

RITA DE CÁSSIA DOS SANTOS NAVARRO DA SILVA

**PERFIL DESCRITIVO OTIMIZADO: NÚMERO DE JULGADORES,  
DELINEAMENTOS E VALIDAÇÃO**

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

VIÇOSA  
MINAS GERAIS – BRASIL  
2013

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

T

S586p  
2013

Silva, Rita de Cássia dos Santos Navarro da, 1986-  
Perfil descritivo otimizado : número de julgadores,  
delineamentos e validação / Rita de Cássia dos Santos Navarro  
da Silva. – Viçosa, MG, 2013.  
xvi, 117 f. : il. ; 29 cm.

Inclui anexos.

Orientador: Valéria Paula Rodrigues Minim.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Alimentos-Avaliação sensorial. 2. Alimentos-Análise.  
3. Tecnologia de alimentos. 4. Tempos, Estudo de.  
I. Universidade Federal de Viçosa. Departamento de Tecnologia  
de Alimentos. Programa de Pós-Graduação em Ciência e  
Tecnologia de Alimentos. II. Título.

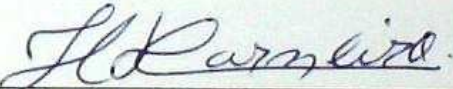
CDD 22 ed. 664.072

PERFIL DESCRITIVO OTIMIZADO: NÚMERO DE JULGADORES,  
DELINEAMENTOS E VALIDAÇÃO

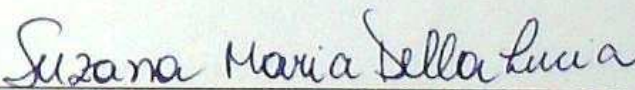
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, para obtenção do título de *Doctor Scientiae*.


APROVADA: 25 de outubro de 2013.

  
Prof.<sup>a</sup> Aline Cristina Arruda Gonçalves

  
Prof. Joel Camilo Souza Carneiro

  
Prof. Moysés Nascimento

  
Prof.<sup>a</sup> Suzana Maria Della Lucia  
(Co-orientadora)

  
Prof.<sup>a</sup> Valéria Paula Rodrigues Minim  
(Orientadora)

*“O coração do homem considera o  
seu caminho, mas o Senhor lhe  
dirige os passos”.*  
*Provérbios 16 v. 9*

## AGRADECIMENTOS

A Deus, pela vida, pela saúde, pelas oportunidades concedidas, pela força na execução deste trabalho e pela vitória de concluí-lo.

À Universidade Federal de Viçosa, em especial, ao Departamento de Tecnologia de Alimentos, pela oportunidade da minha formação acadêmica.

Ao Conselho Nacional Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa, sem ela teria sido impossível!

À professora Valéria Paula Rodrigues Minim. Não tenho palavras para lhe agradecer pela atenção e confiança despendida durante este trabalho. Agradeço por confiar no meu trabalho e nos meus julgamentos. Agradeço pela amizade construída nesses seis anos de trabalho e convivência. Agradeço pela sua rigorosidade em avaliar meu trabalho, a qual, com certeza, contribuiu para me tornar uma profissional melhor. Obrigada pela oportunidade!

Aos meus co-orientadores, professor Luis Antônio Minim e professora Suzana Maria Della Lucia, pela atenção, conselhos, ensinamentos e ajuda fundamental para o desenvolvimento deste trabalho. Agradeço ainda à professora Suzana pela disponibilidade do laboratório para coleta parcial dos dados deste trabalho.

Aos professores Aline Cristina Arruda Gonçalves, Moysés Nascimento e Joel Camilo de Souza Carneiro, pela pronta disponibilidade na participação da banca de defesa. Agradeço ainda aos professores Aline Arruda e João de Deus pela disponibilização dos laboratórios para a realização de parte deste trabalho

Aos meus pais, Virgínia e Erly, por sempre terem sido mais do que presentes na minha vida, me ajudando em tudo que podiam e até mesmo no que não podiam. Deus abençoe muito vocês! Agradeço pela criação, pela paciência comigo e por saber que posso sempre contar com vocês para qualquer dificuldade! Amo muito vocês!

Ao meu esposo, Alexandre, pelo amor, carinho e paciência comigo; por sempre ter sido meu ombro amigo e com quem sempre pude contar e confiar. Agradeço pela ajuda imensa neste trabalho, pelas idéias (que foram muitas!!!), pela ajuda na execução das análises e por ler cada linha que eu escrevi...(rsrs) Agradeço pelo seu companheirismo de sempre! Te amo muito!

À minha princesinha linda, minha irmã Cristina. Agradeço por você existir! Agradeço pela sua amizade e carinho, você é muito importante para mim. Te amo muito!

À minha vózinha, Natália, por sempre ter cuidado de mim com tanto carinho e amor. Ao meu querido avô, José Maria, que tanto amo e nunca esquecerei. Ah, como eu queria que você estivesse aqui.... Saudades!

Aos meus sogros, Carlos e Conceição, pelo apoio, carinho e por sempre torcerem por mim.

À minha segunda família, Luiz Carlos e Liette, por terem me acolhido quando mais precisei, pelo carinho e paciência. Aos meus “pseudo-irmãos”: Talita, Marcus (e Belinha) e Verônica, pela amizade.

À equipe do laboratório: Márcia, Ana Cristina, Andréa, Naiara e Tarcísio, pela amizade e por terem tornado essa fase mais suave e feliz.

Em especial, aos meus estagiários Liliane, Andréa, Raissa, Aline, Pedro, Jociele e Vinícius pela dedicação sem limites, sem vocês teria sido muito mais difícil!

Em especial, agradeço aos meus professores e colegas do Departamento de Estatística. Obrigada pelo excelente convívio, pelos ensinamentos e pela amizade construída.

A todas as equipes de julgadores (Viçosa, Alegre, Sete Lagoas e Lavras) que dedicaram com carinho e responsabilidade à realização das análises sensoriais, meus sinceros agradecimentos.

A todos que de alguma forma contribuíram para que este sonho se tornasse realidade.

## BIOGRAFIA

RITA DE CÁSSIA DOS SANTOS NAVARRO DA SILVA, filha de Erly Cabral dos Santos e Maria Virgínia Felix dos Santos, nasceu em 08 de dezembro de 1986 na capital de São Paulo.

Iniciou seus estudos na escola municipal Maria Yolanda Souza Pinto Hame - SP, em 1990. Em 1992, transferiu-se para a Escola Estadual Júlio Maia. Em 1998 estudou na Escola Estadual Imperatriz Leopoldina onde concluiu o ensino médio.

Em março de 2004, ingressou na Universidade Federal de Viçosa – MG, onde graduou-se em Ciência e Tecnologia de Laticínios em julho de 2008.

Em agosto do mesmo ano, iniciou o curso de Mestrado no programa de Pós-graduação em Ciência e Tecnologia de Alimentos na Universidade Federal de Viçosa, o qual concluiu em Julho de 2010.

Em Agosto do mesmo ano, iniciou o curso de Doutorado no programa de Pós-graduação em Ciência e Tecnologia de Alimentos na Universidade Federal de Viçosa. Em 2012, concomitante ao curso de doutorado, foi professora substituta do Departamento de Estatística da Universidade Federal de Viçosa. No presente ano de 2013, no mês de Outubro, conclui o seu doutoramento.

# ÍNDICE

LISTA DE TABELAS .....	x
LISTA DE FIGURAS .....	xii
RESUMO .....	xiii
ABSTRACT .....	xv
INTRODUÇÃO GERAL .....	1
<b>CAPÍTULO 1 – REVISÃO DE LITERATURA: Métodos Descritivos: Evolução e</b>	
<b>Atualidades .....</b>	<b>5</b>
1. Análise sensorial descritiva .....	5
2. Perfil Convencional .....	6
3. Metodologias alternativas para a descrição rápida de alimentos .....	10
3.1. Ranking Descriptive Analysis (RDA) .....	10
3.2. Flash Profile (FP) .....	11
3.3. Napping® .....	12
3.4. Free Sorting Task (FST) .....	14
3.5. Check-all-that-apply (CATA) .....	16
3.6. Pivot Profile (PP) .....	17
3.7. Polarised Sensory Position (PSP) .....	17
3.8. Perfil Descritivo Otimizado .....	18
4. Considerações finais .....	20
Referências Bibliográficas .....	21
<b>CAPÍTULO 2 – ARTIGO: Número de julgadores necessários em testes sensoriais</b>	
<b>descritivos .....</b>	<b>27</b>
<b>Resumo .....</b>	<b>27</b>
1. Introdução .....	28
2. Como determinar o número ideal de julgadores? .....	29
3. Identificando os parâmetros .....	30
3.1. Probabilidades dos erros tipo I e tipo II .....	30
3.2. Diferença desejada entre as médias ( $d'$ ) .....	31
3.3. Estimativa do desvio padrão do erro experimental (Raiz do Quadrado Médio do Erro - RMSE) .....	32

4. Cálculo do número de julgadores em testes descritivos .....	35
5. Conclusão .....	39
Referências Bibliográficas .....	39
<b>CAPÍTULO 3 – ARTIGO: Perfil Descritivo Otimizado: Quantos julgadores são necessários?</b>	
necessários? .....	43
Resumo .....	43
1. Introdução .....	44
2. Materiais e Métodos .....	46
2.1. Estímulos sensoriais .....	46
2.2. Procedimento .....	47
2.3. Análise dos dados .....	48
2.3.1. Simulação computacional .....	48
2.3.2. Critérios .....	48
2.3.3. Critérios de decisão .....	50
3. Resultados .....	51
3.1. ANOVA e mapa sensorial: painel completo .....	51
3.2. Critério I: estimativa do erro experimental .....	53
3.3. Critério II: significância do efeito de tratamento .....	54
3.4. Critério III: similaridades dos mapas sensoriais .....	55
4. Discussão .....	56
5. Conclusão .....	58
Referências Bibliográficas .....	58
<b>CAPÍTULO 4 – ARTIGO: Delineamento em Blocos Incompletos Balanceados: uma alternativa para coleta de dados no Perfil Descritivo Otimizado</b>	
uma alternativa para coleta de dados no Perfil Descritivo Otimizado .....	60
Resumo .....	60
1. Introdução .....	62
2. Apresentação dos Delineamentos .....	63
2.1. Delineamento em Blocos Incompletos Balanceados (DBIB) .....	63
2.2. Delineamento em Blocos Completos Balanceados (DBCB) .....	64
3. Materiais e Métodos .....	65
3.1. Estímulos sensoriais .....	65
3.2. Procedimento da Análise Sensorial .....	66
3.2.1. Perfil Descritivo Otimizado: avaliação por meio do Delineamento em	

Blocos Incompletos Balanceados .....	67
3.2.2. Perfil Descritivo Otimizado: avaliação por meio do Delineamento em Blocos Completos Balanceados .....	68
3.2.3. Perfil Convencional: avaliação por meio do Delineamento em Blocos Completos Balanceados .....	69
3.3. Análises estatísticas .....	70
3.3.1. Comparação entre as técnicas/delineamentos .....	71
3.3.1.1. Comparação entre as configurações espaciais .....	71
3.3.1.2. Comparação entre os modelos de regressão .....	71
3.3.4. Estudo dos sub-grupos de amostras no DBIB .....	72
4. Resultados .....	72
4.1. Comparação entre as técnicas/delineamentos (PDO/DBIB, PDO/DBCB e PC/DBCB) .....	72
4.1.1. Perfil sensorial dos sucos de uva .....	73
4.1.2. ANOVA: DBIB e DBCB .....	75
4.1.3. Poder de discriminação .....	75
4.1.4. Influência dos teores de polpa e açúcar no perfil sensorial: modelos de regressão .....	77
4.2. DBIB: comparação entre os sub-grupos de amostras .....	79
5. Discussão .....	82
6. Conclusão .....	85
Referências Bibliográficas .....	85
<b>CAPÍTULO 5 – ARTIGO: Validação do Perfil Descritivo Otimizado: exatidão, precisão e robustez .....</b>	<b>87</b>
Resumo .....	87
1. Introdução .....	88
2. Materiais e Métodos .....	90
2.1. Estímulos sensoriais .....	91
2.2. Painéis sensoriais .....	91
2.3. Procedimento da análise sensorial .....	92
2.3.1. Perfil Descritivo Otimizado .....	92
2.3.2. Perfil Convencional .....	94
2.4. Mensuração dos critérios de validação .....	94
2.4.1. Exatidão .....	94

2.4.2. Precisão .....	95
2.4.2.1. Repetibilidade .....	95
2.4.2.2. Reprodutibilidade .....	96
2.4.3. Robustez .....	97
3. Resultados .....	98
3.1. Exatidão .....	98
3.2. Precisão .....	100
3.2.1. Repetibilidade .....	100
3.2.2. Reprodutibilidade .....	102
3.3. Robustez .....	104
4. Discussão .....	107
5. Conclusão .....	111
Referências Bibliográficas .....	111
CONCLUSÃO GERAL .....	114
<b>ANEXOS</b> .....	116

# LISTA DE TABELAS

## CAPÍTULO 2

Tabela 2.1 - Erros de decisão tipo I e tipo II .....	30
Tabela 2.2 - Valores de Raiz do Quadrado Médio do Resíduo (RMSE) encontrados na literatura considerando diferentes países, produtos alimentícios, atributos sensoriais e número de julgadores .....	34
Tabela 2.3 - Número de avaliações necessárias para testes descritivos considerando as probabilidades dos erros tipo I ( $\alpha$ ) e tipo II ( $\beta$ ) e RMSEL, para $d'=0,20$ .....	37
Tabela 2.4 - Número de avaliações necessárias para testes descritivos considerando as probabilidades dos erros tipo I ( $\alpha$ ) e tipo II ( $\beta$ ) e RMSEL, para $d'=0,10$ .....	37
Tabela 2.5 - Número de avaliações necessárias para testes descritivos considerando as probabilidades dos erros tipo I ( $\alpha$ ) e tipo II ( $\beta$ ) e RMSEL, para $d'=0,05$ .....	38
Tabela 2.6 - Número de avaliações necessárias para testes descritivos considerando as probabilidades dos erros tipo I ( $\alpha$ ) e tipo II ( $\beta$ ) e RMSEL, para $d'=0,02$ .....	38

## CAPÍTULO 3

Tabela 3.1 - Formulações-teste de iogurte sabor morango. Quantidade de ingredientes adicionados a um litro (1 L) de iogurte comercial .....	47
Tabela 3.2 - Composição dos chocolates em relação ao tipo e quantidade do chocolate utilizado no processamento .....	47
Tabela 3.3 - Atributos sensoriais descritores das formulações-teste .....	48
Tabela 3.4 - Resumo da ANOVA para os dados do painel completo .....	51
Tabela 3.5 - Percentual de atendimento dos sub-grupos ao critério da estimativa do erro experimental .....	54
Tabela 3.6 - Percentual de atendimento dos sub-grupos ao critério de significância do efeito de tratamento .....	55
Tabela 3.7 - Percentual de atendimento dos sub-grupos ao critério de similaridade dos mapas sensoriais .....	55

## CAPÍTULO 4

Tabela 4.1 - Formulações dos sucos de uva .....	66
Tabela 4.2 - Delineamento em Blocos Incompletos Balanceados, considerando $t=5$ , $k=3$ , $r=6$ , $\lambda=3$ , $b=10$ .....	68
Tabela 4.3 - ANOVA para comparação das técnicas/ delineamentos .....	73

Tabela 4.4 - Resumo da ANOVA obtida para cada técnica/delineamento .....	76
Tabela 4.5 - Modelos de regressão múltipla ajustados para cada atributo nas diferentes técnicas/delineamentos .....	77
Tabela 4.6 - Teste de médias Tukey ( $\alpha = 0,05$ ) para os escores médios obtidos nos diferentes blocos do DBIB .....	80
Tabela 4.7 - Comparação entre as técnicas/delineamentos em relação ao número de etapas e tempo requerido .....	84
 <b>CAPÍTULO 5</b>	
Tabela 5.1 - Tipos de produtos alimentícios utilizados para mensurar os critérios de validação .....	91
Tabela 5.2 - Atributos sensoriais avaliados nos diferentes produtos alimentícios .....	93
Tabela 5.3 - Resumo da ANOVA para verificar a significância do efeito das repetições das avaliações ( <i>ej</i> ) .....	101
Tabela 5.4 - Coeficientes RV na comparação das configurações individuais das repetições das avaliações dos painéis de julgadores semi-treinados (PDO) .....	102
Tabela 5.5 - Resumo da análise de variância conjunta para avaliação da reprodutibilidade de resultados do PDO (precisão inter-laboratorial) .....	103
Tabela 5.6 - Médias dos atributos que apresentaram significância no efeito de laboratórios ( $\alpha = 0,05$ ) .....	103
Tabela 5.7 - Coeficientes RV na comparação das configurações individuais do perfil sensorial das amostras obtidas nos diferentes laboratórios, utilizando a técnica PDO .....	104
Tabela 5.8 - Percentual de sub-painéis que atenderam ao critério ao intervalo de confiança, considerando a matriz alimentar chocolate .....	105
Tabela 5.9 - Percentual de sub-painéis que atenderam ao critério ao intervalo de confiança, considerando a matriz alimentar iogurte de morango .....	106

# LISTA DE FIGURAS

## CAPÍTULO 1

Figura 1.1 - Exemplo da avaliação de cinco amostras por meio da técnica Napping Profile. (a) alocação das amostras na folha tamanho A3 (60 x 40 cm). (b) medição das coordenadas xy .....	13
Figura 1.2 - Representação do mapa sensorial obtido na técnica <i>Ultra-Flash Profile</i> .....	14
Figura 1.3 - Representação do procedimento de avaliação das amostras no Free Sorting Task (a) e da tabulação dos dados (b) .....	15
Figura 1.4 - Representação do procedimento de associação de atributos às amostras na metodologia Free Sorting Task .....	15
Figura 1.5 - Ficha de avaliação das amostras na metodologia Pivot Profile .....	18

## CAPÍTULO 2

Figura 2.1 - Representação do $d'$ no teste descritivo (a) e do modelo Thurstoniano (b) .....	31
Figura 2.2 - Histograma dos dados de RMSEL da literatura e a distribuição de probabilidades de Weibull ajustada aos dados .....	35

## CAPÍTULO 3

Figura 3.1 - Mapas sensoriais obtidas pela Análise de Componentes Principais. (a) iogurtes sabor morango (b) chocolates .....	53
---	----

## CAPÍTULO 4

Figura 4.1 - Mapas sensoriais dos sucos de uva para o PDO no BIBD (a) e no BCBD (b) e para o Perfil Convencional no DBC com protocolo de avaliação monádico (c) .....	74
Figura 4.2. Representação dos intervalos de confiança (IC95%) dos parâmetros dos modelos lineares para as três técnicas/delineamentos .....	78
Figura 4.3 - PCA considerando os diferentes blocos do DBIB .....	79

## CAPÍTULO 5

Figura 5.1 - Representação gráfica no mapa sensorial consensual entre as equipes do PDO e do PC, obtidas pela APG, para as matrizes requeijão (a), suco de uva (b) e chocolate (c) .....	100
--	-----

## RESUMO

SILVA, Rita de Cássia dos Santos Navarro da, D.Sc., Universidade Federal de Viçosa, outubro de 2013. **Perfil descritivo otimizado: número de julgadores, delineamentos e validação.** Orientadora: Valéria Paula Rodrigues Minim. Co-orientadores: Luis Antônio Minim e Suzana Maria Della Lucia.

