha de prova - cabeçalho**

Sendo o objectivo de cada prova o de verificar se o produto em apreciação corresponde ou não à designação que é solicitada, o cabeçalho da ficha encontra-se já preenchido no momento da prova, servindo de referência aos provadores. Os itens incluídos no cabeçalho são:

- Produto – embora a CP também prove outros produtos (nomeadamente a aguardente vínica usada na beneficiação dos vinhos), neste estudo apenas se considera o vinho do Porto, pelo que este item assume a indicação de "Vinho".

- Tipo – este item refere-se à designação proposta para o vinho, podendo, portanto, assumir uma variedade grande de designações, das quais se salientam as mais comuns e já referidas acima: "*Ruby*", "*Tawny*", "10 Anos", "20 Anos", "30 Anos", "40 Anos", "Colheita", "*LBV*" e "*Vintage*". O campo Tipo pode também não ser preenchido, caso em que se trata de um vinho corrente que apenas tem no rótulo "Porto". Nesta situação assume-se o tipo "Não Especificado".

- Cor – este campo corresponde apenas a duas indicações: Branco ou Tinto.

- Doçura – o item doçura é referido apenas em alguns casos, especialmente em vinhos brancos (p. ex. Doce, Muito Doce, Extra-Seco).

- Ano/Idade – este campo é preenchido nos casos em que o vinho é do tipo Colheita, *LBV*, *Vintage* ou *Tawny* com indicação de Idade e refere-se ao ano de colheita do vinho ou à sua idade média.

- Designação Complementar – a designação complementar apenas é preenchida para alguns vinhos correntes ("Não Especificado", "*Ruby*" ou "*Tawny*") e pode tomar os valores 1 (nível inferior), 2, 3 e 3 Idade. A este último nível deverá corresponder uma idade de, pelo menos, 7 anos.

- **Ficha de prova – apreciação**

A parte da ficha que corresponde à apreciação do vinho inclui vários campos, cada um relacionado com uma característica importante na avaliação do vinho. As escalas utilizadas dependem dos itens em questão.

No que diz respeito às características Limpidez, Cor e Espuma, as escalas são nominais, correspondendo-lhes a descrição da característica em causa – de límpido a turvo, de retinto a branco dourado e de vermelha a branca, respectivamente, passando pelos níveis intermédios considerados relevantes.

A característica Defeito é, igualmente, expressa numa escala nominal, correspondendo-lhe a indicação do defeitos mais frequentes e ainda um campo em aberto para o caso em que o defeito notado não esteja mencionado na lista.

Embora, à primeira vista, não pareça plausível, os campos Aroma e Sabor são avaliados em escalas binárias, às quais apenas estão associados dois estados possíveis ("presente" ou "ausente", "característico" ou "não característico", "sim" ou "não", etc...). Acontece que, dos 6 níveis propostos para estas características, apenas 2 são aplicáveis em cada avaliação, dependendo do tipo de vinho em questão. Assim, o Aroma e o Sabor apresentam os seguintes estados:

- 01 (com defeito perceptível) ou 02 (sem defeito perceptível) – quando o vinho é de tipo Não Especificado e não possui Designação Complementar.

- 03 (característico da designação) ou 04 (não característico da designação) – no caso, mais frequente, de o vinho pertencer a um tipo especificado (e, portanto, campo Tipo estar preenchido).

- 05 (compatível com a designação complementar) ou 06 (não compatível com a designação complementar) – quando o vinho é de tipo Não Especificado e possui Designação Complementar.

O campo Prova de Comparação apenas é preenchido quando se trata de provas deste género. Por uma questão de simplicidade no tratamento dos resultados, e dado que o número de provas de comparação é consideravelmente inferior ao das outras provas, onde este campo não é preenchido, este campo não será considerado no estudo.

A característica Qualidade é avaliada numa escala de intervalo de 1 a 10, que pretende traduzir a qualidade global do vinho do Porto. Ou seja, uma nota de 7 é muito boa para um vinho corrente e é inaceitável para um *Vintage* que, para ser aprovado,

deverá ser classificado com uma nota mínima de 9. Para os vinhos correntes sem designação complementares serem aprovados, a nota mínima é de 5, sendo de 7 para os restantes tipos aqui referidos.

Os campos Idade e Designação são, também, expressos por variáveis binárias. O primeiro, apenas é preenchido nos casos em que tal se justifica, indicando se o vinho em questão corresponde, ou não, à idade indicada no campo Tipo ou Ano/Idade do cabeçalho.

Relativamente à Designação, e tal como acontecia com o Aroma e com o Sabor, apenas é preenchida uma de duas das 6 opções apresentadas, consoante o tipo de vinho em questão. Note-se que é neste item que o provador manifesta a sua opinião final em termos de "aprovação" ou "não aprovação" do vinho.

5.2.3.2 Metodologia Seguida na Análise dos Dados da Ficha de Prova Inicial

A primeira aproximação ao caso estudado consistiu na análise estatística dos dados de prova existentes, na perspectiva de aprofundar o conhecimento sobre o comportamento dos provadores do Instituto do Vinho do Porto. Esta análise foi feita sobre 765 vinhos dos tipos acima referidos, provados ao longo de 5 meses pelos 6 provadores da CP, perfazendo um total de 3735 provas. Na tabela 5.1, apresenta-se a distribuição destas provas pelos provadores e pelos tipos de vinho.

Tabela 5.1 – Distribuição do número de provas pelos provadores e pelos tipos de vinho

Tipos de Vinho	Prov. A	Prov. B	Prov. C	Prov. D	Prov. E	Prov. F	Total
Não Especificado	122	139	165	120	149	137	270
Ruby	51	58	63	53	62	56	116
Tawny	99	107	97	92	104	107	62
10 Anos	45	43	44	41	51	46	615
20 Anos	19	17	23	17	21	19	427
30 Anos	9	11	11	11	12	8	33
40 Anos	6	4	7	6	6	4	343
Colheita	92	94	120	101	107	101	606
Vintage	80	77	75	61	72	66	431
LBV	63	77	78	60	80	69	832
Total	586	627	683	562	664	613	3735

O objecto fundamental do estudo foi a análise da variabilidade existente nos julgamentos que os provadores fazem das várias características dos vinhos. As fontes de variação consideradas foram duas: o vinho e o provador.

As técnicas utilizadas consistiram, essencialmente, na construção de histogramas, na comparação entre valores esperados com base na análise de variância e/ou no teste de Kruskal-Wallis (ANOVA não paramétrica), na comparação das variâncias associadas a cada provador, para cada tipo de vinho e na análise de correlação entre as avaliações dos diversos provadores. O nível de significância adoptado foi o de $\alpha = 5\%$ (exceptuando os casos em que seja referido expressamente outro valor).

A ANOVA clássica foi utilizada sempre que se considerou estarem reunidas as condições mínimas de aplicabilidade desta técnica (ver, por exemplo, Guimarães e Cabral (1998), p. 465): (i) normalidade da distribuição dos erros; (ii) homogeneidade da variância dos erros; (iii) independência mútua dos erros.

Relativamente à normalidade da distribuição dos erros, existe o problema de as escalas serem discretas e, em alguns casos, com apenas 2 valores possíveis. No entanto, e visto que se trata de amostras de dimensões elevadas e que se testam os valores esperados, o teorema do limite central acaba por garantir que as distribuições admitidas para a estatística de teste sejam aproximações razoáveis da normal.

No que diz respeito à homogeneidade da variância dos erros, sucedeu que nos casos analisados praticamente nunca se verificou. Contudo, esta é uma condição em relação à qual a ANOVA é bastante robusta, sendo geralmente reconhecido que só se torna crítica quando as dimensões dos grupos são muito diferentes (normalmente, considera-se $N_{máx} \geq 2 \cdot N_{mín}$). Acontece que, neste estudo, foram detectadas bastantes situações para as quais $N_{máx} \geq 2 \cdot N_{mín}$, pelo que, nestes casos, se optou por realizar testes não paramétricos (Kruskal-Wallis).

Finalmente, considerou-se que a condição de independência dos erros era respeitada nas situações analisadas, tendo-se verificado sistematicamente que o seu desvio padrão não dependia linearmente do valor médio das observações.

5.2.3.3 Resultados Obtidos na Análise Inicial

Seguidamente, apresentam-se e comentam-se os resultados obtidos para as características estudadas. No final da secção, tecem-se alguns comentários e elaboram-se algumas análises mais generalizadas.

A primeira característica a ser analisada é a Qualidade, sendo as restantes estudadas pela ordem por que a ficha é preenchida (e já descrita em 5.2.3.1). Esta escolha deve-se, essencialmente, ao facto de a Qualidade ser uma característica medida numa escala de

intervalo comum a todos os tipos de vinho, permitindo que, através da análise dos correspondentes resultados, se adquira uma visão geral das diferenças entre os diversos tipos de vinho e entre os diversos provadores, o que, por sua vez, se revela útil para a compreensão das análises seguintes.

- **Qualidade**

As classificações atribuídas pelos provadores à característica Qualidade referem-se à qualidade global do vinho e são expressas numa escala de intervalo com 10 níveis (1 a 10). Na figura 5.2, apresenta-se o histograma das referidas classificações para cada provador e para os 765 vinhos considerados. Assume-se que as amostras são de dimensão suficiente para que todos os provadores tenham provado aproximadamente o mesmo número de vinhos e da mesma qualidade (note-se que há sempre pelo menos um provador que não prova em cada sessão). Desta forma, e ao comparar os histogramas, verifica-se que, de um modo geral, as distribuições das classificações atribuídas pelos diversos provadores são bastante semelhantes, embora, aparentemente, se distinga o provador A.

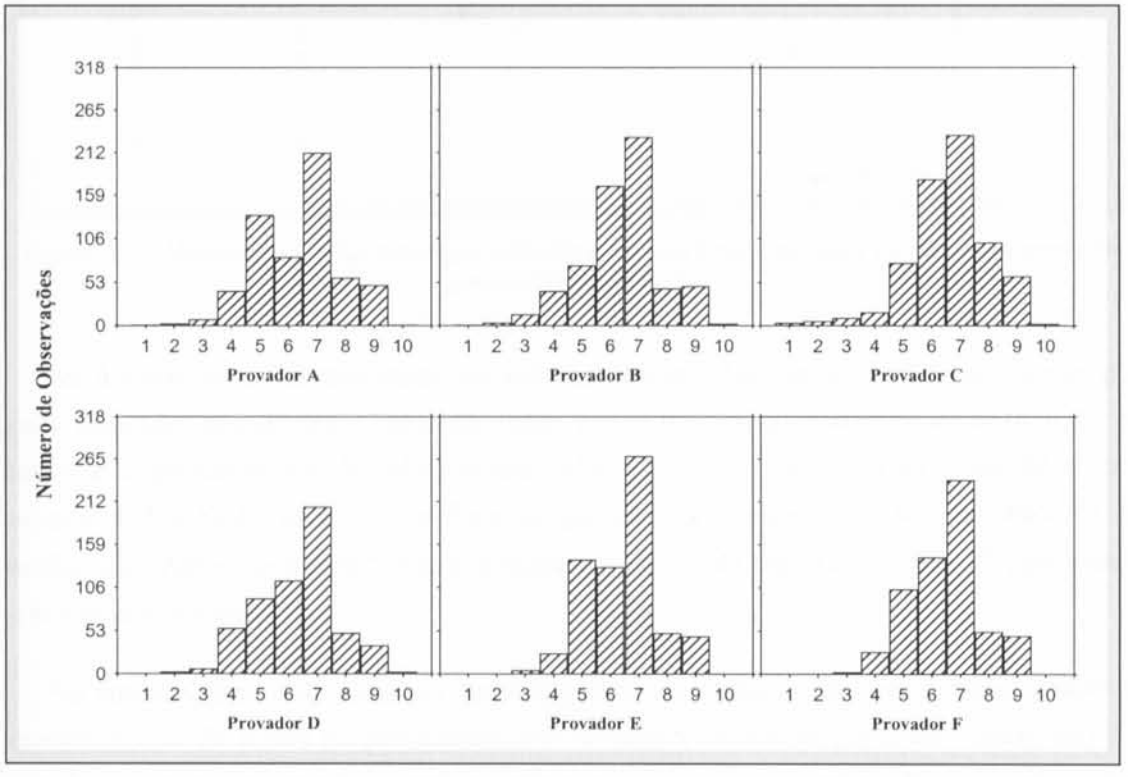

Figura 5.2 – Histogramas da pontuação atribuída, por cada provador, à característica Qualidade

Com o objectivo de explorar um pouco mais a informação contida nos dados de prova, na figura 5.3 representam-se graficamente as médias da variável Qualidade para os vários tipos de vinho em função dos provadores.

Podem distinguir-se claramente três grupos: um primeiro, com piores classificações e composto pelos tipos Não Especificado (NE), *Ruby* e *Tawny*; um segundo, que se situa numa zona intermédia da classificação e que é composto pelos tipos correspondentes às várias Idades, pelo Colheita e pelo *LBV*; e finalmente um terceiro grupo correspondente aos vinhos *Vintage*, que se distingue pelas elevadas classificações.

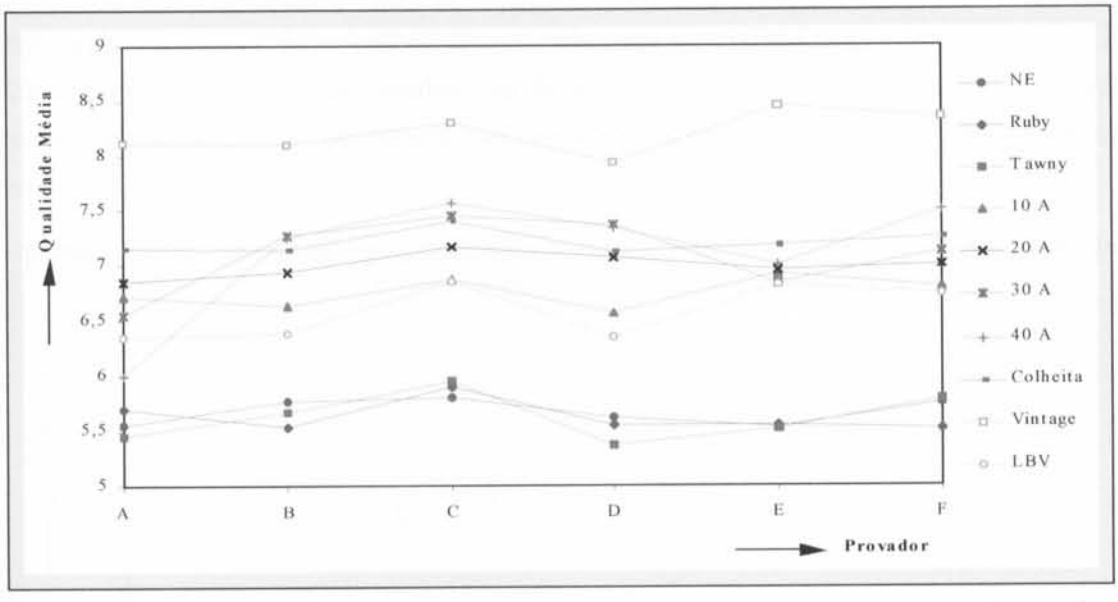

Figura 5.3 – Valores médios da pontuação atribuída à característica Qualidade em função do provador, para cada tipo de vinho

As diferenças existentes entre os valores médios das classificações atribuídas por cada provador aos diversos tipos de vinho foram confirmadas através da realização de testes não paramétricos (Kruskal-Wallis ANOVA) e, em alguns casos, também com recurso à ANOVA clássica. Verificou-se que os três grupos de vinhos encontrados na análise do gráfico da figura 5.3 se distinguem ao nível de significância de 5%, para cada um dos provadores.

No entanto, dentro de cada um dos três grupos de vinhos referidos e considerando as classificações de todos os provadores em conjunto, na maior parte dos casos não foi possível distinguir estatisticamente entre os valores esperados das classificações. Salienta-se a existência de dois sub-grupos no grupo intermédio: o conjunto dos elementos {10 Anos, LBV} que, não se distinguindo entre si, se distinguem de qualquer um dos elementos do sub-grupo {20 Anos, Colheita} os quais, por sua vez, também não

se distinguem entre si. Os vinhos de 30 e 40 Anos, tal como se pode verificar na tabela 5.1, são grupos com um número de observações pequeno quando comparado com o dos outros, pelo que não foram incluídos nesta análise mais pormenorizada.

Analise-se agora a figura 5.4, onde se representam os mesmos resultados da figura 5.3, mas em função dos tipos de vinho e para cada provador. Verifica-se que os provadores exibem um comportamento médio que não difere muito de uns para os outros. Pode, no entanto, constatar-se que o provador C atribui, em média, classificações mais elevadas do que os outros, e que o provador A revela um comportamento diferente no que diz respeito aos vinhos dos tipos 30 Anos e 40 Anos, atribuindo-lhes, em média, classificações inferiores às atribuídas aos vinhos de 10 e 20 Anos. Esta diferença pode resultar do facto de, no período examinado, ter havido poucas provas deste tipo de vinhos, e o provador em causa os ter considerado de pouca qualidade.

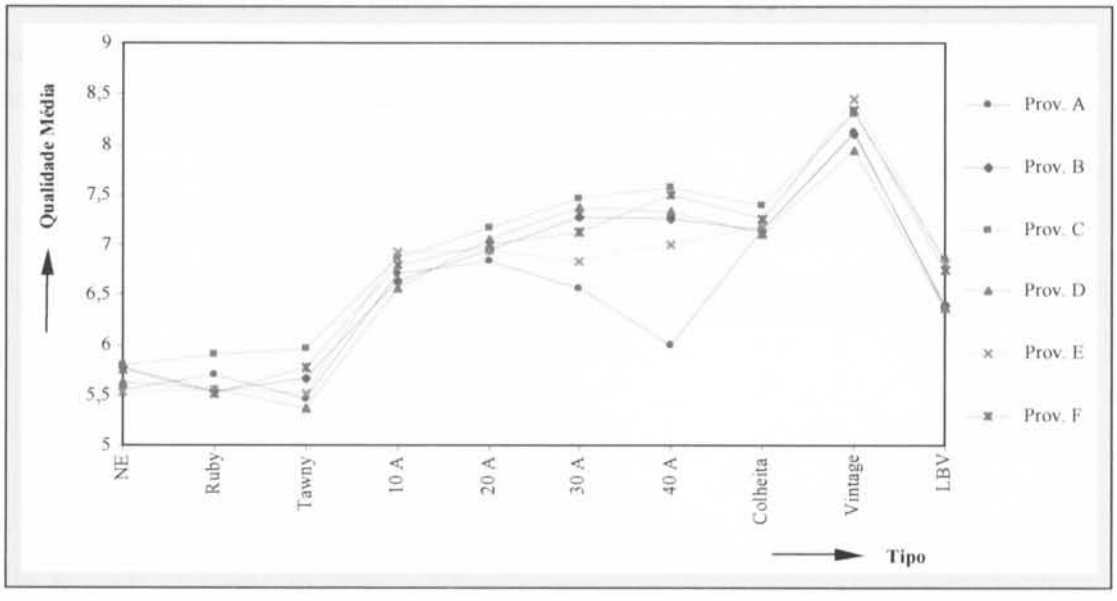

Figura 5.4 – Valores médios da pontuação atribuída à característica Qualidade em função do tipo de vinho, para cada provador

A análise de variância a um factor – provador – revelou diferenças significativas entre o provador C e os provadores A, B, D e E. Fez-se, também, uma ANOVA com os dois factores simultaneamente (provador e tipo de vinho), sendo que, neste caso, foram retirados da análise os vinhos do tipo 30 e 40 Anos. A interacção entre os dois factores não se revelou estatisticamente significativa e a diferença entre os provadores referidos manteve-se, tendo surgido uma outra a diferença – entre o provador D e os provadores E e F.

Além das médias, foram também estudadas as variâncias. Neste contexto, pode considerar-se que a variabilidade tem duas componentes essenciais:

(i) a que advém da diferença entre os vinhos provados que, mesmo quando pertencem a um determinado tipo, não são, de forma nenhuma, iguais;

(ii) a que deriva do facto de o provador poder, em momentos diferentes, avaliar de forma diferente o mesmo vinho.

A primeira componente traduz diferenças reais entre os vinhos que, uma vez analisadas, permitem avaliá-los uns relativamente aos outros. Já no que diz respeito à segunda componente, pode dizer-se que prejudica a avaliação do vinho não só pelo facto de traduzir uma variação no julgamento do mesmo produto, mas também porque é difícil, na prática, separá-la da primeira.

Seguidamente, é feita uma breve abordagem à variabilidade encontrada nas avaliações feitas aos vinhos provados. Saliente-se, no entanto, que é impossível, com os dados que se analisam, distinguir as duas componentes referidas e que as diferenças de avaliações entre provadores não podem ser completamente eliminadas.

Na figura 5.5, estão representadas as variâncias das pontuações atribuídas à característica Qualidade em função do provador, para cada tipo de vinho.

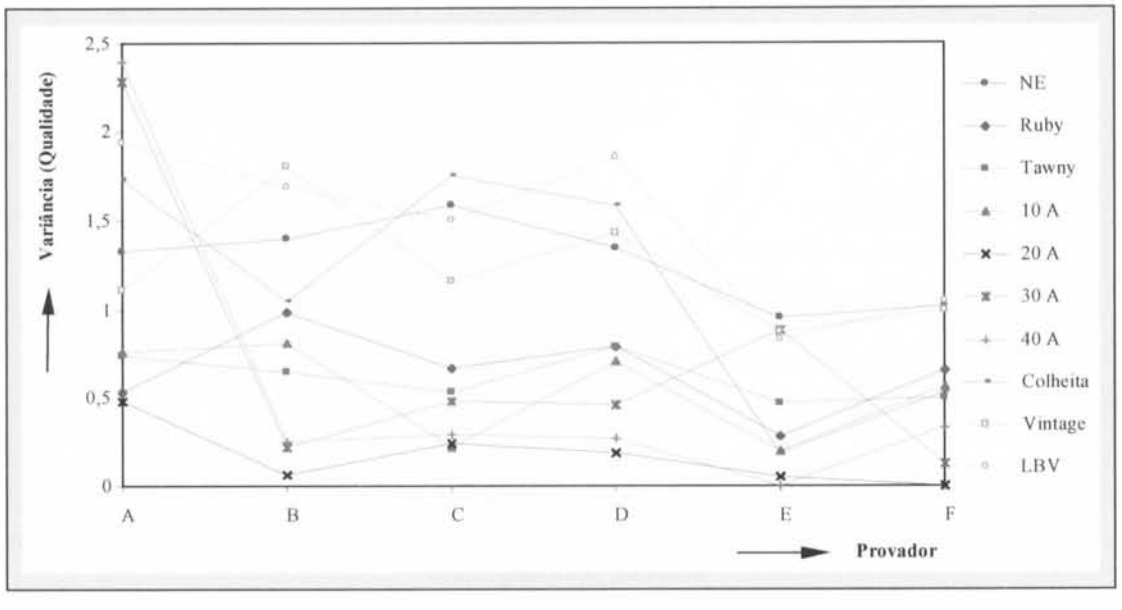

Figura 5.5 – Valores da variância da pontuação atribuída à característica Qualidade em função do provador, para cada tipo de vinho

Pode verificar-se que as variâncias mais baixas são as correspondentes aos tipos 20 Anos, 10 Anos, *Tawny* e *Ruby*. Nos vinhos de 40 Anos, as variâncias são, de um

modo geral, baixas, excepto para o provador A. No caso dos vinhos de tipo Não Especificado, as variâncias situam-se a um nível intermédio, sendo, nos restantes casos, mais elevadas.

O facto de se verificarem valores elevados para as variâncias relativamente a um determinado tipo de vinho, pode significar que os vinhos desse tipo submetidos para prova se revelaram de qualidades bastante diversas. Inversamente, quando as variâncias apresentam valores baixos, é de considerar a hipótese de a qualidade dos vinhos provados ser bastante homogénea.

Numa tentativa de esclarecer um pouco melhor a natureza da variabilidade das classificações atribuídas aos vinhos, atente-se nas diferenças entre as variâncias dos provadores. Note-se que não foi possível elaborar testes de hipóteses à igualdade de variâncias entre os provadores, em virtude de os pressupostos exigidos para tal não serem cumpridos, nomeadamente a normalidade da distribuição base.

A figura 5.6 representa a mesma informação já apresentada na figura 5.5, mas de uma forma diferente: as variâncias das pontuações atribuídas à característica Qualidade encontram-se em função do tipo de vinho, para cada provador.

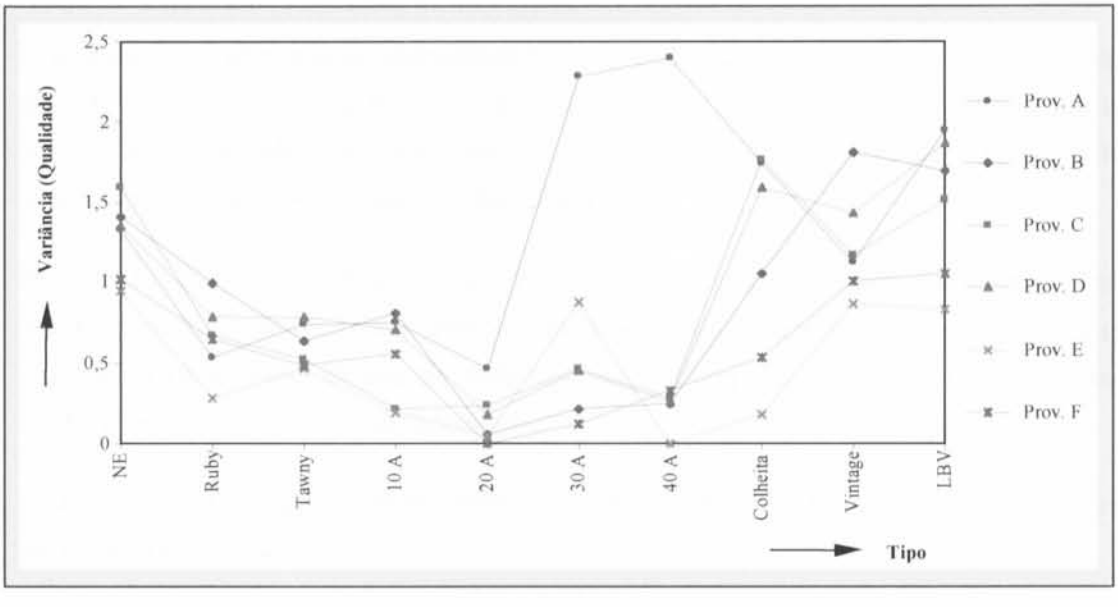

Figura 5.6 – Valores da variância da pontuação atribuída à característica Qualidade em função do tipo de vinho, para cada provador

Verifica-se que, por exemplo, para o tipo *Tawny*, os provadores apresentam variâncias semelhantes, ao passo que para o tipo Colheita as variâncias são muito diferentes. Pode ainda verificar-se que há casos de variância nula: o provador F atribuiu

a mesma classificação a todos os vinhos 20 Anos que provou (19 vinhos), o mesmo tendo acontecido com o provador E em relação aos vinhos 40 Anos (6 vinhos).

Da análise da mesma figura pode também depreender-se que, de uma forma geral, o provador A atribui notas mais variadas aos vinhos do que os outros provadores e que o provador E tende a atribuir notas de um modo mais uniforme do que os outros (com excepção do tipo 30 Anos).

Numa primeira análise, estes resultados apresentam um comportamento bastante homogéneo da parte dos provadores, mostrando que estes profissionais respondem melhor do que seria de esperar àquilo que lhes é solicitado.

Detectam-se, todavia, pequenas diferenças que são naturais e que, com base na informação apresentada até agora, se podem interpretar resumidamente da seguinte forma:

Provador A – É o provador que mais distingue, tanto entre tipos de vinho como dentro de cada tipo. No entanto, para os vinhos de idade, em termos médios, a sua discordância em relação aos outros é clara. Trata-se de uma situação que merece um estudo mais aprofundado no sentido de verificar quais as razões de tais diferenças.

Provador B – Tal como os provadores C e D, distingue bastante entre os vários tipos de vinhos. Se se considerar cada tipo em particular, este provador distingue muito dentro de quase todos os tipos, com excepção dos 20 Anos, 30 Anos e Mais de 40 Anos, onde a sua variabilidade é inferior à média.

Provador C – É o provador que atribui, em média, classificações mais altas, diferenciando-se, por isso, dos outros. Verifica-se que discrimina bastante tanto entre tipos de vinhos como dentro de cada tipo, com excepção para os vinhos 10 Anos, onde apresenta uma variabilidade baixa.

Provador D – Distingue bastante tanto entre os diversos tipos de vinhos como dentro de cada tipo.

Provador E – É o provador que menos distingue os vinhos, tanto entre tipos como dentro de cada tipo. A excepção são os vinhos 30 Anos, onde diferencia mais do que a maioria dos restantes. Verifica-se que atribui sempre a mesma nota (7) aos vinhos 40 Anos e que, no caso dos vinhos 20 Anos, há apenas um vinho ao qual atribui nota diferente.