O Perfil Descritivo Otimizado (PDO) é uma nova metodologia de descrição sensorial que foi proposta recentemente com o objetivo de reduzir o tempo do teste sensorial. No PDO a avaliação dos atributos sensoriais é realizada de forma quantitativa por meio da avaliação de um painel semi-treinado. A avaliação da intensidade dos atributos em uma escala intervalar possibilita a identificação da magnitude de diferença entre os produtos, aumentando a aplicabilidade do método e o destacando entre as técnicas sensoriais rápidas. A proposta do presente estudo foi validar a técnica PDO como metodologia descritiva, além de determinar as condições experimentais ideais para a aplicação do método, sendo abordados estudos de determinação do número de julgadores e de comparação entre delineamentos experimentais. Na determinação do número ideal de julgadores, foi realizado um estudo de simulação computacional, utilizando a técnica de re-amostragem de dados de um painel original, considerando 10.000 sub-grupos com reposição. Os critérios considerados foram: (i) obtenção de um erro experimental menor ou igual ao erro verificado na metodologia referência (Perfil Convencional), (ii) discriminação das amostras semelhantemente à discriminação obtida pela equipe completa, e (iii) mínima perda de informação no mapa sensorial. Para a avaliação do critério (i) foi realizada uma coleta na literatura de valores de erro experimental obtidos em estudos prévios com o Perfil Convencional, sendo estabelecido um “ponto de corte” para a determinação do número ideal de julgadores no PDO. Na análise dos resultados, foi verificado que, no mínimo, dezesseis julgadores devem compor a equipe sensorial para atendimento do PDO ao critério da magnitude do erro experimental, sendo os demais critérios atendidos plenamente quando este número de julgadores foi utilizado. No estudo de delineamentos experimentais, foi avaliada a viabilidade da utilização do Delineamento em Blocos Incompletos Balanceados (DBIB) para coleta de dados segundo o protocolo do PDO. Para tanto, um mesmo conjunto de amostras foi avaliado por três painéis independentes, sendo que cada painel avaliou as amostras segundo uma metodologia e delineamento específico. Dois painéis avaliaram as amostras segundo a

técnica do PDO, sendo uma equipe conduzida sob o Delineamento em Blocos Completos Balanceados (DBCB) e outra equipe sob o DBIB. A terceira equipe avaliou as amostras por meio da técnica Perfil Convencional (PC), utilizando uma equipe treinada. Os escores médios obtidos nas diferentes técnicas/delineamentos não apresentaram diferença significativa pelo teste F ( $p > 0,05$ ), considerando os diferentes atributos sensoriais. Foi verificado elevado grau de similaridade entre os mapas sensoriais originados pelas diferentes técnicas/delineamentos, apresentando coeficiente RV superior a 0,95. O DBIB se apresentou como uma alternativa para coleta de dados no PDO sem perda na qualidade dos resultados. No estudo de validação da metodologia PDO foram mensuradas a exatidão, a precisão (repetibilidade e reprodutibilidade) e a robustez. A exatidão dos dados gerados pelo PDO foi mensurada por meio da comparação com o método referência (PC), considerando diferentes matrizes alimentares. A precisão foi mensurada em nível de painel (repetibilidade) e inter-laboratorial (reprodutibilidade). Na robustez foi avaliada a sensibilidade do método quanto à redução do número de julgadores no painel. Na mensuração dos critérios de validação, a metodologia PDO apresentou atendimento aos critérios estabelecidos, apresentando grau de proximidade com a metodologia referência superior a 95%. Na avaliação da repetibilidade de medições, considerando três repetições das avaliações por um mesmo painel, o PDO apresentou igualdade estatística entre as repetições ( $p > 0,05$ ) e com grau de proximidade superior a 99%. O método apresentou reprodutibilidade dos dados quando os laboratórios foram alterados, apresentando combinação perfeita entre os perfis sensoriais. O PDO se mostrou robusto mesmo reduzindo o número de julgadores, sendo que painéis menores apresentaram variação aleatória ainda menor que o painel completo. O PDO foi validado com sucesso e apresentou as medidas de validação com elevado grau de certeza, considerando diferentes matrizes alimentares.

## ABSTRACT

SILVA, Rita de Cássia dos Santos Navarro da, D.Sc., Universidade Federal de Viçosa, October 2013. **Optimized descriptive profile: number of judges, designs and validation.** Adviser: Valéria Paula Rodrigues Minim. Co-advisers: Luis Antônio Minim and Suzana Maria Della Lucia.

The Optimized Descriptive Profile (ODP) is a new methodology for sensory description that was recently proposed in order to reduce the time of sensory testing. In the ODP, the evaluation of sensory attributes is performed by means of quantitative evaluation by a semi-trained panel. Evaluation of attributed intensity on an interval scale permits identification of the magnitude of difference between products, increasing applicability of the method as well as its relevance among fast sensory techniques. The purpose of this study was to validate the ODP technique as a descriptive methodology, as well as determine the optimal experimental conditions for application of the method, including studies on determining the number of judges and comparison of experimental designs. In determining the ideal number of judges, a computer simulation study was conducted, using the re-sampling technique for data from an original evaluating panel, considering 10,000 sub-groups with replacement. The criteria considered were: (i) obtaining an experimental error less than or equal to the error found in the reference method (Conventional Profile), (ii) discrimination of samples as discriminated obtained by the complete panel, and (iii) minimal loss of information in the sensory map. For evaluation of criterion (i) a literature review was performed to collect experimental error values obtained in previous studies with the Conventional Profile, establishing a “cut-off” for determining the optimal number of judges in the ODP. In analyzing the results, it was found that at least sixteen judges should make up the sensory panel so that the ODP meets the criteria for magnitude of the experimental error, where the remaining criteria were fully met when this number of judges was used. In the study of experimental designs, was evaluated the feasibility of using a Balanced Incomplete Block Design (BIBD) to collect data according to the protocol of the ODP. Thus, an identical set of samples was evaluated by three independent teams, where each team evaluated the samples using a specific methodology and design. Two panels evaluated the samples using the ODP technique, where one panel used the Balanced Complete Block Design (BCBD) and the other the BIBD. A third panel evaluated the samples according to the

Conventional Profile technique (CP) using a trained judges panel. The mean scores obtained in the different techniques/designs showed no significant difference by the F-test ( $p > 0.05$ ), considering the different sensory attributes. A high degree of similarity was found between the sensory maps obtained by different techniques/designs, presenting RV coefficients greater than 0.95. The DBIB is presented as an alternative to collecting data in the ODP without loss in quality of the results. In the validation study of the ODP methodology, accuracy, precision (repeatability and reproducibility) and robustness were measured. Accuracy of the data generated by the ODP was measured by comparison with the reference method (CP), considering different food matrices. Precision was measured at the panel (repeatability) and inter-laboratory levels (reproducibility). Robustness considered evaluation of the sensitivity of the method with regards to reducing the number of judges making up the panel. In measuring of the validation criteria, the ODP methodology met the established criteria, presenting a degree of proximity with the reference methodology greater than 95%. In evaluation of measurement repeatability, considering three repetitions of the evaluations by a single panel, the ODP showed statistical equality among the repetitions ( $p > 0.05$ ) with degree of proximity exceeding 99%. This method also showed reproducibility of the data when laboratory were altered, presenting a perfect combination among sensory profiles. The ODP showed to be robust in reducing the number of judges, where smaller teams presented less random variation than the complete panel. It was also successfully validated and presented validation measures with a high degree of certainty, considering different food matrices.

# *Introdução Geral*

Metodologias descritivas convencionais consistem na descrição e quantificação das propriedades sensoriais dos alimentos por meio da avaliação sensorial de uma equipe de julgadores treinados. Nesse tipo de avaliação, após o levantamento e a descrição dos atributos sensoriais, os julgadores treinados também quantificam os estímulos percebidos em uma escala de intensidade (Murray, Delahunty, & Baxter, 2001).

A descrição quantitativa dos alimentos proporciona informações importantes do perfil sensorial possibilitando a sua utilização no controle de qualidade, na comparação de protótipos de alimentos com o intento de entender o comportamento do consumidor em relação aos atributos sensoriais e para o mapeamento sensorial de produtos (Gacula, 1997). Também podem ser utilizadas para acompanhar as mudanças do produto ao longo do tempo em relação aos prazos de validade e aos efeitos de embalagem, ou ainda, para investigar os efeitos dos ingredientes ou variáveis de processamento na qualidade sensorial final de um produto (Murray, Delahunty, & Baxter, 2001). Segundo Bleibaum et al. (2002), uma das principais contribuições da análise descritiva é a sua capacidade de possibilitar a correlação entre as medidas sensoriais e instrumentais, além de permitir segmentar os consumidores de acordo com as suas preferências por meio da correlação dos dados de descrição sensorial com os testes afetivos.

Por outro lado, as metodologias descritivas convencionais apresentam longo tempo de execução da técnica sensorial, uma vez que requerem treinamento extensivo dos julgadores para se certificar de que o vocabulário e as escalas de avaliação são utilizados de forma consistente e que o painel apresenta consenso, habilidade de discriminar as amostras e repetibilidade de resultados. A fase de treinamento é extensiva e o tempo demandado apresenta-se como um limitante para sua aplicação no contexto prático da indústria (Rodrigue et al., 2000; Dairou & Sieffermann, 2002; Delarue & Sieffermann, 2004; Labbe, Ritz, & Rugi, 2004; Cartier et al., 2006; Perrin et al., 2008; Richter et al., 2010; Silva et al., 2012; 2013). Segundo Delarue e Sieffermann (2004), a principal limitação da análise descritiva convencional é o tempo consumido e, muitas

vezes, isto leva a não execução desta valiosa ferramenta de qualidade. Este fato ressalta a necessidade de uma avaliação rápida no contexto prático da indústria.

Diversos pesquisadores têm trabalhado no intuito de desenvolver metodologias alternativas aos métodos descritivos convencionais, as quais necessitem de menor tempo para obtenção do perfil sensorial, de forma a viabilizar sua utilização na indústria de alimentos. Encontram-se na literatura recente metodologias que utilizam o procedimento de ordenação e de classificação para avaliar os produtos em relação aos atributos descritores, objetivando facilitar o procedimento de avaliação pelos julgadores. Nesses métodos é necessária somente a comparação da intensidade dos produtos em vez da “calibração” da memória sensorial. Metodologias que utilizam esse processo de classificação/ordenação foram propostas recentemente, tais como: Perfil *Flash* (Dairou & Sieffermann, 2002; Delarue & Sieffermann, 2004), *Napping* (Pagès, 2005); *Free Sorting Task* (Cartier et al., 2006), *Check-all-that-apply* (Adams et al., 2007; Ares et al., 2010), *Pivot profile* (Thuillier, 2007), *Ultra-Flash* (Perrin et al., 2008), *Polarised Sensory Position* (Teillet et al., 2010), *Ranking Descriptive Analysis* (Richter et al., 2010) e variações destes métodos.

Essas técnicas têm apresentado uma redução satisfatória no tempo do teste de descrição sensorial, porém há limitações na sua aplicabilidade devido à avaliação qualitativa das amostras. Por meio das técnicas alternativas apresentadas, não é possível identificar a magnitude da diferença entre as amostras e, portanto, esses métodos não são recomendados para análises de estabilidade e controle de qualidade.

O Perfil Descritivo Otimizado (PDO), proposto por Silva et al. (2012), tem o objetivo de suprir a demanda por métodos descritivos rápidos e, ao mesmo tempo, fornecer informações quantitativas sobre os atributos sensoriais presentes nos alimentos. Nesta metodologia, foi proposto que julgadores semi-treinados avaliassem as amostras em uma escala de intensidade não estruturada ancorada nos extremos pelos materiais de referência (nenhum/fraco e forte). Foi recomendada a utilização do protocolo atributo-por-atributo para avaliação dos produtos-teste, de forma que fosse avaliado apenas um atributo por vez. Desta forma, no momento da avaliação das amostras, os materiais de referência dos atributos sensoriais são apresentados juntamente com todas as amostras para possibilitar que os julgadores comparem as amostras entre si e também com as referências, facilitando a alocação da intensidade dos atributos na escala não

estruturada. Esta estruturação do teste sensorial possibilitou que julgadores com baixo grau de “calibração” avaliassem a intensidade dos produtos de forma consistente.

Na proposta da metodologia, a técnica foi aplicada na descrição sensorial de cinco formulações de requeijão cremoso, os quais diferiam entre si nas concentrações de gordura e água. Foi verificado que o PDO permitiu a obtenção de um perfil sensorial muito semelhante à avaliação da equipe treinada, utilizando o Perfil Convencional, além de possibilitar a redução de 54% do tempo do teste sensorial, sem perda de informações. No PDO, é enfocada a caracterização quantitativa dos atributos sensoriais, possibilitando realizar análises de estabilidade, controle de qualidade, otimização de formulações e correlação entre medidas sensoriais e instrumentais.

Neste contexto, o objetivo deste trabalho é validar a metodologia PDO, além de determinar o número ideal de julgadores para compor a equipe sensorial e verificar a viabilidade da utilização do Delineamento em Blocos Incompletos Balanceados (DBIB) para coleta dos dados no Perfil Descritivo Otimizado. Assim, esta tese foi dividida em cinco capítulos (artigos), sendo o primeiro capítulo de revisão de literatura. No segundo e terceiro capítulos, a determinação do número ideal de julgadores é abordada, sendo que um estudo global para métodos descritivos convencionais é apresentado no capítulo dois e um estudo específico para o PDO é trazido no capítulo três. O capítulo quatro é apresentado um estudo comparativo de delineamentos, sendo estudada a viabilidade do DBIB na coleta de dados do PDO. E, por fim, no capítulo cinco é abordada a validação da técnica PDO, apresentando medidas de exatidão, precisão (repetibilidade e reprodutibilidade) e robustez.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks. In: *7th Pangborn Sensory Science Symposium*. Minneapolis, USA, 12-16, August, 2007.
- Ares, G., Deliza, R., Barreiro, C., Giménez, A., & Gámbaro, A. (2010). Comparison of two sensory profiling techniques based on consumer perception. *Food Quality and Preference*, 21, 417-426.
- Bleibaum, R.N., Stonea, H., Tanb, T., Labrecheb, S., Saint-Martin, E., & Isz, S. (2002). Comparison of sensory and consumer results with electronic nose and tongue sensors for apple juices. *Food Quality and Preference*, 13, 409–422.
- Cartier, R., Ritz, A., Lecomte, A., Poblete, F., Krystlik, J., Belin, E., & Martin, N. (2006). Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference*, 17, 562–571.

- Dairou, V., & Sieffermann, J. M. (2002). A comparison of 14 jams characterized by conventional profile and a quick original method, the Flash Profile. *Journal of Food Science*, 67, 826–834.
- Delarue, J., & Sieffermann, J.M. (2004). Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference*, 15, 383–392.
- Gacula, M. C. (1997). *Descriptive sensory analysis in practice*. Trumbull, CT: Food and Nutrition Press.
- Labbe, D., Ritz, A., & Hugi, A. (2004). Training is a critical step to obtain reliable product profiles in a real food industry context. *Food Quality and Preference*, 15, 341–348.
- Murray, J.M., Delahunty, C.M., & Baxter, I.A. (2001). Descriptive sensory analysis: past, present and future. *Food Research International*, 34, 461-471.
- Pagès, J. (2005). Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley. *Food Quality and Preference*, 16, 642–649.
- Perrin, L., Symoneaux, R., Maître, I., Asselin, C., Jourjon, F., & Pagès, J. (2008). Comparison of three sensory methods for use with the Napping procedure: Case of tem wines from Loire valley. *Food Quality and Preference*, 19, 1–11.
- Richter, V. B., Almeida, T. C. A., Prudencio, S. H., & Benassi, M.T. (2010). Proposing a ranking descriptive sensory method. *Food Quality and Preference*, 21, 611 - 620.
- Rodrigue, N., Guillet, M., Fortin, J., & Martin, J.F. (2000). Comparing information obtained from ranking and descriptive tests of four sweet corn products. *Food Quality and Preference*, 11 (1), p. 47-54.
- Silva, R. C. S. N., Minim, V. P. R., Simiqueli, A. A., Moraes, L. E. S., Gomide, A. I. & Minim, L. A. (2012). Optimized Descriptive Profile: a rapid methodology for sensory description. *Food Quality and Preference*, 24, 190-200.
- Teillet, E., Schlich, P., Urbano, C., Cordelle, S. & Guichard, E. (2010). Sensory methodologies and the taste of water. *Food Quality and Preference*, 21, 967–976.
- Thuillier, B. (2007). *Rôle du CO<sub>2</sub> dans l'Appréciation Organoleptique des Champagnes – Expérimentation et Apports Méthodologiques*. Reims, France: Thèse de l'URCA.



# *Métodos Descritivos: Evolução e Atualidades*

## **1. Análise sensorial descritiva**

Testes sensoriais descritivos tradicionais envolvem a detecção (discriminação) e a descrição qualitativa e quantitativa dos atributos sensoriais dos alimentos, por meio da avaliação de uma equipe de julgadores treinados (Meilgaard, Civille e Carr, 2006). Os aspectos qualitativos de um alimento incluem os atributos que compõem a aparência, o aroma, o sabor e a textura do produto. Nas metodologias descritivas, após o levantamento e a descrição dos atributos sensoriais, os julgadores treinados também quantificam os estímulos percebidos (Murray, Delahunty e Baxter, 2001).

A principal importância da análise descritiva é a sua capacidade de permitir a correlação entre medidas sensoriais e instrumentais e/ou a segmentação dos consumidores de acordo com a sua preferência por meio da correlação da avaliação descritiva com testes afetivos (Bleibaum et al., 2002; Meigaard, Civille e Carr, 2006). Análises descritivas também são utilizadas no controle de qualidade, na comparação de protótipos de alimentos visando entender o comportamento do consumidor em relação aos atributos sensoriais e para o mapeamento sensorial de produtos (Gacula, 1997). Também podem ser utilizadas para acompanhar as mudanças do produto ao longo do tempo em relação aos prazos de validade e aos efeitos de embalagem, ou ainda, para investigar os efeitos dos ingredientes ou variáveis de processamento na qualidade sensorial final de um produto (Murray, Delahunty e Baxter, 2001).

Existem vários métodos de análise descritiva, incluindo o método do Perfil de Sabor (Cairncross e Sjostrom, 1950), Perfil de Textura (Brandt, Skinner e Coleman, 1963), Análise Descritiva Quantitativa (Stone et al., 1974), Spectrum (Meilgaard, Civille e Carr 1991), Perfil Livre (Langron, 1983; Thompson e MacFie, 1983) e metodologias genéricas. Os métodos específicos refletem diversas abordagens sensoriais (Lawless e Heymann, 1998), no entanto, a análise descritiva genérica, que

pode combinar diferentes abordagens de todos estes métodos, é frequentemente empregada durante a aplicação prática no intuito de atender aos objetivos específicos do estudo (Murray, Delahunty e Baxter, 2001).

## **2. Perfil Convencional**

A metodologia descritiva Perfil Convencional representa uma técnica descritiva genérica, a qual combina diferentes abordagens dos métodos tradicionais, resultando assim em uma combinação destes. Na análise descritiva genérica todas as etapas do teste sensorial, desde o recrutamento dos voluntários até a análise dos dados, devem ser conduzidas de acordo com as recomendações da literatura para métodos descritivos. Metodologias genéricas foram utilizadas por Dairou e Sieffermann (2002), Delarue e Sieffermann (2004), Blancher et al. (2007), Sinésio et al. (2010) e Silva et al. (2012) e foram denominadas “Perfil Convencional”.

As metodologias descritivas exigem, em sua maioria, uma equipe com algum grau de treinamento ou orientação e os julgadores necessitam ter um nível razoável de acuidade sensorial. Para alcançar este objetivo, recruta-se de 2 a 3 vezes o número de julgadores necessários para a realização do teste sensorial. Os voluntários são geralmente recrutados com base em alguns critérios, tais como: disponibilidade de tempo, condições de saúde, não apresentação de alergias, avaliação da personalidade, criatividade verbal, concentração, motivação, não-fumante, hábitos alimentares, experiência prévia, não utilização de próteses dentárias e medicamentos que possam interferir nas análises. Após o recrutamento, estes julgadores são pré-selecionados com base na capacidade discriminatória de produtos (Murray, Delahunty e Baxter, 2001; Stone e Sidel, 2004).

A personalidade do julgador tem um grande impacto sobre o sucesso ou fracasso da equipe sensorial. Piggott e Hunter (1999) discutem os benefícios de elaborar procedimentos de pré-seleção (Lesschaeve e Issanchou, 1996), mas não necessariamente prevêem a habilidade do julgador em executar as avaliações. Testes de concentração e de personalidade, juntamente com a criatividade verbal podem ser os melhores preditores de capacidade futura (Wechsler, 1944). Questionários dietéticos abrangentes (por exemplo, questionários de frequência alimentar) também podem ser úteis no levantamento dos hábitos alimentares dos julgadores. Informações sobre

relutância para comer algo desconhecido (experimental, por exemplo) podem ser relevantes na caracterização sensorial de determinados alimentos, a qual pode ser avaliada por meio da escala de neofobia alimentar (Pliner e Hobden, 1992).

A fase de treinamento das técnicas descritivas inicia com o desenvolvimento de uma linguagem comum que descreva de forma abrangente e precisa os atributos do produto. Geralmente, a nova equipe irá desenvolver a terminologia sensorial por si só, porém a participação de um líder experiente pode auxiliar o processo de aprendizagem. Uma linguagem já existente pode também ser adaptada por um novo painel, embora se esta foi desenvolvida por outro laboratório, ou em um diferente país ou outra região, dificuldades na compreensão e interpretação dos termos podem ocorrer. A solução para este problema é a utilização das definições completas dos atributos e o estabelecimento de materiais de referência para ancorar as intensidades (nenhum/fraco e forte) de cada atributo sensorial (Hunter e McEwan, 1998).

A seleção dos atributos para inclusão na lista final de avaliação é geralmente um procedimento de consenso. O líder do painel também pode influenciar o processo de seleção dos descritores incentivando ou enfatizando certos atributos que têm sido relatados na literatura (no entanto, isso nem sempre é necessário) (Murray, Delahunty e Baxter, 2001). Técnicas para o levantamento de atributos são encontradas na literatura, tais como: o método da lista prévia, associação controlada, discussão aberta (Damásio e Costell, 1991) e o método-rede (Moskowitz, 1983). Murray (1999) sugeriu a utilização de um método menos subjetivo para a seleção dos atributos, o qual procede a uma avaliação quantitativa da adequação dos termos ao produto-teste, eliminando conceitos semelhantes. A linguagem descritiva final deve ser precisamente definida e conter termos o suficiente para incluir todos os atributos susceptíveis de serem encontrados no alimento-teste, mas não deve ser tão grande a ponto de ser inviável (Piggott e Canaway, 1981).

Uma vez selecionados os termos, a equipe de julgadores é treinada para utilizar um conjunto comum de referências, as quais ilustram e definem os atributos e também ancoram sua intensidade. Isso geralmente é conseguido por meio da exposição da equipe a uma gama de produtos na categoria do alimento sob avaliação (Munoz e Civille, 1998). Durante o treinamento, os julgadores descrevem as percepções sensoriais de forma qualitativa com suas próprias palavras e, quantitativamente, utilizando suas experiências anteriores para classificar as intensidades. Julgadores treinados, no entanto, devido ao tempo de treinamento, adquirem uma linguagem qualitativa comum e

utilizam um conjunto quantitativo de referências, o que possibilita a avaliação padronizada dos produtos em uma escala de intensidade (Murray, Delahunty e Baxter, 2001). A escolha da escala de intensidade a ser utilizada depende do procedimento de avaliação a ser utilizado, ou seja, da metodologia a ser adotada. Poderá ir desde escala ordinal para o Perfil de Sabor (Cairncross e Sjostrom, 1950), até escalas não estruturadas para o *Spectrum* (Meilgaard, Civille e Carr 1991) e para a Análise Descritiva Quantitativa (Stone et al., 1974). No caso do Perfil Convencional, metodologia genérica, a escolha da escala deve ser realizada de acordo com os objetivos do estudo a ser realizado.

Muitos autores têm, portanto, recomendado a utilização dos materiais de referência para alcançar o consenso das equipes sensoriais (Civille e Lawless, 1986; Murray e Delahunty, 2000; Ritvanen et al., 2004; Melo, Bolini e Efraim, 2009; Richter et al., 2010; Silva et al., 2012), os quais são padrões tanto em termos quantitativos quanto qualitativos (Meilgaard, Civille e Carr, 2006). Padrões de referência são definidos como "qualquer produto químico, ingrediente ou alimento" (Rainey, 1986). Esta definição pode ser estendida para incluir materiais não alimentares relacionados aos estímulos sensoriais, por exemplo, grama para sabor de "gramado", papelão para aroma de "oxidado", aquarelas para cores e assim por diante (Murray, Delahunty e Baxter, 2001).

Os procedimentos de treinamento para padronização da forma de manipulação das amostras têm por objetivo evitar erros na interpretação dos atributos sensoriais e expor claramente aos julgadores a que estímulo sensorial se refere cada atributo, ou seja, esclarecer "o quê" e "como" deve ser avaliado o estímulo sensorial (Silva et al., 2012). No geral, a fase de treinamento deve ser tão extensa quanto possível. Os procedimentos adotados durante o treinamento dependerão da abordagem do método escolhido, do tempo disponível e da complexidade dos produtos em teste.

O desempenho dos julgadores pode ser avaliado por meio dos critérios de seleção propostos por Damásio e Costell (1991), os quais consistem em avaliar a habilidade dos julgadores em diferenciar os produtos e também de repetir os resultados das avaliações. Para tanto, é recomendado realizar uma simulação da avaliação dos produtos-teste pelos julgadores, devendo ser avaliados todos os atributos sensoriais previamente definidos. Desta forma, o julgador recebe os produtos na cabine individual e deve avaliar a intensidade dos estímulos em uma escala não estruturada ancorada nos extremos pelas intensidades "nenhum" ou "fraco" (âncora esquerda) e "forte" (âncora

direita). O poder de discriminação dos produtos e de repetibilidade de resultados é, então, avaliada por meio da Análise de Variância com duas fontes de variação (produto e repetição), sendo selecionado o julgador que obtiver uma probabilidade máxima de 50% para a discriminação dos produtos/formulações ( $p.F_{\text{FORMULAÇÃO}} < 0,50$ ) e probabilidade mínima de 5% para na repetibilidade ( $p.F_{\text{REPETIÇÃO}} > 0,05$ ). Estes critérios também podem sofrer ajustes conforme o interesse do pesquisador. Silva et al. (2012) utilizaram uma probabilidade máxima para o  $F_{\text{FORMULAÇÃO}}$  de 30%, com o objetivo de ser mais rigoroso na seleção da equipe quanto ao poder de discriminação dos produtos.

A avaliação dos produtos pela equipe treinada e selecionada, com o objetivo de obter o perfil sensorial destes alimentos, deve ser realizada em relação a todos os atributos sensoriais em uma escala de intensidade ancorada nos extremos (“nenhum/fraco” e “forte”). Na literatura verifica-se a recomendação/utilização de escalas de 7 pontos (Cartier et al., 2006), 9 pontos (Melo, Bolini, e Efraim, 2009; Sinésio et al., 2010; Richter et al., 2010), 10 pontos (Rodrigue et al., 2000; Monrozier et al., 2001), 12 pontos (Ginés et al., 2004) e 15 pontos (Stone e Sidel, 2004). O delineamento de apresentação dos produtos aos julgadores deve seguir um delineamento em blocos (julgador), sendo recomendado o Delineamento em Blocos Completos ou o Delineamento em Blocos Incompletos (Stone e Sidel, 2004). A escolha do delineamento deve ser realizada em função do número de amostras presente no estudo. Stone e Sidel (1974) e Ishii, Stampanoni e O’Mahony (2008) recomendam a avaliação de uma única amostra por vez (protocolo monádico serial). Agrupamento de amostras em uma mesma sessão de avaliação tem sido observado na literatura, o que pode levar o julgador a compará-las e ordená-las antes da alocação da intensidade do atributo na escala não estruturada, o que é indesejável para julgadores treinados.

Segundo Murray, Delahunty e Baxter (2001), a análise sensorial descritiva de alimentos apresenta-se como o método mais abrangente, flexível e útil dentre os métodos sensoriais, proporcionando informações detalhadas sobre todos os atributos que compõem um alimento. No decorrer deste milênio, é esperado que a análise descritiva seja cada vez mais utilizada e envolva uma faixa ainda mais ampla de aplicações do que já foi feito anteriormente. Assim, é vital o contínuo investimento no desenvolvimento de técnicas descritivas que desafiem as ideias tradicionais, a fim de assegurar que o potencial ideal seja adquirido com este método no futuro.

### **3. Metodologias alternativas para a descrição rápida de alimentos**

Na situação particular da indústria de alimentos, onde há restrições em termos de tempo, acompanhar a evolução do desempenho de um painel de julgadores acarreta em grandes dificuldades na aplicação da Análise Descritiva como ferramenta de qualidade. Segundo Labbe, Ritz e Hugi (2003), o treinamento dos julgadores é a etapa crítica para obtenção de perfis sensoriais no contexto prático da indústria. Outra situação em que os métodos rápidos são muito bem-vindos, é na caracterização sensorial de alimentos perecíveis, onde o acompanhamento de um painel treinado pode acarretar em variações significativas nas características sensoriais dos alimentos ou até mesmo na perda da vida-de-prateleira do produto antes do término das avaliações descritivas.

Neste sentido, diversos pesquisadores têm trabalhado no intuito de desenvolver metodologias alternativas à análise descritiva convencional, as quais necessitem de menor tempo demandado para obtenção do perfil sensorial dos alimentos, de forma a viabilizar sua utilização na indústria de alimentos. Encontram-se na literatura recente metodologias propostas por Rodrigue et al. (2000), Dairou & Sieffermann (2002), Delarue & Sieffermann (2004), Pagès (2005), Lawles et al. (1995), Adams et al. (2007), Ares et al. (2010), Thuillier (2007), Perrin et al. (2008), Teillet et al. (2010) e Richter et al. (2010).

#### *3.1. Ranking Descriptive Analysis (RDA)*

Rodrigue et al. (2000) propuseram a utilização do teste de ordenação, comumente utilizado como metodologia discriminatória, como teste descritivo. Foi proposto que a equipe de julgadores avaliasse as amostras utilizando a escala de ordenação em relação a cada um dos atributos descritivos. Os dados foram analisados por meio da Análise Generalizada de Procrustes (APG) para obtenção do perfil sensorial. Os autores concluíram que foi possível obter uma representação gráfica semelhante à obtida por meio da avaliação de uma equipe treinada. Porém não foi possível identificar a magnitude da diferença entre as amostras, devido à escala de avaliação ser discreta e assimétrica.

No ano de 2010, Richter et al. desenvolveram uma metodologia baseada no procedimento de ordenação proposto por Rodrigue et al. (2000), a qual foi denominada *Ranking Descriptive Analysis* (Análise Descritiva por Ordenação). Foi proposto neste método que os julgadores pré-selecionados levantassem os atributos sensoriais que

caracterizavam os pudins de chocolate e ordenassem as amostras em relação a esses atributos. Nesta metodologia, foi recomendada a utilização de referências qualitativas para evitar erro na interpretação dos atributos. Referências qualitativas não ancoram os extremos de intensidade de uma escala de avaliação, as servem para padronizar a percepção dos estímulos sensoriais e padronizar a manipulação das amostras e a forma de avaliação dos atributos. Os autores compararam o perfil sensorial obtido por meio da ordenação das amostras utilizando equipe sem treinamento (RDA) e por meio das metodologias clássicas Análise Descritiva Quantitativa e Perfil Livre. Os resultados das três metodologias foram analisados por meio da APG e os dados da RDA também foram analisados pelo teste de Friedman. A caracterização sensorial dos produtos por meio das três metodologias apresentou-se semelhante. O teste de Friedman possibilitou a análise comparativa dos produtos de forma rápida.

Silva et al. (2013) compararam os resultados obtidos por meio da técnica RDA e por meio do Perfil Convencional na caracterização sensorial de chocolates e verificaram a obtenção de perfis sensoriais muito similares. Em relação à discriminação das amostras, os autores verificaram que a metodologia RDA possibilitou menor discriminação entre os produtos, além de não possibilitar a mensuração da magnitude de diferença dos produtos na intensidade dos atributos sensoriais.

### 3.2. *Flash Profile (FP)*

Dairou e Sieffermann (2002) propuseram uma metodologia que combina o Perfil Livre e o método de ordenação, a qual foi denominada Perfil *Flash*. Nesta técnica, cada julgador possui uma lista individual de atributos e são orientados a ordenarem as amostras em relação a estes atributos. A definição de materiais de referência qualitativa é dispensável, uma vez que os julgadores avaliam as amostras somente em relação aos atributos que eles mesmos identificam e descrevem. No desenvolvimento da metodologia, os autores aplicaram o Perfil Convencional (PC), o Perfil Livre e a metodologia proposta (*Flash Profile* ou Perfil *Flash*) para descrever quatorze geléias de frutas. A metodologia descritiva Perfil Convencional representa uma técnica descritiva genérica, a qual combina diferentes abordagens dos métodos tradicionais, resultando assim em uma combinação destes. Nessa pesquisa, o PC representou uma modificação da técnica tradicional Análise Descritiva Quantitativa (Stone et al., 1974) e abordou um treinamento extensivo dos julgadores e avaliação dos produtos por meio da escala não estruturada de 10 cm ancorada nos extremos por materiais de referência. Os resultados

do PC foram analisados por meio da Análise de Componentes Principais e os dados das outras metodologias analisados por meio da APG. A caracterização sensorial obtida pelas diferentes metodologias apresentou-se muito similar. O *Flash Profile* requereu menor tempo para obtenção dos dados, apresentando-se como uma alternativa rápida para a descrição sensorial. É importante ressaltar que esta técnica não é recomendada para estudos de estabilidade e controle de qualidade, uma vez que não indica a magnitude da diferença entre os produtos. Por outro lado, devido à terminologia semântica, esta metodologia apresenta-se como uma ferramenta de comunicação entre a pesquisa, o desenvolvimento e o *marketing*.

Em 2004, Delarue e Sieffermann aplicaram a metodologia do *Flash Profile* em produtos com características sensoriais mais similares com objetivo de verificar como seria o comportamento da técnica na caracterização de produtos com diferenças de menores magnitudes. A técnica foi aplicada para caracterizar seis iogurtes de diferentes marcas comerciais. O *Flash Profile* foi comparado com as metodologias clássicas Perfil Livre e Perfil Convencional. A configuração espacial obtida pela Análise de Componentes Principais (ACP) para os dados do Perfil Convencional apresentou pouca discrepância da configuração obtida pela APG, considerando os dados do *Flash Profile*. Na comparação da metodologia proposta com o Perfil Livre, por meio da APG, o mapa descritivo apresentou-se idêntico. Este resultado confirma que esta técnica é uma alternativa interessante para a descrição sensorial de alimentos.

Após estas publicações (Dairou & Sieffermann, 2002; Delarue & Sieffermann, 2004), a metodologia tem sido aplicada para descrever goma de mascar (Delarue & Loescher, 2004), geléias de frutas (Blancher et al., 2007), pães (Lassoued et al., 2007), purês de frutas (Tarea et al., 2007), chá gelado (Veinand et al., 2011) e café solúvel (Kobayashi & Benassi, 2012).

### 3.3. *Napping*®

O *Napping Profile* é uma técnica rápida de descrição sensorial, a qual utiliza julgadores com determinado nível de acuidade sensorial. Nesta técnica, para a descrição dos produtos, os julgadores são orientados a alocar as amostras em uma folha tamanho A3 (60 x 40 cm) de acordo com suas similaridades e dissimilaridades (Figura 1.1a). Os julgadores são instruídos que dois produtos deverão ser alocados próximos se forem identificados como iguais e alocados distantes se forem diferentes. Não há mais instruções de como proceder à alocação das amostras no papel, cada julgador utiliza seu

próprio critério (Pagès, 2005). Para análise dos resultados, a alocação de cada produto é medida no sistema de coordenadas  $xy$  (Figura 1.1b). As avaliações individuais dos julgadores são submetidas à uma análise multivariada para obtenção do mapa sensorial.

A técnica apresenta quatro variantes, sendo: *Global Napping* (GN), *Partial Napping* (PN), *Sorted Napping* (SN) e, por fim, uma variante que foi denominada como uma metodologia específica, o *Ultra-Flash Profile*. No *Global Napping*, apenas uma sessão de avaliação das amostras é realizada, sendo as amostras alocadas na folha sem nenhuma restrição, representando a proposta original do método. Já na técnica *Partial Napping* (PN), são realizadas três avaliações distintas, sendo uma para a aparência dos produtos, outra para o sabor e uma última para a textura (Delholm et al., 2012).

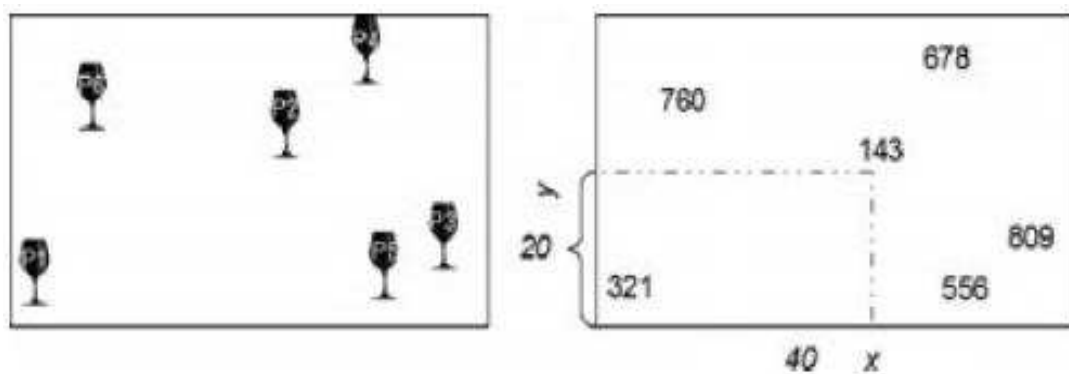


Figura 1.1 - Exemplo da avaliação de cinco amostras por meio da técnica *Napping Profile*. (a) alocação das amostras na folha tamanho A3 (60 x 40 cm). (b) medição das coordenadas  $xy$ .