Provador F – Distingue pouco os diversos tipos de vinho entre si, assim como também distingue pouco dentro de cada tipo de vinho. De notar que este provador atribui sistematicamente a mesma nota (7) aos vinhos 20 Anos.

Esta abordagem às diferenças entre provadores, além de muito geral, apenas versou a comparação entre os valores médios e as variâncias das classificações dadas por cada provador aos vários tipos de vinho. No entanto, mais do que as diferenças em termos médios, interessa analisar se os provadores concordam entre si em relação ao que dizem dos vinhos. Para tal, foram calculados os valores das correlações entre as classificações atribuídas pelos vários provadores, que se apresentam na tabela 5.2 (note-se que estes valores são estatisticamente diferentes de zero ($\alpha = 5\%$)).

Tabela 5.2 – Correlações entre as classificações atribuídas pelos vários provadores

$\rho_{prov.i,\ prov.j}$	Prov. A	Prov. B	Prov. C	Prov. D	Prov. E	Prov. F
Prov. A	1	0.8	0.79	0.77	0.81	0.8
Prov. B	0.8	1	0.81	0.85	0.78	0.79
Prov. C	0.79	0.81	1	0.82	0.81	0.78
Prov. D	0.77	0.85	0.82	1	0.73	0.81
Prov. E	0.81	0.78	0.81	0.73	1	0.83
Prov. F	0.8	0.79	0.78	0.81	0.83	1

Constata-se que, de uma forma geral, a correlação entre as classificações atribuídas pelos diversos provadores é alta, sendo o máximo de 0.85 (entre os provadores B e D) e o mínimo de 0.73 (entre os provadores D e E). O facto de o provador A não apresentar valores baixos de correlação com os outros provadores, o que seria esperado da análise feita anteriormente, poderá advir de o número de vinhos provados dos tipos para os quais este provador discorda mais (20 Anos, 30 Anos e 40 Anos) ser relativamente pequeno quando comparado com o número de vinhos dos outros tipos. Pelo mesmo motivo, não foi possível interpretar as correlações entre as classificações atribuídas pelos vários provadores a estes tipos de vinho.

Na Figura 5.7, são apresentados diagramas para os pares de provadores (D,E) e (B,D), onde se pode verificar o número de vezes que os provadores em causa atribuíram classificações concordantes ou discordantes para cada nível da variável, sendo, portanto, possível ter uma percepção de entre que níveis se verificaram as maiores discrepâncias. Os níveis variam de 2 a 10, já que nos vinhos estudados não houve classificações de 1.

Classific.	Provador D								
Prov. E	2	3	4	5	6	7	8	9	10
2									
3									
4		2	15		1	2			
5	1	1	15	47	37	5			
6		1	5	22	49	12	1		
7			11	3	12	142	26	1	
8			2		1	8	8	14	1
9						8	7	11	1
10									

Classific.	Provador B								
Prov. D	2	3	4	5	6	7	8	9	10
2									
3		2							
4		1	9	5	3	6			
5			10	51	17				
6			1	42	72	7			
7			3	5	14	177	11	3	
8						24	18	4	
9						1	10	27	
10									

Figura 5.7 – Diagramas das classificações atribuídas à variável Qualidade pelos pares de provadores (D,E) e (B,D)

Pode confirmar-se que entre os provadores D e E há um menor número de casos de concordância (272) do que entre os provadores B e D (290), embora tenham sido provados simultaneamente pelo primeiro par mais 37 vinhos do que pelo segundo. Observa-se, igualmente, que, mesmo quando há discordância entre os elementos do par (B,D), esta não se verifica entre classificações tão distantes como com os elementos do par (D,E), ao qual corresponde uma mancha de números na tabela mais dispersa.

- **Limpidez**

Os resultados dos testes Kruskal-Wallis realizados permitiram concluir que as classificações atribuídas à característica Limpidez diferem consoante o tipo de vinho, sendo que os vinhos dos tipos Não Especificado, *LBV* e 40 Anos têm tendência para se apresentarem menos límpidos do que os outros. Já quanto ao factor provador, não foi possível detectar diferenças significativas.

Na figura 5.8, apresentam-se os diagramas de concordância/discordância entre os provadores que concordaram menos (B e E) e os que concordaram mais (D e F). Pode verificar-se que o número de discordância é pequeno, mesmo para o caso do par (B,E).

Limpidez	Provador B			
Provador E	1	2	3	4
1	513	3	12	
2	4		1	
3	1		1	1
4				

Limpidez	Provador D			
Provador F	1	2	3	4
1	391	6		
2	8	10	6	
3			6	
4				

Figura 5.8 – Diagramas das classificações atribuídas à variável Limpidez pelos pares de provadores (B,E) e (D,F)

Note-se que, tanto no caso desta característica, como das analisadas seguidamente, não faz sentido elaborar análises de correlações, visto que as escalas não o permitem. Mesmo no que diz respeito aos coeficientes de correlação não paramétricos, como o de Spearman, que pressupõem a existência de uma escala ordinal, não faz grande sentido ordenar cerca de 500 vinhos quando apenas se utilizam 4 níveis e a maioria dos casos se concentra num desses níveis.

- **Cor / Espuma**

As características Cor e Espuma são praticamente equivalentes, verificando-se que, na quase totalidade das avaliações, elas recebem a mesma classificação ou classificação equivalente. Desta forma, os comentários aqui referidos são válidos tanto para a Cor como para a Espuma, embora, por uma questão de simplificação, apenas se refira a Cor.

Os resultados dos testes Kruskal-Wallis com o factor Tipo não são de modo nenhum surpreendentes, tendo revelado uma diferença significativa entre a cor dos diversos tipos de vinho, e não tendo sido possível encontrar diferenças entre provadores. A suposição, não contrariada pelos testes, de que os provadores concordam, de uma forma geral, em relação à cor dos vinhos, é confirmada pelos próprios provadores.

- **Aroma / Sabor / Designação**

Tendo sido realizadas análises estatísticas separadas a cada uma destas características, verificou-se que os resultados obtidos foram muito semelhantes, tendo sido mesmo iguais no caso das características Aroma e Sabor. Este facto revela que, por um lado, existe uma forte interacção entre o aroma e o sabor do vinho e, por outro, que a aprovação ou não dos vinhos depende fortemente destes dois factores. Assim, optou-se, tal como no caso anterior, por apresentar os resultados em conjunto

(utilizaram-se os resultados obtidos com a Designação, visto que é a característica que evidencia a aprovação ou a não aprovação do vinho pelo provador).

Nas figuras 5.9 e 5.10, apresentam-se, graficamente, as percentagens de vinhos aprovados por provador e por tipo, respectivamente.

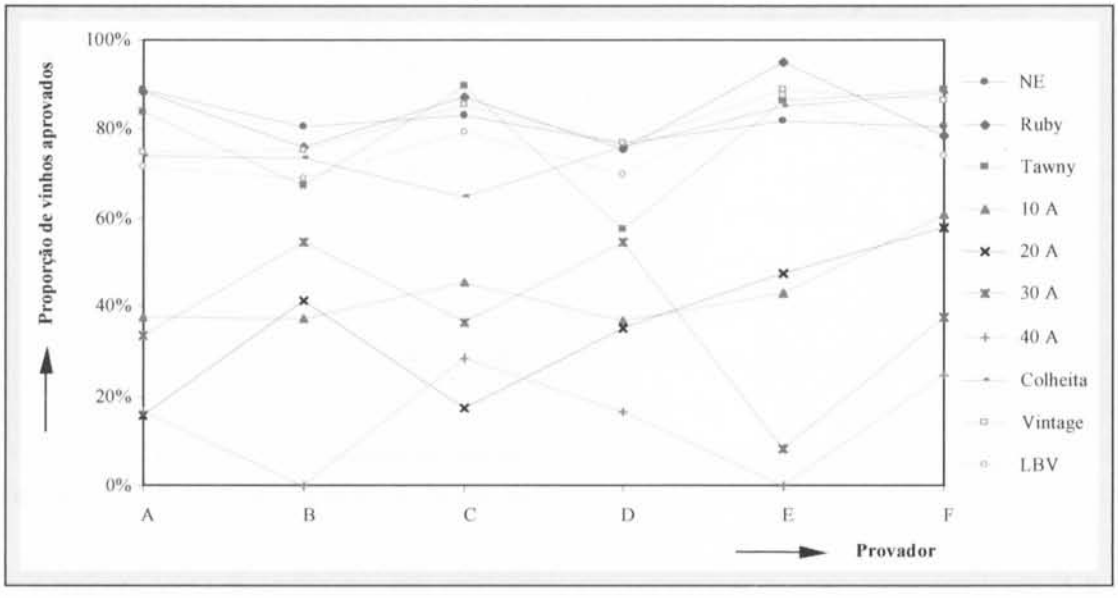

Figura 5.9 – Valores da proporção de vinhos aprovados em função do provador, para cada tipo de vinho

Na figura 5.9, verifica-se, facilmente, que os tipos de vinhos onde a percentagem de aprovação é menor são os vinhos com indicação de idade (10 Anos, 20 Anos, 30 Anos e 40 Anos). Foram feitos, para cada provador, testes à diferença entre proporções binomiais relativas aos dois grupos identificados no gráfico: (i) vinhos com indicação de idade e (ii) outros vinhos. Verificou-se que a diferença entre os dois grupos foi significativa para todos os provadores. Esta análise revela que, em geral, os vinhos com indicação de idade são aprovados menos frequentemente do que os outros, facto que deve merecer uma análise posterior, no sentido de averiguar, se possível, as causas que o originam.

Verifica-se, também, que, em média, o provador F aprova mais vinhos do que os restantes, o mesmo acontecendo com o provador E, se não se considerarem os vinhos com indicação de Idade.

No que diz respeito à figura 5.10, observa-se que há diferenças entre as proporções de aprovação para os vários provadores. Essas diferenças revelam-se estatisticamente significativas entre o provador com valor de proporção mais alto e o provador com valor de proporção mais baixo, no caso de todos os tipos de vinhos, excepto no

40 Anos, devido ao reduzido número de vinhos desse tipo provados no período em estudo.

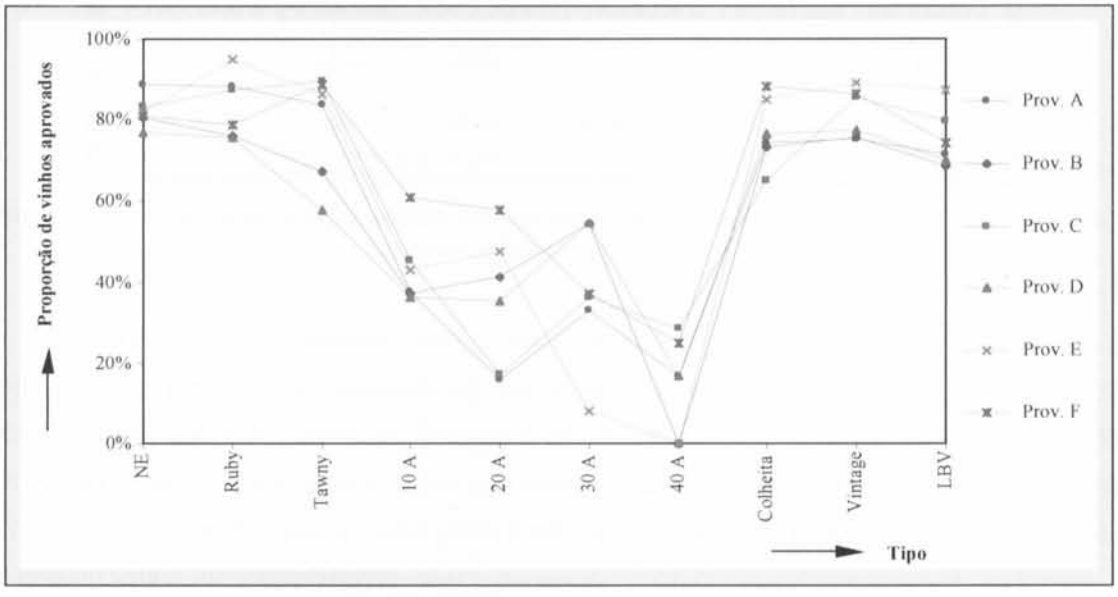

Figura 5.10 – Valores da proporção de vinhos aprovados em função do tipo de vinho, para cada provador

Analisando as duas figuras em conjunto, pode verificar-se que os provadores B e D apresentam, em média, um comportamento semelhante no que diz respeito à aprovação dos vinhos. De facto, não foi possível rejeitar a hipótese nula de igualdade entre as correspondentes proporções binomiais para nenhum dos tipos de vinho. Observa-se ainda que o provador D é o que apresenta maior uniformidade nas proporções de aprovação para os diversos tipos de vinho.

Na figura 5.11, apresentam-se os diagramas de concordância entre os provadores que concordaram menos vezes (provadores B e E) e os que concordaram mais vezes (B e D) no que respeita aos vinhos Aprovados e Não Aprovados. A concordância entre os provadores dois a dois varia, portanto, entre 78.4% (caso dos provadores B e E) e 89.6% (provadores B e D).

É curioso notar que os resultados relativos aos provadores B e D confirmam que estes provadores têm um comportamento semelhante, agora já não apenas em termos médios, mas sim no que diz respeito à opinião relativamente aos mesmos vinhos.

Designação	Provador B	
Provador E	Aprovado	Não Aprov.
Aprovado	63,1%	14,7%
Não Aprov.	6,9%	15,3%

Designação	Provador B	
Provador D	Aprovado	Não Aprov.
Aprovado	61,2%	2,4%
Não Aprov.	8,0%	28,4%

Figura 5.11 – Diagramas de concordância entre os provadores (B,E) e (B,D), relativamente à aprovação ou não aprovação dos vinhos

Tal como já foi referido atrás, os resultados apresentados são relativos à característica Designação. Saliente-se, no entanto, que as poucas diferenças existentes entre as avaliações do Sabor (e Aroma) e da Designação se referem a vinhos que não foram aprovados, tendo sido o sabor e o aroma classificados como característicos. Este facto apenas significa que a causa para a não aprovação do vinho foi outra, diversa do aroma ou sabor incaracterístico.

Nesta primeira análise não foi considerada a característica Idade, que apenas é comum a alguns tipos de vinho. No entanto, verifica-se que nos vinhos com indicação de Idade há mais situações de diferenças entre a avaliação da Qualidade (e/ou do Sabor e Aroma) e da Designação do que nos outros tipos. Pode assim concluir-se que, por vezes, a não aprovação do vinho se deve ao facto de este não ter a Idade a que se candidata, mesmo quando a característica Qualidade (e/ou Sabor e Aroma) se revela a um nível compatível com o tipo de vinho.

Seguidamente, destacam-se alguns dos resultados que se consideram mais importantes.

- Existe uma afinidade grande entre as apreciações dos provadores B e D, que se manifesta tanto a nível das médias das classificações atribuídas à Qualidade (ver figura 5.4), como da sua variância (ver figura 5.6) e das correlações entre essas mesmas classificações. Também ao nível da Designação, se verifica a existência de um elevado grau de concordância entre estes dois provadores (ver figura 5.11).

- O provador C atribuiu à Qualidade classificações superiores às atribuídas pelos outros provadores (em média). Apesar disso, não foi quem aprovou mais vinhos, o que poderá significar que, mais frequentemente do que os outros, as razões que o levaram a não aprovar os vinhos foram diferentes da falta de Qualidade.

- Os provadores F e E (este último com excepção dos vinhos com indicação de Idade) foram os que mais vinhos aprovaram, sendo que as classificações atribuídas à Qualidade por estes provadores foram das que exibiram menores variâncias, tendo-se os seus valores médios situado ligeiramente abaixo dos das classificações atribuídas por C.

- O provador A utiliza mais amplamente a escala da característica Qualidade, atribuindo notas mais variadas aos vinhos, ao passo que os provadores E e F concentram mais as classificações em alguns valores particulares da escala. Uma interpretação possível destes resultados é a de que o provador A distingue mais os diversos vinhos do que os outros, enquanto os provadores E e F os distinguem menos. Mas a variabilidade pode também ser consequência de erros introduzidos pelo próprio provador que, perante vinhos semelhantes, os diferencia. Para distinguir estes dois tipos de variabilidade, é necessário utilizar outras técnicas, recorrendo a ensaios nos quais os provadores provam o mesmo vinho várias vezes, de forma a poder estimar a variabilidade associada ao provador e distingui-la daquela que resulta de se analisarem diferentes vinhos.

No estudo que acabou de ser descrito, evidenciam-se as diferenças entre os vários tipos de vinho e faz-se uma primeira abordagem às diferenças existentes entre os provadores. Da análise efectuada ressalta a evidência de que é necessário aprofundar o estudo da variabilidade dos julgamentos dos provadores. A mesma evidência é reforçada, se se tiver em consideração que os provadores são instrumentos de medida e, tal como foi referido no Capítulo 4, a calibração dos instrumentos e o controlo da qualidade dos resultados que produzem, encontram-se intimamente relacionados com a variabilidade desses mesmos resultados.

No entanto, e também através deste estudo, fica patente a complexidade do problema que se pretende tratar. A existência de diversidade natural de opiniões e de julgamentos entre provadores e o facto de o vinho do Porto ser um produto com um grau de complexidade sensorial muito elevado, não sendo possível definir "padrões", dificultam a calibração dos provadores e/ou o seu controlo estatístico.

5.3 A Complexidade do Problema

A natureza do problema da calibração e do controlo de um painel de provadores é, tal como já foi salientado noutras secções, muito diversa da natureza das situações nas quais se utilizam as técnicas de calibração e controlo tradicionais.

Inicialmente, foi considerada uma abordagem próxima da calibração tradicional, onde se admitiu que as variáveis apreciadas pelos provadores eram expressas em escalas discretas de intervalo (escalas multinomiais). Na ficha de prova em vigor, tal só acontecia no caso da Qualidade, mas os próprios provadores foram da opinião de que poderiam, com alguma facilidade, avaliar, numa escala de intervalo de 1 a 5, outras características como o Aroma, o Sabor, a Idade e a Designação.

Assim, numa primeira fase, admitiu-se que seria possível utilizar o procedimento descrito em 4.4.1, para calibrar cada provador do painel para cada uma das características do vinho passíveis de serem avaliadas em escalas multinomiais. O mesmo procedimento permitiria, a partir dos resultados da calibração dos provadores, calibrar o próprio painel.

A primeira dificuldade encontrada na tentativa de planificar uma experiência como esta foi a definição dos padrões que, neste contexto, é muito complexa. Uma hipótese seria a de tentar obter diferentes lotes de vinho do Porto que fossem diferenciada e unanimemente classificados pelos provadores e que, portanto, servissem como padrão, um para cada nível da escala.

Mesmo admitindo que seria possível essa definição, colocam-se outras dificuldades na implementação deste procedimento, das quais se salientam:

- O elevado número de lotes padrão requerido. Note-se que seria necessário obter 5 lotes padrão (um para cada nível da escala) para cada característica avaliada (Aroma ou Designação, por exemplo) e para cada tipo de vinho. Assumindo que os provadores seriam calibrados para 5 características avaliadas em escalas de 1 a 5 e para os 10 tipos de vinho mais comuns, seriam necessários 250 lotes padrão. Este problema seria minimizado se, ao seleccionar os lotes, se tivesse em atenção que cada lote poderia servir de padrão para várias características. No entanto, o volume de vinho necessário manter-se-ia em níveis muito elevados.

- O facto de uma amostra servir para avaliar várias características de uma forma independente. O provador é calibrado para cada característica, pelo que deverá apreciar uma característica de cada vez. Aliás, ao ser calibrado no Aroma, por exemplo, o provador deverá apreciar o Aroma de vários padrões sem ter que se preocupar com outras características dos mesmos vinhos. No entanto, se os mesmos vinhos forem utilizados para padrão de várias características, o provador poderá ter que os provar demasiadas vezes, o que provocará, eventualmente, cansaço e, também, alguma confusão. A apreciação simultânea de várias características não foi considerada, visto que não é compatível com a natureza do conceito de calibração, que deve ser feita para cada característica individualmente, evitando, assim, ao máximo os possíveis efeitos de interacção.

- O elevado número de provas a que os provadores seriam sujeitos, independentemente de se tratar ou não dos mesmos vinhos. Em cada calibração cada provador deveria provar cada vinho padrão de uma forma cega, pelo menos três vezes, havendo o risco de ocorrerem efeitos de adaptação acentuados.

- Os provadores poderem reconhecer alguns dos vinhos e atribuir a classificação de memória, sem executarem a rotina da prova. Este risco é mais acentuado nos padrões correspondentes aos níveis extremos da escala (classificações demasiado baixas ou demasiado elevadas), que são mais facilmente recordados.

- A impossibilidade de repetir a calibração dentro de períodos de tempo aceitáveis, dado o volume de provas requerido para calibrar os provadores para todos os níveis de todas as características de todos os tipos de vinho.

- A possibilidade de os vinhos padrão se alterarem ao longo do tempo e, portanto, deixarem de poder ser considerados padrão.

Todas estas questões apontam para que a implementação, na prática, do procedimento referido ou de outros baseados nos conceitos tradicionais seja, se não impossível, pelo menos muito difícil de concretizar. Acresce que o período de tempo necessário para testar o procedimento seria demasiado longo e que o teste acarretaria custos significativos.

No próximo capítulo, apresenta-se uma proposta de abordagem ao problema da avaliação e do controlo das decisões dos provadores. Esta proposta, aplicada ao caso do IVP, tem a vantagem de alterar muito pouco a rotina de prova e de apenas exigir um número adicional de provas perfeitamente comportável de um ponto de vista prático.

6

Avaliação e Controlo dos Julgamentos de Câmaras de Provadores

Conteúdo:

- Introdução
- Estudo da Ficha de Prova
- Caracterização do Comportamento dos Provadores e Controlo da Sua Estabilidade
- Avaliação do Desempenho Diário dos Provadores
- Avaliação Periódica do Desempenho Global dos Provadores
- Aferição Periódica de Capacidades Específicas dos Provadores
- Decisão Final do Júri de Prova

6.1 Introdução

O problema do controlo da qualidade dos resultados emitidos pelos provadores profissionais na sua qualidade de instrumentos de medida foi sendo colocado ao longo deste trabalho.

Numa primeira fase, verificou-se que na área da Análise Sensorial os métodos propostos são insuficientes. Na realidade, existem várias abordagens para a selecção e o treino de provadores, mas, relativamente ao controlo do seu desempenho, as propostas são escassas. Na maior parte das vezes, limitam-se a aconselhar a repetição periódica dos testes de selecção, por forma a garantir que os provadores mantêm as capacidades com que foram contratados.

Numa segunda fase, fez-se uma tentativa de aplicação dos conceitos tradicionais de calibração de instrumentos e de controlo estatístico aos processos de avaliação sensorial, tendo-se verificado que é muito difícil, na prática, reunir as condições necessárias para que tal seja possível.

Neste capítulo, propõe-se uma abordagem global ao problema, incidindo em quatro vertentes distintas:

(i) **Estabilidade do processo de decisão individual do provador**. Quando o parecer final do provador sobre uma dada amostra é passível de ser traduzido numa variável binária ("0" – "Não aprovado" e "1" – "Aprovado", por exemplo), é natural que o resultado que lhe está associado traduza a apreciação que ele faz das características organolépticas da amostra. É importante que o provador tenha um processo de decisão estruturado e estável ao longo do tempo. Se duas amostras obtêm classificações idênticas nas várias características apreciadas, é de esperar que a decisão final seja a mesma nos dois casos. Assim, sugere-se que o processo de decisão individual de cada provador seja modelizado através de uma regressão logística (para o caso citado da decisão binária) e que se controle a sua estabilidade ao longo do tempo através de uma carta de controlo CUSUM.

(ii) **Monitorização diária da fiabilidade do provador**. Com base nos resultados obtidos a partir da apreciação de réplicas de amostras já provadas e dos dados relativos à dispersão das classificações atribuídas num dado período de tempo, desenvolveu-se um índice de fiabilidade, que deve ser actualizado diariamente e

que pode ser utilizado como indicador da capacidade do provador para reproduzir resultados. Este procedimento pode ser utilizado para qualquer característica avaliada numa escala multinomial.

(iii) **Avaliação periódica do desempenho global do provador**. O método proposto utiliza a análise ROC (*Receiver Operating Characteristic*) para definir um indicador da sensibilidade do provador à qualidade global das amostras durante o período de tempo em apreciação. Para que seja possível utilizar este método, é necessário que, à característica em análise (neste caso, a qualidade global das amostras), esteja associada uma decisão entre duas alternativas (por exemplo, aprovar ou não as amostras). Simultaneamente, essa característica deverá também ser avaliada numa outra escala (que poderá ser de intervalo ou apenas ordinal).

(iv) **Aferição da capacidade para identificar características e/ou defeitos**. Na maior parte dos produtos alimentares há algumas características e/ou defeitos cuja presença (em determinadas quantidades) é necessário detectar. Recomenda-se que, periodicamente, se faça a aferição das aptidões dos provadores para discriminar e classificar diferentes intensidades das características organolépticas que se revelem importantes. Há numerosos ensaios sensoriais que se podem utilizar para este fim. O ensaio seleccionado permite obter um indicador da capacidade de discriminação do provador entre as diferentes intensidades da característica em avaliação.

Esta parte do trabalho foi desenvolvida em colaboração estreita com a Câmara de Provadores do IVP, o que constituiu uma excelente oportunidade de aplicação a um caso real. Assim, foi possível, por um lado, avaliar a exequibilidade da implementação, na prática, dos procedimentos propostos e, por outro, proceder à introdução de algumas alterações que facilitaram essa implementação.

Na próxima secção, analisam-se os resultados do estudo relativo à ficha de prova e apresenta-se a nova ficha utilizada na abordagem efectuada ao longo de todo este capítulo. Nas secções seguintes (6.3 a 6.6), aborda-se cada uma das vertentes mencionadas através da descrição dos procedimentos envolvidos e da análise e interpretação dos resultados obtidos com os dados de prova da Câmara.

Após a análise dos comportamentos dos provadores individualmente, é necessário definir o modo como os resultados das suas avaliações se conjugam, de forma a obter-se a avaliação final do painel. Na secção 6.7, aborda-se, então, o problema de como

articular as decisões dos membros de um painel de provadores, traduzindo, o mais fielmente possível, a avaliação dos produtos em análise.

Note-se que, embora a abordagem que se propõe seja baseada no caso do IVP, ela é facilmente aplicável a outras situações para as quais o problema é estruturalmente semelhante.

6.2 Estudo da Ficha de Prova

No capítulo anterior, foi apresentada a ficha de prova inicialmente utilizada no IVP. Pode constatar-se que as escalas utilizadas são dos tipos nominal ou binário (exceptuando a escala correspondente à Qualidade). Verifica-se que, particularmente neste último caso, a informação resultante é muito limitada ("presente" ou "ausente", "característico ou "não característico", etc.), dificultando abordagens mais elaboradas.

A situação foi analisada com todo o cuidado, envolvendo, entre outros agentes, os próprios provadores. Seguidamente, salientam-se algumas considerações que resultaram dessa análise e que serviram de base às decisões tomadas na elaboração da nova ficha de prova.

- No caso de algumas das características avaliadas, é possível utilizar escalas de intervalo com 5 níveis, sem que tal implique um esforço excessivo para o provador. É o caso do Aroma, do Sabor, da Idade e da Designação.

- Relativamente ao Defeito, considerou-se que o campo (expresso em escala nominal) apenas especifica o tipo de defeito que, quando existe, se manifesta sempre no aroma ou no sabor do vinho.

- A Limpidez, a Cor e a Espuma são características difíceis de avaliar em escalas de intervalo.

- A Limpidez e a Cor são características passíveis de serem medidas objectivamente com recurso a instrumentos de medida convencionais.

- A Limpidez é uma variável determinante na aprovação do vinho: um vinho que não esteja límpido não pode ir para o mercado. No entanto, trata-se de uma característica que é normalmente mais consensual entre os provadores do que as restantes e verifica-se que, dos vinhos não aprovados pela Câmara, apenas cerca

de 13% não se apresentam límpidos, baixando essa percentagem para cerca de 3% quando se considera o total de vinhos submetidos à Câmara.

- A Cor e a Espuma estão normalmente associadas à Idade, ao Aroma ou ao Sabor do vinho. Ou seja, quando a cor não é a correspondente à categoria do vinho, esse facto reflecte-se nas outras características. Exceptuando os tipos "*Ruby*" e "*Tawny*" (que, como se verá mais adiante, por vezes geram alguma confusão), não é comum que a razão da não aprovação de um vinho se relacione exclusivamente com a sua cor.

- O Corpo foi considerado uma característica importante na avaliação global do vinho. Verificou-se que, também neste caso, se poderia utilizar uma escala de intervalo com 5 níveis.

A introdução das modificações à ficha de prova foi feita de uma forma cuidadosa: a ficha inicial continuou a ser preenchida tal como foi referido no capítulo anterior, e apenas foram acrescentados, no final, cinco campos referentes às características Aroma, Sabor, Corpo, Idade e Designação. Estas últimas são avaliadas numa escala de intervalo de "1" a "5", em que "1" significa que a característica se encontra num nível muito abaixo do esperado (para o tipo de vinho em questão) e "5" significa um nível acima do esperado.