Fonte: Valetin et al. (2012) modificado.

No *Sorted Napping*, após a alocação das amostras na folha de papel, os julgadores são orientados a circular as amostras, agrupando-as. Nesta técnica, a equipe determina qual distância entre as amostras pode ser considerada como significativa na formação do agrupamento das amostras (Pagès et al., 2010). Na variação da técnica denominada como *Ultra-Flash Profile*, após os julgadores terem posicionado os produtos na folha de papel, são solicitados a descrever cada produto escrevendo algumas palavras (atributos) que estão associadas aos produtos (Perrin et al., 2008). Estas palavras são escritas diretamente na folha A3 (Figura 1.2).

A técnica *Napping Profile* e suas variações têm sido utilizadas para a caracterização de diversos produtos, tais como chocolate (Kennedy & Heymann, 2009), *snacks* (Kennedy, 2010), sucos (Nestrud & Lawless, 2008), vinhos (Morand & Pagès, 2006; Pagès, 2005; Perrin & Pagès, 2009; Perrin et al., 2008), bebidas quentes

(Moussaoui & Varela, 2010), peixes (Albert et al., 2011) e maçãs (Nestrud & Lawless, 2010).

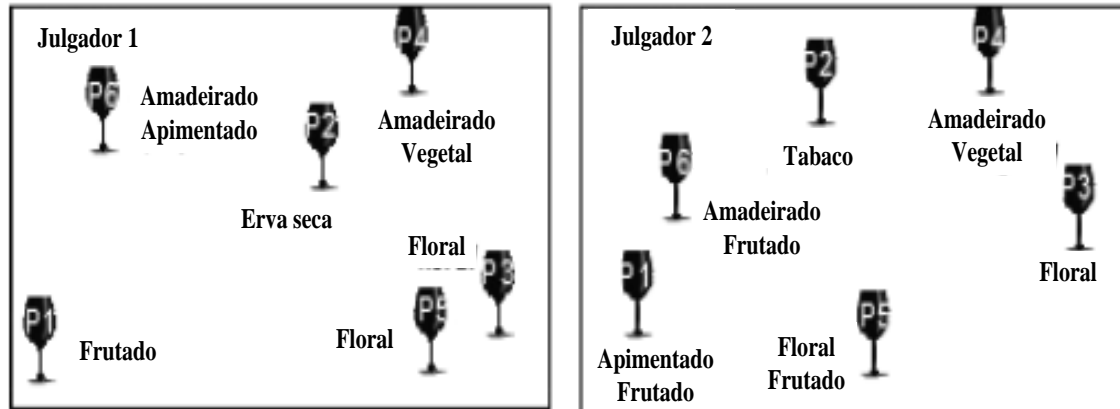


Figura 1.2 - Representação do mapa sensorial obtido na técnica *Ultra-Flash Profile*.  
Fonte: Valentin et al. (2012) modificado.

### 3.4. Free Sorting Task (FST)

A técnica *Free Sorting Task* (FST) originada da psicologia (Hulin & Katz, 1935) foi utilizada pela primeira vez na caracterização sensorial de alimentos por Lawless et al. (1995) e têm sido amplamente utilizada até os dias atuais (Lawless et al., 1995; Tang & Heymann, 1999; Saint-Eve et al., 2004; Faye et al., 2004; Lim & Lawless, 2005; Faye et al., 2006; Cartier et al., 2006; Blancher et al., 2007; Lelièvre et al., 2008; Bécue-Bertaut & Lê, 2011). Esta metodologia recomenda a utilização de um procedimento de classificação para agrupar as amostras semelhantes, consistindo de uma simples sessão. Todas as amostras são apresentadas aos julgadores de forma simultânea, os julgadores são orientados a olhar, cheirar e degustar os produtos e agrupá-los de acordo com a similaridade deles (Figura 1.3a).

Na tabulação dos resultados, é construída uma matriz de dados (Figura 1.3b), sendo atribuído o valor 1 quando dois produtos encontram-se no mesmo grupo e o valor 0 é atribuído se as amostras não estão situadas no mesmo grupo. Na Figura 1.3b observa-se que foi atribuído o valor 1 à célula que representa a junção do produto P1 e P6, indicando que estes estão alocados no mesmo grupo. Na Figura 1.3a, pode-se observar que os produtos P1 e P6 estão no mesmo grupo. Na análise dos dados, as categorias (1 e 0) atribuídas às amostras pelos diferentes julgadores são submetidas à análise de correspondência para obtenção do mapa sensorial.

Cada julgador utiliza seu próprio critério para agrupar as amostras, sendo o número de grupos formados definido individualmente, não necessitando ser definido em

consenso com a equipe (Valentin et al., 2012). Uma vez formados os grupos, os julgadores são solicitados a associarem alguns termos (atributos sensoriais) aos grupos de amostras, caracterizando-os (Figura 1.4).

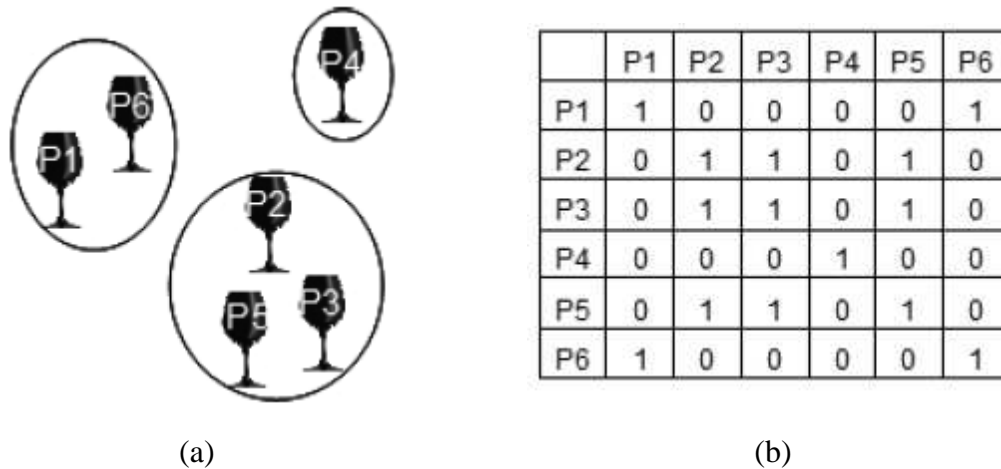


Figura 1.3 - Representação do procedimento de avaliação das amostras no *Free Sorting Task* (a) e da tabulação dos dados (b).  
 Fonte: Valentin et al. (2012) modificado.

Esta técnica têm sido utilizada na caracterização sensorial de feijões (Heymann, 1994), queijos (Lawless et al., 1995), bebidas (Falahee & MacRae, 1995, 1997; Teillet et al., 2010), geléias de frutas (Tang & Heymann, 1999; Blancher et al., 2007), cervejas (Chollet & Valentin, 2001; Abdi et al., 2007; Lelièvre et al., 2008; 2009), vinhos (Piombino et al., 2004; Ballester et al., 2005; Bécue-Bertaut & Lê, 2011), iogurtes (Saint-Eve et al., 2004), amoras (Derndorfer & Baiertl, 2006) e tomates (Deegan et al., 2010).

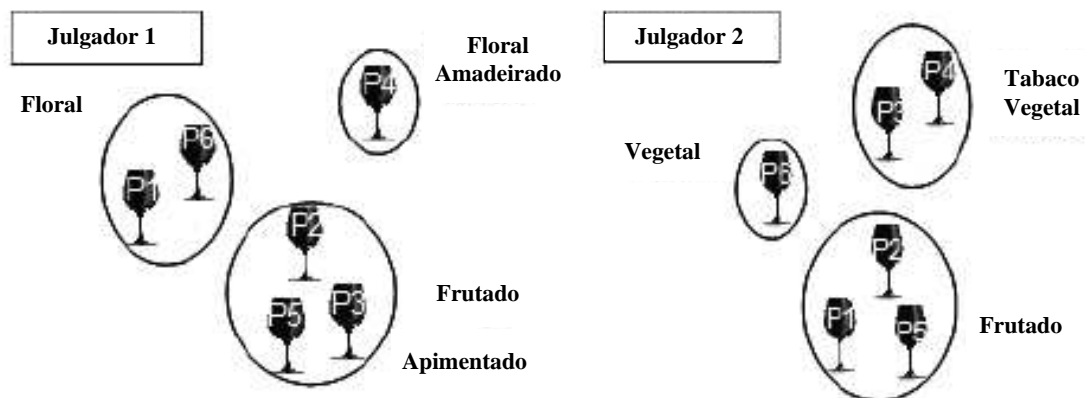


Figura 1.4 - Representação do procedimento de associação de atributos às amostras na metodologia *Free Sorting Task*.  
 Fonte: Valentin et al. (2012) modificado.

No estudo de Cartier et al. (2006) a metodologia FST foi conduzida por uma equipe sem treinamento e por uma equipe treinada, sendo obtido um mapa sensorial para cada equipe. A Análise Descritiva Quantitativa (ADQ) também foi aplicada para comparação dos dados. No estudo, foram descritos quatorze tipos de cereais matinais. Os autores observaram que ocorreram pequenas alterações na representação gráfica gerada pela ADQ e pelo método de classificação (FST). A equipe treinada e a não treinada apresentaram desempenho semelhante, sendo verificado que este procedimento pode ser realizado por equipes de julgadores não treinados sem perda de informações. Esta metodologia, dentre outras, apresenta-se como uma alternativa rápida para a descrição sensorial.

### 3.5. *Check-all-that-apply (CATA)*

A técnica *Check-all-that-apply*, comumente conhecida como CATA, é uma abordagem tradicionalmente utilizada na pesquisa de *marketing* para determinar como é realizada a percepção dos consumidores frente aos produtos, identificando quais associações os consumidores realizam ao ver/consumir um determinado produto (Driesener & Romaniuk, 2006). Recentemente esta técnica tem sido introduzida na avaliação sensorial, objetivando a otimização de formulações (Adams et al., 2007; Lancaster & Foley, 2007).

O procedimento do CATA consiste na elaboração de uma lista de atributos (palavras ou frases) que a equipe de julgadores considera apropriada para descrever os produtos em avaliação. Os produtos são apresentados monadicamente aos julgadores, os quais são solicitados a marcar os atributos que eles consideram presentes naquela amostra específica. O número de atributos selecionados é definido pelo julgador, por seu próprio critério (Valentim, 2012). Os atributos não são restringidos a aspectos sensoriais, mas também podem ser relacionados a aspectos emocionais e hedônicos (Dooley et al., 2010), procedimento que possibilita a otimização de formulações (Ares et al., 2011).

Na análise dos dados, é possível identificar a frequência que cada termo foi associado a um determinado produto e também confeccionar um mapa sensorial por meio da análise de correspondência (Valentin et al., 2012). A utilização desta técnica na caracterização descritiva de alimentos ainda é recente. Encontra-se na literatura estudos descritivos por meio do CATA para caracterizar *snacks* (Adams et al., 2007; Popper et al., 2011), morangos (Lado et al., 2010), sorvetes de baunilha (Dooley et al., 2010),

chocolate (Ares et al., 2010a,b), sucos de laranja (Ares et al., 2011) e derivados lácteos (Bruzzone et al., 2011).

### 3.6. *Pivot Profile (PP)*

A técnica *Pivot Profile* (PP) foi proposta por Thuillier (2007) na descrição sensorial de vinhos e foi construída na ideia da descrição livre, a qual consiste em permitir que os julgadores associem livremente termos sensoriais (atributos) aos produtos, sem alguma restrição.

Nesta técnica, é proposta a elaboração de um produto referência, denominado “pivot”. Assim, os julgadores são solicitados a comparar os produtos-teste com o “pivot” em relação a uma lista de atributos pré-elaborada individualmente. Na comparação são atribuídos termos de intensidade (menos intenso e mais intenso) para cada termo sensorial. As amostras são apresentadas de forma monádica aos julgadores de acordo com um delineamento balanceado. Os julgadores devem observar, cheirar ou degustar cada produto (dependendo dos objetivos específicos do estudo) e a referência, devendo escrever se o produto apresenta maior ou menor intensidade que a referência, em relação a cada atributo sensorial. Os julgadores são instruídos a utilizarem somente termos descritos, não sentenças (Valentin et al., 2012).

O produto referência é escolhido/elaborado dentro de uma gama de produtos a serem avaliados. Na análise dos dados, é realizado um re-agrupamento dos termos sensoriais, sendo agrupados termos semelhantes e, opcionalmente, agrupados em categorias. As frequências de utilização de cada termo é computada, sendo a matriz final submetida à análise de correspondência, sendo obtido o mapa sensorial. A técnica apresenta somente uma publicação científica (Thuillier, 2007), além de uma revisão de literatura (Valentin et al., 2012) e, portanto, não tem sido explorada.

### 3.7. *Polarised Sensory Position (PSP)*

A técnica *Polarised Sensory Position* (PSP) foi inicialmente proposta por Teillet et al. (2010) para definir as características sensoriais da água utilizando consumidores. A proposta da técnica é a substituição de elevado número de atributos sensoriais por poucos produtos protótipos, os quais seriam compostos por “meta-atributos”.

No PSP, os julgadores primeiramente recebem três produtos referência, sendo solicitados a avaliar os produtos-teste monadicamente em uma escala contínua em relação ao grau de similaridade existente entre os produtos-teste e os três produtos

referências (Figura 1.5). Os resultados são submetidos à análise fatorial para obtenção do mapa sensorial.

Esta metodologia foi proposta recentemente e apresenta apenas duas publicações científicas (Teillet et al., 2010; Saldamando et al., 2013), nas quais a técnica foi utilizada para caracterizar águas minerais.



Figura 1.5 - Ficha de avaliação das amostras na metodologia *PSP*.  
Fonte: Valentin et al. (2012) modificado.

É importante ressaltar que todas as metodologias alternativas supracitadas atendem a demanda atual por métodos mais rápidos para o contexto da indústria, porém há limitações na aplicabilidade destes métodos devido à avaliação qualitativa dos produtos. Por meio das técnicas apresentadas, não é possível identificar a magnitude da diferença entre as amostras e, portanto, estes métodos não são recomendados para análises de estabilidade e controle de qualidade, com exceção da *PSP*.

### 3.8. Perfil Descritivo Otimizado (PDO)

O Perfil Descritivo Otimizado (PDO), proposto por Silva et al. (2012), teve o objetivo de suprir a demanda por métodos descritivos rápidos e, ao mesmo tempo, fornecer informações quantitativas sobre os atributos sensoriais presentes nos alimentos. Esta metodologia compreende as seguintes etapas: recrutamento e pré-seleção dos candidatos, levantamento dos atributos sensoriais e definição dos materiais de referência, familiarização da equipe de julgadores com as referências e, por último, a avaliação dos produtos por meio da escala de intensidade. É recomendada, neste método, a utilização da escala não estruturada de 9 cm ancorada nos extremos pelos termos “nenhum/fraco” e “forte”, os quais são representados por materiais de referência.

No PDO, o recrutamento é realizado por meio de questionários estruturados, os quais abordam informações sobre frequência alimentar, familiarização com termos

sensoriais e habilidade de trabalhar com escalas não-estruturadas (Murray, Delahunty e Baxter, 2001; Meilgaard, Civille e Carr, 2006). A pré-seleção é realizada por meio de testes discriminatórios e tem o objetivo de verificar a acuidade sensorial dos voluntários (Stone e Sidel, 2004; Meilgaard, Civille e Carr, 2006). Os atributos sensoriais que descrevem o alimento-teste são levantados por meio de técnicas existentes na literatura, tais como: método rede (Moskowitz, 1983), lista prévia, associação controlada e discussão aberta (Damásio e Costell, 1991). Os materiais de referência que ancoram os extremos da escala de avaliação são definidos com o auxílio da equipe de julgadores, conforme descrito em Ritvanen et al. (2005), Melo, Bolini e Efraim (2009) e Richter et al. (2010).

Na etapa de familiarização dos julgadores com os atributos, os materiais âncoras são apresentados a eles em uma única sessão com o objetivo de padronizar a forma de avaliação dos atributos. Em seguida, os julgadores iniciam a avaliação dos produtos, utilizando o protocolo atributo-por-atributo. Nesta forma de avaliação, os julgadores recebem todas as amostras de uma só vez e são orientados a comparar as amostras entre si, em relação a um único atributo, e indicar a intensidade das amostras em uma escala não estruturada ancorada nos extremos pelos materiais de referência (fraco e forte). Nesse protocolo de avaliação a re-degustação dos produtos e re-avaliação da pontuação pode ser realizada até o convencimento de que as pontuações representam o "espaçamento" correto entre as intensidades, o que possibilita que julgadores com baixo grau de "calibração" avaliem a intensidade dos produtos de forma consistente (Ishii, Stampanoni e O'Mahony, 2008).

Na avaliação dos produtos, os materiais de referência dos atributos sensoriais são apresentados juntamente com as amostras para possibilitar que os julgadores comparem as amostras com as referências, facilitando a alocação da intensidade dos atributos na escala não estruturada. Como é utilizado o protocolo atributo-por-atributo, a ficha de avaliação é organizada por atributos e não por amostras. Assim, apenas um atributo é avaliado por sessão.

No desenvolvimento da metodologia, a técnica foi aplicada na descrição sensorial de cinco formulações de requeijão cremoso, os quais diferiam entre si nas concentrações de gordura e água. Os produtos também foram avaliados por meio da técnica descritiva convencional (Perfil Convencional) para comparação dos dados. Foi verificado que o PDO permitiu a obtenção de um perfil sensorial muito semelhante à avaliação da equipe treinada em relação à configuração gráfica das amostras e da

correlação dos atributos com os componentes da Análise de Componentes Principais, além de possibilitar a redução de 54% do tempo do teste sensorial, sem perda de informações. Também foi observada nesse estudo a diminuição do efeito de interação entre a avaliação dos julgadores e os produtos avaliados. O efeito de interação indica que pelo menos um julgador está avaliando os produtos de forma diferente dos demais membros da equipe, o qual é um efeito indesejado na avaliação sensorial (Stone e Sidel, 2004). Embora este efeito seja indesejado, é comum em testes sensoriais e apresenta difícil controle, uma vez que este efeito pode ocorrer simplesmente pelo fato de os julgadores estarem utilizando diferentes partes da escala para expressarem as diferenças nas intensidades sensoriais e não devido a uma deficiência no treinamento (Silva & Damasio, 1994; O'Neill, Nicklauss, & Sauvageot, 2003; Silva et al., 2012). Este efeito de interação tem sido observado em diversos trabalhos utilizando equipes com alto grau de treinamento, tais como: Cardello e Faria (1998; 2000), Monrozier e Danzart (2001), Rocha et al. (2003), Richter et al. (2010) e Silva et al. (2012).