Pode verificar-se que quatro destas características (Aroma, Sabor, Idade e Designação) são avaliadas duas vezes, uma primeira na ficha inicial numa escala binária e uma segunda na escala multinomial referida no parágrafo anterior. O trabalho que se descreve nas próximas secções baseia-se, essencialmente, na segunda parte da ficha, pelo que, salvo indicação em contrário, a referência ao Aroma, ao Sabor, à Idade e à Designação terá em consideração a escala multinomial e não a binária.

No entanto, como a escala binária referente ao campo Designação traduz a decisão final do provador relativamente ao vinho, por uma questão de simplificação, decidiu-se designar a característica que lhe está associada por Decisão ("0" – "Não Aprovado" ou "1" – "Aprovado").

6.3 Caracterização do Comportamento dos Provadores e Controlo da Sua Estabilidade

A impressão global que o provador retém de um determinado produto resulta da interacção entre as muitas sensações percebidas ao longo do processo de prova. Quando se trata de provadores profissionais e, portanto, com formação e experiência nesta área, torna-se mais fácil isolar algumas das referidas sensações e apreciar separadamente as características fundamentais do produto. É, aliás, o que se passa no preenchimento da ficha de prova. Nestas condições, a apreciação global do produto não é mais do que uma síntese das apreciações das suas várias características, que resulta de processos de avaliação e decisão complexos, mais ou menos conscientes.

Em determinadas circunstâncias, é possível explicitar a síntese elaborada pelo provador através de modelos matemáticos, nomeadamente regressões múltiplas, onde a qualidade global do produto é a variável dependente e as outras características são as variáveis independentes.

No caso prático estudado, o parecer final do provador é traduzido na variável binária Decisão que, por sua vez, deverá reflectir a apreciação das outras características avaliadas. Em 6.3.1 analisa-se um modelo – o de regressão logística – que permite relacionar uma variável dependente binária com variáveis independentes expressas em escalas de intervalo. A equação resultante (uma para cada provador) traduz o processo de decisão do provador em termos de aprovação ou não dos vinhos.

Com base numa medida de adequação entre os resultados obtidos com a equação da regressão e os resultados reais, é proposto, em 6.3.2. um método para controlar a estabilidade do processo de decisão de cada provador ao longo do tempo.

6.3.1 Modelo para o Processo de Decisão Individual dos Provadores

O modelo proposto para o processo de decisão de cada provador parte do princípio de que a variável binária Decisão exprime a sua apreciação global do produto analisado, que resulta da avaliação individual de cada característica relevante do vinho.

Este tipo de processos é comum a muitas organizações. Como exemplos, podem apontar-se decisões como a do lançamento, ou não, de um produto no mercado ou a

verificação da conformidade do produto com os requisitos que lhe são exigidos (conforme/não conforme).

Admite-se, então, que a variável dependente é a Decisão que corresponde à aprovação, ou não, do vinho ("0" – "Não Aprovado" ou "1" – "Aprovado"), e que as variáveis independentes são o Aroma, o Sabor, o Corpo e a Idade (esta última apenas nos casos em que o tipo do vinho o justifica). As razões que levaram à escolha destas características foram já apresentadas na secção 6.2.

O modelo escolhido – o da regressão logística múltipla – adequa-se particularmente ao caso em questão, dado que se trata de uma variável dependente que é binária e de várias variáveis independentes expressas numa escala de intervalo.

6.3.1.1 Regressão Logística

A técnica de regressão logística é muito utilizada na área das ciências bio-médicas como modelo de previsão de fenómenos aos quais se associam dois estados (infectado ou não infectado, sobrevive ou não sobrevive, etc...). Por exemplo, em Hunter et al. (2000) e Melvin et al. (2000), os resultados obtidos com esta técnica são comparados com os obtidos com outras técnicas disponíveis nos casos da previsão de sobrevivência após um acidente e da rejeição em transplantes de fígado, respectivamente.

Na regressão logística simples, a relação entre a variável dependente, binária, e a independente é dada por:

$$\mu_{Y|x} = \frac{e^{\beta_0+\beta_1 \cdot x}}{1+e^{\beta_0+\beta_1 \cdot x}}$$

onde:

$\mu_{Y|x}$: valor esperado de Y dado que X toma o valor x;

β_0 e β_1: parâmetros da regressão.

Note-se que, no caso de Y ser codificado com os valores 0 e 1, $\mu_{Y|x}$ corresponde à probabilidade de Y tomar o valor 1 dado x e, portanto, a probabilidade de Y tomar o valor 0 dado x é dada por (1 - $\mu_{Y|x}$).

Em Freund e Wilson (1998, p. 383) são salientadas algumas das propriedades deste modelo, nomeadamente:

- O valor de $\mu_{Y|x}$ varia entre 0 e 1. No caso de $\beta_1 > 0$, $\mu_{Y|x}$ aproxima-se de 1 quando o valor de x aumenta e de 0 quando este diminui. Inversamente, quando

$\beta_1 < 0$, $\mu_{Y|x}$ aproxima-se de 1 quando o valor de x diminui, e de 0 quando este aumenta. Note-se que esta propriedade está em conformidade com o facto de a variável Y apenas tomar os valores 0 ou 1.

- $\mu_{Y|x} = \frac{1}{2}$ quando $x = -\frac{\beta_0}{\beta_1}$.
- A curva que representa graficamente $\mu_{y|x}$ em função de x é monótona e, quando $\beta_1 > 0$, tem a forma de S (*S shaped*).

O motivo pelo qual os parâmetros da regressão logística são representados pelos mesmos símbolos dos da regressão linear simples deve-se ao facto de, no processo de estimação dos referidos parâmetros, se utilizar uma transformação de variáveis – transformação logit – que resulta numa regressão linear simples. A transformação logit (aqui representada por $g()$), quando aplicada à equação da regressão logística, assume a seguinte forma:

$$g(\mu_{Y|x}) = \ln\left[\frac{\mu_{Y|x}}{1-\mu_{Y|x}}\right] = \beta_0 + \beta_1 \cdot x$$

A grandeza representada entre parênteses rectos denomina-se *odds* em x, pelo que a transformação logit pode também ser interpretada como o logaritmo de *odds* em x.

Uma outra grandeza referente à regressão logística, importante para a interpretação do significado do coeficiente β_1, é a *odds ratio*. Na regressão logística simples, a *odds ratio* é o quociente entre a *odds* em $(x+1)$ e a *odds* em x, podendo demostrar-se facilmente que:

$$odds\ ratio = \frac{odds \text{ em } x+1}{odds \text{ em } x} = e^{\beta_1}$$

A *odds ratio* é uma medida de associação entre a variável dependente e a independente, que se pode interpretar como sendo o aumento na proporção de elementos do tipo $Y = 1$, quando ocorre um aumento unitário em X.

No caso do cancro do pulmão tratado por Hosmer e Lemeshow (1989, p. 41), assume-se que Y representa a presença (1) ou ausência (0) de cancro de pulmão e que X representa o facto de a pessoa fumar (1) ou não (0). Se, por exemplo, a *odds ratio* assumir o valor 2, tal indica que o cancro de pulmão ocorre duas vezes mais nos fumadores do que nos não fumadores.

Quando há mais do que uma variável independente que influencia a variável em estudo, o modelo utilizado é o da regressão logística múltipla, expresso por:

$$\mu_{Y|x_1 \ldots x_i} = \frac{e^{\beta_0 + \beta_1 \cdot x_1 + \ldots + \beta_i \cdot x_i}}{1 + e^{\beta_0 + \beta_1 \cdot x_1 + \ldots + \beta_i \cdot x_i}}$$

onde:

$\mu_{Y|x_1 \ldots x_i}$: valor esperado de Y dado que os vários X_i tomam os valores x_i;

i: número de variáveis independentes;

x_j: variáveis independentes;

β_0: coeficiente na origem;

β_j ($j = 1..i$): coeficientes relativos às variáveis independentes.

A regressão logística múltipla é uma generalização da regressão logística simples, pelo que se aplica igualmente a transformação *logit* e a definição de *odds ratio* já referida. Note-se que, neste caso, é possível definir tantas *odds ratio* quantas as variáveis independentes.

Um estudo exaustivo da regressão logística e dos métodos alternativos que se poderiam utilizar em situações semelhantes encontra-se fora do âmbito deste trabalho (para aprofundar o tema ver, por exemplo, Amemiya (1981), Hosmer e Lemeshow (1989) ou Gessner et al. (1988)). O objectivo é a apresentação de uma proposta de modelização do processo de decisão dos provadores que permita explicitar e compreender melhor o processo em si e cuja estabilidade seja passível de ser controlada.

No caso da Câmara de Provadores do IVP, e tal como será referido nas próximas secções, o modelo da regressão logística múltipla cumpriu os requisitos referidos. No entanto, não se exclui a possibilidade de utilização de outras ferramentas que, dependendo do contexto, poderão eventualmente revelar-se mais apropriadas.

6.3.1.2 Modelos Obtidos para o Processo de Decisão dos Provadores do IVP

As variáveis independentes consideradas no estudo do processo de decisão dos provadores do IVP foram já referidas: o Aroma (*A*), o Sabor (*S*), o Corpo (*C*) e a Idade (*I*). A Decisão do provador (*D*) é a variável dependente, binária, que toma o valor "1" quando o provador decide aprovar o vinho, e "0" no caso contrário.

Um dos primeiros problemas que se colocou nesta fase do trabalho foi o da característica Idade, que não é relevante para a Decisão em alguns dos tipos de vinho (ver capítulo 5). Nestes casos, a Idade nem sequer é avaliada na ficha de prova. Verifica-se, no entanto, que os outros parâmetros são apreciados para todos os tipos de vinho, e que o tipo de critério utilizado na apreciação é o mesmo (note-se que o provador atribui a classificação consoante a conformidade entre as características da amostra e as características que um vinho daquele tipo deverá possuir).

Desta forma, para cada provador foram definidos dois modelos: um para os vinhos em que a característica Idade é apreciada onde são consideradas as quatro variáveis independentes já mencionadas, e o outro para os vinhos em que a característica Idade não é relevante e onde são consideradas apenas as três variáveis independentes – Aroma, Sabor e Corpo.

Os modelos são definidos pelas seguintes expressões:

$$\mu_{D|a,s,c,i} = \frac{e^{\beta_0+\beta_A \cdot a+\beta_S \cdot s+\beta_C \cdot c+\beta_I \cdot i}}{1+e^{\beta_0+\beta_A \cdot a+\beta_S \cdot s+\beta_C \cdot c+\beta_I \cdot i}}$$, para os vinhos com indicação de idade e

$$\mu_{D|a,s,c} = \frac{e^{\beta_0+\beta_A \cdot a+\beta_S \cdot s+\beta_C \cdot c}}{1+e^{\beta_0+\beta_A \cdot a+\beta_S \cdot s+\beta_C \cdot c}}$$, para os vinhos sem indicação de idade,

onde:

$\mu_{D|a,s,c,i}$: valor esperado da variável D dados os valores das variáveis A, S, C e I;

$\mu_{D|a,s,c}$: valor esperado da variável D dados os valores das variáveis A, S e C;

β_0: coeficiente na origem;

β_A, β_S, β_C e β_I: coeficientes das variáveis A, S, C e I, respectivamente;

a, s, c e i: valores assumidos pelas variáveis A, S, C e I, respectivamente.

Na tabela 6.1, apresentam-se os coeficientes dos modelos de regressão logística múltipla obtidos a partir de dados correspondentes a 6 meses de prova para cada um dos provadores em estudo. Estes dados foram previamente examinados, tendo sido retirados da análise os casos suspeitos de enganos no preenchimento da ficha e os casos em que os vinhos não foram aprovados por não se encontrarem límpidos. O número de amostras incluídas no estudo para cada elemento do painel em cada um dos grupos (sem indicação de idade ou com indicação de idade) durante este período variou entre 320 e 478. Os modelos foram estimados utilizando o método da máxima verosimilhança, incluído nos programas informáticos Statistica (1999) e SPSS (1996).

Os provadores estão identificados com letras diferentes das utilizadas no capítulo anterior. Esta opção foi tomada com o intuito de evitar a identificação das pessoas em causa, a fim de se manter a devida confidencialidade.

Nos casos dos provadores G e H, a variável Sabor não foi incluída na regressão pelo facto de apresentar uma forte correlação (superior a 90%) com a variável Aroma, o que dava origem a valores muito elevados para as variâncias de β_A e β_S, que denunciavam instabilidade no modelo.

Tabela 6.1 – Coeficientes da regressão logística múltipla obtidos para os vários provadores

Provador	Vinhos com Idade					Vinhos sem Idade			
	β_0	β_A	β_S	β_C	β_I	β_0	β_A	β_S	β_C
G	-17	4.2	--	0.7*	1.3	-10.5	4.2	--	0.5*
H	-20.7	5.8	--	1.6	0.2*	-11.5	5.1	--	-0.1*
I	-29.4	4,5	1.6*	1,6	2.2	-14.9	6.1	0.6*	-0.3*
J	-34.9	6.2	3.7	2.9	0*	-13.6	0.6*	5.1	-0.4*
K	-28.3	2.5	5.5	1.4*	2.1	-12.5	1.1*	4.1	0.4*
L	-24	3.3	2.4	2.3	0.6*	-19.6	3.7	2.7	1.1*

* - este valor não é significativamente diferente de 0 ($\alpha = 5\%$)

A análise dos coeficientes apresentados na tabela 6.1 permite concluir que, nos vinhos com indicação de idade, o Aroma é, de uma forma geral, a característica mais determinante na aprovação do vinho, seguido do Sabor e do Corpo. Curiosamente, o factor Idade não se revela tão determinante como os outros, havendo mesmo três casos nos quais não fica provada a existência de relação entre a classificação atribuída à idade do vinho e a sua aprovação ou não. Uma explicação possível para este fenómeno é a de que a Idade se projecta nas outras características em apreciação. Note-se que no modelo univariado só com a variável independente Idade, o respectivo coeficiente (β_I) é significativamente diferente de zero.

Ainda relativamente aos vinhos com indicação de idade, verifica-se que, por exemplo, o provador H atribui maior importância ao Corpo do vinho do que à sua Idade, ao passo que com o provador I sucede o oposto.

No que diz respeito aos vinhos sem indicação de idade, é interessante notar que o Corpo é uma característica que não parece interferir na decisão final dos provadores, contrariamente ao que acontecia nos vinhos com indicação de idade. O Aroma e o Sabor continuam a ser características que influenciam fortemente a decisão final para a maioria dos provadores.

A informação contida na tabela 6.1 não fornece qualquer indicação sobre a qualidade do ajuste de cada modelo aos dados. Uma das medidas de qualidade de ajuste que, na prática, é mais facilmente apreendida é o número (ou a proporção) de vezes que o modelo acerta e/ou se engana. Essa informação é normalmente sintetizada na tabela da classificação de casos, que é fornecida pelos programas informáticos que processam os dados.

A título de ilustração, na figura 6.1 apresentam-se quatro tabelas de classificação de casos, seleccionadas de entre as 12 correspondentes aos modelos apresentados na tabela 6.1. Foram escolhidos, para cada tipo de modelo (vinhos com indicação de idade e vinhos sem indicação de idade), os provadores para os quais os modelos se ajustam pior e melhor, respectivamente.

Vinhos com Indicação de Idade

Provador L

	Previsto: 0	*Previsto: 1*	*Concordância [%]*
Decisão: 0	35	6	85.4%
Decisão: 1	4	275	98.6%

Provador I

	Previsto: 0	*Previsto: 1*	*Concordância [%]*
Decisão: 0	100	4	96.2%
Decisão: 1	1	369	99.7%

Vinhos sem Indicação de Idade

Provador G

	Previsto: 0	*Previsto: 1*	*Concordância [%]*
Decisão: 0	36	10	78.2%
Decisão: 1	14	289	95.4%

Provador I

	Previsto: 0	*Previsto: 1*	*Concordância [%]*
Decisão: 0	56	2	96.6%
Decisão: 1	1	413	99.8%

Figura 6.1 – Tabelas de classificação de casos

Nas tabelas da figura 6.1, pode observar-se o número de previsões correctas e incorrectas, quer para as decisões "0" quer para as decisões "1". Pode ainda verificar-se que a proporção de vezes que os modelos acertaram variou entre 78.2% e 99.8%. No entanto, na maior parte dos casos que não se representam aqui, o ajuste é superior a 93%. É interessante notar que o provador para o qual os modelos se ajustam melhor é o mesmo no caso dos vinhos com e sem indicação de idade (provador I).

A utilização da regressão logística permitiu obter modelos estruturados que possibilitam uma análise cuidada e fundamentada das características mais importantes e/ou decisivas para cada provador e das suas importâncias relativas, assim como a comparação dos processos de decisão dos provadores entre si.

Um dos problemas que esta abordagem coloca é o de reconhecer quando é que o modelo de regressão, ou os seus coeficientes, se alteram. Nos casos em que é plausível assumir que o modelo encontrado descreve de uma forma correcta o comportamento normal do provador, é possível, com base na proporção do número de vezes que o modelo erra $(1-p)$, controlar a estabilidade do processo de decisão ao longo do tempo. Na secção seguinte apresenta-se uma proposta para controlar essa estabilidade, exemplificada com alguns dos resultados obtidos para provadores do IVP.

6.3.2 Controlo da Estabilidade do Processo de Decisão dos Provadores

Ao admitir que o processo de decisão de um dado provador é passível de ser modelizado por uma determinada regressão logística, admite-se, implicitamente, que, enquanto o provador mantiver o mesmo processo de decisão, os coeficientes da regressão e as proporções do número de discordâncias (ou concordâncias) se mantêm.

Tomando, por exemplo, o provador I e os vinhos com indicação de idade (ver figura 6.1), espera-se que, para os vinhos provados após o estabelecimento do modelo, a discordância (concordância) entre o modelo e a realidade ocorra com uma frequência de cerca de 3.8% (96.2%) no caso dos vinhos "Não Aprovados" e de 0.3% (99.7%) no caso dos vinhos "Aprovados".

Para cada tipo de decisão ("0" ou "1"), o número de vezes que o modelo erra pode ser interpretado como uma variável aleatória binomial (as várias decisões são mutuamente independentes) com parâmetros (N_0, p_0) (decisões "0") e (N_1, p_1) (decisões "1"). Desta forma, é possível controlar a estabilidade do processo de decisão do provador, através de dois testes de hipóteses, cujas hipóteses nulas são, respectivamente:

$H_0: p_{na} = p_0$

onde:

p_{na}: proporção de discordâncias para os vinhos não aprovados;

p_0 : proporção de discordâncias para os vinhos não aprovados relativa ao modelo;

e $H_0: p_a = p_1$

onde:

p_a: proporção de discordâncias para os vinhos aprovados;

p_0 : proporção de discordâncias para os vinhos aprovados relativa ao modelo.

No caso de rejeição de alguma das hipóteses nulas, pode concluir-se que a respectiva proporção se alterou de uma forma estatisticamente significativa e, portanto, que o modelo já não representa o processo de decisão do provador.

Os referidos testes de hipóteses podem ser efectuados ao longo do tempo através da utilização de cartas de controlo. Quando N_i (N_0 ou N_1) é fixo, as cartas de controlo mais utilizadas são a carta p, a np e a CUSUM modificada (ver figura 4.4).

As cartas CUSUM detectam mais facilmente desvios pequenos em p do que as cartas p e np (ver, por exemplo, Harrison et al. (1986). Reynolds e Stoumbos (2000) referem mesmo que, surpreendentemente, obtiveram melhores resultados com a carta CUSUM do que com a p, mesmo para desvios grandes em p.

Há vários tipos de cartas CUSUM para detectar alterações na proporção binomial. Harrison et al. (1986) referem a carta tradicionalmente mais utilizada – a CUSUM binomial – que baseia na estatística que representa o número de ocorrências de um tipo de uma variável com distribuição binomial numa amostra de dimensão N. Referem, no entanto, que, para alguns valores dos parâmetros N e p, a variável em causa pode ser aproximada a uma de Poisson ou a uma Normal, devendo, nestes casos, ser utilizada uma carta CUSUM baseada nas respectivas estatísticas. Reynolds e Stoumbos (1999) aconselham a carta CUSUM de Bernoulli que é semelhante à Binomial, mas utiliza $N = 1$. Os mesmos autores (Reynolds e Stoumbos (2000)), ao compararem o desempenho das duas cartas, concluem que a de Bernoulli detecta melhor do que a binomial desvios grandes em p, sendo o desempenho semelhante para desvios pequenos e moderados.

Quando a proporção p é muito pequena, Sun e Zhang (2000) aconselham a utilização da carta CUSUM que se baseia na estatística "número de unidades conformes entre duas não-conformes" e que segue uma distribuição geométrica.

Neste trabalho optou-se por utilizar a carta CUSUM binomial. As razões desta escolha são essencialmente de natureza prática. Trata-se de uma carta mais utilizada do que as outras, cujos parâmetros se encontram tabelados e cujo procedimento é aconselhado pela British Standards Institution (ver Harrison et al. (1986)).

Note-se que, quando a proporção do número de vezes que o modelo erra (p_0 ou p_1) é demasiado elevada (por exemplo, superior a 20%) ou demasiado baixa (por exemplo inferior a 0.5%), não se recomenda a utilização de cartas de controlo para monitorizar a estabilidade do processo de decisão desses provadores. No primeiro caso, torna-se óbvio que o modelo não se ajusta adequadamente ao processo de decisão do provador, pelo

que se deve investigar porquê e tentar encontrar um outro modelo. No segundo caso, o ajustamento entre o modelo e o processo é próximo do perfeito, pelo que a ocorrência de um só caso de discordância justifica uma análise da situação, e, portanto, o uso de cartas de controlo torna-se desnecessário.

Quando se justifica a utilização da carta CUSUM, a ocorrência de um ponto fora de controlo é considerada uma forte evidência de que o provador não se encontra a utilizar o mesmo processo de decisão. A situação deve ser devidamente analisada com o objectivo de detectar as causas do desvio e, perante estas, deve tomar-se a decisão de manter o modelo inicial ou de recalcular os seus parâmetros por forma a que reflictam o novo processo de decisão.

Finalmente, e independentemente de ocorrerem sinais de alarme na carta, recomenda-se que os referidos parâmetros sejam recalculados periodicamente, no sentido de manter actual o modelo do processo de decisão que, naturalmente, sofrerá algumas alterações ao longo do tempo.

No capítulo 4, foram abordados alguns conceitos relacionados com as cartas de controlo de uma forma geral e, na secção 4.3.3, com a carta CUSUM em particular. No caso da Câmara de Provadores do IVP, utilizou-se a versão tabelada da carta CUSUM para proporções binomiais, cujo procedimento se encontra publicado em Harrison et al. (1986) e se apresenta resumidamente na figura 6.2. As tabelas referidas na figura (B1 e B2) encontram-se reproduzidas no Apêndice B e correspondem a um dos esquemas CUSUM apresentados pelos referidos autores. Optou-se por utilizar o esquema onde a alteração em p é detectada mais rapidamente (esquema C2), visto que o respectivo valor da taxa de falsos alarmes é, na prática, aceitável (α toma valores entre 0.4% e 0.5%).

O procedimento descrito na figura 6.2 foi definido para detectar tanto incrementos como decréscimos do valor de p. Neste procedimento a expressão "*Reset*" significa que o cálculo da soma acumulada correspondente à estatística CUSUM se deve reiniciar. Se se pretender detectar apenas uma das tendências, bastará calcular somente o valor de C correspondente (C^1 para a detecção de incrementos ou C^2 para detecção de decréscimos).

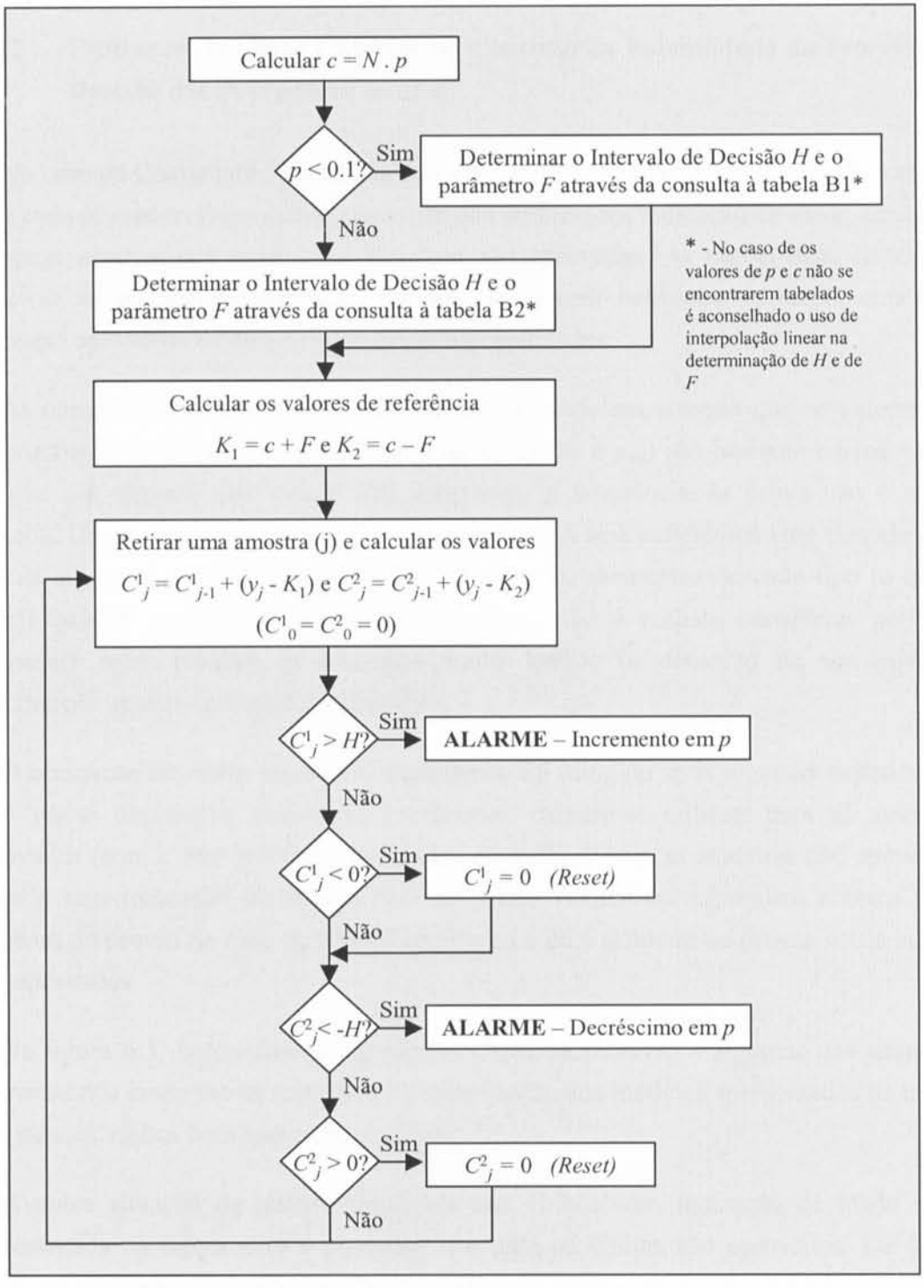

Figura 6.2 – Esquema do procedimento a seguir para construir tabelas CUSUM para a proporção binomial (adaptado de Harrison et al. (1996))

Na próxima secção, ilustra-se a utilização deste procedimento com alguns dados relativos aos provadores do IVP.

6.3.2.1 Utilização da Carta CUSUM no Controlo da Estabilidade do Processo de Decisão dos Provadores do IVP

No caso da Câmara de Provadores do IVP, foram utilizadas quatro cartas de controlo para cada provador. Duas delas referem-se aos vinhos com indicação de idade, uma para amostras aprovadas e outra para amostras não aprovadas. As outras duas cartas são relativas ao modelo determinado para os vinhos sem indicação de idade, uma para amostras aprovadas e outra para amostras não aprovadas.

As dimensões das amostras foram escolhidas tendo em atenção que os valores das proporções binomiais relativos às hipóteses nulas (p_a e p_{na}) são bastante baixos e que, no que diz respeito aos vinhos não aprovados, a frequência de prova não é muito elevada. Desta forma, se, por um lado, convém que N seja suficientemente elevado para que as amostras contenham um número razoável de elementos de cada tipo (o que é difícil quando p é muito pequeno), por outro, não é realista considerar períodos temporais entre recolha de amostras muito longos (a detecção de um possível descontrolo atrasar-se-ia demasiadamente).

O compromisso entre estes dois argumentos foi atingido após algumas experiências com várias dimensões amostrais. Finalmente, decidiu-se utilizar, para as amostras aprovadas (com e sem indicação de idade), $N_1 = 50$, e, para as amostras não aprovadas (com e sem indicação de idade), $N_0 = 25$. Estes valores correspondem a cerca de 2 semanas de provas no caso de vinhos aprovados e de 6 semanas de provas no de vinhos não aprovados.

Na figura 6.3, apresentam-se as tabelas CUSUM relativas a algumas das situações ocorridas nos cinco meses seguintes à determinação dos modelos apresentados na tabela 6.1, para os vinhos com indicação de idade.

A única situação de alarme verificada nos vinhos com indicação de idade foi a representada na figura para o provador L e para os vinhos não aprovados. De facto, pode observar-se que o número de discordâncias relativamente ao modelo aumentou bruscamente na amostra 3, o que foi imediatamente detectado pela tabela CUSUM. Este caso merece uma análise mais profunda, de forma a que possam ser tomadas medidas correctivas. No entanto, e visto que se trata de uma questão interna do IVP, não serão aqui abordadas as possíveis causas de sinais de "fora de controlo" e respectivas medidas correctivas.

Vinhos com Indicação de Idade - Aprovados

Provador G

$p = 0.03$ $N = 50$ $c = 1.5$ $H = 4.3$ $F = 1.2$ $K_1 = 2.9$ $K_2 = 0.5$

Nº. da amostra (j)	Nº. de discord. (y_j)	C^1_j	C^2_j
1	0	-2.91 (*Reset*)→ 0	-0,51
2	0	-2.91 (*Reset*)→ 0	-1,02
3	0	-2.91 (*Reset*)→ 0	-1,53
4	0	-2.91 (*Reset*)→ 0	-2,04
5	0	-2.91 (*Reset*)→ 0	-2,55
6	0	-2.91 (*Reset*)→ 0	-3,06

Provador H

$p = 0.016$ $N = 50$ $c = 0.8$ $H = 3.5$ $F = 0.7$ $K_1 = 1.48$ $K_2 = 0.08$

Nº. da amostra (j)	Nº. de discord. (y_j)	C^1_j	C^2_j
1	0	-1.48 (*Reset*)→ 0	-0.17
2	3	1.51	2.74 (*Reset*)→ 0
3	0	0.03	-0.08
4	1	-0.45 (*Reset*)→ 0	0.83 (*Reset*)→ 0
5	1	-0.48 (*Reset*)→ 0	0.91 (*Reset*)→ 0
6	0	-1.48 (*Reset*)→ 0	-0.08

Vinhos com Indicação de Idade – Não Aprovados

Provador G

$p = 0.067$ $N = 25$ $c = 1.65$ $H = 4$ $F = 1.4$ $K_1 = 3$ $K_2 = 0.41$

Nº. da amostra (j)	Nº. de discord. (y_j)	C^1_j	C^2_j
1	2	-1.05 (*Reset*)→ 0	1.5 (*Reset*) → 0
2	2	-1.05 (*Reset*)→ 0	1.75 (*Reset*)→ 0
3	3	-0.05 (*Reset*)→ 0	2.75 (*Reset*)→ 0

Provador L

$p = 0.15$ $N = 25$ $c = 3.75$ $H = 4.5$ $F = 1.75$ $K_1 = 5.4$ $K_2 = 1.9$

Nº. da amostra (j)	Nº. de discord. (y_j)	C^1_j	C^2_j
1	2	-3.41 (*Reset*)→ 0	0.09 (*Reset*) → 0
2	4	-1.41 (*Reset*)→ 0	2.09 (*Reset*) → 0
3	15	9.59 **ALARME**	13.09 (*Reset*)→ 0

Figura 6.3 – Exemplos de tabelas CUSUM para alguns provadores e para vinhos com indicação de idade

Ainda no que diz respeito às tabelas CUSUM apresentadas na figura 6.3, pode constatar-se que o provador G, para os vinhos aprovados, apresentou sempre valores nulos do número de discordâncias. A manter-se esta tendência, a carta assinalará um decréscimo no valor de p na amostra nº 9. Os casos do provador H para os vinhos aprovados e do provador G para os não aprovados são típicos de processos sob controlo estatístico, não evidenciando qualquer alteração no processo de decisão dos provadores.

Na figura 6.4, apresentam-se as tabelas CUSUM para algumas situações ocorridas no caso dos vinhos sem indicação de idade. O período de tempo analisado foi o já referido para os vinhos com indicação de idade (5 meses). As situações escolhidas foram todas aquelas cujas tabelas CUSUM produziram sinais de alarme.

Note-se que, no período de tempo em causa foram provados mais vinhos sem indicação de idade do que com indicação de idade. Mesmo assim, o maior número de não-aprovações registou-se no grupo dos vinhos com indicação de idade. Este facto comprova que há, realmente, mais problemas na aprovação dos vinhos com indicação de idade do que nos restantes.

Vinhos sem Indicação de Idade - Aprovados

Provador G

$p = 0.046$ $N = 50$ $c = 2.3$ $H = 5$ $F = 1.3$ $K_1 = 3.6$ $K_2 = 1$

Nº. da amostra (j)	Nº. de discord. (y_j)	C^1_j	C^2_j
1	3	-0.61 (*Reset*)→ 0	1.98 (Reset)→ 0
2	0	-3.61 (*Reset*)→ 0	-1.01
3	1	-2.61 (*Reset*)→ 0	-1.02
4	0	-3.61 (*Reset*)→ 0	-2.03
5	0	-3.61 (*Reset*)→ 0	-3,04
6	0	-3.61 (*Reset*)→ 0	-4,05
7	1	-2.61 (*Reset*)→ 0	-4,06
8	0	-3.61 (*Reset*)→ 0	-5.07 **ALARME**

Provador K

$p = 0.035$ $N = 50$ $c = 1.75$ $H = 4.4$ $F = 1.25$ $K_1 = 3$ $K_2 = 0.51$

Nº. da amostra (j)	Nº. de discord. (y_j)	C^1_j	C^2_j
1	0	-3.01 (*Reset*)→ 0	-0.51
2	0	-3.01 (*Reset*)→ 0	-1.03
3	0	-3.01 (*Reset*)→ 0	-1.54
4	0	-3.01 (*Reset*)→ 0	-2.06
5	0	-3.01 (*Reset*)→ 0	-2.57
6	0	-3.01 (*Reset*)→ 0	-3.09
7	0	-3.01 (*Reset*)→ 0	-3.6
8	0	-3.01 (*Reset*)→ 0	-4.12
9	0	-3.01 (*Reset*)→ 0	-4.63 **ALARME**

Vinhos sem Indicação de Idade – Não Aprovados

Provador G

$p = 0.217$ $N = 25$ $c = 5.43$ $H = 5.6$ $F = 1.9$ $K_1 = 7.33$ $K_2 = 3.53$

Nº. da amostra (j)	Nº. de discord. (y_j)	C^1_j	C^2_j
1	2	-5.33 (*Reset*)→ 0	-1.53
2	2	-5.33 (*Reset*)→ 0	-3.07
3	0	-7.33 (*Reset*)→ 0	-6.6 **ALARME**

Figura 6.4 – Exemplos de tabelas CUSUM para alguns provadores e para vinhos sem indicação de idade

Na figura 6.4, pode verificar-se que ambas as tabelas CUSUM relativas ao provador G emitiram um sinal de alarme correspondente a um decréscimo de p. Uma explicação possível reside no facto de se tratar do provador para o qual o modelo se encontra mais desajustado (o valor de p é superior a 20%). Por outro lado, e analisando pormenorizadamente as situações de discordância, constatou-se que o provador, na fase de calibração do modelo, aprovava frequentemente vinhos aos quais atribuía as classificações de 2 no Aroma e 3 no Corpo, tendo, a partir de uma certa altura, passado a não aprovar nenhum desses vinhos. Trata-se de um exemplo típico de alteração do processo de decisão, pelo que será um caso a analisar com o próprio provador e, caso se justifique, o seu modelo deverá ser refeito com base em novos dados de prova.

Com o provador K sucedeu um fenómeno semelhante, tendo-se verificado que durante o período de calibração este provador aprovava a maioria de vinhos com a classificação Aroma – 2, Sabor – 2 e Corpo – 3, tendo, posteriormente, passado a não os aprovar.

Embora o período de calibração do modelo tenha sido iniciado 3 meses após a introdução das novas escalas, estas alterações podem ser devidas a uma adaptação dos provadores a essas mesmas escalas, a qual poderia ainda não estar concluída.

O caso dos provadores do IVP foi aqui apresentado apenas numa perspectiva de exemplo de aplicação da metodologia proposta. Esta metodologia implica, no entanto, que algumas condições sejam satisfeitas. Para além daquelas que foram sendo referidas ao longo desta secção, realça-se o facto de, quando se trata de processos complexos, o modelo de regressão logística dever incluir todos os regressores que sejam importantes para a decisão final.

Por outro lado, verifica-se que a utilização das cartas de controlo nos processos de decisão exige dimensões amostrais consideráveis (visto que o valor de p é baixo), o que pode dilatar o intervalo de tempo entre observações. Assim, o controlo será tanto mais eficaz quanto maior for a frequência de prova.

6.4 Avaliação do Desempenho Diário dos Provadores

Os métodos propostos na secção anterior permitem explicitar o processo de decisão dos provadores e controlar a sua estabilidade. No entanto, não detectam directamente situações de fraca reprodutibilidade. Por outras palavras, o facto de o provador se encontrar sob controlo no que respeita ao seu processo de decisão não garante que ele seja capaz de reproduzir devidamente as suas apreciações.

Na secção 6.4.1, propõe-se uma metodologia para avaliar o desempenho dos provadores no que diz respeito à reprodutibilidade dos julgamentos, numa base diária. A ilustração da sua aplicação, feita na secção 6.4.2, baseia-se nas classificações atribuídas pelos provadores do IVP à característica Designação.

6.4.1 Metodologia Proposta para a Avaliação da Reprodutibilidade dos Provadores Inseridos num Painel

A reprodutibilidade dos provadores foi já definida no capítulo 3, onde também se apontaram algumas das suas condicionantes. A sua avaliação implica que a prova das amostras seja repetida uma ou mais vezes e que os resultados das várias provas das mesmas amostras sejam comparados nas características apreciadas.

A metodologia que aqui se propõe pode ser utilizada relativamente a uma ou várias características avaliadas durante as sessões de prova, desde que medidas numa escala adequada (no mínimo, escala de intervalo com 5 níveis). Pressupõe, igualmente, a existência de um painel de prova composto, no mínimo, por 4 elementos.

Seguidamente, descreve-se a metodologia para o caso de uma característica expressa numa escala com L níveis e avaliada simultaneamente por K provadores para S amostras diferentes.

6.4.1.1 Índice de Reprodutibilidade (*RI*)

Admita-se que uma determinada característica é avaliada pelos K provadores duas vezes para cada amostra em condições de reprodutibilidade (ver secção 3.4.2.1). Uma das formas de exprimir a reprodutibilidade do provador k ($k = 1, 2, ..., K$) relativamente à característica da amostra s ($s = 1, 2, ..., S$) é através do valor absoluto da diferença entre a pontuação que ele lhe atribuiu da primeira e da segunda vez que a provou, tal como indica a expressão:

$$R_{ks} = | C^1_{ks} - C^2_{ks} |,$$

onde:

R_{ks}: reprodutibilidade do provador k relativamente à amostra s;

C^1_{ks}: classificação atribuída pelo provador k à característica da amostra s na primeira vez que a provou;

C^2_{ks}: classificação atribuída pelo provador k à característica da amostra s na segunda vez que a provou.

A reprodutibilidade do provador k relativamente à característica em avaliação pode ser expressa pelo valor médio de R para as S amostras provadas, $\overline{R}_k$. Com base nesta grandeza, foi desenvolvido um índice a que se chamou índice de reprodutibilidade (RI_k), caracterizado por:

$$RI_k = 1 - \sqrt{\frac{\overline{R}_k}{E(R)}} \quad \text{se } \overline{R}_k \leq E(R)$$

$$RI_k = 0 \quad \text{se } \overline{R}_k > E(R)$$

onde:

RI_k: índice de reprodutibilidade;

$\overline{R}_k$: valor médio de R_k;

$E(R)$: valor esperado de R_k admitindo que as classificações (C^1_{ks} e C^2_{ks}) provêm de uma distribuição uniforme discreta com L níveis, o que equivale a considerar que o provador atribui as classificações às amostras de uma forma aleatória sem ter em conta a apreciação da característica em causa. Nestas condições, $E(R) = \frac{L-1}{2}$.

Este índice tem a vantagem de variar sempre entre 0 e 1, independentemente do número de níveis da escala utilizada. Toma o valor 0, quando o provador tem um desempenho de tal forma mau que classifica aleatoriamente as amostras ou atribui propositadamente classificações opostas às amostras e às suas réplicas. Assume o valor 1, quando o provador reproduz exactamente os seus julgamentos em relação a cada par de amostras do mesmo produto.

Foram experimentadas funções alternativas na construção deste índice, tanto com valores simulados como com valores reais. Por exemplo, testou-se a expressão $RI_k = 1 - \frac{\overline{R}_k}{E(R)}$, tendo-se verificado que os valores dela resultantes tinham pouco poder de discriminação. Uma outra hipótese ensaiada foi a de utilizar a média móvel amortecida de R em vez de $\overline{R}$. Verificou-se que, para factores de amortecimento (α) superiores a 0.1, o índice *RI* apresentava variações muito bruscas. Quando, pelo contrário, α assumia valores inferiores, o índice distinguia-se relativamente pouco do calculado com $\overline{R}$. Assim, tendo em conta que das experiências efectuadas não resultou nenhum benefício sensível, decidiu-se utilizar o índice *RI* já apresentado.

6.4.1.2 Índice de Estratégia Defensiva (*DI*)

A utilização de qualquer dos índices de reprodutibilidade que se baseiam apenas em $\overline{R}_k$ está sujeita a um problema: se um provador atribuir sempre a mesma classificação a todas as amostras que prova, independentemente do julgamento que faz da característica, a sua reprodutibilidade é considerada excelente. Por outras palavras, o facto de o índice *RI* de um dado provador ser próximo de 1 pode não ser um indício de uma boa reprodutibilidade, mas da adopção, por parte do provador em causa, de uma estratégia "defensiva", que se caracteriza pela utilização de poucos níveis da escala e se traduz numa dispersão baixa das classificações por ele atribuídas.

Outra estratégia "defensiva" possível é a utilização de níveis extremos da escala. Se o provador decidir que apenas atribui classificações "muito bom" e "muito mau", é-lhe

mais fácil reproduzi-las, na medida em as dúvidas apenas se lhe colocam entre 2 níveis, e não entre L níveis, como seria natural.

Para avaliar a extensão da utilização de tais estratégias, foi desenvolvido um outro índice, DI, definido por:

$$DI_k = 1 - \left| \frac{\sigma_{TP} - \sigma_k}{\sigma_{TP}} \right| \qquad \text{se } \sigma_k \leq 2 \cdot \sigma_{TP}$$

$$DI_k = 0 \qquad \text{se } \sigma_k > 2 \cdot \sigma_{TP}$$

onde:

DI_k: índice de estratégia defensiva relativa ao provador k;

σ_{PP}: desvio padrão referente ao painel de provadores;

σ_k: desvio padrão referente ao provador k.

O desvio padrão referente ao painel de provadores é o desvio padrão das classificações atribuídas por todos os elementos do painel a todas as amostras provadas durante o período de tempo em análise (os julgamentos de réplicas de amostras devem ser excluídos). O desvio padrão referente ao provador k é a mesma grandeza, considerando apenas os julgamentos do provador k. O período de tempo considerado deve ser suficientemente longo para que as classificações atribuídas à característica avaliada nas amostras provadas sejam variadas. De preferência, deverão cobrir a gama completa da escala.

Esta abordagem parte do princípio de que a dispersão dos valores das classificações atribuídas pelo conjunto de todos os provadores à característica em estudo é a correcta, ou seja, que é aquela que se obteria se fosse possível medir objectivamente a característica em causa. Assim, o índice DI_k assume o valor máximo, 1, quando o provador k utiliza a escala do mesmo modo que o painel de provadores, e assume o valor mínimo, 0, quando o provador utiliza a escala de formas muito diferentes do painel: ou sem dispersão ($\sigma_k = 0$) ou com uma dispersão igual ou superior ao dobro da do painel ($\sigma_k \geq 2 \cdot \sigma_{TP}$).

6.4.1.3 Índice de Fiabilidade

O índice proposto para avaliar a reprodutibilidade real dos provadores, que se denominou de índice de fiabilidade (I_k), consiste no produto de RI_k por DI_k:

$$I_k = RI_k \cdot DI_k$$

O índice I_k, tal como os seus componentes, varia entre 0 e 1. O valor de 0 ocorre quando o provador tem uma má reprodutibilidade ou usa uma estratégia defensiva muito intensa. O valor de $I_k = 1$ corresponde ao caso em que o provador atribui sempre a mesma pontuação aos pares de amostras provados e, simultaneamente, apresenta um valor de dispersão igual à do painel. A ocorrência simultânea destas condições é praticamente impossível, tal como ficou explícito no capítulo 2, quando se analisou o tema da percepção sensorial e das suas condicionantes, pelo que se deve ter em conta que o valor 1 é apenas um limite superior teórico que, em princípio, é inatingível.

Os cálculos inerentes à metodologia proposta são relativos a dados que são recolhidos ao longo do tempo. Note-se que no cálculo de RI_k se utiliza o valor médio dos módulos das amplitudes das diferenças entre as classificações atribuídas à primeira amostra provada e à sua réplica, e que no cálculo de DI_k se entra em consideração com desvios-padrão. No primeiro caso, é necessário que haja um número suficiente de repetições para que faça sentido considerar o seu valor médio, ao passo que no segundo se exige um número de provas tal que as classificações obtidas cubram a gama completa da escala (ou, pelo menos, uma gama considerável). O período de tempo a considerar (d dias), que deverá ser o mesmo para o cálculo dos dois índices, depende, portanto, de serem satisfeitas estas condições.

Admitindo que todos os dias se efectuam provas de réplicas de amostras, o índice I_k pode ser actualizado diariamente, tendo em conta os resultados obtidos nos últimos d dias.

6.4.2 Aplicação da Metodologia Proposta ao Caso do Painel de Provadores do IVP

No IVP aplicou-se a metodologia proposta à característica Designação que, como já foi referido na secção 6.2, é avaliada numa escala multinomial com 5 níveis e corresponde à qualidade global da amostra apreciada, tendo em conta o tipo de vinho que para ela é proposto.

O período de tempo considerado foi de 50 dias de trabalho e, portanto, tanto RI_k como DI_k foram calculados diariamente com os dados correspondentes ao dia em causa e aos 49 dias de trabalho que o precederam. O número de vinhos provados por cada provador no período em questão é de cerca de 500, sendo o número de réplicas de cerca de 50 (10% do número total de vinhos).

Seguidamente apresenta-se um conjunto de gráficos e tabelas com o objectivo de ilustrar a composição do índice I_k e, simultaneamente, clarificar alguns conceitos que poderão não ter ficado suficientemente explícitos na apresentação teórica do tema.

A figura 6.5 representa os histogramas de frequências relativas do módulo da diferença entre as classificações atribuídas à Designação das duas vezes que a mesma amostra é provada ($|C^1 - C^2|$), para quatro dos provadores, ao longo de 50 dias de trabalho. Os restantes provadores apresentam resultados semelhantes aos representados na figura.

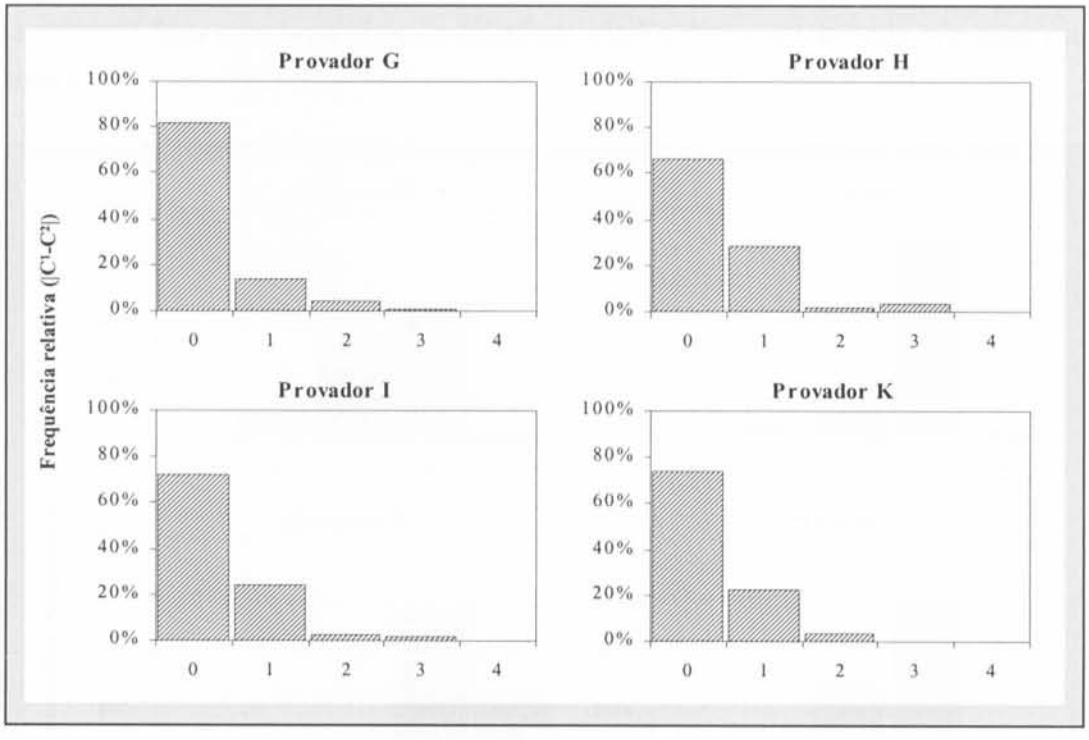

Figura 6.5 – Histogramas de frequências relativas do módulo das diferenças das classificações atribuídas às réplicas pelos provadores G, H, I e K

Ao analisar a figura 6.5, pode verificar-se que os quatro provadores replicaram bastante bem os julgamentos. A percentagem de julgamentos iguais encontra-se entre 65% e 82%. Se a estes valores se adicionar a proporção correspondente aos julgamentos que diferiram em apenas 1 classe, as percentagens passam a variar entre 94% e 97%.

Note-se que os próprios histogramas apresentados na figura 6.5 já fornecem alguma informação sobre o comportamento dos provadores nos dias considerados. Na realidade, os histogramas de $|C^1 - C^2|$ não são mais do que uma "fotografia" do que se passou entre as datas analisadas. Aconselha-se, assim, a actualização diária destes histogramas.

Paralelamente, e como decisão política, podem ser estabelecidas algumas regras relativamente às percentagens do número de provas admissíveis em cada um dos níveis: por exemplo, se, no nível 3 ou superior, o provador exceder os 5%, deve ser avisado e, se exceder os 7%, deverá ser excluído do júri. (O IVP implementou, nos seus procedimentos internos, regras deste género).

Na figura 6.6, apresentam-se outros histogramas referentes ao mesmo período de tempo considerado nos da figura 6.5: os histogramas das classificações atribuídas pelos mesmos provadores e pela Câmara no seu conjunto à característica Designação relativamente a todos os vinhos provados (exceptuando as réplicas). Registe-se que aqui é já possível ficar com uma ideia da diferença entre a dispersão de cada provador e a da Câmara.

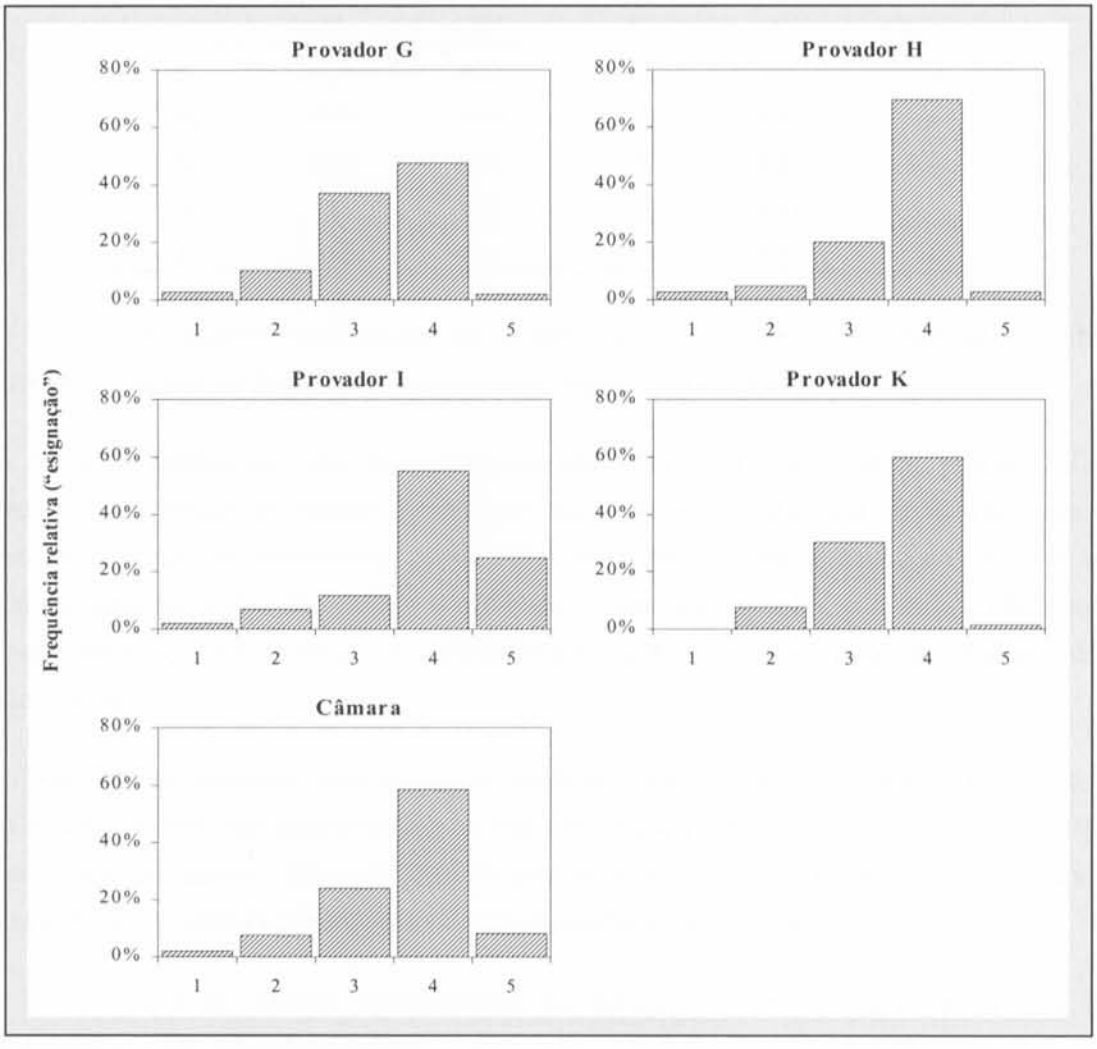

Figura 6.6 – Histogramas de frequências relativas das classificações atribuídas à Designação pelos provadores G, H, I e K e pela Câmara no seu conjunto

BIBLIOTECA F.E. PORTO

Através da análise da figura 6.6 verifica-se que, por exemplo, o provador K tem uma tendência para usar os níveis intermédios da escala maior do que os seus colegas, o que pode ser indicativo da adopção de uma estratégia mais "defensiva". Importa referir que este provador foi o único que não apresentou amplitudes com o valor 3 nos histogramas da figura 6.5. O provador I revela-se o mais benevolente relativamente às classificações, concentrando-as muito nos níveis 4 e 5 da escala.

Os valores dos desvios padrão correspondentes a esta figura encontram-se, tal como os valores médios $\overline{R}_k$, na tabela 6.2, que apresenta os valores das várias grandezas envolvidas no cálculo de I_k para o dia em análise.

Tabela 6.2 – Cálculo dos vários componentes de I_k para quatro dos provadores

	Provador				
	G	**H**	**I**	**K**	**Câmara**
$\overline{R}_k$	0.24	0.43	0.33	0.3	---
RI_k	0.65	0.54	0.59	0.62	---
σ_k	0.81	0.74	0.89	0.66	0.81
DI_k	1	0.92	0.90	0.81	---
$\boldsymbol{I_k}$	**0.65**	**0.49**	**0.53**	**0.50**	---

No dia em análise verificou-se que o provador G teve um comportamento (medido através do índice de fiabilidade) claramente superior aos restantes.

Além da análise do valor dos índices no dia a dia, interessa, também, perceber a sua evolução ao longo do tempo e verificar, na prática, se esta está de acordo com os fenómenos que se pretendem medir. Por exemplo, se há a sensação de que um determinado provador tem vindo a melhorar o seu desempenho, e o índice não evolui positivamente, deve analisar-se a situação para determinar as causas da disparidade e eliminá-las.

Nas figuras seguintes apresentam-se gráficos da evolução do índice *RI* (figura 6.7), do desvio padrão dos provadores e da Câmara (figura 6.8), do índice *DI* (figura 6.9) e, finalmente do índice *I* (figura 6.10). Os provadores escolhidos foram os quatro para os quais incidiu o estudo precedente e foram considerados 137 dias de trabalho.