No PDO, é focada a caracterização quantitativa dos atributos sensoriais, possibilitando a segmentação dos consumidores de acordo com as suas preferências por meio da correlação dos dados com testes afetivos, o controle de qualidade, a otimização de formulações e a correlação entre medidas sensoriais e instrumentais. Silva et al. (2012) realizaram análises de correlação entre as medidas sensoriais e instrumentais da textura de requeijão cremoso por meio do coeficiente de correlação de *Pearson* (dados contínuos) e verificaram correlação significativa entre as medidas.

A técnica também foi aplicada na caracterização sensorial de chocolates (Silva et al., 2013), sendo comparada com a metodologia referência (Perfil Convencional). Os autores verificaram perfis sensoriais muito similares entre o PDO e o PC. O poder de discriminação das amostras também foi elevado, com poder de decisão de 99%. Os autores também verificaram correlação entre os dados sensoriais obtidos pelo PDO e medidas instrumentais de cor e textura. O PDO possibilitou a redução de 50% do tempo do teste sensorial, além de fornecer medidas quantitativas dos atributos sensoriais, ampliando sua aplicação no contexto descritivo.

#### **4. Considerações finais**

Métodos sensoriais descritivos rápidos são a tendência da última década. Diversas metodologias alternativas aos métodos tradicionais foram desenvolvidas

recentemente, as quais estão sendo amplamente utilizadas e exploradas pela comunidade científica. As técnicas rápidas de descrição sensorial apresentam um futuro promissor tanto na obtenção do perfil sensorial dos alimentos quanto nos estudos com consumidores na otimização de formulações. O Perfil Descritivo Otimizado (PDO) destaca-se entre os métodos rápidos, pois além de reduzir o tempo do teste sensorial também fornece informações sobre a intensidade dos atributos sensoriais. O PDO é uma metodologia mais complexa que os demais métodos rápidos, exigindo maior envolvimento da equipe de julgadores. As informações obtidas no PDO são muito semelhantes à técnica convencional e sua resposta em termos quantitativos amplia a aplicabilidade da metodologia.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Abdi, H., Valentin, D., Chrea, C. & Chollet, S. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and application. *Food Quality and Preference*, 18, 627–640.
- Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of check-all-that-apply response compared to tradicional scaling of attributes for salty snacks. In: *7th Pangborn Sensory Science Symposium*. Minneapolis, USA, 12-16, August, 2007.
- Ares, G., Giménez, A., Barreiro, C. & Gámbaro, A. (2010a). Use of an open-ended question to identify drivers of liking of milk desserts: comparison with preference mapping techniques. *Food Quality and Preference*, 21, 286–294.
- Ares, G., Deliza, R., Barreiro, C., Giménez, A., & Gámbaro, A. (2010b). Comparison of two sensory profiling techniques based on consumer perception. *Food Quality and Preference*, 21, 417-426.
- Ares, G., Varela, P., Rado, G. & Giménez, A. (2011a). Identifying ideal products using three different consumer profiling methodologies. Comparison with external preference mapping. *Food Quality and Preference*, 22, 581–591.
- Bécue-Bertaut, M. & Lê, S. (2011). Analysis of multilingual labeled sorting tasks: application to a cross-cultural study in wine industry. *Journal of Sensory Studies*, 26, 299–310.
- Ballester, J., Dacremont, C., Le Fur, Y. & Etievant, P. (2005). The role of olfaction in the elaboration and use of the Chardonnay wine concept. *Food Quality and Preference*, 16, 351–359.
- Bécue-Bertaut, M. & Lê, S. (2011). Analysis of multilingual labeled sorting tasks: application to a cross-cultural study in wine industry. *Journal of Sensory Studies*, 26, 299–310.
- Blancher, G., Chollet, S., Kesteloot, R., Nguyen Hoang, D., Cuvelier, G., & Sieffermann, J.-M. (2007). French and Vietnamese: How do they describe texture characteristics of the same food? A case study with jellies. *Food Quality and Preference*, 18, 560–575.

- Bleibaum, R.N., Stonea, H., Tanb, T., Labrecheb, S., Saint-Martin, E., & Isz, S. (2002). Comparison of sensory and consumer results with electronic nose and tongue sensors for apple juices. *Food Quality and Preference*, 13, 409–422.
- Brandt, M. A., Skinner, E. Z., & Coleman, J. A. (1963). Texture profile method. *Journal of Food Science*, 28, 404–409.
- Bruzzone, F., Ares, G., & Giménez, A. (2011). Consumers' texture perception of milk desserts. ii – comparison with trained assessors' data. *Journal of Texture Studies*, 43(3), 214-226.
- Cairncross, S. E., & Sjostrom, L. B. (1950). Flavour profiles: a new approach to flavour problems. *Food Technology*, 4, 308–311.
- Cardello, H., & Faria, J. B. (1998). Análise descritiva quantitativa da aguardente de cana durante o envelhecimento em tonel de carvalho (*Quercus alba* L.). *Ciência Tecnologia de Alimentos*, 18(2), 169-175.
- Cardello, H., & Faria, J. B. (2000). Sensory Profile and Physicochemical Characteristics of Commercial Brazilian Sugar Cane Spirits, Both Aged and Non-Aged. *Brazilian Journal of Food Technology*, 3, 31-40.
- Cartier, R., Ritz, A., Lecomte, A., Poblete, F., Krystlik, J., Belin, E., & Martin, N. (2006). Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference*, 17, 562–571.
- Chollet, S. & Valentin, D. (2001). Impact of training on beer flavor perception and description: are trained and untrained assessors really different? *Journal of Sensory Studies*, 16, 601–618.
- Civille, G. V., & Lawless, H. T. (1986). The importance of language in describing perception. *Journal of Sensory Studies*, 1, 203–215.
- Dairou, V., & Sieffermann, J. M. (2002). A comparison of 14 jams characterized by conventional profile and a quick original method, the Flash Profile. *Journal of Food Science*, 67, 826–834.
- Damásio, M. H., & Costell, E. (1991). Análisis sensorial descriptivo: generación de descriptores y selección de catadores. *Revista Agroquímica de Tecnología de Alimentos*, 31 (2), 165-178.
- Deegan, K.C., Koivisto, L., Näkkilä, L.J., Hyvönen, L. & Tuorila, H. (2010). Application of a sorting procedure to greenhouse-grown cucumbers and tomatoes. *Food Science and Technology*, 43, 393–400.
- Delarue, J. & Loescher, E. (2004). Dynamics of food preferences: a case study with chewing gums. *Food Quality and Preference*, 15, 771–779.
- Delaure, J., & Sieffermann, J.M. (2004). Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference*, 15, 383–392.
- Delholm, C., Brockhoff, P. B., Meinert, L., Aaslung, M. D., & Bredie, W. L. P. (2012). Rapid descriptive sensory methods – Comparison of Free Multiple Sorting, Partial Napping, Napping, Flash Profiling and conventional profiling. *Food Quality and Preference*, 26, 267–277.
- Driesener, C. & Romaniuk, J. (2006). Comparing methods of brand image measurement. *International Journal of Market Research*, 48, 681–698.

- Derndorfer, E. & Baierl, A. (2006). Development of an aroma map of spices by multidimensional scaling. *Journal of Herbs, Spices & Medicinal Plants*, 12, 39–50.
- Dooley, L., Lee, Y.S. & Meullenet, J.F. (2010). The application of check-all-that-apply (CATA) consumer profiling to preference mapping of vanilla ice cream and its comparison to classical external preference mapping. *Food Quality and Preference*, 21, 394–401.
- Falahee, M. & MacRae, A.W. (1997). Perceptual variation among drinking waters: the reliability of sorting and ranking data for multidimensional scaling. *Food Quality and Preference*, 8, 389–394.
- Faye, P., Bre´maud, D., Teillet, E., Courcoux, P., Giboreau, A. & Nicod, H. (2006). An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality and Preference*, 17, 604–614.
- Gacula, M. C. (1997). *Descriptive sensory analysis in practice*. Trumbull, CT: Food and Nutrition Press.
- Ginés, R., Valdimarsdottir, T., Sveinsdottir, K., Thorarensen, H. (2004). Effects of rearing temperature and strain on sensory characteristics, texture, colour and fat of Arctic charr (*Salvelinus alpinus*). *Food Quality and Preference*, 15, 177–185.
- Heymann, H. (1994). A comparison of free choice profiling and multidimensional scaling of vanilla samples. *Journal of Sensory Studies*, 9, 445–453.
- Hulin, W.S. & Katz, D. (1935). The Frois-Wittmann pictures of facial expression. *Journal of Experimental Psychology*, 18, 482–498.
- Hunter, E. A., & McEwan, J. A. (1998). Evaluation of an international ring trial for sensory profiling of hard cheese. *Food Quality and Preference*, 9(5), 343–354.
- Ishii, R., Chang, H., & O’Mahony, M. (2007). A comparison of serial monadic and attribute-by-attribute protocols for simple descriptive analysis with untrained judges. *Food Quality and Preference*, 18(2), 440–449.
- Ishii, R., Stampanoni, C., & O’Mahony, M. (2008). A comparison of serial monadic and attribute-by-attribute descriptive analysis protocols for trained judges. *Food Quality and Preference*, 19(3), 277–285.
- Kennedy, J. (2010). Evaluation of replicated projective mapping of granola bars. *Journal of Sensory Studies*, 25, 672–684.
- Kennedy, J., & Heymann, H. (2009). Projective mapping and descriptive analysis of milk and dark chocolates. *Journal of Sensory Studies*, 24, 220–233.
- Kobayashi, M. L., Benassi, M.T. (2012). Caracterização sensorial de cafés solúveis comerciais por *Perfil Flash*. *Semina - Ciências Agrárias (Online)*, 33, 3069-3074, 2012.
- Labbe, D., Ritz, A., & Hugi, A. (2004). Training is a critical step to obtain reliable product profiles in a real food industry context. *Food Quality and Preference*, 15, 341–348.
- Lado, J., Vicente, E., Manzsoni, A. & Ares, G. (2010). Application of a check-all-that-apply question for the evaluation of strawberry cultivars from a breeding program. *Journal of the science of Food and Agriculture*, 90, 2268–2275.

- Lancaster, B. & Foley, M. (2007). Determining statistical significance for choose-all that-apply question responses. In: *7th Pangborn Sensory Science Symposium*. August 16th 2007, Minneapolis, MN, USA.
- Langron, S. P. (1983). *The application of Procrustes statistics to sensory profiling*. In A. A. Williams, R. K. Atkin, Sensory quality in food and beverages: definition, measurement and control. Chichester: Ellis Horwood Ltd.
- Lassoued, N., Delarue, J., Launay, B. & Michon, C. (2007). Baked product texture: correlations between instrumental and sensory characterization using Flash Profile. *Journal of Cereal Science*, 48, 133–143.
- Lawless, H. T., & Heymann, H. (1998). *Sensory evaluation of food: principles and practices*. New York: Chapman and Hall.
- Lawless, H.T., Sheng, N. & Knoops, S.S.C.P. (1995). Multidimensional scaling of sorting data applied to cheese perception. *Food Quality and Preference*, 6, 91–98.
- Lelièvre, M., Chollet, S., Abdi, H. & Valentin, D. (2008). What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality and Preference*, 19, 697–703.
- Lelièvre, M., Chollet, S., Abdi, H. & Valentin, D. (2009). Beer trained and untrained assessors rely more on vision than on taste when they categorize beers. *Chemosensory Perception*, 2, 143–153.
- Lesschaeve, I., & Issanchou, S. (1996). Could selection tests detect the future performance of descriptive panels? *Food Quality and Preference*, 7, 177–183.
- Lim, J. & Lawless, H.T. (2005). Qualitative differences of divalent salts: multi-dimensional scaling and cluster analysis. *Chemical Senses*, 30, 719–726.
- Meilgaard, M. C., Civille, G. V., & Carr, B. T. (1991). *Sensory evaluation techniques* (2nd ed.). Boca Raton, FL: CRC Press.
- Melo, L. L. M. M., Bolini, H. M. A., & Efraim, P. (2009). Sensory profile, acceptability, and their relationship for diabetic/reduced calorie chocolates. *Food Quality and Preference*, 20(2), 138–143.
- Monrozier, R., Danzart, M. (2001). A quality measurement for sensory profile analysis: The contribution of extended cross-validation and re-sampling techniques. *Food Quality and Preference*, 12, 393–406.
- Morand, E. & Pages, J. (2006). Procrustes multiple factor analysis to analyze the overall perception of food products. *Food Quality and Preference*, 17, 36–42.
- Moskowitz, H. R. (1983). *Product Testing and Sensory Evaluation of Foods-Marketing and R&D Approaches*. Westport: Food and Nutrition Press.
- Munoz, A. M., & Civille, G. V. (1998). Universal, product and attribute specific scaling and the development of common lexicons in descriptive analysis. *Journal of Sensory Studies*, 13, 57–75.
- Murray, J. M. (1999). *Determination of consumer preference for cheese using descriptive sensory analysis*. PhD thesis, University College Cork, National University of Ireland.
- Murray, J.M., Delahunty, C.M., & Baxter, I.A. (2001). Descriptive sensory analysis: past, present and future. *Food Research International*, 34, 461–471.

- Nestrud, M.A. & Lawless, H.T. (2008). Perceptual mapping of citrus juices using projective mapping and profiling data from culinary professionals and consumers. *Food Quality and Preference*, 19, 431–438.
- O'Neill, L., Nicklauss, S., & Sauvageot, F. (2003). A matching task as a potential technique for descriptive profile validation. *Food Quality and Preference*, 14(7), 539–547.
- Pagès, J. (2005). Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley. *Food Quality and Preference*, 16, 642–649.
- Pagès, J., Cadoret, M. & Lê, S. (2010). The sorted napping: a new holistic approach in sensory evaluation. *Journal of Sensory Studies*, 25, 637–658.
- Perrin, L., Symoneaux, R., Maître, I., Asselin, C., Jourjon, F., & Pagès, J. (2008). Comparison of three sensory methods for use with the Napping procedure: Case of tem wines from Loire valley. *Food Quality and Preference*, 19, 1–11.
- Pliner, P., & Hobden, K. (1992). Development of a scale to measure the trait of food neophobia in humans. *Appetite*, 19, 105–120.
- Piggott, J. R., & Canaway, P. R. (1981). *Finding the word for it: methods and uses of descriptive sensory analysis*. In Flavour. Berlin: Walter de Gruyter & Co, Berlin.
- Piombino, P., Nicklaus, S., LeFur, Y., Moio, L. & Le Quéré, J. (2004). Selection of products presenting given flavor characteristics: na application to wine. *American Journal of Enology and Viticulture*, 55, 27–34.
- Popper, R., Abdi, H., Williams, A. & Kroll, B.J. (2011). Multi-block Hellinger analysis for creating perceptual maps from check-all-thatapply questions. In: *9th Pangborn Sensory Science Symposium*. September 7th, 2011, Toronto, CA.
- Rainey, B. A. (1986). Importance of reference standards in training panelists. *Journal of Sensory Studies*, 1, 149–154.
- Richter, V. B., Almeida, T. C. A., Prudencio, S. H., & Benassi, M.T. (2010). Proposing a ranking descriptive sensory method. *Food Quality and Preference*, 21, 611 - 620.
- Ritvanen, T., Lampolahti, S., Lilleberg, L., Tupasela, T., Isoniemi, M., Appelbye, U., Lyytikäinen, T., Eerola, S., & Uusi-Rauva, E. (2005). Sensory evaluation, chemical composition and consumer acceptance of full fat and reduced fat cheeses in the Finnish market. *Food Quality and Preference*, 16(6), 479–492.
- Rocha, F. L., Minim, V. P. R., Della Lucia, F. D., Minim, L. A., & Coimbra, J. S. R. (2003). Avaliação da influência dos milhos QPM nas características sensoriais de bolo. *Ciência e Tecnologia de Alimentos*, 23(2), 129 – 134.
- Rodrigue, N., Guillet, M., Fortin, J., & Martin, J.F. (2000). Comparing information obtained from ranking and descriptive tests of four sweet corn products. *Food Quality and Preference*, 11 (1), p. 47-54.
- Saint-Eve, A., Pac, i Kora, E. & Martin, N. (2004). Impact of the olfactory quality and chemical complexity of the flavouring agent on the texture of low fat stirred yogurts assessed by three different sensory methodologies. *Food Quality and Preference*, 15, 655–668.

- Saldamando, L., Delgado, J., Herencia, P., Giménez, A., Ares, G. (2013). Polarized sensory positioning: Do conclusions depend on the poles? *Food Quality and Preference*, 29, 25–32.
- Silva, M. P. A., & Damásio, M. H. (1994). *Análise sensorial descritiva*. Campinas: Fundação Tropical de Pesquisas e Tecnologia “André Tosello”.
- Silva, R. C. S. N., Minim, V. P. R., Simiqueli, A. A., Moraes, L. E. S., Gomide, A. I., Minim, L. A. (2012). Optimized Descriptive Profile: a rapid methodology for sensory description. *Food Quality and Preference*, 24(1), 190-200.
- Silva, R.C.S.N., Minim, V.P.R., Carneiro, J.D.S., Nascimento, M., Della Lucia, S.M., Minim, L.A. (2013). Quantitative sensory description using the Optimized Descriptive Profile: comparison with conventional and alternative methods for evaluation of chocolate. *Food Quality and Preference*, 30, 169-179.
- Sinésio, F., Peparario, M., Moneta, E., & Comendador, F.J. (2010). Perceptive maps of dishes varying in glutamate content with professional and naive subjects. *Food Quality and Preference*, 21, 1034–1041.
- Stone, H., Sidel, J. L., Oliver, S., Woolsey, A., & Singleton, R. C. (1974). Sensory evaluation by Quantitative Descriptive Analysis. *Food Technology*, 28(11), 24–33.
- Stone, H., & Sidel, J.L. (2004). *Sensory evaluation practices*. (3rd ed.) New York: Academic.
- Tang, C. & Heymann, H. (1999). Multidimensional sorting, similarity scaling and free choice profiling of grape jellies. *Journal of Sensory Studies*, 17, 493–509.
- Tarea, S., Cuvelier, G. & Sieffermann, J.-M. (2007). Sensory evaluation of the texture of 49 commercial apple and pear purees. *Journal of Food Quality*, 30, 1121–1131.
- Teillet, E., Schlich, P., Urbano, C., Cordelle, S. & Guichard, E. (2010). Sensory methodologies and the taste of water. *Food Quality and Preference*, 21, 967–976.
- Thompson, D. M. H., & MacFie, H. J. H. (1983). *Is there an alternative to descriptive sensory assessment?* In A. A. Williams, & R. K. Atkin, Sensory quality in food and beverages: Definition, measurement and control. Chichester: Ellis Horwood Ltd.
- Thuillier, B. (2007). *Rôle du CO<sub>2</sub> dans l’Appréciation Organoleptique des Champagnes – Expérimentation et Apports Méthodologiques*. Reims, France: Thèse de l’URCA.
- Valentin, D., Chollet, S., Lelièvre, M., & Hervé Abdi. (2012). Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science and Technology*, 47, 1563–1578.
- Veinand, B., Godefroy, C., Adam, C. & Delarue, J. (2011). Highlight of important product characteristics for consumers. Comparison of three sensory descriptive methods performed by consumers. *Food Quality and Preference*, 22, 474–485.
- Wechsler, D. (1944). *The measurement of adult intelligence* (3rd ed.). Baltimore, MD: Williams and Wilkins.