Note-se que os valores apresentados na tabela 6.2 correspondem ao último dia representado nas referidas figuras.

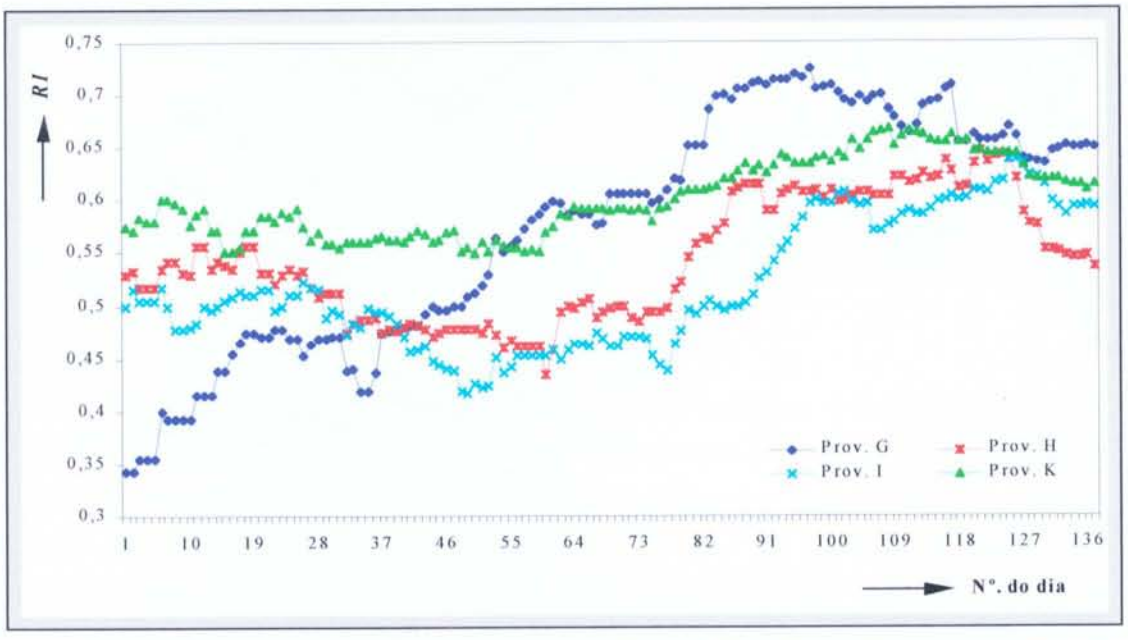

Figura 6.7 – Evolução do índice *RI* para os provadores G, H, I e K, ao longo de 137 dias

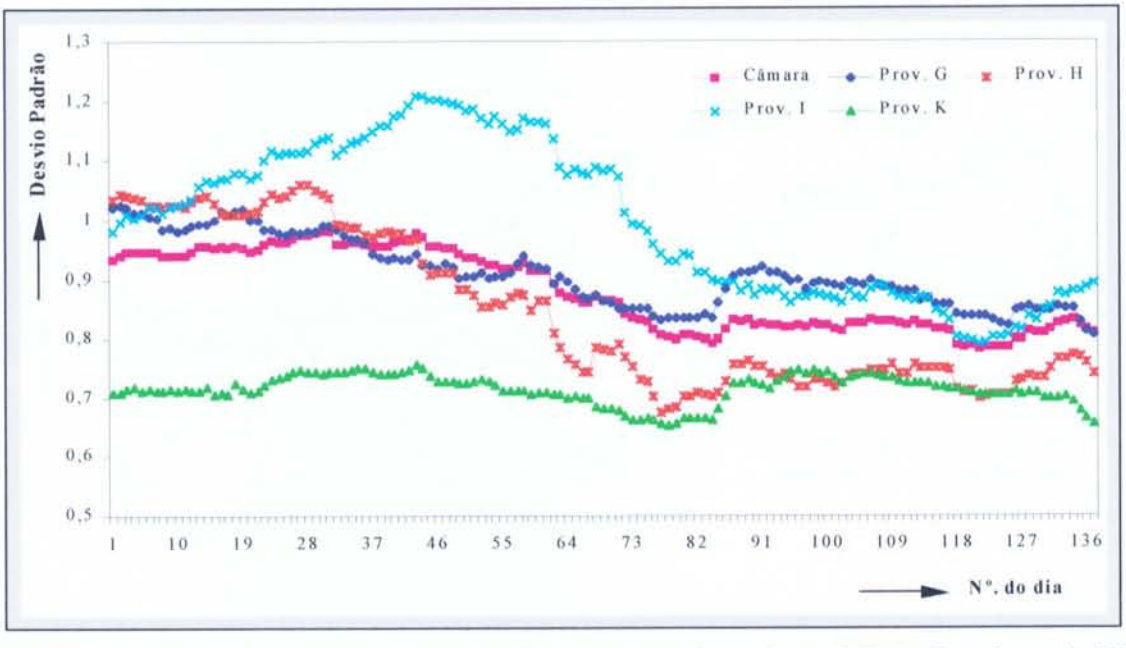

Figura 6.8 – Evolução do desvio padrão para a Câmara e para os provadores G, H, I e K, ao longo de 137 dias

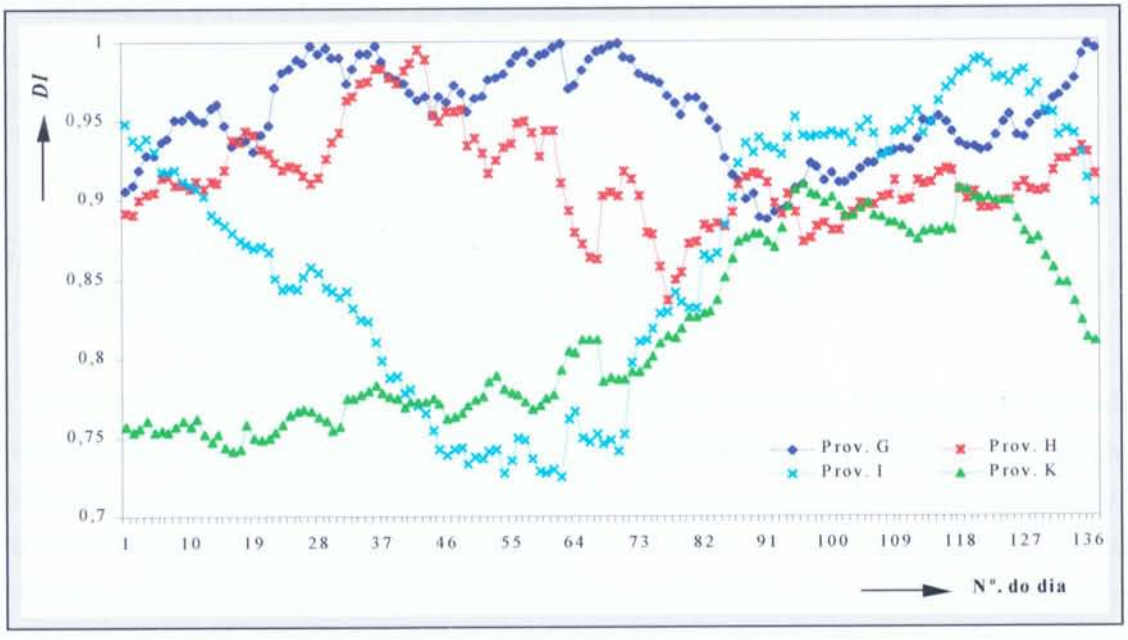

Figura 6.9 – Evolução do índice *DI* para os provadores G, H, I e K, ao longo de 137 dias

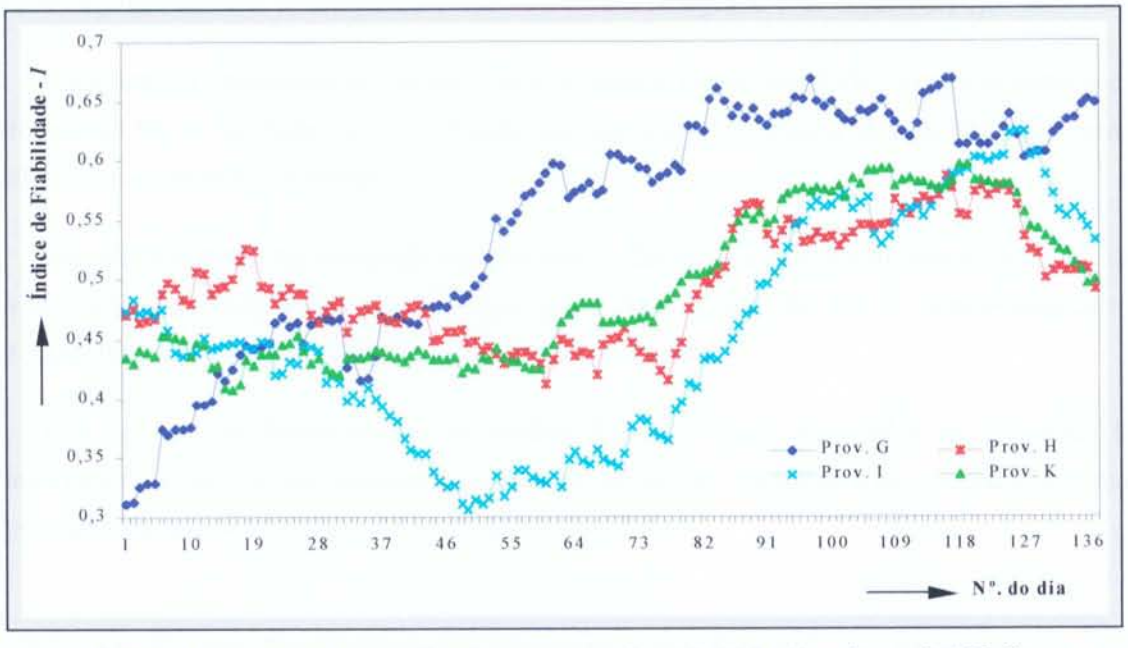

Figura 6.10 – Evolução do índice *I* para os provadores G, H, I e K, ao longo de 137 dias

Uma das conclusões gerais que se podem retirar da apreciação destes gráficos é a de que, na realidade, ao longo do tempo o desempenho dos provadores difere bastante e se vai alterando. Assim, por exemplo, o provador com pior desempenho no início do período é o que apresenta melhor desempenho no final.

Verifica-se, por outro lado, que há provadores mais constantes do que outros. Neste caso, o comportamento dos provadores H e K varia menos ao longo do período do que o dos outros dois provadores.

6.5 Avaliação Periódica do Desempenho Global dos Provadores

Os métodos propostos na secção anterior têm como objectivo a medição do grau de desempenho dos provadores em termos de reprodutibilidade e fiabilidade numa base diária. Quando se trata de avaliar e comparar o desempenho global dos provadores num período de tempo mais alargado (6 meses ou 1 ano), os mesmos métodos não se revelam muito úteis. De facto, por um lado, os índices são calculados para uma determinada característica, não sendo indicados para a medição do desempenho global, e, por outro, podem variar muito ao longo do período de tempo em análise, tornando difícil o julgamento e a comparação de uma forma integrada (ver figura 6.10).

Nesta secção propõe-se um método que se baseia na análise ROC e cujo objectivo é, precisamente, o de permitir a avaliação do desempenho global dos provadores num determinado período de tempo.

Este método pode ser utilizado sempre que as decisões cujos resultados se pretendem avaliar sejam do tipo binário e exige que, para além da decisão em si, os provadores se pronunciem sobre o grau de certeza com que a tomaram.

Uma vez que os fundamentos da análise ROC já foram abordados no Capítulo 3, nesta secção optou-se por apresentar o método proposto, à medida que se descreve a sua aplicação ao caso da Câmara de Provadores do IVP.

6.5.1 Avaliação Periódica do Desempenho Global do Provadores do IVP

O desempenho global dos provadores do IVP num determinado período de tempo será avaliado através de uma medida da sua sensibilidade à qualidade do vinho: o desempenho do provador será tanto melhor quanto mais sensível à qualidade ele se revelar. A análise ROC, tal como foi referido na secção 3.2.1.2, possibilita a obtenção de indicadores da sensibilidade independentes do critério usado pelo provador. Este

factor, aliado ao facto de não ser necessário alterar a rotina normal de prova, permitiu que esta técnica fosse ensaiada no contexto da Câmara de Provadores do IVP.

A variável binária considerada é a Decisão da Câmara (*FD*) – variável que exprime o resultado final do julgamento da Câmara como um todo e que é aqui entendida como sendo a "verdade" acerca da qualidade do vinho. Considera-se "vinho bom" o vinho que é aprovado pela Câmara e "vinho mau" o que não é aprovado. Ao "vinho bom" corresponde o "sinal + ruído", ao passo que ao "vinho mau" corresponde apenas o "ruído".

Note-se que a decisão final da Câmara é determinada pela maioria das decisões individuais dos provadores que a compõem (expressas pela variável Decisão, já considerada em 6.3).

O procedimento de utilizar a decisão da Câmara como a "verdade" sobre o vinho é discutível. Efectivamente, dado o tipo de produto e de medições em questão, não se afigura possível encontrar uma outra forma de verificar se o vinho tem, ou não, a qualidade para ser aprovado. No entanto, devem tomar-se algumas precauções no sentido de evitar erros nessa decisão: a Câmara deverá ser constituída pelo maior número possível de provadores (assume-se aqui que, no mínimo, devem ser 4) e estes deverão estar em condições físicas e psicológicas para exercer as suas funções de instrumentos de medida. Por outras palavras, não deverão ser considerados os provadores que se encontrem "fora de controlo" ou que tenham índices de fiabilidade inaceitáveis.

As percepções dos provadores sobre a qualidade global das amostras são expressas nas classificações que estes atribuem à variável Designação, numa escala de 1 a 5. Esta foi, aliás, a variável que serviu de base à exemplificação do método proposto em 6.4.

As referida escala pode ser interpretada como uma forma de controlar a variação do critério do provador, sendo, portanto, possível construir curvas ROC baseadas na informação simultânea da qualidade do vinho ("bom" ou "mau") e na classificação atribuída pelo provador à Designação numa escala de cotação (1 a 5).

De facto, as classificações atribuídas por cada provador às amostras dão uma indicação da sua sensibilidade à qualidade do vinho: se um provador atribui uma boa classificação a um "vinho mau" e uma má classificação a um "vinho bom", não revela uma boa sensibilidade à qualidade do vinho.

Contudo, e como foi já referido no Capítulo 3, a sensibilidade é independente do critério utilizado. Para ilustrar esta independência no contexto em análise, considerem-se dois provadores que têm a mesma sensibilidade e que atribuem as mesmas classificações à Designação de todas as amostras que provam. No entanto, exibem critérios diferentes: enquanto um deles usa um critério lato e assume que todas as amostras com classificação igual ou superior a "2" devem ser aprovadas, o outro, utilizando um critério restrito, assume que apenas as amostras com classificação de "4" ou "5" devem ser aprovadas. Embora a sensibilidade seja a mesma, os critérios de aprovação diferem.

Na figura 6.11, apresentam-se os dados que serviram de base à construção das curvas ROC para os quatro provadores já estudados na secção 6.4, para um período aproximado de um ano.

Provador G

Classificação	Vinhos "Maus"	Vinhos "Bons"
1	35	13
2	78	117
3	8	477
4	4	549
5	2	69
Total	127	1225

Provador H

Classificação	Vinhos "Maus"	Vinhos "Bons"
1	49	19
2	81	61
3	22	352
4	10	991
5	1	46
Total	163	1469

Provador I

Classificação	Vinhos "Maus"	Vinhos "Bons"
1	78	16
2	67	42
3	12	221
4	5	830
5	2	243
Total	164	1352

Provador K

Classificação	Vinhos "Maus"	Vinhos "Bons"
1	2	0
2	90	79
3	45	470
4	11	717
5	1	34
Total	149	1300

Figura 6.11 – Classificações atribuídas por cada provador às diversas amostras de vinhos "maus" e "bons" provadas ao longo de um ano

A partir da informação contida nas tabelas da figura 6.11, elaboraram-se as tabelas apresentadas na figura 6.12, de onde é possível construir directamente as curvas ROC.

Estas tabelas apresentam os valores das proporções de falsos positivos (p_{FP} – proporção de vinhos "maus" considerados "bons") e de verdadeiros positivos (p_{TP} – proporção de vinhos "bons" considerados "bons") que cada provador apresentou no período de tempo em estudo para cada critério considerado, que não são mais do que as coordenadas (x,y) dos vários pontos das curvas ROC. Note-se que se trata de critérios

"artificiais": a partir das classificações atribuídas pelos provadores às amostras criam-se os critérios que vão do mais restrito ao mais lato.

Provador G

Critério	p_{FP}	p_{TP}
Aprovado - 5 Não Aprovado - 1,2,3,4	0.016	0.056
Aprovado - 4,5 Não Aprovado - 1,2,3	0.047	0.504
Aprovado - 3,4,5 Não Aprovado - 1,2	0.110	0.894
Aprovado - 2,3,4,5 Não Aprovado - 1	0.724	0.989

Provador H

Critério	p_{FP}	p_{TP}
Aprovado - 5 Não Aprovado - 1,2,3,4	0.006	0.031
Aprovado - 4,5 Não Aprovado - 1,2,3	0.067	0.706
Aprovado - 3,4,5 Não Aprovado - 1,2	0.202	0.946
Aprovado - 2,3,4,5 Não Aprovado - 1	0.699	0.987

Provador I

Critério	p_{FP}	p_{TP}
Aprovado - 5 Não Aprovado - 1,2,3,4	0.012	0.180
Aprovado - 4,5 Não Aprovado - 1,2,3	0.042	0.794
Aprovado - 3,4,5 Não Aprovado - 1,2	0.116	0.957
Aprovado - 2,3,4,5 Não Aprovado - 1	0.524	0.988

Provador K

Critério	p_{FP}	p_{TP}
Aprovado - 5 Não Aprovado - 1,2,3,4	0.007	0.026
Aprovado - 4,5 Não Aprovado - 1,2,3	0.081	0.578
Aprovado - 3,4,5 Não Aprovado - 1,2	0.383	0.939
Aprovado - 2,3,4,5 Não Aprovado - 1	0.987	1

Figura 6.12 – Coordenadas dos pontos das curvas ROC calculadas a partir das tabelas da figura 6.11

Os critérios utilizados nas tabelas apresentadas na figura 6.12 são 4. O mais restrito é o que corresponde à primeira linha das tabelas, onde se considera que o provador aprova apenas as amostras que considera muito boas (com classificação de "5"), não aprovando nenhuma das outras (com classificações de "1", "2", "3" ou "4"). O critério mais lato é o oposto, que está representado na última linha e onde se considera que o provador aprova as amostras a que atribui as classificações de "2", "3", "4" ou "5" e apenas não aprova aquelas que classifica como muito más ("1"). Entre estes dois critérios surgem outros dois, mais moderados, os quais correspondem às situações intermédias.

Para cada critério, as proporções p_{FP} e p_{TP} calculam-se da seguinte forma:

$$p_{FP} = \frac{\text{número de vinhos "maus" aprovados pelo provador}}{\text{número total de vinhos "maus"}} \quad \text{e}$$

$$p_{TP} = \frac{\text{número de vinhos "bons" aprovados pelo provador}}{\text{número total de vinhos "bons"}}$$

Note-se que o provador é tanto mais sensível quanto mais p_{FP} se aproxima do valor 0 e, simultaneamente, p_{TP} se aproxima do valor 1.

Na figura 6.13, apresentam-se as curvas ROC construídas a partir dos dados apresentados na figura 6.12 e com os pontos unidos por segmentos de recta. Esta abordagem, tal como foi referido no Capítulo 3, é conservadora, na medida em que conduz sempre a áreas (e, portanto, a medidas da sensibilidade) ligeiramente menores do que as que se obteriam ao aproximar as ROCs por linhas curvas.

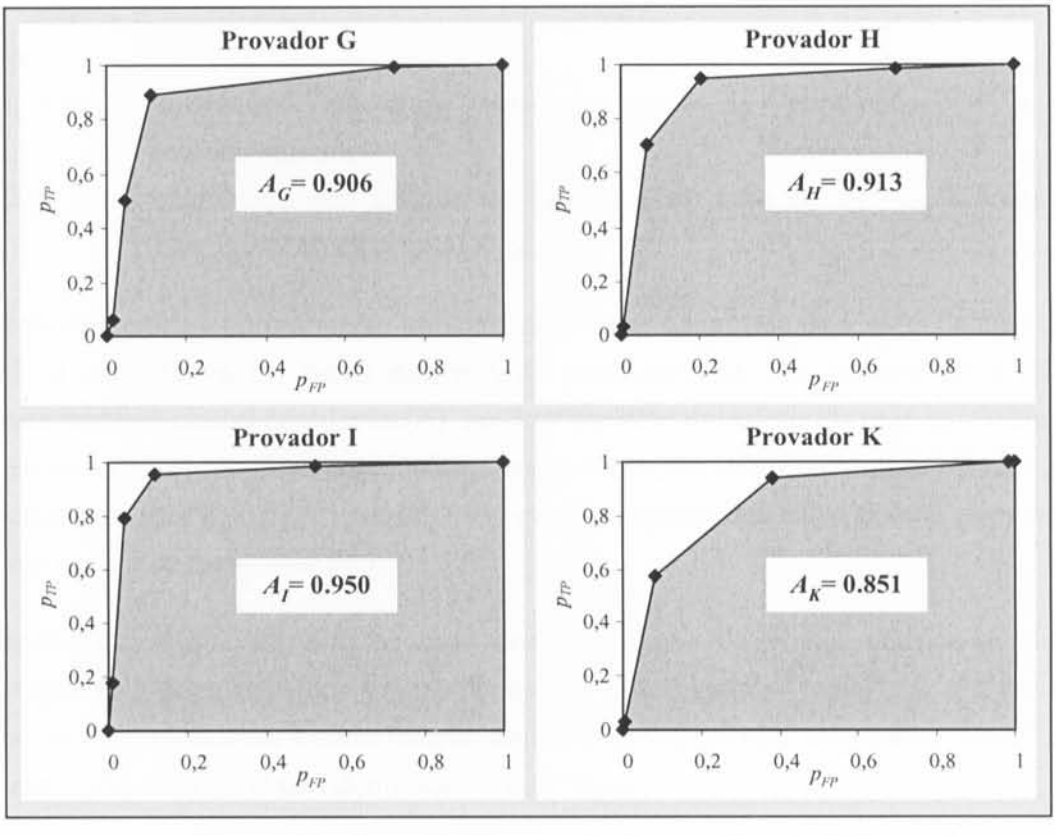

Figura 6.13 – Curvas ROC para quatro dos provadores do IVP

Na figura 6.13, além das curvas ROC propriamente ditas, apresentam-se, também, os valores das áreas debaixo de cada curva. O significado dessa área neste caso particular é uma estimativa da probabilidade de o provador identificar correctamente o vinho "bom" quando este é provado em conjunto com um vinho "mau" num teste de comparação por pares.

Verifica-se que todas as curvas se situam bastante acima da diagonal [(0,0) (1,1)] e que, consequentemente, as correspondentes áreas são bastante superiores a 0.5, o que, em princípio, significa que os provadores se encontram a discriminar bem entre os vinhos "maus" e os vinhos "bons".

Na secção 3.2.1.2, foram apresentadas várias medidas de sensibilidade baseadas nas curvas ROC, tendo-se verificado que muitas delas se referiam à área debaixo da curva. Uma outra abordagem, que conduz igualmente à referida área, e que não assume qualquer distribuição estatística para as classificações atribuídas às amostras "boas" ou "más", é a da estatística de Wilcoxon (W). Esta estatística é referida em Hanley e McNeil (1982) e é dada pela seguinte expressão:

$$W = \frac{1}{n_A \cdot n_N} \sum_{1}^{n_A} \sum_{1}^{n_N} S(x_A, x_N)$$

onde:

n_A e n_N: dimensões amostrais correspondentes às populações A e N, respectivamente;

$S(x_A, x_N)$: quantidade que assume os valores 1 no caso de $x_A > x_N$, 0.5 quando $x_A = x_N$ e 0 no caso de $x_A < x_N$.

Esta estatística é normalmente utilizada para testar a hipótese de o nível de uma dada variável quantitativa (x) numa determinada população (A) ser superior ao nível da mesma variável noutra população (N), sendo a hipótese nula definida pela igualdade dos níveis da variável nas duas populações. Nestas condições, a probabilidade (θ) de x_A ser superior a x_N será de 0.5. A variável x discriminará tanto mais entre as duas populações quanto mais θ se aproximar de 1.

Aplicando o que foi dito ao caso dos provadores de vinho, verifica-se que A corresponde à população dos vinhos "bons", N à dos vinhos "maus", x_A é a variável aleatória correspondente à classificação atribuída pelo provador aos vinhos "bons" e x_N a mesma grandeza correspondente aos vinhos "maus".

No mesmo artigo (Hanley e McNeil (1982)), os autores referem que W é uma estimativa da área debaixo da curva ROC (θ) e que, quando θ excede 0.5, ou seja, quando a característica x é, na realidade, discriminatória, W passa a ter uma distribuição estatística conhecida cujo desvio padrão se pode determinar. O cálculo do desvio padrão de W (S_W) é feito através da expressão:

$$S_W = \sqrt{\frac{\theta(1-\theta) + (n_A - 1) \cdot (Q_1 - \theta^2) + (n_N - 1) \cdot (Q_2 - \theta^2)}{n_A \cdot n_N}}$$

onde:

θ, n_A e n_N : têm o significado já referido anteriormente;

Q_1: probabilidade de dois elementos de N (retirados ao acaso) serem classificados com um nível inferior a um elemento de A (igualmente retirado ao acaso);

Q_2: probabilidade de um elemento retirado ao acaso de N ser classificado com um nível inferior a dois elementos retirados ao acaso de A.

Assume-se aqui que, em média, os elementos de A possuem um nível da característica em causa superior aos elementos de N.

Q_1 e Q_2 são funções bastante complexas das distribuições de x_A e x_N. No entanto, no artigo já citado, os autores concluem que estas quantidades variam essencialmente com o valor de W e que, nos casos em que a ordem de grandeza das áreas é superior a 80%, as distribuições exponenciais negativas conduzem a valores de S_W ligeiramente mais conservadores (ou seja, superiores) do que as outras distribuições testadas (gaussiana e gamma). Assumindo então que as referidas distribuições se podem considerar exponenciais negativas, os valores de Q_1 e Q_2 são dados pelas seguintes expressões:

$$Q_1 = \frac{\theta}{2-\theta} \text{ e } Q_2 = \frac{2\cdot\theta^2}{1+\theta}$$

Na tabela 6.3, apresentam-se os valores de W para cada um dos quatro provadores considerados (que, tal como foi referido em 3.2.1.2, coincidem com os valores das áreas assinaladas na figura 6.12), os respectivos valores de S_W e os intervalos de confiança a 95% para W. No cálculo destes intervalos, assumiu-se que a proporções W seguiam distribuições aproximadamente normais, dado que para todos os casos apresentados se verificou que $n \cdot W > 7$ (sendo n o número de casos de vinhos "maus" provados pelo provador).

Tabela 6.3 – Estatística de Wilcoxon e respectivo desvio padrão calculados para 4 dos provadores do IVP

	Provador			
	G	**H**	**I**	**K**
W	0.906	0.913	0.950	0.851
S_W	0.010	0.008	0.006	0.013
IC_W *(95%)*	[0.887,0.924]	[0.897,0.929]	[0.939,0.961]	[0.826,0.875]

Analisando os resultados apresentados na tabela 6.3, verifica-se que os valores médios de W se encontram entre 0.851 (provador K) e 0.95 (provador I), sendo, portanto, todos bastante superiores a 0.5. Ao analisar os intervalos de confiança, verifica-se que os valores de W são todos estatisticamente diferentes de 0.5 (pois

nenhum dos intervalos inclui este valor), sendo que o valor mínimo do intervalo mais próximo de 0.5 se encontra ainda muito distante deste valor. Pode, desta forma afirmar-se com mais segurança que os provadores não provam ao acaso e que possuem sensibilidades bastante altas à qualidade do vinho (valores acima de 0.8 são, normalmente, considerados bons).

Para detectar diferenças significativas entre os provadores, é possível elaborar testes de hipóteses à diferença entre as áreas debaixo das curvas. Assumindo, mais uma vez, que se trata de amostras de grandes dimensões e de distribuições aproximadamente simétricas, pode utilizar-se o teste Z à diferença de proporções binomiais, sendo a hipótese nula a igualdade das áreas e a estatística de teste dada por:

$$z = \frac{W_1 - W_2}{\sqrt{{S_{W_1}}^2 + {S_{W_2}}^2}}$$

Hanley e McNeil (1983) argumentam que, quando as áreas são obtidas a partir da leitura dos mesmos casos (na situação, prova dos mesmos vinhos), o desvio padrão que figura no denominador deve ser corrigido devido à correlação existente entre os julgamentos dos vários provadores, utilizando-se então a seguinte estatística de teste:

$$z = \frac{W_1 - W_2}{\sqrt{{S_{W_1}}^2 + {S_{W_2}}^2 - 2 \cdot r \cdot S_{W_1} \cdot S_{W_2}}}$$

Sucede que, no caso do IVP, e, tal como foi referido no Capítulo 5, em cada semana há um provador que está encarregue da organização da prova e que, portanto, não prova. Como consequência, ao estudar o comportamento dos provadores dois a dois, não se pode considerar que eles tenham provado sempre os mesmos vinhos: em média, cerca de 2/3 das amostras serão referentes aos mesmos vinhos, sendo 1/3 de vinhos diferentes.

No entanto, e analisando as duas fórmulas anteriores, conclui-se que a primeira conduz a um valor de z inferior, já que não se prevê a existência de correlações negativas entre as classificações que os provadores atribuem aos vinhos. Nestas condições torna-se mais difícil rejeitar a hipótese nula no caso da utilização da primeira fórmula. A sua utilização numa situação como a descrita é, mais uma vez, conservadora relativamente à utilização da segunda expressão.

Optou-se, assim, por elaborar os testes de hipóteses utilizando a primeira expressão. Uma outra alternativa teria sido considerar, para cada par de provadores, apenas as amostras que tinham provado em comum. Esta hipótese não foi considerada pelo facto

de diminuir substancialmente o número de amostras provadas, conduzindo, em alguns casos, a distribuições pouco simétricas ($N \cdot p < 7$).

Na tabela 6.4, apresentam-se os valores da estatística de teste (z) obtidos para os vários testes de hipóteses à diferença entre as áreas debaixo das curvas ROC.

Tabela 6.4 – Valores da estatística de teste (z) para os testes à diferença entre as áreas debaixo das curvas ROC (W) considerando os vários pares de provadores

	Par de Provadores					
	G-H	**G-I**	**G-K**	**H-I**	**H-K**	**I-K**
z	0.580	4.015	3.512	3.791	4.202	7.283

Analisando os valores apresentados na tabela e considerando que, para um teste bilateral ao nível de significância de 5%, $z = 1.96$, conclui-se pela rejeição de todas as hipóteses nulas, excepto da relativa aos provadores G e H. Admite-se, assim, que, com os dados disponíveis, não foi possível encontrar diferenças estatisticamente significativas entre a sensibilidade à qualidade do vinho do provador G e a do provador H.

Finalmente, e ainda relativamente à avaliação periódica do desempenho global dos provadores, decidiu-se verificar a compatibilidade entre o valor médio do índice I calculado para a qualidade global do vinho e os valores de W obtidos para os mesmos provadores e para o mesmo período de tempo. Considerados os 6 provadores do painel, verifica-se que, no geral, há uma razoável correspondência entre os valores de W e de $\bar{I}$. Ao considerar apenas os quatro provadores referidos ao longo deste capítulo, ressalta bastante a diferença existente no provador I que, sendo aquele a que corresponde o valor mais elevado de W, é, simultaneamente, aquele a que corresponde o valor mais baixo de $\bar{I}$. Os valores de W e de $\bar{I}$ para os 4 provadores em causa são apresentados na tabela 6.5.

Tabela 6.5 – Valores de W e de $\bar{I}$ para os 4 provadores considerados

	Provador			
	G	**H**	**I**	**K**
W	0.906	0.913	0.950	0.851
$\bar{I}$	0.549	0.496	0.455	0.492

A principal razão que se supõe estar na base das discrepâncias encontradas é a de que, na realidade, estes dois índices não medem exactamente a mesma grandeza. Enquanto W mede o grau de sensibilidade do provador à qualidade do vinho, $\bar{I}$ mede a capacidade do provador de reproduzir o seu julgamento quando confrontado com

amostras repetidas (embora levando em consideração a forma como este utiliza a escala). De facto, o provador pode ser bastante sensível, mas verificar-se que não reproduz os seus julgamentos tão fielmente como outros provadores menos sensíveis (a sua falta de reprodutibilidade pode dever-se a outras razões que não a falta de sensibilidade). É o que, aparentemente, acontece com o provador I.

Por outro lado, o índice *I* foi definido de uma forma empírica, tendo sido realçado, na secção 6.4, que poderiam ter sido utilizadas outras expressões no seu cálculo. Admite-se, então, a possibilidade de os resultados de outras expressões traduzirem de uma forma mais adequada o desempenho dos provadores em termos de reprodutibilidade.

Quando o objectivo é avaliar o desempenho global dos provadores numa base periódica, devem ter-se em conta as limitações apontadas, pelo que se aconselha a utilização dos dois indicadores referidos (W e $\bar{I}$), de forma a considerar, nessa avaliação, o máximo de informação disponível.

6.6 Aferição Periódica de Capacidades Específicas dos Provadores

No Capítulo 3, foram referidos alguns aspectos a ter em conta quando se pretende seleccionar e treinar um painel de provadores profissional. Um dos procedimentos mais comuns é o de aferir as capacidades dos provadores com amostras padrão. No entanto, tal só é praticável em algumas situações, nomeadamente quando é possível definir tais padrões.

Para os sabores elementares, por exemplo, encontram-se definidas na literatura as substâncias e as respectivas quantidades que se devem utilizar na preparação de soluções padrão (ver Hough et al. (1995)). O mesmo se passa relativamente a outros aromas e odores de natureza mais ou menos complexa, como os aromas de baunilha, cacau ou limão (ver Zamora e Calvino (1996)). Um exemplo de selecção de provadores para um painel com base neste tipo de testes pode encontrar-se em González-Viñas et al. (1996), que seleccionaram provadores para um painel de prova de vinagre de vinho com base no resultados dos limiares de detecção que os candidatos exibiram para várias soluções com diferentes concentrações de ácido acético e de ácido cítrico.

A complexidade de muitos produtos, de entre os quais o vinho, não permite a definição de padrões. No caso do vinho do Porto em particular, a definição de padrões é

virtualmente impossível. De facto, a variedade de tipos de vinhos do Porto existente, aliada à variabilidade inerente a cada tipo, que se deve a factores como o estilo próprio de cada produtor ou as condições climatéricas ao longo do ano de produção da uva, dificultam extraordinariamente a definição de um padrão único de vinho do Porto para cada tipo.

No entanto, os produtos referidos possuem muitas vezes características particulares que podem e devem ser identificadas e para cuja detecção os provadores deverão estar preparados. Embora tais características sejam sensoriais, pode dizer-se que derivam de um conjunto de compostos químicos. Nessa medida, a busca de qualquer tipo de padrão passa, necessariamente, pela identificação dos compostos químicos que dão origem aos aromas e aos sabores dos produtos em questão.

Pode verificar-se que na área da Química Alimentar há muitos estudos que versam esta temática. Por exemplo, Kallithraka et al.(1997) analisam o efeito dos ácidos láctico e málico na percepção que os provadores têm da adstringência, do amargor e da acidez de vinhos tintos. Kallithraka e Bakker (1997) avaliam os efeitos de outras duas substâncias químicas nas mesmas características sensoriais do vinho tinto e de uma solução modelo com uma composição química similar à de um vinho de mesa seco.

Benn e Peppard (1996) identificaram mais de 175 compostos químicos em amostras de tequila. Porém, quando pretenderam reconstituir a tequila a partir dos componentes identificados, não foram bem sucedidos, o que indica que a constituição desta bebida inclui outros compostos que não foram identificados no estudo.

Steinhaus e Schierberle (2000) identificaram algumas substâncias de aroma num fermento utilizado na indústria da cerveja, mas salientam que, para confirmar os resultados sobre o impacto dos compostos identificados no aroma do fermento, são necessários estudos posteriores, nomeadamente experiências de recombinação dos referidos compostos.

No que diz respeito a estudos específicos sobre a identificação de compostos responsáveis por aromas dos vinhos, salientam-se: Freitas et al. (1999), onde são identificados alguns dos compostos responsáveis pelo aroma floral dos vinhos do Porto tipo *Ruby*; Kotseridis e Baumes (2000), que identificam algumas substâncias com impacto no aroma dos vinhos de Bordeaux; Marchand et al. (2000), que identificam a presença de algumas moléculas no vinho pela primeira vez. Neste último artigo, os autores salientam a influência que a matriz base tem no limiar de percepção dos

provadores, tendo sido estudados 3 meios: um sintético, um vinho branco e um vinho tinto.

Relativamente ao vinho do Porto em particular, têm sido conduzidos alguns estudos sobre os seus compostos voláteis. Em Rogerson e Freitas (2001a), Rogerson e Freitas (2001b) e Rogerson et al. (2001), podem encontrar-se os resultados de alguns desses estudos.

Seguidamente, apresenta-se um estudo piloto realizado em colaboração com o Instituto do Vinho do Porto e com o Departamento de Química da Universidade de Aveiro, no qual se utilizam os resultados das análises químicas e espectrofotométricas feitas aos vinhos no IVP e, simultaneamente, os resultados de análises feitas ao aroma dos vinhos por um "nariz electrónico". O objectivo deste estudo foi o de fazer uma primeira abordagem a um método que permitisse a caracterização e a distinção entre alguns tipos de vinho do Porto.

6.6.1 Caracterização e Distinção entre Alguns Tipos de Vinho do Porto

Do que já foi dito ao longo deste trabalho, pode depreender-se que uma das dificuldades que surgem ao tentar calibrar a Câmara de Provadores do IVP é a ausência de fronteiras definidas entre os diversos tipos de vinho do Porto. Particularmente no que diz respeito à idade do vinho, este é um problema que se coloca frequentemente. De um ponto de vista sensorial, é fácil distinguir um vinho de 3 Anos de um de 10, mas já é mais complicado distinguir um de 9 de um de 11 Anos, o que torna, muitas vezes, controversa a decisão de aprovar ou reprovar um vinho por causa da sua idade.

Com este estudo pretendeu-se dar uma contribuição no sentido de ensaiar um método que possa, futuramente, ser utilizado para ajudar a definir mais claramente a fronteira entre os diversos tipos de vinho. Foram analisadas 42 amostras, tendo sido escolhidos os tipos de vinho do Porto mais comuns: *Tawny* novos (8), *Tawny* com indicação de idade - 10 Anos (9) e 20 Anos (7), *Ruby* (8) e *Vintage* novos (com aproximadamente 3 anos) (10). As amostras foram seleccionadas pelo Serviço de Prova do IVP, tendo sido aprovadas, tanto no que diz respeito às análises químicas como às análises sensoriais (feitas pelos provadores da Câmara).

6.6.1.1 Estudo das Análises Químicas e de Cor

Numa primeira fase, foram processados os resultados das análises químicas e de cor feitas às amostras. As análises químicas consideradas foram: *massa volúmica*, *etanol*, *grau Baumé*, *pH*, *acidez volátil*, *acidez fixa*, *acidez total*, *SO_2 total*, *extracto seco*, *açúcares redutores*, *extracto não redutor*, *5-hidroximetilfurfural* (*5-HMF*), *etanal*, *acetato de etilo*, *metanol*, *2-butanol*, *1-propanol*, *isobutanol*, *álcool alílico*, *1-butanol*, *álcoois amílicos*, *álcoois superiores totais*, *relação 1-propanol/isobutanol* e *relação álcoois amílicos/isobutanol*. As análises espectrofotométricas foram feitas no sistema CIELAB e fornecem três coordenadas de cor: *L*, *a* e *b*.

Os resultados destas análises para todas as amostras foram sujeitos a uma análise de componentes principais (ACP). A ACP é uma das técnicas mais utilizadas quando se pretende efectuar uma análise factorial (ver capítulo 3). A ideia base da ACP é, portanto, a de descrever a variação presente num conjunto de dados multivariados em termos de um conjunto de variáveis independentes entre si (componentes). Os componentes são extraídos por ordem decrescente de importância (ou de variância explicada). Se, entretanto, se verificar que os primeiros 2 ou 3 componentes são responsáveis por uma grande parte da variância presente nos dados originais, podem ser utilizados para sintetizar esses dados sem grandes perdas de informação, e com a vantagem de reduzir o número de dimensões, facilitando a sua compreensão (para mais pormenores sobre a ACP, ver, por exemplo, Everitt e Dunn (1991) ou Hair et al. (1998)).

Na figura 6.14, apresentam-se, de uma forma gráfica, os valores obtidos para as pontuações que cada amostra obteve nos dois componentes principais resultantes da análise feita aos resultados das análises químicas e de cor já referidas (ACP_1). O primeiro componente, representado no eixo dos *XX*, explica 49.2% da variância dos dados originais, enquanto o segundo, representado no eixo dos *YY*, explica 18.5% da mesma variância.

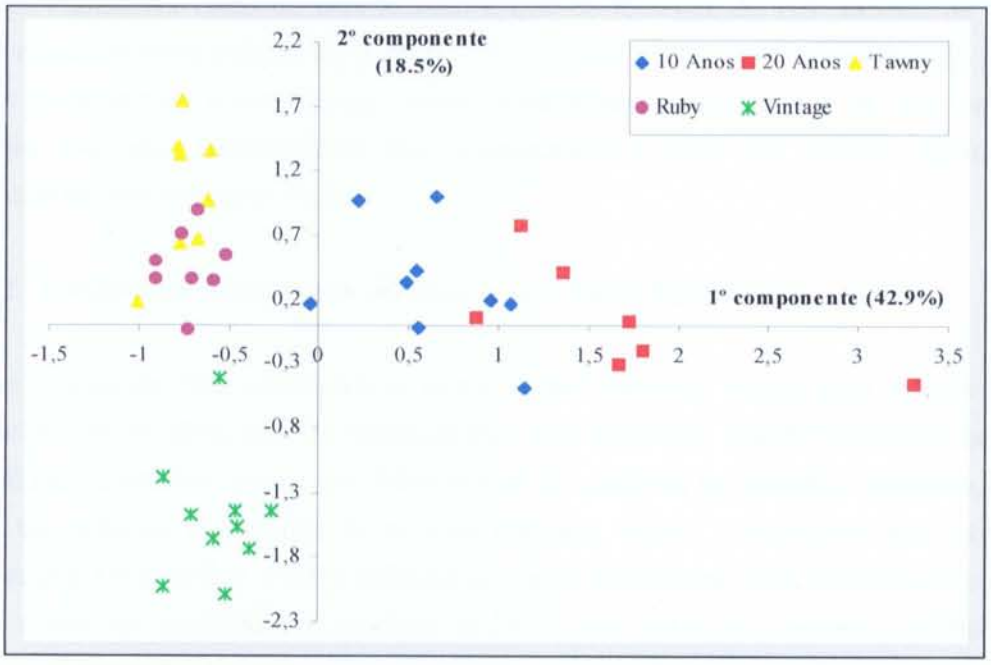

Figura 6.14 – Representação gráfica das pontuações obtidas por cada amostra no 1° e 2° componentes principais da ACP_1

Através da observação da figura 6.14 pode verificar-se que o primeiro componente se relaciona com a idade dos vinhos. De facto, todos os vinhos novos possuem abcissas negativas, os vinhos 10 Anos possuem abcissas superiores (todas positivas, excepto um caso), e os 20 Anos possuem, em média, abcissas ainda mais elevadas. Quanto ao segundo eixo, parece claro que é o responsável pela separação entre os *Vintage* e os restantes tipos. Os tipos *Ruby* e *Tawny* são aqueles que se confundem mais, sendo que entre os 10 e 20 Anos também se verifica alguma sobreposição.

No que diz respeito à interpretação dos componentes, foi já referido que o primeiro tem a ver com a idade do vinho. As análises químicas que se relacionam mais com este componente são, por ordem decrescente do valor do peso no componente, o *5-HMF* (0.94), o *acetato de etilo* (0.94), a *acidez volátil* (0.92), o *álcool alílico* (0.88), os *álcoois amílicos* (0.85), o *extracto seco* (0.85), a *massa volúmica* (0.85), os *álcoois superiores totais* (0.85) e o *grau Baumé* (0.84).

O segundo componente é mais difícil de interpretar, recorrendo apenas à informação contida na figura 6.14. No entanto, ao verificar quais as análises químicas que se relacionam mais com este componente (*acidez fixa* (0.86), *cor-L* (0.8), *cor-b* (0.72) e *acidez total* (0.71)), pode afirmar-se que, das análises feitas, as que mais distinguem os *Vintage* dos outros vinhos são a sua acidez e a sua cor.

Estes resultados estão de acordo com o que os técnicos do IVP sabem: os vinhos mais velhos revelam valores de 5-*HMF* e de *acetato de etilo* mais elevados do que os vinhos mais novos, e os *Vintage* novos apresentam menos acidez do que os *Ruby* (aqueles com que normalmente são comparados) e uma cor retinta, muito mais acentuada do que os outros vinhos.

6.6.1.2 Estudo dos Resultados Obtidos com o Nariz Electrónico

Numa segunda fase, estudaram-se amostras dos mesmos vinhos com recurso a uma nariz electrónico. Este tipo de equipamento tem suscitado grande interesse, tanto na comunidade científica como nos laboratórios de análises de grandes empresas, como sendo um método de análise de aromas simples, rápido e objectivo. Por exemplo, Chatonnet e Dubourdieu (1999) utilizam um nariz electrónico para discriminar entre os vários níveis de queimado da madeira utilizada nos barris de carvalho (os barris de madeira de carvalho são, muitas vezes, queimados por dentro), Zonderban et al. (1999) usam um sistema semelhante para controlar o processo de aquecimento de uma mistura de leite em pó e açúcar utilizada na produção de chocolate, Aparicio et al. (2000) recorrem a um nariz electrónico para detectar o aroma a "ranço" no azeite e Rocha et al. (1998) utilizam-no no controlo da qualidade de rolhas. O funcionamento do equipamento baseia-se, tal como o olfacto humano, em três fases: detecção dos compostos de aroma, processamento do sinal recebido e reconhecimento e interpretação dos dados.

O equipamento utilizado possui uma matriz de 32 sensores que respondem à presença das várias componentes voláteis alterando as suas propriedades eléctricas. As alterações nas resistências dos sensores podem ser visualizadas através de histogramas, característicos do aroma que está a ser analisado.

Os dados estudados foram os referentes ao valor médio do *output* de cada sensor, calculado entre os 30 e os 110 segundos após a introdução da amostra no equipamento. Mais pormenores sobre o processo de análise podem encontrar-se em Rocha et al. (1999).

Estes resultados foram, tal como os das análises químicas e de cor, sujeitos a uma Análise de Componentes Principais (ACP_2), cujos resultados, em termos de pontuações que cada amostra obteve nos dois componentes principais, se apresentam na figura 6.15.

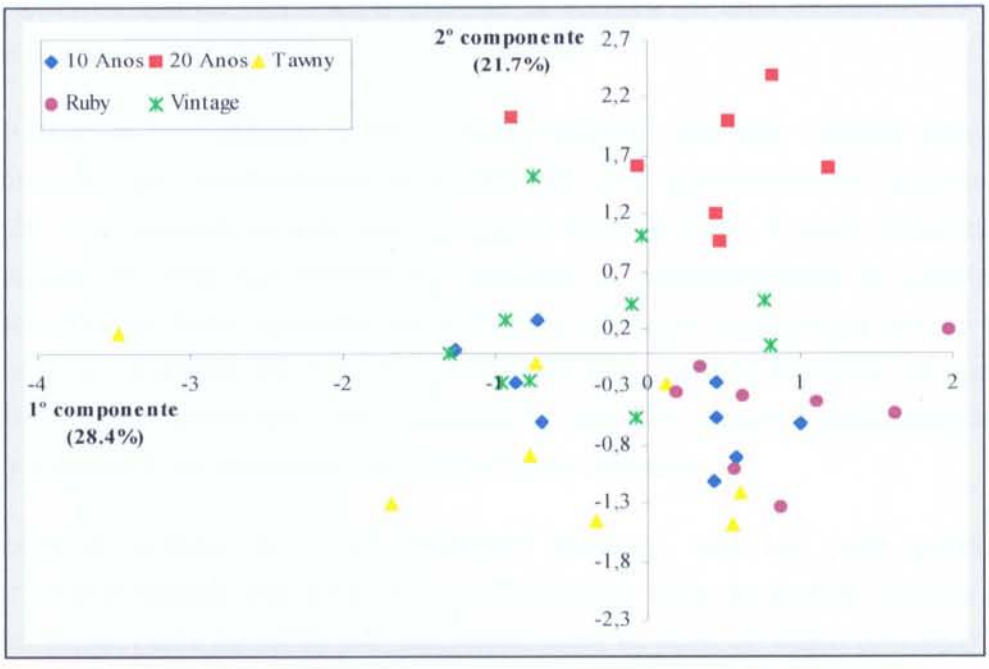

Figura 6.15 – Representação gráfica das pontuações obtidas por cada amostra no 1° e 2° componentes principais da ACP_2

Aparentemente, os resultados da ACP_2 não são animadores: verifica-se que o tipo de vinho melhor discriminado é o 20 Anos; os *Vintage* parecem "misturar-se" com todos os outros; os *Tawny* e *Ruby* são ligeiramente melhor discriminados do que na ACP_1, mas os 10 Anos sobrepõem-se a estes dois tipos.

Uma outra dificuldade que se coloca é a da interpretação dos eixos: observando a figura 6.15, é difícil interpretá-los, e os resultados dos sensores que mais se relacionam com cada factor também não ajuda muito, visto não haver informação sobre a que é que cada sensor é sensível. Verifica-se, no entanto, que aqueles que mais contribuem para a constituição do primeiro factor são os sensores números 31 e 32, enquanto para o segundo factor são os números 1 e 24.

6.6.1.3 Estudo das Análises Químicas, de Cor e dos Sensores em Conjunto

Tendo em consideração os resultados obtidos com as duas análises apresentadas em 6.6.1.1 e 6.6.1.2 (ACP_1 e ACP_2), e com o objectivo de melhorar a discriminação obtida com a ACP_1, foram feitas outras análises de componentes principais com variáveis seleccionadas dos dois conjuntos considerados (análises químicas e de cor e sensores). Saliente-se que não se fez uma análise com todas as variáveis em conjunto devido ao facto de, para a utilização desta técnica, ser necessário um maior número de

observações do que de variáveis, o que não se verifica (nº total de variáveis = 61; nº total de observações = 42).

Utilizaram-se dois critérios na selecção das variáveis para estas análises posteriores. Em primeiro lugar, seleccionaram-se as variáveis que, individualmente, apresentaram um poder discriminante elevado entre os vários tipos de vinho. O poder discriminante foi medido com base nos valores da estatística *F* correspondentes às Análises de Variância com um factor aplicadas aos 42 vinhos estudados (distribuídos pelos 5 tipos). Elaboraram-se, portanto, 61 ANOVAs, uma para cada variável estudada. As variáveis escolhidas foram aquelas que, dentro de cada um dos três conjuntos (análises químicas, de cor e sensores), apresentaram valores de *F* mais elevados.

O segundo critério foi o de considerar variáveis com um peso grande nos componentes principais que evidenciaram discriminar entre os vinhos. Note-se que o primeiro componente da ACP_2 não discrimina entre os tipos de vinho e, portanto, não foram considerados os sensores com pesos elevados nesse componente.

Após várias análises, decidiu-se considerar, no estudo final, e com base no primeiro critério, as variáveis *acidez volátil* (*AV*), *5-hidroximetilfurfural* (*5-HMF*), *acetato de etilo* (*AE*), *parâmetro de cor L* (*CL*), *parâmetro de cor a* (*Ca*) e sensores números 1, 23, 24 e 28 (*S*1, *S*23, *S*24 e *S*28, respectivamente). A única variável acrescentada pelo segundo critério foi a *acidez fixa* (*AF*). Saliente-se que a maior parte das outras variáveis com pesos altos nos componentes já tinham sido consideradas no primeiro critério.

Os resultados desta análise de componentes principais (ACP_3) apresentam-se, em termos das pontuações obtidas por cada amostra nos dois componentes principais, na figura 6.16.

Através da observação desta figura, verifica-se que o primeiro componente, que explica 48.8% da variância contida nos dados em estudo, continua a relacionar-se com a idade dos vinhos. Nota-se, inclusivamente, que não há sobreposições entre os três conjuntos de idades diferentes. O segundo componente, responsável por 29.4% da mesma variância, continua a ser mais difícil de interpretar. Constata-se, no entanto, que ajuda a discriminar entre os vinhos novos, embora os tipos *Ruby* e *Tawny* continuem a exibir alguma sobreposição (menor do que em ACP_1).

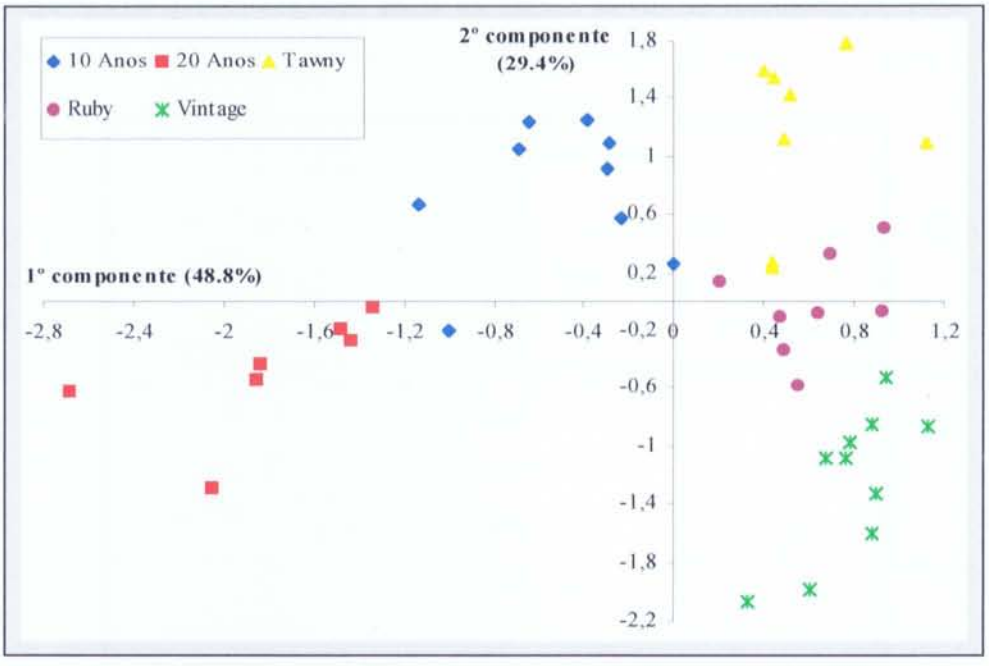

Figura 6.16 – Representação gráfica das pontuações obtidas por cada amostra no 1° e 2° componentes principais da ACP_3

Convém referir que se compararam os resultados apresentados na figura 6.16 com os do painel de provadores para os mesmos vinhos. As conclusões foram interessantes: o vinho 10 Anos, que aparece em cima do eixo dos *YY* e, portanto, mais perto dos vinhos novos, foi reprovado pela Câmara, pelo que não é, na realidade, um vinho do Porto de 10 Anos (de facto, constatou-se que houve um engano no envio das amostras que era suposto serem todas de vinhos aprovados) e o *Vintage* que se situa mais próximo dos *Ruby* foi aprovado com três votos contra dois, o que não aconteceu a mais nenhuma das outras amostras.

Com o objectivo de ajudar a compreender melhor o significado de cada componente e quais as variáveis que os influenciam, apresenta-se, na figura 6.17, a representação gráfica dos pesos das variáveis em cada um dos dois componentes principais já representados na figura 6.16.

Através do gráfico representado na figura 6.17, verifica-se que as variáveis *acidez volátil*, *acetato de etilo* e *5-hidrometilfurfural* são as que se relacionam apenas com o componente que distingue entre as idades dos vinhos. Por outro lado, a *acidez fixa* e o sensor 28 relacionam-se apenas com o 2° componente, que distingue entre os vinhos novos. Finalmente, as variáveis referentes à cor (*CL* e *Ca*) e aos sensores 1, 23 e 24, relacionam-se com os dois componentes, o que revela que tanto a idade do vinho como

as características que distinguem entre os vinhos novos se traduzem, de certa forma, nestas variáveis.

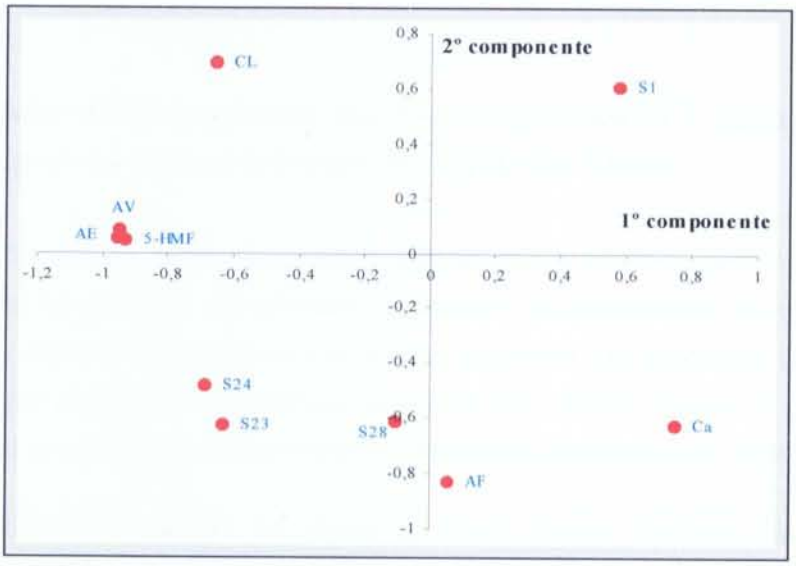

Figura 6.17 – Representação gráfica dos pesos das variáveis consideradas na ACP_3 em cada um dos dois componentes principais

Deve realçar-se, aqui, que os resultados desta análise são promissores. A confirmar-se a possibilidade de distinção entre os tipos de vinho estudados através dos resultados de análises químicas, de cor e do nariz electrónico, está aberto um caminho que poderá levar à definição de fronteiras entre os vários tipos de vinho e, consequentemente, à possibilidade de calibrar periodicamente os provadores com vinhos aferidos de uma forma mais objectiva.