## Número de julgadores necessários em testes sensoriais descritivos

### RESUMO

Para a determinação do número de julgadores necessários em testes descritivos quatro parâmetros são necessários: probabilidade do erro tipo I ( $\alpha$ ), probabilidade do erro tipo II ( $\beta$ ), diferença em média que é procurada no experimento ( $d'$ ) e o desvio padrão do erro experimental ( $s$ ). As probabilidades dos erros experimentais e a diferença que se quer detectar entre as médias devem ser estipuladas pelo pesquisador. Portanto, somente a estimativa do erro experimental não pode ser obtida previamente, sendo necessária a obtenção desta experimentalmente ou por meio de trabalhos similares realizados previamente. Como em análise sensorial descritiva a abordagem mais comum para análise dos dados é a análise de variância, a estimativa do desvio padrão do erro experimental é obtida pela raiz do quadrado médio do erro (RMSE). Portanto, foram coletados na literatura 574 valores de RMSE de estudos prévios. Aos dados coletados foi ajustada uma distribuição de probabilidades conhecida *Weibull* (1,8081; 0,11419), sendo cinco percentis da distribuição considerados nos cálculos. A determinação do número de avaliações necessárias foi realizada utilizando o procedimento “*sample size and power analysis*” do software JMP/SAS. Foram estipulados três níveis de probabilidade dos erros tipo I e II, quatro níveis de diferença média a ser detectada no experimento e cinco percentis da distribuição de probabilidades do RMSE. Os números necessários de avaliações em testes descritivos, considerando essas diferentes condições experimentais, foram calculados totalizando 180 cenários. Considerando a mediana do erro experimental, um valor de alfa (erro tipo I) de 1%, um valor de beta (erro tipo II) de 5% e uma diferença entre as médias de 10% da escala sensorial, são necessárias 33 avaliações nos testes descritivos. Considerando, por exemplo, ainda que cada julgador avalie as amostras em três repetições, 11 julgadores serão necessários, para este conjunto específico de parâmetros. Outros cenários também foram abordados.

**Palavras-chave:** tamanho da amostra; poder do teste; tamanho do painel; RMSE.

---

**Publicação:** Silva et al. (2014). Number of judges necessary for descriptive sensory tests. *Food Quality and Preference*, 31(1), 22-27.

## **1. Introdução**

A análise sensorial descritiva de alimentos consiste na avaliação das características sensoriais por meio de uma equipe de julgadores, os quais irão identificar e quantificar a intensidade dos estímulos sensoriais presentes nos alimentos, utilizando os cinco sentidos humanos (visão, olfato, audição, tato e paladar) (Murray, Delahunty e Baxter, 2001). A avaliação descritiva é uma ferramenta valiosa nas diversas etapas do processamento de alimentos: no desenvolvimento de novos produtos, no controle de qualidade, no armazenamento e vida-de-prateleira (Meilgaard et al., 2006; Stone e Sidel, 2012).

As técnicas de avaliação descritiva tradicionais requerem a formação de equipes de julgadores com alto grau de treinamento, sendo que o elevado número de julgadores pode se apresentar como um obstáculo na aplicação desta valiosa ferramenta na indústria (Heymann et al., 2012). A recomendação do número ideal de julgadores para compor uma equipe não é muito clara na literatura. São encontradas diferentes recomendações em função da técnica utilizada, por exemplo: seis julgadores para o Perfil de Sabor (Cairncross & Sjostrom, 1950); dez julgadores para o Perfil de Textura (Brandt, Skinner, & Coleman, 1963) e dez a doze julgadores para a Análise Descritiva Quantitativa (Stone & Sidel, 1985). Porém, os critérios para determinação dos números de julgadores necessários não são apresentados.

Por outro lado, a utilização de diferentes tamanhos de equipes é verificada nas técnicas genéricas denominadas “Perfil Convencional” ou “Análise Descritiva”. Em trabalhos na literatura, que utilizaram técnicas genéricas, verifica-se a utilização de seis julgadores (Lee & Chambers, 2010; Warmund et al., 2011), sete julgadores (Guàrdia et al., 2010; Chueamchaitrakum et al., 2011), oito julgadores (Perrin & Pagès, 2009; Anyango et al., 2011), nove julgadores (Westad et al., 2003; Campo et al., 2010; Tesfaye et al., 2010; Silva et al., 2012), dez julgadores (Leighton et al., 2010; Speziale et al., 2010; Silva et al., 2013), onze julgadores (Bitnes et al., 2009; Plaehn, 2009; Moussaoui & Varela, 2010; Parr et al., 2010), doze julgadores (Sinesio et al., 2009; Bitnes et al., 2009; Biasoto et al., 2010), quinze julgadores (Garcia-Carpintero et al., 2011), dezessete julgadores (Guinard et al., 1999) e vinte julgadores (Delgado e Guinard, 2011). Em geral, o limite inferior é de seis julgadores e o limite superior de vinte julgadores.

Dado que o custo associado à avaliação sensorial aumenta com o número de julgadores participantes, é importante determinar o número ótimo de avaliadores que é necessário para os testes sensoriais. Segundo Heymann et al. (2012), é óbvio que treinar um número menor de julgadores requer menor tempo, dinheiro e esforços, mas isto pode acarretar em uma “falsa economia” devido à possibilidade de obtenção de dados “pobres”. Desta forma, o desafio é determinar o número ideal de julgadores necessários nas avaliações descritivas que possibilite a redução do tamanho da equipe, mas sem que ocorram perdas de informações sobre o perfil sensorial dos alimentos, da discriminação dos produtos e que possibilite a realização de testes estatísticos poderosos.

## 2. Como determinar o número ideal de julgadores?

O cálculo do número de julgadores em testes sensoriais tem sido pouco explorado na literatura. Alguns estudos recentes foram realizados para testes afetivos, nos quais foi determinado o número de consumidores ideal para testes de aceitação sensorial. Nestas pesquisas, os autores calcularam o tamanho da amostra por meio de dados obtidos experimentalmente (Gacula & Rutenbeck, 2006; Mammasse & Schlich, 2013), por meio de dados obtidos de revisão de literatura (Hough et al., 2006) e por meio da análise de sobrevivência (Hough et al., 2007; Libertino, Osornio, Hough, 2011). Para testes descritivos, técnicas de re-amostragem com dados experimentais foram utilizadas por King et al. (1995), Pagès e Perinel (2003), Gacula e Rutenbeck (2006) e Heymann et al. (2013).

No cálculo do número de julgadores, quatro parâmetros devem ser conhecidos, utilizando o conceito “*sample size and power of analysis*”, descrito em Kraemer e Thiemann (1987), Montgomery (2001) e Walpole et al. (2011), os quais são: (i) nível  $\alpha$  – probabilidade do erro tipo I; (ii) nível  $\beta$  – probabilidade do erro tipo II, também expresso como poder do teste ( $1 - \beta$ ); (iii)  $d'$  – diferença em média que é procurada no experimento; (iv)  $s$  – desvio padrão do erro experimental. As probabilidades dos erros experimentais ( $\alpha$  e  $\beta$ ) e a diferença que se quer detectar entre as médias devem ser estipuladas pelo pesquisador. Portanto, somente o erro experimental não pode ser obtido previamente, sendo necessária a obtenção deste experimentalmente ou por meio de trabalhos similares realizados previamente.

### 3. Identificando os parâmetros

O conhecimento da definição e interpretação dos quatro parâmetros que devem ser conhecidos para o cálculo do número de julgadores é de extrema importância no processo de determinação do número necessário de avaliações nos testes descritivos. De forma a estabelecer o critério e o modo de obtenção destes parâmetros, uma breve explicação e conceituação destes será realizada.

#### 3.1. Probabilidades dos erros tipo I e tipo II

Os erros de decisão denominados tipo I e tipo II estão associados às hipóteses do teste estatístico utilizado para verificar a existência de diferença significativa entre médias dos tratamentos. A hipótese de nulidade ( $H_0$ ) representa a igualdade entre as médias, a qual é testada contra a hipótese alternativa ( $H_a$ ) que contrapõe a decisão de que pelo menos uma média difere das demais (Montgomery, 2001). Uma vez que são utilizadas amostragens e as decisões são tomadas com base em distribuições de probabilidades, há a probabilidade de se cometer erros nas decisões. Assim, quando se detecta, em um teste estatístico, que as médias dos tratamentos são diferentes, há uma probabilidade de que estas sejam iguais considerando a população, o que caracteriza o erro de decisão tipo I, o qual apresenta uma probabilidade  $\alpha$  de ocorrência. Similarmente, ao se detectar que as médias dos tratamentos são iguais entre si, esta conclusão também está sujeita a um erro de decisão, o qual é denominado como o erro tipo II com probabilidade  $\beta$  de ocorrência. Na Tabela 2.1 estão ilustrados os erros de decisão associados aos testes estatísticos.

Tabela 2.1 - Erros de decisão tipo I e tipo II.

		Decisão	
		Rejeitar $H_0$	Não rejeitar $H_0$
Verdade	$H_0$ verdadeira	Erro tipo I (probabilidade $\alpha$ )	Decisão correta
	$H_0$ falsa	Decisão correta	Erro tipo II (probabilidade $\beta$ )

As probabilidades de cometer os erros tipo I e tipo II devem ser estabelecidas antes da condução do experimento (Stone & Sidel, 2012). Estas probabilidades também são utilizadas nos cálculos de poder do teste e tamanho de amostra (Montgomery, 2001; Walpole et al., 2011), conforme o objetivo desta pesquisa.

### 3.2. Diferença desejada entre as médias ( $d'$ )

O valor de  $d'$  determina qual a magnitude de diferença que deve existir entre as médias dos tratamentos para que seja detectada significância entre as mesmas, a um dado valor de  $\alpha$  e  $\beta$  (Hough et al., 2006). Esta magnitude é fornecida em percentual do tamanho da escala. Assim, se utilizado uma escala de 9 cm e um valor de  $d'$  de 0,02 for estipulado, as médias dos tratamentos deverão apresentar uma diferença mínima de 0,18 cm (2% de 9 cm) para que sejam consideradas diferentes entre si (Figura 2.1a). Por exemplo, para um valor de  $d'$  igual a 2% na escala de 9 cm (diferença mínima de 0,18cm), isto significa que, se a média do produto A é igual a 6,0 e a média do produto B é de 6,3, os produtos (A e B) diferem significativamente uma vez que a diferença entre eles é de 0,3 cm ( $> 0,18$  cm).

Para evitar má interpretação sobre o parâmetro  $d'$  utilizado neste trabalho e o valor de delta do modelo Thurstoniano, esta teoria será abordada rapidamente, realizando uma comparação. Em testes sensoriais de diferença, o modelo Thurstoniano é utilizado para estimar a distância ( $\delta$ ) entre as médias de intensidade ( $\mu_A$  e  $\mu_B$ ) do estímulo de dois produtos, a qual é dada pelo número de desvios padrões pelos quais as duas distribuições são separadas (Thurstone, 1927), Figura 2.1b. O valor de  $d'$  neste caso representa a estimativa do parâmetro  $\delta$ . Na teoria Thurstoniana, é assumido que a intensidade de percepção dos estímulos sensoriais segue distribuição normal e ainda que ambos os produtos (A e B) apresentam o mesmo desvio padrão (Ennis, 1993; Ishii et al., 2007; Jesionka et al., 2013).

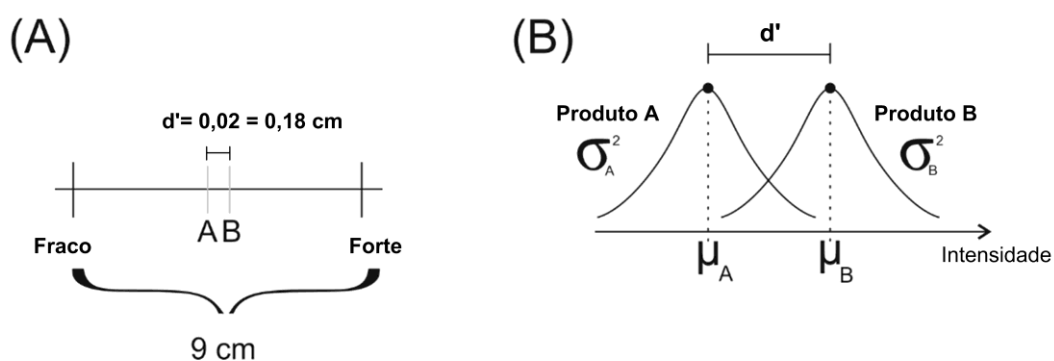


Figura 2.1 - Representação do  $d'$  no teste descritivo (a) e do modelo Thurstoniano (b).

Desta forma, fica claro que o valor de  $d'$  do modelo Thurstoniano é uma medida de número de desvios padrão, a qual depende estritamente da distribuição de

probabilidades e da variância da percepção do estímulo sensorial. Paralelamente, o valor de  $d'$  assumido nesta pesquisa se refere a um valor de percentagem da escala, ou seja, escore sensorial. Neste caso, nenhuma suposição de variância é requerida.

Assim, um valor de  $d'$  igual 1 na teoria Thurstoniana (testes de diferença) significa que há uma distância entre as médias igual a um desvio padrão  $\sigma$ , portanto, os dois estímulos podem ser confundíveis (Meilgaard, Civille & Carr, 2006). Por outro lado, nos testes descritivos, o valor de  $d'$  igual a 1 indica que as médias devem apresentar 100% de diferença sensorial, em relação ao tamanho da escala, para que seja detectada diferença significativa. Desta forma, um tratamento estará ancorado no extremo “fraco” da escala e o outro tratamento ancorado no extremo “forte” e, portanto, facilmente discriminados.

O conceito do valor de  $d'$  desta pesquisa difere do conceito apresentado pela teoria Thurstoniana e segue a mesma descrição proposta por Hough et al. (2006).

### *3.3. Estimativa do desvio padrão do erro experimental (Raiz do Quadrado Médio do Erro)*

Em análise sensorial descritiva a abordagem mais comum para análise dos dados é a análise de variância, sendo a estimativa do desvio padrão do erro experimental obtida pela raiz do quadrado médio do erro (*Root Mean Square Error* - RMSE). O modelo da ANOVA adotado nesta pesquisa foi o modelo com apenas uma fonte de variação (Atributo = média + produto + erro).

O cálculo do RMSE depende da coleta dos dados experimentais e, portanto, é impossível de ser obtido *a priori* dos testes. Como a determinação do número necessário de avaliações deve ser realizada antes da condução dos testes, uma estimativa deste valor é necessária. Essa estimativa pode ser obtida por meio da revisão de literatura considerando estudos similares (mesmo tipo de produto, mesma metodologia) ou por meio da experiência prévia do pesquisador em relação aos testes anteriores realizados com a mesma equipe ou com produtos similares. Alternativamente, testes preliminares de avaliação individual de desempenho dos julgadores também podem ser úteis na obtenção de um valor esperado para o RMSE.

A determinação de um valor de RMSE para posterior cálculo do número de avaliações está sujeita a um erro de decisão, portanto, uma faixa de segurança é recomendada, até mesmo por precaução de desistências. De toda forma, é recomendada sempre a utilização de um valor de RMSE maior do que o esperado, o que

proporcionará ao pesquisador trabalhar com uma faixa de segurança na determinação do número necessário de avaliadores.

Para obtenção da estimativa do desvio padrão do erro experimental foi realizada uma coleta de valores de RMSE obtidos em estudos prévios de caracterização sensorial por meio de metodologias genéricas (Perfil Convencional ou Análise Descritiva). Desta forma, foram coletados 574 valores de RMSE de 34 estudos prévios publicados nos periódicos *Food Quality and Preference* e *Journal of Sensory Studies* nos anos compreendidos entre 1993 e 2012. Na Tabela 2.2 estão apresentados os valores de raiz do quadrado médio do erro obtidos de diferentes estudos de análise descritiva de alimentos, considerando diferentes países, produtos alimentícios, tamanhos de escalas não estruturadas e números de julgadores na equipe. A demografia dos julgadores e os tipos de produtos alimentícios foram suficientemente variados para cobrir diversas situações envolvendo testes descritivos.

Como foram coletados dados de estudos que utilizaram metodologias genéricas, diferentes tamanhos de escalas podem ter sido utilizados. Para padronizar os valores, a raiz do quadrado médio do erro de cada medida foi dividida pelo comprimento da escala, obtendo o *Root Mean Square Error Length - RMSEL* (Raiz do Quadrado Médio do Erro corrigida pelo tamanho da escala), conforme recomendado por Hough et al. (2006). Assim, se um RMSE de 1,2 foi obtido utilizando uma escala de 15 cm, o valor do RMSEL será  $1,2/15 = 0,08$ . No cálculo da análise de variância, se um conjunto de dados for dividido por um número, o resultado da raiz do quadrado médio do erro também estará dividido pelo mesmo número. Portanto, dividindo o RMSE pelo tamanho da escala é equivalente a ter previamente padronizado o comprimento de todas as escalas de avaliação para uma escala de 0-1.

Hough et al. (2006) calcularam o número de consumidores necessários para testes de aceitação utilizando dados coletados na literatura. No estudo, os autores assumiram que os valores do RMSEL apresentavam distribuição simétrica e, portanto, efetuaram os cálculos de tamanho da amostra considerando a média e o desvio padrão dos RMSEL obtidos na literatura. A observação da frequência de ocorrência de RMSEL em estudos anteriores possibilita inferir um modelo probabilístico que representa esses dados e, assim, obter um maior conhecimento sobre este parâmetro. O ajuste de uma distribuição de probabilidades conhecida pode possibilitar o cálculo do número de julgadores de forma mais realista, levando em consideração a probabilidade de obtenção experimentalmente de um valor de RMSEL de determinada magnitude.

Tabela 2.2 - Valores de Raiz do Quadrado Médio do Resíduo (RMSE) encontrados na literatura considerando diferentes países, produtos alimentícios, atributos sensoriais e número de julgadores.

País	Produtos	Número de julgadores	Número de medidas	RMSE m ± s	RMSEL m ± s
Noruega	Sorvete, suco, soluções de gostos primários, sopa de tomate, chá gelado.	9 - 12	256	1,1739 ± 0,8053	0,0784 ± 0,0537
EUA	Cerveja	11 - 17	12	1,4123 ± 0,4160	0,0942 ± 0,0277
França	Café, drinks, vinho	8 - 11	54	1,4507 ± 0,5864	0,1220 ± 0,0591
Espanha	Presunto	7 - 11	12	0,9607 ± 0,1760	0,0640 ± 0,0117
Itália	Tender, soluções aquosas dos gostos primários	10 - 12	60	0,7475 ± 0,2280	0,0831 ± 0,0253
Brasil	Queijo, chocolate, peixe, suco, iogurte, taro, cachaça, flan, salame.	6 - 18	180	1,5393 ± 0,6655	0,1497 ± 0,0588

RMSE: Raiz do Quadrado Médio do Erro; RMSEL: Raiz do Quadrado Médio do Erro corrigida pelo tamanho da escala; m: média; s: desvio padrão.