Convém registar, no entanto, que se tratou apenas de um estudo exploratório. A quantidade de amostras analisadas foi pequena e, portanto, os resultados das análises de componentes principais e a selecção de variáveis deverão ser confirmadas posteriormente, através de um estudo mais alargado.

6.6.1.4 Caracterização da Idade do Vinho do Porto através de Análises Químicas

Embora a tentativa de caracterização da Idade do vinho do Porto não se enquadre no âmbito deste trabalho, não deve deixar de se referir um estudo feito na Universidade Victor Segalen Bordéus 2 (ver Ferreira (1998)), em que o autor analisa vários modelos de regressão para previsão da idade de vinhos com data de Colheita e *Tawny* com indicação de Idade, com base nos resultados de uma série de análises químicas. O modelo que melhor se adapta conduz a intervalos de confiança a 95% para a idade do vinho com uma amplitude de cerca de 4 anos.

O autor conclui que os resultados deste estudo podem ser particularmente úteis, se utilizados como um meio complementar da certificação da idade dos vinhos do Porto envelhecidos por oxidação.

6.6.2 Aferição da Capacidade dos Provadores do IVP para a Detecção de Algumas Características do Vinho do Porto

Tal como ficou claro na secção anterior, a definição de padrões para o vinho do Porto (e para o vinho em geral) é, actualmente, impossível de concretizar. No entanto, com o conhecimento que existe de alguns compostos químicos (ou conjuntos de compostos) responsáveis por algumas características de aroma dos vinhos, parece ser possível aferir as capacidades de percepção dos provadores para essas características específicas.

Nesse sentido, no Instituto do Vinho do Porto deu-se início a um estudo cujo objectivo é o de aferir, periodicamente, as capacidades dos provadores para detectar alguns defeitos no vinho e para identificar as características que distinguem entre vinhos *Tawny* de diversas idades.

Os testes sensoriais que se podem utilizar são muitos (ver Capítulos 2 e 3), tendo-se optado por realizar testes de ordenação simples de amostras. Este tipo de testes é fácil de executar, permite avaliar várias amostras ao mesmo tempo e é indicado para estudos exploratórios (Pangborn (1984)).

A experiência, ainda em fase inicial, começou pelo estudo da capacidade dos provadores para a detecção do defeito "acetificado" e da sua sensibilidade à idade dos vinhos *Tawny*.

No caso do defeito "acetificado", foi identificado um vinho com um grau de defeito muito acentuado e um outro sem defeito e, a partir destes, foram feitas três soluções com teores diferentes de ácido acético e de acetato de etilo. Foi, então, pedido aos provadores que ordenassem as 5 amostras (as 2 "puras" e as 3 "misturas") da menos acetificada para a mais acetificada.

Relativamente à característica Idade, a técnica utilizada foi a mesma, tendo sido considerados dois vinhos (um de aproximadamente 7 Anos e outro de aproximadamente 20 Anos) e três misturas destes. Note-se que esta prova foi feita com luz vermelha de modo a que a cor do vinho não influenciasse o julgamento.

Numa primeira fase, houve que "afinar" as misturas, de forma a que a ordenação não fosse uma tarefa demasiado óbvia nem demasiado difícil. Uma outra precaução que se tomou foi a de garantir que entre as várias amostras, quando ordenadas, não existissem grandes "degraus" relativamente à intensidade da característica em causa.

No que diz respeito ao processamento dos resultados (ver sub-secção 3.2.1.3), optou-se por utilizar o coeficiente de correlação ordinal de Spearman, tal como aconselham Cardinal et al. (1994). Este coeficiente é definido por:

$$Sp_k = 1 - 6 \cdot \sum \frac{d_i^{\,2}}{N \cdot (N^2 - 1)}$$

onde:

Sp_k: coeficiente de correlação de Spearman relativo ao provador k;

d_i: diferença entre a ordem de referência (da concentração real) da característica em causa e a ordem atribuída pelo provador;

N: número de amostras ordenadas (5, no caso em análise).

Este coeficiente varia entre os valores –1 e 1, correspondendo o valor –1 à situação em que a ordenação feita pelo provador é completamente inversa da ordem referência, e o valor 1, à situação em que as duas ordenações são iguais. Neste sentido, para uma determinada característica, a capacidade de discriminação do provador é tanto maior quanto mais próximo do valor 1 for o seu coeficiente de correlação de Spearman.

Como o estudo ainda se encontra numa fase preliminar de testes, optou-se por não incluir, neste trabalho, quaisquer resultados.

Note-se que os resultados obtidos com este tipo de análises, são, mais uma vez, relativos. É possível determinar qual (ou quais) o(s) provador(es) que detecta(m) melhor ou pior a característica em estudo, mas não é possível perceber se tais resultados são tão bons, melhores ou piores do que aquilo que deveria ser exigido. Para se chegar a este género de conclusões, é necessário elaborar experiências muito mais alargadas, tanto ao nível do número de testes como ao do número de provadores. Foi, aliás, o que foi feito relativamente aos sabores fundamentais, cujos valores dos limiares se encontram tabelados.

6.7 Decisão Final do Júri de Prova

Os métodos e técnicas referidos neste capítulo tiveram como objectivo a avaliação do desempenho de cada provador, individualmente. Embora essa avaliação permita estabelecer comparações e, em alguns casos, seja mesmo dependente dos julgamentos do painel como um todo (caso do índice *I* e da curva ROC), até este ponto ainda não foi discutida a questão de como incorporar a informação obtida sobre cada provador na decisão final do júri de prova.

O método utilizado pelo IVP é aquele que se utiliza normalmente em Análise Sensorial. Uma vez seleccionados os provadores, todos têm o mesmo peso na decisão final do painel e, portanto, a decisão de aprovar, ou não, um vinho é tomada consoante a opinião da maioria dos provadores envolvidos na prova.

Neste trabalho propõe-se outra abordagem. Entende-se que será vantajoso considerar o nível de desempenho diário de cada provador na decisão final do júri de prova. Assim, a opinião dos provadores que, num determinado dia, apresentem um nível de desempenho superior pesará mais na decisão final da Câmara do que a dos provadores cujo desempenho se tenha revelado de nível inferior.

Por outro lado, a fronteira entre as amostras "aprovadas" e as "não aprovadas" é, por vezes, ténue e difícil de definir. O procedimento proposto permite estabelecer que, em determinadas condições nas quais poderão surgir dúvidas em relação à aprovação ou não das amostras, não seja tomada nenhuma decisão, sendo recomendada uma segunda prova (cujos resultados devem ser conjugados com os da primeira).

A decisão final do painel de provadores será, assim, baseada na soma pesada dos valores das decisões individuais dos provadores. Os pesos serão proporcionais aos índices I_k já referidos em 6.4, que são actualizados diariamente e medem a fiabilidade dos provadores. A decisão final (*FD*) será, então, expressa por um número dado pelas seguintes expressões:

$$FD_A = \sum_k W_k \cdot DA_k \qquad \text{e} \qquad FD_R = \sum_k W_k \cdot DR_k$$

onde:

FD_A: variável que representa a decisão final relativamente às aprovações;

FD_R: variável que representa a decisão final relativamente às não aprovações;

DA_k: assume o valor "1" se o provador k aprova o vinho e "0" se não aprova;
DR_k: assume o valor "-1" se o provador k não aprova o vinho e "0" se aprova;
W_k: peso relativo do provador: $W_k = I_k \cdot K / \sum I_k$.

Note-se que os pesos W_k são os índices I_k ajustados de forma a que $\sum I_k = K$. Portanto, se todos os provadores exibirem o mesmo grau de desempenho, o peso de cada um assumirá o valor 1.

Uma determinada amostra só será "aprovada" no caso de FD_A ser igual ou superior a um valor pré-estabelecido e só será "não aprovada" se FD_R for igual ou inferior a um outro valor, também pré-estabelecido. Uma proposta razoável é a de que os valores pré-estabelecidos sejam simétricos e de valor absoluto igual a $(k+1)/2$.

Assim, uma amostra será aprovada se $FD_A \geq (k+1)/2$ e não será aprovada se $FD_R \leq -(k+1)/2$. No caso de se verificar, simultaneamente, que $FD_A < (k+1)/2$ e $FD_R > -(k+1)/2$, a prova deverá ser repetida, considerando-se, para o resultado final, os resultados das duas provas (K passa, portanto, a ser igual a K_1 (primeira prova) + K_2 (segunda prova)).

Seguidamente, apresenta-se um exemplo do procedimento proposto para 5 provadores, recorrendo a valores de I_k simulados e, portanto, não correspondentes a nenhuma situação real.

Na tabela 6.6, apresentam-se os referidos valores simulados (I_k) e os correspondentes valores para o peso de cada provador (W_k).

Tabela 6.6 – Valores simulados de I_k e de W_k para 5 provadores

	Provador				
	U	**V**	**X**	**Y**	**Z**
I_k	0.4	0.73	0.52	0.6	0.78
W_k	0.66	1.2	0.86	0.99	1.29

Admitindo os pesos indicados na tabela 6.6 e a regra atrás enunciada (que, para este exemplo, se traduz por "uma amostra é aprovada se $FD_A > 3$ e não é aprovada se $FD_R < -3$), pode verificar-se que, nas situações em que apenas um provador não aprova, a decisão final é de aprovação, seja qual for o provador discordante. Nos casos em que há 3 provadores a aprovar a amostra e, portanto, 2 a não aprovar, a decisão final será a de aprovar a amostra, se os provadores que aprovam tiverem um peso relativamente elevado, e de repetir a prova, no caso contrário.

Com o objectivo de ilustrar o que acabou de ser dito, apresentam-se, na tabela 6.7, todas as combinações possíveis das decisões em que, dos 5 provadores, 3 aprovam a amostra, e o valor correspondente das duas variáveis de decisão final (FD_A e FD_R) e da própria decisão final do painel.

Tabela 6.7 – Valores de FD_A, de FD_R e da *Decisão Final* para as situações nas quais 3 dos 5 provadores aprovam a amostra

Provadores que aprovaram a amostra	FD_A	FD_R	***Decisão Final***
U, V, X	2.72	-2.28	**Repetir**
U, V, Y	2.85	-2.15	**Repetir**
U, V, Z	3.15	-1.85	**Aprovada**
U, X, Y	2.51	-2.49	**Repetir**
U, X, Z	2.81	-2.19	**Repetir**
U, Y, Z	2.94	-2.06	**Repetir**
V, X, Y	3.05	-1.95	**Aprovada**
V, X, Z	3.35	-1.65	**Aprovada**
V, Y, Z	3.48	-1.52	**Aprovada**
X, Y, Z	3.14	-1.86	**Aprovada**

O procedimento descrito é geral e aplicável a diversas situações. Os valores pré-estabelecidos para decidir a aprovação ou não das amostras podem, inclusivamente, ser definidos consoante os objectivos do painel de provadores e os custos dos dois tipos de erros associados à prova (aprovar uma amostra que não devia ser aprovada e não aprovar uma amostra que devia sê-lo). No caso de os custos destes dois tipos de erros serem diferentes, os limites estabelecidos não deverão ser simétricos.

Chama-se ainda a atenção para o facto de o procedimento de aprovação por maioria não ser mais do que um caso particular do proposto, em que o número de provadores é obrigatoriamente ímpar e o peso de cada provador é constante e igual a 1.

Finalmente, recorda-se que os índices I_k devem ser actualizados diariamente, o que permite que as decisões finais do painel sejam tomadas com base numa medida actual do desempenho dos provadores. Outras vantagens que se atribuem à utilização deste procedimento são: (i) não ser necessário um número ímpar de provadores na constituição do painel e (ii) ser possível estabelecer limites de aprovação/não aprovação que permitam, de uma forma sistemática, definir as situações em que a prova deve ser repetida.

7

Conclusões

Conteúdo:

- **Conclusões Gerais**
- **Perspectivas de Desenvolvimento Futuro**

7.1 Conclusões Gerais

No trabalho que acaba de se apresentar, abordou-se o problema do controlo dos painéis de provadores em Análise Sensorial. Tornou-se evidente a sua complexidade, agravada pelo facto de o tema envolver conhecimentos de áreas distintas, tais como a Análise Sensorial, a Estatística e a Qualidade (mais especificamente, o Controlo Estatístico de Processos).

Tal como se esperava, os métodos "clássicos" utilizados habitualmente nestas áreas não são fácil nem directamente aplicáveis ao controlo dos julgamentos em Análise Sensorial. Concluiu-se, assim, que o problema exigia que se efectuassem desenvolvimentos específicos e que se testassem empiricamente os respectivos desempenhos.

Nestas condições, para avaliar a prestação individual dos provadores pertencentes a painéis de prova, foi necessário desenvolver um conjunto de métodos estatísticos. Estes métodos, em larga medida inspirados nas diversas vertentes do problema do controlo da análise sensorial no Instituto do Vinho do Porto, foram extensivamente testados recorrendo à sua Câmara de Provadores. Mencionam-se, em seguida, os quatro tópicos fundamentais dos métodos propostos.

- **Caracterização do Processo de Decisão do Provador e Controlo da sua Estabilidade**

  A caracterização do processo de decisão de cada provador foi efectuada com recurso à técnica de regressão logística. Para o controlo da estabilidade das decisões recorreu-se a cartas de controlo CUSUM de proporções binomiais. Os resultados obtidos revelaram-se promissores, na medida em que os próprios provadores se identificaram com os respectivos modelos e que as raras ocasiões em que os processos se afastaram das situações de controlo puderam ser justificadas na prática.

- **Avaliação Diária do Desempenho dos Provadores em Termos de Reprodutibilidade dos Julgamentos**

  Para a avaliação diária da capacidade de os provadores reproduzirem as suas decisões, utilizou-se um índice de fiabilidade desenvolvido com base na

comparação das apreciações efectuadas em provas repetidas duas vezes (réplicas). Os resultados obtidos com o índice escolhido (de entre outros que foram ensaiados) foram animadores, sendo compatíveis com as opiniões emitidas por responsáveis pela Câmara de Provadores do IVP sobre a evolução dos seus vários provadores.

- **Avaliação Periódica do Desempenho Global de Cada Provador**

  O tratamento dos resultados dos ensaios sensoriais é, em muitos casos, problemático. De facto, a aplicação directa de técnicas como os testes de hipóteses e a utilização de indicadores como a proporção de respostas correctas, podem não se revelar adequados em determinadas circunstâncias. A análise ROC é aconselhada em muitas destas situações, visto que conduz a resultados que são independentes dos critérios utilizados pelos provadores.

  Esta técnica foi utilizada para medir a sensibilidade dos provadores do IVP à qualidade global do vinho num período de tempo alargado (1 ano). Mais uma vez se verificou que, no geral, os resultados foram compatíveis com as avaliações sobre o desempenho dos provadores efectuadas por responsáveis pela Câmara de Provadores do IVP.

- **Aferição Periódica de Capacidades Específicas dos Provadores**

  Nenhuma das três abordagens anteriores tem em consideração, de uma forma directa, a capacidade específica do provador para detectar determinadas características ou substâncias nos produtos. É recomendado que, quando possível, seja feita periodicamente a aferição de tais capacidades. A técnica proposta baseia-se num ensaio de ordenação de amostras e permite obter um indicador da capacidade do provador para detectar a característica analisada. A implementação deste procedimento no IVP encontra-se numa fase inicial, pelo que ainda não é possível analisar os seus resultados.

Ao longo do processo de desenvolvimento e teste dos métodos apresentados, foram sendo detectadas algumas condições para que o sistema proposto tenha um bom desempenho. De entre estas, salientam-se: (i) as boas condições físicas para a realização dos ensaios; (ii) a existência de um número razoável de provadores que integram o painel (o mínimo foi considerado de 4, mas preferencialmente deverão ser mais); (iii) uma elevada frequência de prova (que permita obter um grande volume de dados).

Por outro lado, dada a complexidade do processo de prova e as limitações que foram sendo apontadas ao longo de todo o trabalho, aconselha-se a utilização simultânea dos quatro procedimentos propostos, por forma a que o controlo dos provadores seja encarado com a devida precaução, e que as decisões que dele advêm sejam tomadas com base no máximo de informação disponível.

O objectivo final da avaliação do desempenho individual dos provadores é o de que o Painel por eles constituído produza resultados o mais fiáveis possível. É, portanto, essencial que as decisões finais do Painel de Provadores sejam tomadas tendo em consideração os resultados dessa avaliação. Assim, no final do trabalho foi proposta uma técnica baseada no índice de fiabilidade (utilizado para a avaliação diária da reprodutibilidade dos provadores) que permite integrar de forma adaptativa os julgamentos dos vários provadores na decisão final do Painel.

Finalmente, sublinha-se que os métodos desenvolvidos e apresentados ao longo do trabalho não são específicos do Instituto do Vinho do Porto, podendo ser aplicados em muitas outras situações comuns em Análise Sensorial. O facto de o Instituto do Vinho do Porto lidar com uma situação de uma complexidade invulgar, permite confiar que não deverão surgir problemas adicionais, quando, eventualmente, se pretender transferir os procedimentos propostos para outros contextos.

7.2 Perspectivas de Desenvolvimento Futuro

A abordagem dos problemas tratados neste trabalho encontra-se ainda numa fase inicial, pelo que são muitas as perspectivas de desenvolvimento futuro.

No geral, pode dizer-se que haverá, certamente, métodos alternativos aos que foram aqui desenvolvidos e utilizados, e que seria interessante promover a comparação entre eles, quer no contexto do vinho do Porto, quer no de outras aplicações da Análise Sensorial ao ramo alimentar.

Em particular, foram sendo tecidos comentários ao longo do texto, que evidenciam algumas matérias que merecem ser analisadas e/ou apuradas:

(i) O aprofundamento do estudo da variabilidade inerente aos diversos tipos de vinhos e a análise da possibilidade de se vir a utilizar um modelo do tipo beta-binomial para processar resultados de ensaios de diferenciação.

(ii) A continuação do estudo da caracterização das amostras dos vários tipos de vinho do Porto com recurso ao nariz electrónico, às análises químicas e às análises de cor.

(iii) A continuação das experiências para a determinação das concentrações a utilizar nos ensaios de ordenação de amostras, que têm como objectivo a aferição periódica das capacidades dos provadores.

(iv) A extensão do tipo de experiências referidas em (ii) a outras características que não apenas o defeito "acetificado" e a Idade do vinho.

Os aspectos referidos relacionam-se todos, de uma maneira ou de outra, com a tentativa de estabelecer padrões para os vinhos em geral (caso de (i) e (ii)) ou para algumas das suas características em particular (caso de (iii) e (iv)). A principal dificuldade apontada ao longo de todo o trabalho – a ausência de padrões para os produtos alimentares de uma forma geral – continua, portanto, a merecer a maior atenção. A procura de soluções para este problema (ainda que por vezes apenas parciais) é considerada da maior importância para futuros desenvolvimentos nesta área.

Referências Bibliográficas

AFNOR (1991). *Recueil de Normes Françaises - Contrôle de la Qualité des Produits Alimentaires: Analyse Sensorielle* (4ª ed.). Paris: AFNOR.

Almeida, T., E. Cubero e M. O'Mahony (1999). Same-Different Discrimination Tests with Interstimulus Delays up to One Day. *Journal of Sensory Studies, 14,* 1-18.

Alvelos, H. (1995). *Garantia da Qualidade no Desenvolvimento de Produtos de Consumo: Análise de uma Metodologia*. Dissertação de Mestrado. Intituto Superior de Estudos Empresariais, Universidade do Porto.

Amemiya, T. (1981). Qualitative Response Models: a Survey. *Journal of Economic Literature, 19,* 481-536.

ASTM (1981). *Guidelines for the Selection and Training of Sensory Panel Members, ASTM Special Technical Publication, 758*. Philadelphia: American Society for Testing and Materials.

ASTM (1996). *ASTM Manual 26: Sensory Testing Methods* (2ª ed.). Edgar Chambers IV e Mona Baker Wolf (Ed.), Philadelphia: American Society for Testing and Materials.

ASTM (1997a). *Standard Practice E679-91, Standard Practice for Determination of Odor and Taste Thresholds By a Forced-Choice Ascending Concentration Series Method of Limits*. Philadelphia: American Society for Testing and Materials.

ASTM (1997b). *Standard Practice E1432-91, Standard Practice for Defining and Calculating Individual and Group Sensory Thresholds from Forced-Choice Data Sets of Intermediate Sizes*. Philadelphia: American Society for Testing and Materials.

Antoniou, K. D., D. Petridis, S. Raphaelides, Z. Ben Omar e R. Kesteloot (2000). Texture Assessment of French Cheeses. *Journal of Food Science, 65,* 168-172.

Aparicio, R. S. Rocha, I. Delgadillo e M. Morales (2000). Detection of Rancid Defect in Virgin Olive Oil by the Electronic Nose. *Journal of Agriculture and Food Chemistry, 46,* 853-860.

Arnold, G. (1990). Problems with Procrustes Analysis. *Journal of Applied Statistics, 17,* 449-451.

Bai, D. e I. Choi (1995). $\overline{X}$ and R Control Charts for Skewed Populations. *Journal of Quality Technology, 27,* 120-131.

Beneke, M., L. M. Leemis, R. E. Schlegel e B. L. Foote (1988). Spectral Analysis in Quality Control: A Control Chart Based on the Periodogram. *Technometrics, 30,* 63-70.

Benn, S. e T. Peppard (1996). Characterization of Tequila Flavor by Instrumental and Sensory Analysis. *Journal of Agriculture and Food Chemistry, 44,* 557-566.

Benneyan, J. C. e Kaminsky, F. C. (1994). Modeling Discrete Data in SPC: The g and h Charts. *ASQC 48th Annual Quality Congress Proceedings,* 32-42.

Bi, J. (1995). Nonparametric Models for Discrimination Methods and Sensivity Analysis for Triads. *Journal of Sensory Studies, 10,* 325-340.

Bi, J. e D. Ennis (1998a). Sensory Thresholds: Concepts and Methods. *Journal of Sensory Studies, 13,* 133-148.

Bi, J. e D. Ennis (1998b). A Thurstonian Variant of the Beta-Binomial Model for Replicated Difference Tests. *Journal of Sensory Studies, 13,* 461-466.

Bi, J. e D. M. Ennis (1999). The Power of Sensory Discrimination Methods Used in Replicated Difference and Preference Tests. *Journal of Sensory Studies, 14,* 289-302.

Bi, J. e M. O'Mahony (1995). Table for Testing the Significance of the R-Index. *Journal of Sensory Studies, 10,* 341-347.

Bi, J., D. Ennis e M. O'Mahony (1997). How to Estimate and use the Variance of d' from Difference Tests. *Journal of Sensory Studies, 12,* 87-104.

Bi, J., L. Templeton-Janik, J. Ennis e D. Ennis (2000). Replicated Difference and Preference Tests: How to Account for Inter-trial Variation. *Food Quality and Preference, 11,* 269-273.

Bissel, A. F. (1969). Cusum Techniques for Quality Control. *Applied Statistics, 18,* 1-30.

Bodden, K. e S. Rigdon (1999). A Program for Approximating the In-Control ARL for the MEWMA Chart. *Journal of Quality Technology, 31,* 120-123.

Borror, C., D. Montgomery e G. Runger (1999). Robustness of the EWMA Control Chart to Non-Normality. *Journal of Quality Technology, 31,* 309-316.

Bothe, D. R. (1986). Quantifying the Defects. *Quality, 25,* 71-72.

Bowrring-Carr, C. e J. West-Burnham (1994). *Managing Quality in Schools - A Training Manual.* Essex: Longman.

Braga, A. C., P. Oliveira e A. Gomes (1997). Comparação entre Unidades de Cuidados Intensivos Neonatais Baseada na Análise ROC. *Actas do V Congresso Anual da Sociedade Portuguesa de Estatística,* Curia, 117-126.

Braga, Ana Cristina da Silva (2001). *Curvas ROC: Aspectos Funcionais e Aplicações.* Tese de Doutoramento. Universidade do Minho.

Branger, E. B., C. A. Sims, R. H. Schmidt, S. F. O'Keefe e J. A. Cornell (1999). Sensory Characteristics of Cottage Cheese Whey and Grapefruit Juice Blends and Changes During Processing. *Journal of Food Science, 64,* 180-184.

Burr, I. W. (1967). The Effect of Non-Normality on Constants for $\overline{X}$ and R Charts. *Industrial Quality Control, 23,* 563-569.

Burr, I. W. (1969). Control Charts for Measurements with Varying Sample Sizes. *Journal of Quality Technology, 1,* 163-167.

Burr, I. W. (1976). *Statistical Quality Control Methods.* New York: Marcel Dekker.

Cabral, J. A. (1990). *Análise, Desenvolvimento e Teste de Regras de Compra de Mercadorias Sujeitas a Flutuações de Preço.* Tese de Doutoramento. Faculdade de Engenharia da Universidade do Porto.

Cabral, J. A. (2001). *Notas de Apoio à Disciplina de Gestão de Qualidade.* Não publicadas. Faculdade de Engenharia da Universidade do Porto.

Calvino, A., M. Zamora e M. Sarchi (1996). Principal Components and Cluster Analysis for Descriptive Sensory Assessment of Instant Coffee. *Journal of Sensory Studies, 11,* 191-210.

Cardinal, M., J. Cornet, A. Qannari e M. Qannari (1994). Performances d'un Groupe d'Évaluation Sensorielle: Exemples de Traitements Statistiques des Données. *Sciences des Aliments, 14,* 251-263.

Carola, R., J. P. Harley e C. R. Noback (1992). *Human Anatomy*. New York: McGraw Hill.

Cassady, C. R., R. Bowden, L. Liew e E. Pohl (2000). Combining Preventive Maintenance and Statistical Process Control: a Preliminary Investigation. *IIE Transactions, 32,* 471-478.

Castillo, E. del (1995). Relations between $\overline{X}$ Control Chart Design Variables and Production Control. *International Journal of Production Research, 33,* 2709-2721.

Chan, L. Y., M. Xie e T. N. Coh (2000). Cumulative Quantity Control Charts for Monitoring Production Processes. *International Journal of Production Research, 38,* 397-408.

Chang, S. I. e C. A. Aw (1996). A Neural Fuzzy Control Chart for Detecting and Classifying Process Mean Shifts. *International Journal of Production Research, 34,* 2265-2278.

Chao, M. T. e S. W. Cheng (1996). Semicircle Control Chart for Variables Data. *Quality Engineering, 8,* 441-446.

Chatonnet, P. e D. Dubourdieu (1999). Using Electronic Odor Sensors to Discriminate among Oak Barrel Toasting Levels. *Journal of Agriculture and Food Chemistry, 47,* 4319-4312.

Cheng, A., R. Liu e J. Luxhoj (2000). Monitoring Multivariate Aviation Safety Data by Data Depth: Control Charts and Threshold Systems. *IIE Transactions, 32,* 861-872.

Chinna, K., S. Kadir e M. Abdullah (2000). Box-chart: Combining $\overline{X}$ and S Charts. *Total Quality Management, 11,* S857-S862.

Christensen, C. M. (1983). Effects of Color on Aroma, Flavor and Texture Judgements of Foods. *Journal of Food Science, 48,* 787-790.

Cliff, M., M. King, C. Scaman e B. Edwards (1997). Evaluation of R-Indices for Preference Testing of Apple Juices. *Food Quality and Preference, 8,* 241-246.

Cochran, W. G. (1954). Some Methods for Strengthening the Common χ^2 Tests. *Biometrics, 10,* 417-451.

Costell, E., M. Pastor, L. Izquierdo e L. Durán (1994). Comparison of Simplified Methods for Evaluation of Sensory Thresholds of Added Substances. *Journal of Sensory Studies, 9,* 365-382.

Cubero, E., T. Almeida e M. O'Mahony (1995). Cognitive Aspects of Difference Testing: Memory and Interstimulus Delay. *Journal of Sensory Studies, 10,* 307-324.

Dacremont, C., C. Roger e F. Sauvageot (1998). Replicated Triangle Tests: Effect of Feed-Back and Product Comparison on Performance. *Journal of Sensory Studies, 13,* 413-433.

Dijksterhuis, G. (1994). Procrustes Analysis in Studying Sensory-Instrumental Relations. *Food Quality and Preference, 5,* 115-120.

Dijksterhuis, G. (1995). Assessing Panel Consonance. *Food Quality and Preference, 6,* 7-14.

Dijksterhuis, G. e J. Gower (1992). The Interpretation of Generalised Procrustes Analysis and Allied Methods. *Food Quality and Preference, 3,* 67-87.

Dijksterhuis, G., M. Flipsen e P. Punter (1994). Principal Component Analysis of TI-Curves: Three Methods Compared. *Food Quality and Preference, 5,* 121-127.

Dijksterhuis, G. e W. Heiser (1995). The Role of Permutation Tests in Exploratory Multivariate Data Analysis. *Food Quality and Preference, 6,* 263-270.

Dorfman, D. D. e K. S. Berbaum (1986). RSCORE-J: Pooled Rating-Method Data: A Computer Program for Analysing Pooled ROC Curves. *Behavior Research Methods, Instruments and Computers, 18,* 452-462.

DuBose, C. N., A. V. Cardello e O. Maller (1980). Effects of Colorants and Flavorants on Identification, Perceived Flavor Intensity and Hedonic Quality of Fruit-Flavored Beverages and Cake. *Journal of Food Science, 45,* 1393-1415.

Duncan, A. J. (1950). A Chi-Square Chart for Controlling a Set of Percentages. *Industrial Quality Control, 7,* 11-15.

Duncan, A. J. (1986). *Quality Control and Industrial Statistics* (5ª ed.). Homewood, Illinois: Irwin.

Durier, C., H. Monod e A. Bruetschy (1997). Design and Analysis of Factorial Sensory Experiments with Carry-Over Effects. *Food Quality and Preference, 8,* 141-149.

Ennis, D. (1993). The Power of Sensory Discrimination Methods. *Journal of Sensory Studies, 8,* 353-370.

Ennis, D. e J. Bi (1998). The Beta-Binomial Model: Accounting for Inter-trial Variation in Replicated Difference and Preference Tests. *Journal of Sensory Studies, 13,* 389-412.

Ennis, D., H. Boelens, H. Haring e P. Bowman (1982). Multivariate Analysis in Sensory Evaluation. *Food Technology, ,* 83-90.

EFQM (1999). *The EFQM Excellence Model.* Bruxelas: European Foundation for Quality Management.

Everitt, B. e G. Dunn (1991). *Applied Multivariate Data Analysis.* London: Edward Arnold.

Ferreira, A. (1998). *Caracterisation du Vieillissement du Vin de Porto: Approche Chimique et Statistique Role Aromatique du Sotolon.* Tese de Doutoramento. Université Victor Segalen Bordeaux 2.

Freitas, V., P. Ramalho, Z. Azevedo e A. Macedo (1999). Identification of Some Volatile Descriptors of the Rock-Rose-Like Aroma of Fortified Red Wines from Douro Demarcated Region. *Journal of Agriculture and Food Chemistry, 47,* 4327-4331.

Freund, R. e W. Wilson (1998). *Regression Analysis: Statistical Modeling of a Response Variable.* London: Academic Press.

Gatchalian, M., S. Leon e T. Yano (1990). Quantified Approach to Sensory Panelist Selection. *Food Quality and Preference, 2,* 233-241.

Gay, C. e R. Mead (1992). A Statistical Appraisal of the Problem of Sensory Measurement. *Journal of Sensory Studies, 7,* 205-228.

Gessner, G., N. Maholta, W. Kamakura e M. Zmijewski (1988). Estimating Models with Binary Dependent Variables: Some Theoretical and Empirical Observations. *Journal of Business Research, 16,* 49-65.

Geyer, P.L., S. H. Steiner e G. O. Wesolowsky (1996). Optimal SPRT and CUSUM Procedures with Compressed-Limit Gauges. *IIE Transactions, 28,* 489-496.

González-Viñas, M. Salvador e M. Cabezudo (1996). Taste Group Thresholds and Sensory Evaluation of Spanish Wine Vinegars. *Journal of Sensory Studies, 11,* 129-140.

González-Viñas, M. Salvador e P. Martin-Alvarez (1998). Comparison of Two Simple Methods for the Measurement of Detection Thresholds for Basic, Umami and Metallic Tastes. *Journal of Sensory Studies, 13,* 299-314.

Greenwood, M. e H. Gaunt (1994). *Total Quality Management for Schools*. New York: Cassel.

Guimarães, R. e J. Cabral (1998). *Estatística*. Lisboa: Mc Graw Hill.

Hair, J., R. Anderson, R. Tatham e W. Black (1998). *Multivariate Data Analysis* (5ª ed.). New Jersey: Prentice Hall.

Hanley, J. e B. McNeil (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology, 143,* 29-36.

Hanley, J. e B. McNeil (1983). A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. *Radiology, 148,* 839-843.

Harrison, M., K. Stephens e A. Godfrey (1986). *Modern Methods for Quality Control and Improvement*. New York: John Wiley & Sons.

Hautus, M. e R. Irwin (1995). Two Models for Estimating the Discriminability of Foods and Beverages. *Journal of Sensory Studies, 10,* 203-215.

Hayes, B. E. (1992). *Measuring Customer Satisfaction: Development and Use of Questionnaires. Livro*. Milwaukee: ASQC Quality Press.

Hoop, W. e M. Sturgis (2000). Quiting Manufacturing Due Dates Subject to a Service Level Constraint. *IIE Transactions, 32,* 771-784.

Hosmer, D. W. e S. Lemeshow (1989). *Applied Logistic Regression*. New York: Wiley.

Hough, G., E. Martinez, A. Contarini, T. Barbieri e M. Vega (1995). Selection of Assessors Based on Their Skill in Identifying Basic Tastes in Low Concentration Solutions. *Journal of Sensory Studies, 10,* 1-14.

Huang, Y. e H. Lawless (1998). Sensivity of the ABX Discrimination Test. *Journal of Sensory Studies, 13,* 229-239.

Huitson, A. (1989). Problems with Procrustes Analysis. *Journal of Applied Statistics, 16,* 39-45.

Hunter, A. (1996). Experimental Design. Em Naes, T. e E. Risvik (Eds.), *Multivariate Analysis of Data in Sensory Science* (p. 37-69). Amsterdam: Elsevier.

Hunter, A., L. Kennedy, J. Henry e I. Ferguson (2000). Application of Neural Networks and Sensitivity Analysis to Improved Prediction of Trauma Survival. *Computer Methods and Programs in Biomedicine, 62,* 11-19.

Hunter, A., G. Dijksterhuis, E. Qannari e J. MacFie (1995). Second Sensometrics Meeting - Edinburgh, 16-18 September 1994: Introduction on Behalf of the Organising Committee. *Food Quality and Preference, 6,* 215-216.

Hunter, J. S. (1996). Just What Does an EWMA Do? (Part 2). *ASQC Statistics Division Newsletter, 16,* 4-12.

IPQ (1994). *Projecto de Norma Portuguesa - prNP 4263 - Análise Sensorial Vocabulário.* Lisboa: Instituto Português da Qualidade.

IPQ (1996a). *Vocabulário Internacional de Metrologia. Termos Fundamentais e Gerais* (2ª ed.). Lisboa: Instituto Português da Qualidade.

IPQ (1996b). *Guia para a Expressão da Incerteza de Medição nos Laboratórios de Calibração.* Lisboa: Instituto Português da Qualidade.

ISO (2000). *Final Draft International Standard ISO/FDIS 9001:2000.* Geneva: International Organization for Standardization.

Irwin, R. e J. Stillman (1996). Note on the Independence of Sensivity and Response Bias in Detection Theory. *Journal of Sensory Studies, 11,* 81-84.

Ishii, R., A. Vié e M. O'Mahony (1992). Sensory Difference Testing: Ranking R-Indices Are Greater than Rating R-Indices. *Journal of Sensory Studies, 7,* 57-61.

Jiang, W., K. Tsui e W. Woodall (2000). A New SPC Monitoring Method: The ARMA Chart. *Technometrics, 42,* 399-410.

Juran, J. M. e F. Gryna (1988). *Juran's Quality Control Handbook* (4ª ed.). New York: McGraw Hill.

Kallithraka, S. e J. Bakker (1997). Evaluation of Bitterness and Astringency of (+)-Catechin and (-)-Epicatechin in Red Wine and in Model Solution. *Journal of Sensory Studies, 12,* 25-37.

Kallithraka, S., J. Bakker e M. Clifford (1997). Red Wine and Model Wine Astringency as Affected by Malic and Lactic Acid. *Journal of Food Science, 62,* 416-420.

Kelley, D. (1999). *How to Use Control Charts for Healthcare.* Milwaukee, Wisconsin: ASQ Quality Press.

Kim, K. e M. O'Mahony (1998). A New Approach to Category Scales of Intensity I: Traditional Versus Rank-Rating. *Journal of Sensory Studies, 13,* 241-249.

Kim, K., D. Ennis e M. O'Mahony (1998). A New Approach to Category Scales of Intensity II: Use of *d'* Values. *Journal of Sensory Studies, 13,* 251-267.

King, B. e P. Arents (1991). A Statistical Test of Consensus Obtained from Generalized Procrustes Analysis of Sensory Data. *Journal of Sensory Studies, 6,* 37-48.

Koning, A. e R. Does (2000). CUSUM Charts for Preliminary Analysis of Individual Observations. *Journal of Quality Technology, 32,* 122-132.

Kotseridis, Y. e R. Baumes (2000). Identification of Impact Odorants in Bordeaux Red Grape Juice, in the Commercial Yeast Used for Its Fermentation, and in the Produced Wine. *Journal of Agriculture and Food Chemistry, 48,* 400-406.

Kydoniefs, L. e C. Lindblom (2000). Using Calibration In an Office Environment. *Quality Progress, 33,* 67-73.

Lawless, H. e H. Heymann (1999). *Sensory Evaluation of Food: Principles and Practices.* Maryland, USA: Aspen Publishers.

Lim, P. C. e N. K. Tang (2000). A Study of Patients' Expectations and Satisfaction in Singapore Hospitals. *International Journal of Health Care Quality, 13,* 290-299.

Liu, R. (1995). Control Charts for Multivariate Processes. *Journal of the American Statistical Association, 90,* 1380-1387.

Lowry, C. A. e D. C. Montgomery (1995). A Review of Multivariate Control Charts. *IIE Transactions, 27,* 800-810.

Lu, Chao-Wen e M. Reynolds, Jr. (1999). Control Charts for Monitoring the Mean and Variance of Autocorrelated Processes. *Journal of Quality Technology, 31,* 259-274.

Lucas, J. e R. Crosier (2000). Fast Initial Response for CUSUM Quality-Control Schemes: Give Your CUSUM a Head Start. *Technometrics, 42,* 102-107.

Luceño, A. e J. Puig-Pey (2000). Evaluation of the Run-Length Probability Distribution for CUSUM Charts: Assessing Chart Performance. *Technometrics, 42,* 411-416.

Lundahl, D. e M. McDaniel (1988). The Panelist Effect - Fixed or Random?. *Journal of Sensory Studies, 3,* 113-121.

Maga, J. A. (1974). Influence of Color on Taste Thresholds. *Chemical Senses and Flavor, 1,* 115-119.

Marchand, S., G. de Revel e A. Bertrand (2000). Approaches to Wine Aroma: Release of Aroma Compounds from Reactions between Cystein and Carbonyl Compounds in Wine. *Journal of Agriculture and Food Chemistry, 48,* 4890-4895.

Marcucci, M. (1985). Monitoring Multinomial Processes. *Journal of Quality Technology, 17,* 86-91.

Mason, R., C. Champ, N. Tracy, S. Wierda e J. Young (1997). Assessment of Multivariate Process Control Techniques. *Journal of Quality Technology, 29,* 140-143.

McEwan, J. e P. Schlich (1992). Correspondence Analysis in Sensory Evaluation. *Food Quality and Preference, 3,* 23-36.

Meilgaard, M. (1991). Testing for Sensory Threshold of Added Substances. *American Society of Brewing Chemists Journal, 49,* 128-135.

Meilgaard, M., G. Civille e B. Carr (1987a). *Sensory Evaluation Techniques, Vol.* 1. Boca Raton, Florida: CRC Press.

Meilgaard, M., G. Civille e B. Carr (1987b). *Sensory Evaluation Techniques, Vol.* 2. Boca Raton, Florida: CRC Press.

Meiselman, H. (1993). Critical Evaluation of Sensory Techniques. *Food Quality and Preference, 4,* 33-40.

Melvin, D., M. Niranjan, R. Prager, A. Trull e V. Hughes (2000). Neuro-Computing Versus Linear Statistical Techniques Applied to Liver Transplant Monitoring: A Comparative Study. *IEEE Transactions on Biomedical Engineering, 47,* 1036-1043.

Metz, C. (1978). Basic Principles of ROC Analysis. *Seminars in Nuclear Medicine, 8,* 283-298.

Montgomery, D. (2000). *Introduction to Statistical Quality Control* (4ª ed.). New York: John Wiley & Sons.

Moskowitz, H. e B. Jacobs (1988). Magnitude Estimation: Scientific Background and Use in Sensory Analysis. Em Moskowitz, H. (Ed.), *Applied Sensory Analysis of Food, Vol.* I (p. 193-223). Boca Raton, Florida: CRC Press.

Naes, T. e E. Risvik (Eds.) (1996). *Multivariate Analysis of Data in Sensory Science.* Amsterdam: Elsevier.

Neilson, A., V. Ferguson e D. Kendall (1988). Profile Methods: Flavor Profile and Profile Attribute Analysis. Em Moskowitz, H. (Ed.), *Applied Sensory Analysis of Food, Vol.* I (p. 21-41). Boca Raton, Florida: CRC Press.

O'Mahony, M. (1982). Some Assumptions and Difficulties with Common Statistics for Sensory Analysis. *Food Technology, ,* 75-82.

O'Mahony, M. (1988). Sensory Difference and Preference Testing: The Use of Signal Detection Measures. Em Moskowitz, H. (Ed.), *Applied Sensory Analysis of Food, Vol.* I (p. 145-175). Boca Raton, Florida: CRC Press.

O'Mahony, M. (1992). Understanding Discrimination Tests: a User-Friendly Treatment of Response Bias, Rating and Ranking R-Index Tests and Their Relationship to Signal Detection. *Journal of Sensory Studies, 7,* 1-47.

O'Mahony, M. (1995). Who Told You the Triangle Test Was Simple?. *Food Quality and Preference, 6,* 227-238.

O'Mahony, M., S. Masouka e R. Ishii (1994). A theoretical note on Difference Tests: Models, Paradoxes and Cognitive Strategies. *Journal of Sensory Studies, 9,* 247-272.

Pangborn, R. M. (1984). Sensory Techniques of Food Analysis. Em D. W. Gruenwedel e J. R. Whitacker (Eds), *Food Analysis Principles and Techniques, Vol.* 1 (p. 37-93), New York: Marcel Dekker.

Peynaud, E. e J. Blouin (1997). *O Gosto do Vinho.* Lisboa: Litexa Editora.

Piggott, J. (1995). Design Questions in Sensory and Consumer Science. *Food Quality and Preference, 6,* 217-220.

Priso, H., M. Danzart e J. Hossenlopp (1994). A Statistical Analysis of Difference Tests With Replications. *Journal of Sensory Studies, 9,* 121-130.

Reynolds, M. R. Jr. (1996). Variable-Sampling-Interval Control Charts with Sampling at Fixed Times. *IIE Transactions, 28,* 497-510.

Reynolds, M. e Z. Stoumbos (1999). A CUSUM Chart for Monitoring a Proportion When Inspecting Continuously. *Journal of Quality Technology, 31,* 87-108.

Reynolds, M. e Z. Stoumbos (2000). A General Approach to Modeling CUSUM Charts for a Proportion. *IIE Transactions, 32,* 515-535.

Rifkin, M., E. Zerhouni, C. Gatsonis, L. Quint, D. Paushter, J. Epstein, U. Hamper, P. Walsh e B. McNeil (1990). Comparison of Magnetic Resonance Imaging and Ultrasonography in Staging Early Prostate Cancer. *The New England Journal of Medicine, 323,* 621-626.

Riskey, D. (1988). Uses and Abuses of Category Scales in Sensory Measurement. Em Moskowitz, H. (Ed.), *Applied Sensory Analysis of Food, Vol.* I (p. 177-191). Boca Raton, Florida: CRC Press.

Risvik, E. (1996). Understanding Latent Phenomena. Em Naes, T. e E. Risvik (Eds.), *Multivariate Analysis of Data in Sensory Science* (p. 5-35). Amsterdam: Elsevier.

Roberts, S. W. (2000). Control Chart Tests Based on Geometric Moving Averages. *Technometrics, 42,* 97-101.

Rocha, S., I. Delgadillo, A. Correia, A. Barros e P. Wells (1998). Application of an Electronic Aroma Sensing System to Cork Stopper Quality Control. *Journal of Agricuture and Food Chemistry, 46,* 145-151.

Rocha, S., I. Delgadillo, P. Barros, H. Alvelos e J. A. Cabral (1999). Characterising Different Types of Port Wine Using an Electronic Aroma Sensing Equipment and Chemical and Colour Analyses. *Proceedings of the 24th OIV Mundial Congress, Section II, Vol. 2*, Mainz, 217-224.

Roes, K. e D. Dorr (1997). Implementing Statistical Process Control in Service Processes. *International Journal of Quality Science, 2,* 149-166.

Rogerson, F. e V. Freitas (2001a). Grape Spirit Aroma Contribution to the Complexity of a Single Varietal Port Wine. *Abstracts of ASEV 52nd Annual Meeting, San Diego, California,* 58.

Rogerson, F. e V. Freitas (2001b). Quantitative Analysis of Volatiles Present in Wine Spirits Used for Port Production. *Abstracts of ASEV 52nd Annual Meeting, San Diego, California,* 59.

Rogerson, F., H. Castro, N. Fortunato, Z. Azevedo, A. Macedo e V. Freitas (2001). Chemicals with Sweet Aroma Descriptors Found in Portuguese Wines from the Douro Region: 2,6,6-Trimetylcyclohex-2-ene-1,4-dione and Diacetyl. *Journal of Agriculture and Food Chemistry, 49,* 263-269.

Rousseau, B. e M. O'Mahony (1997). Sensory Difference Tests: Thurstonian and SSA Predictions for Vanilla Flavored Yogurts. *Journal of Sensory Studies, 12,* 127-146.

Rousseau, B., A. Meyer e M. O'Mahony (1998). Power and Sensitivity of the Same-Different Test: Comparison with Triangle and Duo-Trio Methods. *Journal of Sensory Studies, 13,* 149-173.

Runger, G. C. (1996). Multivariate Statistical Process Control for Autocorrelated Processes. *International Journal of Production Research, 34,* 1715-1724.

Runger, G. C. e T. Willemain (1996). Batch-Means Control Charts for Autocorrelated Data. *IIE Transactions, 28,* 483-487.

Sastri, T., J. B. Valdés e B. Flores (1996). Nonlinear Control Charts for Jump Detection. *International Journal of Production Research, 34,* 1023-1044.

Scheffé, H. (1952). An Analysis of Variance for Paired Comparisons. *Journal of the American Statistical Association, 47,* 381.

Schlich, P. (1993). Use of Change-Over Designs and Repeated Measurements in Sensory and Consumer Studies. *Food Quality and Preference, 4,* 223-235.

Serel, D. A., H. Moskowitz e J. Tang (2000). Univariate $\overline{X}$ Control Charts for Individual Characteristics in a Multinormal Model. *IIE Transactions, 32,* 1115-1125.

Shore, H. (2000a). General Control Charts for Variables. *International Journal of Production Research, 38,* 1875-1897.

Shore, H. (2000b). General Control Charts for Attributes. *IIE Transactions, 32,* 1149-1160.

Sparks, R. (2000). CUSUM Charts for Signalling Varying Location Shifts. *Journal of Quality Technology, 32,* 157-171.

SPSS for Windows Standard Version (1996). Versão 7.5.1. SPSS Inc., USA.

Statistica (1999). Versão 5.5. StatSoft, Inc., USA.

Steiner, S. (1999). EWMA Control Charts with Time-Varying Control Limits and Fast Initial Response. *Journal of Quality Technology, 31,* 75-86.

Steiner, S. H., P. L. Geyer e G. O. Wesolowsky (1996). Grouped Data-Sequential Probability Ratio Tests and Cumulative Sum Control Charts. *Technometrics, 38,* 230-237.

Steinhaus, M. e P. Schieberle (2000). Comparison of the Most Odor-Active Compounds in Fresh and Dried Hop Cones (Humulus lupulus L. Variety Spalter Select) Based on GC-Olfactometryand Odor Dilution Techniques. *Journal of Agriculture and Food Chemistry, 48,* 1776-1783.

Stillman, J. e R. Irwin (1995). Advantages of the Same-different Method over the Triangular Method for the Measurement of Taste Discrimination. *Journal of Sensory Studies, 10,* 261-272.

Stone, H. (1999). Sensory Evaluation: Science and Mythology. *Food Tecnology, 53,* 124.

Stone, H. e L. Sidel (1993). *Sensory Evaluation Practices*. San Diego: Academic Press.

Sun, J. e G. Zhang (2000). Control Charts Based on the Number of Consecutive Conforming Items Between Two Successive Nonconforming Items for the Near Zero-Nonconforming Processes. *Total Quality Management, 11,* 235-250.

Swets, J. (1996a). The Relative Operating Characteristic in Psycology. Em Swets, J. (Ed.), *Signal Detection Theory and ROC Analysis in Psycology and Diagnostics - Collected Papers* (p. 7-30). New Jersey: Lawrence Erlbaum Associates.

Swets, J. (1996b). Form of Empirical ROCs in Discrimination and Diagnostic Tasks. Em Swets, J. (Ed.), *Signal Detection Theory and ROC Analysis in Psycology and Diagnostics - Collected Papers* (p. 31-58). New Jersey: Lawrence Erlbaum Associates.

Swets, J. (1996c). Indices of Discrimination or Diagnostic Accuracy. Em Swets, J. (Ed.), *Signal Detection Theory and ROC Analysis in Psycology and Diagnostics - Collected Papers* (p. 59-96). New Jersey: Lawrence Erlbaum Associates.

Swets, J. (1996d). Computer Programs for Fitting ROCs. Em Swets, J. (Ed.), *Signal Detection Theory and ROC Analysis in Psycology and Diagnostics - Collected Papers* (p. 293-294). New Jersey: Lawrence Erlbaum Associates.

Swets, J.A., R. Pickett, S. Whitehead, D. Getty, J. Schnur, J.B. Swets e B. Freeman (1979). Assesment of Diagnostic Technologies. *Science, 205,* 753-759.

Tatum, L. G. (1996). Control Charts for the Detection of a Periodic Component. *Technometrics, 38,* 152-160.

Tepper, B. e T. Kuang (1996). Perception of Fat in a Milk Model System Using Multidimensional Scaling. *Journal of Sensory Studies, 11,* 175-190.

Vavra, T. G. (1997). *Improving Your Measurement of Customer Satisfaction: A Guide to Creating, Conducting, Analysing and Reporting Customer Satisfaction Measurement Programs*. Milwaukee, Wisconsin: ASQ Quality Press.

Wakeling, I., M. Raats e H. MacFie (1992). A New Significance Test for Consensus in Generalized Procrustes Analysis. *Journal of Sensory Studies, 7,* 91-96.

Wardell, D. G. e M. R. Candia (1996). Statistical Process Monitoring of Customer Satisfaction Survey Data. *Quality Management Journal, 3,* 36-50.

Wheeler, D. (1995). Just What Does an EWMA Do? (Part 1). *ASQC Statistics Division Newsletter, 15,* 6-13.

Williams, A. e G. Arnold (1985). A Comparison of the Aromas of Six Coffes Characterised by Conventional Profiling, Free-Choice Profiling and Similarity Scaling Methods. *Journal of Science, Food and Agriculture, 36,* 204-214.

Williams, A. e S. Langron (1984). The Use of Free-Choice Profiling for the Evaluation of Commercial Ports. *Journal of Science, Food and Agriculture, 35,* 558-568.

Woodall, W. H. e E. V. Thomas (1995). Statistical Process Control with Several Components of Common Cause Variability. *IIE Transactions, 27,* 757-764.

Wu, Z. e T. Spedding (2000). A Synthetic Control Chart for Detecting Small Shifts in the Process Mean. *Journal of Quality Technology, 32,* 32-38.

Yarnold, J. K. (1970). The Minimum Expectation in χ^2 Goodness of Fit Tests and the Accuracy of Approximations for the Null Distribution. *Journal of the American Statistical Association, 65,* 864-886.

Zamora, M. e A. Calvino (1996). A Comparison of Methodology Applied to the Selection of a Panel for Sensory Evaluation of Instant Coffee. *Journal of Sensory Studies, 11,* 211-226.

Zanella, A. (1998). A Statistical Model for the Analysis of Customer Satisfaction: Some Theoretical and Simulation Results. *Total Quality Management, 9,* 599-609.

Zondervan, C., S. Muresan, H. de Jonge, E. Van Velzen, C. Wilkinson, H. Nijhuis e T. Leguijt (1999). Controlling Maillard Reactions in the Heating Process of Blockmilk Using an Electronic Nose. *Journal of Agriculture and Food Chemistry, 47,* 4746-4749.

Apêndice A

Ficha de Prova Utilizada no IVP no Início do Estudo

Instituto do Vinho do Porto

Direcção de Serviços Técnicos

FICHA DE PROVA INDIVIDUAL

Processo Nº / /

Produto: Vinho ---------- Outros ----------		**Tipo**:	
Cor:	**Doçura:**	**Ano/Idade:**	**Designação Complementar:**

01 Limpidez
- 01 límpido
- 02 Ligeiramente Enevoado
- 03 Enevoado
- 04 Turvo

02 Cor
- 01 Retinto
- 02 Tinto
- 03 Tinto Alourado
- 04 Alourado
- 05 Alourado Claro
- 06 Branco Pálido
- 07 Branco Palha
- 08 Branco Dourado
- 09 Outra:

03 Espuma
- 01 Vermelha
- 02 Avermelhada
- 03 Rosada
- 04 Esbranquiçada
- 05 Branca

04 Aroma
- 01 Com defeito perceptível
- 02 Sem defeito perceptível
- 03 Característico da designação
- 04 Não característico da designação
- 05 Compatível com a designação complementar
- 06 Não compatível com a designação complementar

05 Sabor
- 01 Com defeito perceptível
- 02 Sem defeito perceptível
- 03 Característico da designação
- 04 Não característico da designação
- 05 Compatível com a designação complementar
- 06 Não compatível com a designação complementar

06 Defeito
- 01 Acetificado
- 02 Mofo
- 03 Prensas
- 04 Sabão
- 05 Ranço
- 06 Casco
- 07 Aroma atípico
- 08 Sabor atípico
- 09 Sulfídrico
- 10 Apetrolado
- 11 Defeito não caracterizado
- 12 Notado dióxido de enxofre
- 98 Outro: _____

07 Prova de comparação

Referência _____
- 01 Semelhante
- 02 Diferente devido a: _____
- 03 Não se torna possível a comparação

08 Qualidade

01 Muito Boa	**9**	**10**
02 Boa	**7**	**8**
03 Suficiente	**5**	**6**
04 Insuficiente	**3**	**4**
05 Má	**2**	
06 Muito Má	**1**	

09 Idade
- 01 Correspondente à indicada
- 02 Não corresponde à indicada

10 Designação
- 01 Corresponde a Vinho do Porto
- 02 Não corresponde a Vinho do Porto devido a: _____
- 03 Corresponde à designação indicada
- 04 Não corresponde à designação indicada devido a: _____
- 05 Compatível com a desig. Complementar indicada
- 06 Não compatível com a desig. Complementar indicada devido a: _____

Observações:

_____

_____

Data: / /	**Assinatura**:	**Nº**

Apêndice B

Tabelas para os Parâmetros *H* e *F* da Versão Tabelada das Cartas CUSUM

Tabela BI – Esquemas CUSUM para número de não conformidades (adaptada de Harrison et al. (1986), considerando o esquema C2)

Taxa Média de Ocorrências	Intervalo de Decisão	Parâmetro F
c	H	
0.25	3.0	0.25
0.32	4.0	0.18
0.40	3.0	0.60
0.50	2.0	1.00
0.64	2.0	1.36
0.80	3.5	0.70
1.00	5.0	0.50
1.25	5.0	0.75
1.60	4.0	1.40
2.00	5.0	1.00
2.50	5.0	1.50
3.20	5.0	1.80
4.00	6.0	2.00
5.00	7.0	2.00

Tabela BII – Esquemas CUSUM para número de unidades não conformes (adaptada de Harrison et al. (1986), considerando o esquema C2)

Tamanho da Amostra	Proporção	Intervalo de Decisão	Parâmetro F
N	p	H	
20	0.1	3	2
20	0.2	7	1
20	0.3	5	2
25	0.1	4	1.5
25	0.2	5	2
25	0.3	8	1.5
35	0.1	6	1.5
35	0.2	7	2
35	0.3	10	1.5
50	0.1	6	2
50	0.2	10	2
50	0.3	11	2
80	0.1	8	6
80	0.2	13	2
80	0.3	12	3

BIBLIOTECA F.E. PORTO

FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO
BIBLIOTECA
0000085099