Neste estudo, foi proposto o cálculo de número de julgadores para testes descritivos por meio do procedimento “*sample size and power of analysis*”, sendo utilizada a distribuição de probabilidade das estimativas do erro experimental, obtidas em estudos prévios disponíveis em literatura especializada.

#### 4. Cálculo do número de julgadores em testes descritivos

Conhecidos os valores de RMSEL obtidos na literatura (Tabela 2.1), verificou-se a existência de nove *outliers* por meio da técnica Box-plot, considerando o limite superior como três vezes a distância inter-quartil acima do terceiro quartil. Não foram observados *outliers* inferiores, uma vez que a distribuição dos valores apresentou assimetria positiva de 0,449 e valores somente positivos. O ajuste de uma distribuição de probabilidades aos dados foi realizado por meio do teste de aderência de Kolmogorov-Smirnov ( $p > 0,10$ ), sendo ajustada a distribuição de *Weibull* (1,8081; 0,11419), representada na Figura 2.2.

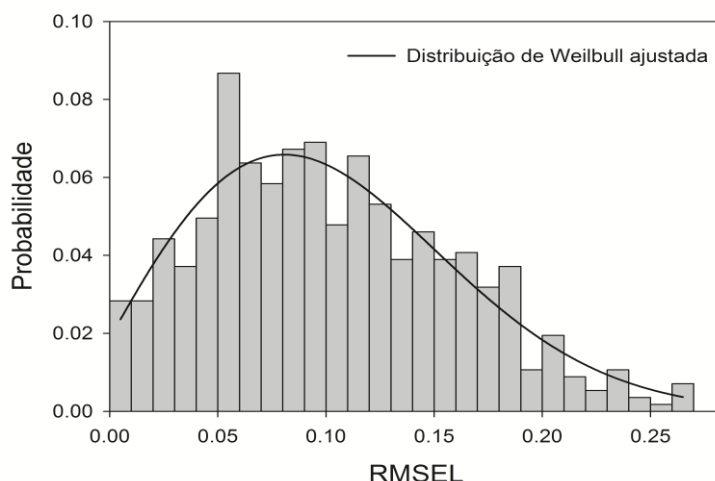


Figura 2.2 - Histograma dos dados de RMSEL da literatura e a distribuição de probabilidades de *Weibull* ajustada aos dados.

No cálculo do número de julgadores, foram considerados cinco percentis da distribuição de probabilidade (10%, 25%, 50%, 75% e 90%). Assim, os valores de RMSEL utilizados foram: 0,0329; 0,0573; 0,0932; 0,1368 e 0,1811.

A determinação do número ideal de julgadores em testes descritivos foi realizada utilizando o procedimento “*sample size and power analysis*” do software JMP/SAS. Foram determinados os números de avaliações necessárias considerando os cinco valores de RMSEL descritos acima, três níveis de probabilidade do erro tipo I ( $\alpha = 0,10; 0,05; 0,01$ ), três níveis de probabilidade do erro tipo II ( $\beta = 0,20; 0,10; 0,05$ ), e quatro níveis de diferença a ser detectada entre as médias do experimento ( $d' = 0,20; 0,10; 0,05; 0,02$ ).

Nas Tabelas 2.3 a 2.6 é apresentado o número de avaliações ( $N_A$ ) necessárias nos testes descritivos, considerando quatro níveis de  $d'$ . Assim, para  $d'=0,20$  (Tabela 2.3),

$d'=0,10$  (Tabela 2.4),  $d'=0,05$  (Tabela 2.5) e  $d'=0,02$  (Tabela 2.6). Cada um dos componentes da equação apresenta uma contribuição para determinar o número de julgadores, mas a variação de alguns componentes tem efeito drasticamente diferente. Foi observado maior efeito no número de julgadores para os componentes de diferença procurada no experimento ( $d'$ ) e para o desvio padrão do erro experimental (RMSEL). O número de avaliações necessárias nos testes descritivos aumenta conforme se aumenta o valor do RMSEL e se diminuem os valores de alfa, beta e  $d'$ .

A recomendação do número de avaliações nos testes descritivos apresentou uma grande variação em função dos parâmetros considerados, sendo verificado desde 2 a 2932 medidas. Em ambos os casos, não seria ideal trabalhar com painéis nem tão pequenos nem tão grandes. A recomendação de um número extremamente elevado de avaliações leva a crer que as condições experimentais destes cenários não estão bem definidas, sendo estes associados a valores elevados de probabilidade dos erros de decisão e elevada variabilidade aleatória (RMSE), não sendo, portanto, recomendados. Em relação aos painéis muito pequenos, estudos prévios de determinação de número de julgadores em testes descritivos, realizados por meio de simulação computacional com dados experimentais, também verificaram que poucos julgadores poderiam ser suficientes para as avaliações descritivas, sendo verificada a necessidade de somente dois julgadores (Pagès & Perinel, 2003), cinco julgadores (Gacula & Rutenbeck, 2006) e oito julgadores (Heymann et al., 2013).

É importante ressaltar que nos testes descritivos são obtidas diversas medidas de RMSE e não uma única medida. Desta forma, o maior valor esperado deve ser considerado na determinação do número ideal de avaliações. Portanto, o número de avaliações será determinado com base no atributo mais crítico, ou seja, o que apresentar a maior variância aleatória. Assim, o número de julgadores participantes da equipe deve ser suficientemente grande para possibilitar que o atributo mais crítico seja avaliado corretamente.

Como nas tabelas (Tabelas 2.3 a 2.6) estão apresentados os números necessários de avaliações ( $N_A$ ), o número de repetições ( $r$ ) das avaliações realizadas por cada julgador também deve ser considerado. Assim, se o julgador avalia cada amostra somente uma vez ( $r = 1$ ), o número de julgadores ( $N$ ) será igual ao número de avaliações ( $N_A$ ). Porém, se cada julgador avaliar cada amostra mais de uma vez ( $r > 1$ ), como comumente utilizado em testes sensoriais descritivos, o número de julgadores será

dado pela razão entre o número de avaliações necessárias e o número de repetições ( $N = N_A/r$ ).

Tabela 2.3 - Número de avaliações necessárias para testes descritivos considerando as probabilidades dos erros tipo I ( $\alpha$ ) e tipo II ( $\beta$ ) e RMSEL, para  $d'=0,20$ .

$\alpha$	RMSEL				
	0,0329	0,0573	0,0932	0,1368	0,1811
$\beta = 0,20$					
<b>0,10</b>	2	3	4	7	11
<b>0,05</b>	2	3	5	9	14
<b>0,01</b>	3	4	7	13	21
$\beta = 0,10$					
<b>0,10</b>	2	3	5	9	15
<b>0,05</b>	3	4	6	8	18
<b>0,01</b>	3	5	9	16	27
$\beta = 0,05$					
<b>0,10</b>	2	3	6	11	19
<b>0,05</b>	3	4	7	14	23
<b>0,01</b>	4	5	10	19	31

Tabela 2.4 - Número de avaliações necessárias para testes descritivos considerando as probabilidades dos erros tipo I ( $\alpha$ ) e tipo II ( $\beta$ ) e RMSEL, para  $d'=0,10$ .

$\alpha$	RMSEL				
	0,0329	0,0573	0,0932	0,1368	0,1811
$\beta = 0,20$					
<b>0,10</b>	3	5	12	24	42
<b>0,05</b>	4	7	15	32	53
<b>0,01</b>	5	10	23	46	79
$\beta = 0,10$					
<b>0,10</b>	3	7	16	33	57
<b>0,05</b>	4	8	20	41	70
<b>0,01</b>	6	12	28	58	100
$\beta = 0,05$					
<b>0,10</b>	4	8	20	42	72
<b>0,05</b>	5	10	24	50	87
<b>0,01</b>	6	14	33	69	119

Para exemplificação, supondo que sejam estipulados pelo pesquisador os critérios esperados em seu experimento como sendo  $\alpha = 0,01$ ;  $\beta = 0,05$ ;  $d'= 0,10$  e  $RMSEL = 0,0932$  (mediana), o número de avaliações necessárias será igual a 33. Considerando que cada julgador realizará três repetições ( $r$ ) da avaliação sensorial, o número de julgadores ideal será de  $33/3 = 11$  julgadores. É importante ressaltar que as

repetições devem ser realizadas com intervalo de tempo suficiente para que o julgador não se lembre do teste anterior.

Tabela 2.5 - Número de avaliações necessárias para testes descritivos considerando as probabilidades dos erros tipo I ( $\alpha$ ) e tipo II ( $\beta$ ) e RMSEL, para  $d'=0,05$ .

$\alpha$	RMSEL				
	0,0329	0,0573	0,0932	0,1368	0,1811
$\beta = 0,20$					
0,10	7	17	44	94	163
0,05	8	22	56	118	207
0,01	12	33	83	117	308
$\beta = 0,10$					
0,10	9	24	61	129	226
0,05	11	29	74	159	277
0,01	15	41	106	225	393
$\beta = 0,05$					
0,10	11	30	76	163	285
0,05	13	36	92	192	342
0,01	18	49	126	268	470

Tabela 2.6 - Número de avaliações necessárias para testes descritivos considerando as probabilidades dos erros tipo I ( $\alpha$ ) e tipo II ( $\beta$ ) e RMSEL, para  $d'=0,02$ .

$\alpha$	RMSEL				
	0,0329	0,0573	0,0932	0,1368	0,1811
$\beta = 0,20$					
0,10	35	103	270	580	1015
0,05	44	130	342	736	1289
0,01	65	194	509	1095	1917
$\beta = 0,10$					
0,10	48	142	373	802	1406
0,05	58	174	458	985	1725
0,01	83	246	648	1395	2442
$\beta = 0,05$					
0,10	60	179	471	1014	1775
0,05	75	215	566	1217	2132
0,01	99	295	776	1669	2932

A faixa dos valores descritos nas Tabelas 2.3 a 2.6 é limitada ao que os autores consideram necessários para a maioria das aplicações práticas. Por exemplo, para um teste muito rigoroso pode ser necessário considerar o valor de  $d'$  menor do que 0,02. Neste caso, o pesquisador pode utilizar um pacote estatístico para calcular o número de avaliações ( $N_A$ ) necessárias no teste descritivo. Por outro lado, os valores de RMSEL também podem ser alterados, segundo a probabilidade de obter uma variância aleatória

experimental considerada pelo pesquisador. Neste ponto, o ajuste de uma distribuição de probabilidades aos valores da literatura mostra-se bastante útil, pois o pesquisador pode obter outros valores de RMSEL por meio da distribuição de *Weibull* (1,8081; 0,11419). Para determinação do número de avaliações necessárias em outras condições experimentais, o procedimento “*sample size and power analysis*” do software JMP/SAS pode ser utilizado.

## 5. Conclusão

A coleta de estimativas do erro experimental de estudos de descrição sensorial de alimentos realizada em literatura especializada possibilitou o cálculo do número de julgadores necessários para as avaliações descritivas. A determinação do número ideal de julgadores nos testes sensoriais é de extrema importância, pois quanto menor o tamanho da equipe, menores são os custos para a realização dos testes sensoriais. Por outro lado, este cálculo deve ser cuidadosamente realizado para que informações sobre o perfil sensorial dos produtos não sejam ocultadas e para que testes estatísticos robustos e poderosos possam ser aplicados.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Anyango, J.O., De Kock, H.L., & Taylor, J.R.N. (2011). Evaluation of the functional quality of cowpea-fortified traditional African sorghum foods using instrumental and descriptive sensory analysis. *LWT – Food Science and Technology*, 44, 2126-2133.
- Biasoto, A.C.T., Catharino, R.R., Sanvido, G.B., Eberlin, M.N., Aparecida, M., & Da Silva, A.P. (2010). Flavour characterization of red wines by descriptive analysis and ESI mass spectrometry. *Food Quality and Preference*, 21, 755-762.
- Bitnes, J., Ueland, Ø., Møller, P., & Martens, M. (2009). Reliability of sensory assessors: issues of complexity. *Journal of Sensory Studies*, 24, 25–40.
- Brandt, M. A., Skinner, E. Z., & Coleman, J. A. (1963). Texture profile method. *Journal of Food Science*, 28, 404–409.
- Cairncross, S. E., & Sjostrom, L. B. (1950). Flavour profiles: a new approach to flavor problems. *Food Technology*, 4, 308–311.
- Campo, E., Ballester, J., Langlois, J., Dacremont, C., & Valentin, D. (2010). Comparison of conventional descriptive analysis and a citation frequency-based descriptive method for odor profiling: An application to Burgundy Pinot noir wines. *Food Quality and Preference*, 21, 755-762.
- Chueamchaitrakun, P., Chompreeda, P., Haruthaithanasan, V., Suwonsichon, T., Kaswmsamran, S., & Prinyawiwatkui, W. (2011). Sensory descriptive and texture

- profile analyses of butter cakes made from composite rice flours. *International of Journal of Food Science and Technology*, 45, 2358-2365.
- Delgado, C., & Guinard, J.X. (2011). How do consumer hedonic ratings for extra olive oil relate to quality ratings by experts and descriptive analysis ratings? *Food Quality and Preference*, 22, 213-322.
- Ennis, D. (1993). The power of sensory discrimination methods. *Journal of Sensory Studies*, 8(4), 353–370.
- Gacula, M., & Rutenbeck, S. (2006). Sample size in consumer test and descriptive analysis. *Journal of Sensory Studies*, 21, 129-145.
- Garcia-Carpintero, E.G., Gomez-Gallego, M.A., Sanchez-Palomo, E., & Gonzales Viñas, M.A. (2011). Sensory descriptive analysis of Bobal red wines treated with oak chips at different stages of winemaking. *Australian Journal of Grape Wine Research*, 17, 368-377.
- Guàrdia, M.D., Aguiar, A.P.S., Claret, A., Arnau, J., & Guerrero, L. (2010). Sensory characterization of dry-cured ham using free-choice profiling. *Food Quality and Preference*, 21,148–155.
- Guinard, J.X., Yip, D., Cubero, E., & Mazzucchelli, R. (1999). Quality ratings by experts, and relation with descriptive analysis ratings: a case study with beer. *Food Quality and Preference*, 10, 59-67.
- Heymann, H., Machado, B., Torri, L., & Robinson, A.L. (2012). How many judges should one use for sensory descriptive analysis? *Journal of Sensory Studies*, 27, 111-122.
- Hough, G., Wakeling, I., Mucci, A., Chambers, E., Gallardo, I.M., & Alves, L.R. (2006). Number of consumers necessary for sensory acceptability tests. *Food Quality and Preference*, 17, 522-526.
- Hough, G., Calle, M.L., Serrat, C., & Curia, A. (2007). Number of consumers necessary for shelf life estimations based on survival analysis statistics. *Food Quality and Preference*, 18, 771-775.
- Ishii, R., Kawaguchi, H., O'Mahony, M., & Rousseau, B. (2007) Relating consumer and trained panels discriminative sensitivities using vanilla flavored ice cream as a médium. *Food Quality and Preference*, 18, 89-96.
- Jesionka, V., et al. Transitioning from proportion of discriminators to a more meaningful measure of sensory difference. *Food Quality and Preference* (2013), <http://dx.doi.org/10.1016/j.foodqual.2013.04.007>
- King, B.M., Arents, P., & Moreau, N. (1995). Cost / Efficiency evaluation of descriptive analysis panels – I. Panel size. *Journal of Sensory Studies*, 6, 245-261.
- Kraemer, H.C., & Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park: Sage Publications, pp.37-22.
- Lee, J., & Chambers, D.H (2010). Descriptive analysis and US consumer acceptability of 6 green tea samples from China, Japan, Korea. *Journal of Food Science*, 75, S141-S147.
- Leighton, C.S., Schonefeldt, H.C., & Kruger, R. (2010). Quantitative descriptive sensory analysis of Five different cultivars of sweet potato to determine sensory na textural profiles. *Journal of Sensory Studies*, 25, 2-18.

- Libertino, L.M., Osornio, M.M.L., & Hough, G. (2011). Number of consumers necessary for survival analysis estimations based on each consumer evaluating a single sample. *Food Quality and Preference*, 22, 24-30.
- Mammasse, N., & Schlich, P. Adequate number of consumers in a liking test. Insights from resampling in seven studies. *Food Quality and Preference* (2012), doi: 10.1016/j.foodqual.2012.01.009
- Meilgaard, M.C, Civille, G.V., & Carr, B.T. (2006). *Sensory Evaluation Techniques*. (4th ed.). Boca Raton: CRC Press.
- Montgomery, D.C. (2001). *Design and analysis of experiments*. (5th Ed.). New York: John Wiley and Sons. p.699.
- Moussaoui, K.A., & Varela, P. (2010). Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality and Preference*, 21, 1088-1099.
- Murray, J.M., Delahunty, C.M., & Baxter, I.A. (2001). Descriptive sensory analysis: past, present and future. *Food Research International*, 34, 461-471.
- Pagès, J., & Périnel, E. (2003). Panel performance and number of evaluations in a descriptive sensory study. *Journal of Sensory Studies*, 19, 273-291.
- Parr, W.V., Valentin, D., Green, J.A., & Dacremont, C. (2010). Evaluation of French and New Zealand Sauvignon wines by experienced French wine assessors. *Food Quality and Preference*, 21, 56–64.
- Perrin, L., & Pagès, J. (2009). A methodology for the analysis of sensory typicality judgments. *Journal of Sensory Studies*, 24, 749–773.
- Plaehn, D. (2009). A variation on external preference mapping. *Food Quality and Preference*, 20, 427–439
- Silva, R. C. S. N., Minim, V. P. R., Simiqueli, A. A., Moraes, L. E. S., Gomide, A. I., & Minim, L. A. (2012). Optimized Descriptive Profile: a rapid methodology for sensory description. *Food Quality and Preference*, 24, 190-200.
- Silva, R.C.S.N., Minim, V.P.R., Carneiro, J.D.S., Nascimento, M., Della Lucia, S.M., Minim, L.A. (2013). Quantitative sensory description using the Optimized Descriptive Profile: comparison with conventional and alternative methods for evaluation of chocolate. *Food Quality and Preference*, 30, 169-179.
- Sinéσιο, F., Comendador, F.J., Peparaio, M., & Moneta, E. (2007). Taste perception of umami-rich dishes in italian culinary tradition. *Journal of Sensory Studies*, 24, 554–580.
- Speziale, M., Vásquez-Araujo, L., Mincione, A., & Carbonell-Barrachina, A.A. (2010). Volatile composition and descriptive sensory analysis of Italian vanilla Torrone. *International of Food Science and Technology*, 45, 1586-1593.
- Stone, H., & Sidel, J.L. (1985). *Sensory evaluation practices*. (1st. ed.) New York: Academic.
- Stone, H., & Sidel, J.L. (2012). *Sensory evaluation practices*. (4th. ed.) New York: Academic.
- Tesfaye, W., Morales, M.I., Callejon, R.M., Gonzales, A.G., Garcia-Parrila, M.C., & Troncoso, A.M. (2010). Descriptive sensory analysis of wine vinegar: Tasting procedure and reliability of new attributes. *Journal of Sensory Studies*, 25, 216-230.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286.

Walpole, R.H., Myers, R. H., Myers, S.L., & Ye, K. (2011). *Probability & Statistics for Engineers & Scientists*. (9th ed.). New York: Pearson. p.816.

Warmund, M.R., Elmore, J.R., Adhikari, K., & McGraw, S. (2011). Descriptive sensory and free sugar contents of chestnut cultivars grown in North American. *Journal of Science of Food Agriculture*, 91, 1940-1945.

Westad, F., Hersleth, M., Lea, P., & Martens, H. (2003). Variable selection in PCA in sensory descriptive and consumer data. *Food Quality and Preference*, 14, 463–472.

## *Perfil Descritivo Otimizado: Quantos julgadores são necessários?*

### **RESUMO**

O custo associado aos testes sensoriais descritivos pode ser proveniente principalmente de duas fontes: (i) tempo de execução do teste e (ii) número de julgadores participantes. A técnica Perfil Descritivo Otimizado (PDO) é uma metodologia nova que propõe a redução do tempo do teste sensorial por meio de um protocolo de avaliação otimizado. Este estudo tem por objetivo determinar o número ideal de julgadores para avaliação sensorial descritiva utilizando o PDO, com o intuito de possibilitar que a técnica apresente redução de tempo, dinheiro e esforços em relação da condução da metodologia e também em relação ao número de participantes da equipe. O estudo de determinação do número ideal de julgadores foi realizado utilizando a técnica de re-amostragem dos dados de um painel original composto por 26 julgadores, por meio de simulação computacional. Foram re-amostrados dados do painel completo considerando 10.000 sub-grupos com reposição. Os critérios para a determinação do número ideal de julgadores foram: (i) obtenção de um erro experimental menor ou igual ao erro verificado na metodologia referência (Perfil Convencional), (ii) discriminação das amostras semelhantemente à discriminação obtida pela equipe completa, e (iii) mínima perda de informação no mapa sensorial. O critério da magnitude da estimativa do erro experimental mostrou-se como uma medida mais robusta na determinação do número de julgadores necessários para a técnica PDO. Como esta técnica descritiva requer baixo nível de treinamento dos julgadores, a avaliação deste critério é de extrema importância, pois uma variação aleatória residual maior pode ser normalmente observada. Para o atendimento ao critério da magnitude do erro experimental, dezesseis julgadores devem compor a equipe sensorial. Os critérios de discriminação das amostras e mapeamento sensorial também foram atendidos plenamente quando este número de avaliadores foi utilizado.

**Palavras-chave:** re-amostragem; simulação computacional; discriminação de amostras; mapa sensorial.

## **1. Introdução**

O custo associado à avaliação sensorial descritiva aumenta com o número de julgadores participantes. Portanto, a determinação do número ideal de avaliadores é de extrema importância. Segundo Heymann et al. (2012), é óbvio que treinar um número menor de julgadores requer menos tempo, custos e esforços, mas isto pode acarretar em uma “falsa economia” devido à possibilidade de obtenção de dados “pobres”. Desta forma, o desafio é determinar o número ideal de julgadores necessários nas avaliações descritivas que possibilite a redução do tamanho da equipe, mas sem que ocorram perdas de informações sobre o perfil sensorial dos alimentos.

A recomendação do número ideal de julgadores para compor uma equipe não é muito clara na literatura. São encontradas diferentes recomendações em função da técnica utilizada, por exemplo, seis julgadores para o Perfil de Sabor (Cairncross & Sjostrom, 1950); dez julgadores para o Perfil de Textura (Brandt, Skinner, & Coleman, 1963); e dez a doze julgadores para a Análise Descritiva Quantitativa (Stone & Sidel, 1985). Porém, os critérios para determinação dos números de julgadores necessários não são apresentados.

O cálculo do número de julgadores em testes sensoriais descritivos tem sido pouco explorado na literatura. Alguns estudos foram realizados na determinação do número ideal de julgadores considerando metodologias genéricas, tais como o “Perfil Convencional” ou a “Análise Descritiva”. Na maioria dos estudos, foi abordada a redução do tamanho da equipe de julgadores por meio de re-amostragem de dados proveniente de um painel maior (King et al., 1995; Pagès & Perinel, 2003; Gacula & Rutenbeck, 2006; Heymann et al., 2013; Silva et al., 2013b).

King et al. (1995) realizaram a descrição sensorial de sorvetes por meio da avaliação de uma equipe de 20 julgadores (painel completo). Os dados da equipe completa foram re-amostrados originando 20 novos painéis, os quais apresentavam de 3 a 13 julgadores. Para os novos painéis formados foi avaliada a significância do efeito de tratamento para cada um dos atributos sensoriais por meio da Análise de Variância (ANOVA). O painel completo apresentou maior explicação da variação do efeito de tratamento, sendo que 76% dos atributos avaliados foram significativos ( $p < 0,05$ ) na ANOVA. Quando o painel foi reduzido pela metade ( $N=10$ ), 67% dos atributos apresentam significância. Reduzindo o painel para um quarto ( $N=5$ ), somente 34% dos atributos foram significativos. Os autores concluíram que a redução do número de

julgadores na equipe acarretou perda significativa de informação sobre o efeito de tratamentos.

No estudo de Pagès e Perinel (2003) foi realizada a descrição sensorial de oito amostras de água mineral gaseificada por meio da avaliação de uma equipe de dezesseis julgadores. Os dados da equipe total foram removidos do conjunto de dados de dois em dois, até o número mínimo de duas avaliações no painel. Para os sub-painéis formados foram avaliados os critérios da magnitude da razão de F e similaridade dos mapas sensoriais obtidos pela ACP (Análise de Componentes Principais). Não foi verificada diferença entre os painéis, considerando esses critérios.

Gacula e Rutenbeck (2006) determinaram o número de julgadores em testes sensoriais descritivos por meio de simulação computacional utilizando dados experimentais. Um painel de seis julgadores treinados procedeu às avaliações das amostras na obtenção dos dados. Na simulação foram consideradas duas medidas experimentais: diferença a ser detectada entre as médias ( $d'$ ) e a variabilidade do experimento (Raiz do Quadrado Médio do Erro / *Root Mean Square Error - RMSE*). Foi determinado um número mínimo de cinco julgadores na equipe sensorial.

Heymann et al. (2013) realizaram um estudo do número de julgadores para testes descritivos por meio de re-amostragem dos dados originais. Foram utilizados dados de três estudos de caracterização sensorial de vinhos tintos, os quais apresentavam de 14 a 22 julgadores. Os dados dos painéis completos foram re-amostrados originando equipes com 4, 6, 8, 12 e 14 julgadores. Os novos painéis formados foram avaliados quanto à significância (p-valor) dos atributos descritores por meio da ANOVA e similaridade dos mapas sensoriais obtidos pela análise fatorial. Os resultados mostraram que são necessários pelo menos oito julgadores na equipe sensorial.

No estudo de Silva et al. (2013b) foi determinado o número ideal de julgadores por meio do cálculo de "*Power analysis and Sample size*". Foram determinados três diferentes níveis de probabilidade dos erros tipo I e tipo II, quatro níveis de diferença a ser detectada entre as médias do experimento ( $d'$ ). Os valores de desvio padrão do erro experimental foram determinados com base na literatura. Foram buscados 574 valores de raiz do quadrado médio do erro (RMSE) em estudos prévios. Aos dados da literatura foi ajustada uma distribuição de probabilidades conhecida, sendo utilizados cinco percentis dessa distribuição no cálculo do número de julgadores. Os números necessários de avaliações em testes descritivos, considerando essas diferentes condições experimentais, foram calculados totalizando 180 cenários.

A metodologia Perfil Descritivo Otimizado (PDO) não apresenta nenhum estudo prévio em relação à determinação de número de julgadores necessário na equipe sensorial. O PDO foi proposto como método descritivo recentemente e, portanto, são poucos os estudos que tratam desta nova técnica sensorial (Silva et al., 2012; 2013a). A técnica utiliza um protocolo de avaliação otimizado, apresentando uma avaliação comparativa entre as amostras e, posteriormente, quantitativa utilizando uma escala intervalar. Uma vez que a técnica recomenda a participação de julgadores com baixo grau de treinamento e o protocolo de avaliação dos alimentos é diferenciado, é necessário que um estudo particular deste método seja realizado para determinar o número ideal de julgadores.

Este estudo tem por objetivo determinar o número ideal de julgadores para a avaliação sensorial descritiva de alimentos utilizando a técnica, com o intuito de possibilitar que a técnica apresente redução de tempo, dinheiro e esforços em relação da condução da metodologia e também em relação ao número de participantes na equipe.

## **2. Materiais e Métodos**

A determinação do número necessário de julgadores na técnica PDO foi realizada utilizando a técnica de re-amostragem dos dados de um painel original composto por 26 julgadores, por meio de simulação computacional. Foram re-amostrados dados do painel completo considerando 10.000 iterações com reposição. Os dados experimentais foram obtidos por meio da caracterização sensorial, utilizando a técnica descritiva PDO, de duas matrizes alimentares: iogurte sabor morango (Experimento A) e chocolate (Experimento B). Os critérios para a determinação do número ideal de julgadores foram: (i) obtenção de um erro experimental menor ou igual ao erro verificado para a metodologia referência (Perfil Convencional), (ii) discriminação das amostras semelhantemente à discriminação obtida pela equipe completa, e (iii) mínima perda de informação no mapa sensorial.

### *2.1. Estímulos sensoriais*

Foram utilizados dois tipos de matrizes alimentar (iogurte e chocolate) na caracterização sensorial. As formulações foram definidas com base em testes

triangulares preliminares, de forma que as amostras apresentassem pequena magnitude de diferença ( $p \leq 0,10$ ) nas características sensoriais.

**Experimento A)** Cinco formulações de iogurte sabor morango foram utilizadas. Para o preparo das formulações um iogurte de marca comercial foi utilizado. Neste foram adicionadas diferentes concentrações de leite, açúcar, leite em pó e corante rosa (Tabela 3.1).

Tabela 3.1 - Formulações-teste de iogurte sabor morango. Quantidade de ingredientes adicionados a um litro (1 L) de iogurte comercial.

Iogurtes	Leite (mL)	Sacarose (g)	Leite em pó (g)	Corante (mL)
Iog-1	100,0	60,0	-	-
Iog-2	50,0	30,0	-	-
Iog-3	-	-	-	-
Iog-4	-	-	90,0	0,2
Iog-5	-	-	180,0	0,4

Iog-3: amostra comercial.

**Experimento B)** Quatro formulações de chocolate foram utilizadas. As formulações de chocolate foram elaboradas com três diferentes tipos de chocolate da mesma marca comercial e cada unidade apresentava 29 mm de diâmetro e 20 mm de altura. Na elaboração das formulações foram utilizadas diferentes misturas de chocolate ao leite, meio amargo e amargo. As quantidades de cada tipo de chocolate utilizadas no processamento estão descritas na Tabela 3.2.

Tabela 3.2 - Composição dos chocolates em relação ao tipo e quantidade do chocolate utilizado no processamento.

Chocolates *	Tipo e quantidade (g) de chocolate		
	Ao leite	Meio Amargo	Amargo
Choc-1	9,6	2,4	-
Choc-2	9,6	-	2,4
Choc-3	-	12,0	-
Choc-4	6,0	-	6,0

\* cada unidade continha 12g de chocolate.

## 2.2. Procedimento

A avaliação sensorial das formulações-teste (iogurte e chocolate) foi realizada por meio do protocolo de avaliação da técnica Perfil Descritivo Otimizado, PDO (Silva et al., 2012; 2013a). Assim, duas equipes de 26 julgadores participaram das avaliações da intensidade dos atributos descritores (Tabela 3.3). Uma equipe de 26 participantes

realizou a avaliação sensorial das cinco formulações de iogurte e a outra equipe de 26 julgadores avaliou sensorialmente as quatro formulações de chocolate.

Tabela 3.3 - Atributos sensoriais descritores das formulações-teste.

<b>Experimento A – Iogurte</b>	<b>Experimento B – Chocolates</b>
Cor rosa (Cr)	Cor marrom (Cm)
Gosto doce (Gd-iog)	Aroma de massa de cacau (Amc)
Sabor de morango (Sm)	Gosto doce (Gd-choc)
Sabor de creme (Sc)	Gosto amargo residual (Gar)
Creiosidade (Cr)	Sabor de massa de cacau (Smc)
Viscosidade (Vis)	Dureza (Dur)
Resistência ao escoamento (Re)	Espalhabilidade (Esp)
Textura Farinácea (Tf)	

Os julgadores semi-treinados (recrutados, pré-selecionados e participantes da reunião de definição dos atributos sensoriais e da apresentação das referências) avaliaram os produtos, em três repetições, em relação aos atributos sensoriais utilizando uma escala de avaliação não-estruturada de 9 cm, conforme o protocolo do PDO.

### 2.3. Análise dos dados

#### 2.3.1. Simulação computacional

O método utilizado para comparar diferentes tamanhos de painel consistiu em medir a perda de informação quando obtidos sub-grupos com  $k$  julgadores a partir do painel original constituído por 26 julgadores. Para medir a perda de informação, 10.000 sub-painéis foram simulados para cada  $k$  número de julgadores por re-amostragem com reposição,  $k= 2, 4, 6, \dots, 22$ . As simulações foram realizadas por meio de programações desenvolvidas no software R (R Core Team, 2012), armazenando as informações dos critérios avaliados para a determinação do número ideal de julgadores.

#### 2.3.2. Critérios

Os critérios considerados na determinação do número ideal de julgadores no PDO foram: estimativa do erro experimental dada pela raiz do quadrado médio do erro corrigida para o tamanho da escala de avaliação (RMSEL), valor da significância (p-valor) da fonte de variação de tratamento por meio da ANOVA com duas fontes de variação e coeficiente RV obtido na Análise de Procrustes Generalizada (Robert & Escoufier, 1973).

*I) Estimativa do erro experimental:*

A variabilidade aleatória do experimento foi obtida por meio da estimativa do erro experimental na ANOVA com dois fatores. Os fatores avaliados na ANOVA foram: formulações (efeito fixo) e julgadores (efeito aleatório), além do efeito de interação entre as formulações e os julgadores, conforme recomendado por Stone e Sidel (2012) para métodos descritivos. A estimativa do erro experimental possibilita que a variação aleatória (não passível de controle) seja estimada, trazendo informações sobre a variabilidade das respostas dadas pelos julgadores nas repetições das avaliações.

A estimativa da variabilidade do erro experimental é obtida por meio do quadrado médio do erro (*Mean Square Error - MSE*) na ANOVA. O desvio padrão do erro experimental é comumente utilizado para medir o efeito aleatório do experimento, uma vez que ao contrário da variância, este apresenta uma medida padronizada. O desvio padrão é obtido por meio da raiz do quadrado médio do erro (*Root Mean Square Error - RMSE*).

Para padronizar os valores obtidos neste estudo de forma que pudessem ser realizadas comparações com outros estudos, os quais tenham utilizado escalas de avaliação de tamanhos diferentes, a raiz do quadrado médio do erro de cada medida foi dividida pelo comprimento da escala, obtendo o RMSEL (*Root Mean Square Length*), conforme recomendado por Hough et al. (2006). Assim, se um RMSE de 1,8 foi obtido utilizando uma escala de 9 cm, o valor do RMSEL será  $1,8/9 = 0,20$ . No cálculo da análise de variância, se um conjunto de dados for dividido por um número, o resultado da raiz do quadrado médio do erro também estará dividido pelo mesmo número. Portanto, dividindo o RMSE pelo tamanho da escala é equivalente a ter previamente padronizado o comprimento de todas as escalas de avaliação para uma escala de 0-1.

O RMSEL foi obtido para 10.000 sub-grupos dentro de cada  $k$  número de julgadores. Os valores foram armazenados e foi verificado se estes atendiam ao requisito definido para este critério.

*II) Significância do efeito de tratamento*

A significância do efeito de tratamento foi verificada por meio do p-valor do  $F_{\text{FORMULAÇÕES}}$  na ANOVA com dois fatores. O p-valor indica o nível mínimo de significância da hipótese  $H_a$  (existe diferença entre as amostras). Quanto menor este valor, menor é o risco de errar ao concluir que as amostras apresentam diferença, portanto, maior é a probabilidade de estas serem diferentes entre si (Montgomery,

2001). Foram realizadas ANOVAs para 10.000 sub-grupos dentro de cada  $k$  número de julgadores no software R, sendo os  $p$ -valores do  $F_{\text{FORMULAÇÕES}}$  armazenados e posteriormente comparados com o requisito definido neste estudo.

### III) Similaridade dos mapas sensoriais

O mapa sensorial pode ser obtido por diferentes análises estatísticas multivariadas, tal como a Análise de Componentes Principais (ACP) e representam de forma gráfica o perfil sensorial dos produtos caracterizados pelo painel sensorial (Meilgaard et al., 2006). A comparação entre duas configurações espaciais pode ser realizada pela Análise de Procrustes Generalizada (APG), obtendo-se o coeficiente de correlação multivariado RV (Robert & Escoufier, 1973). Este coeficiente mede a concordância entre duas ou mais configurações espaciais e varia de 0 (discordância total entre as configurações) a 1 (concordância perfeita). A configuração espacial de cada um dos 10.000 sub-grupos formados dentro de cada  $k$  número de julgadores foi comparada com a configuração espacial do painel completo, um a um. O coeficiente RV foi calculado no software R, armazenado e comparado com a regra de decisão descrita abaixo.

#### 2.3.3. Critérios de decisão

Para cada  $k$  número de julgadores, dez mil valores foram obtidos por critério, correspondendo a 10.000 simulações. Para cada critério, foi definida uma regra decisória para determinar o número ideal de julgadores na equipe do PDO, sendo:

(i) *Estimativa do erro experimental*: não mais que 10% dos sub-grupos com valores de RMSEL superior ao “ponto de corte” (PC = 0,1811). Este “ponto de corte” foi considerado como sendo o Percentil 90 da distribuição de RMSEL apresentada em Silva et al. (2013b). A distribuição de RMSEL considerada foi obtida por meio de uma revisão de literatura, onde foram coletados 574 valores de RMSEL de estudos prévios de descrição de alimentos utilizando a técnica tradicional “Perfil Convencional”.

(ii) *Significância do efeito de tratamento*: não mais que 10% dos sub-grupos com significância do efeito de tratamento em discordância do grupo completo. Por exemplo: sendo que para o atributo  $X$  foi verificado efeito significativo de tratamento ( $p < 0,0001$ ) no grupo completo, então, no máximo 10% dos sub-

grupos, para  $k$  julgadores, podem apresentar efeito não significativo, ao mesmo nível de probabilidade.

(iii) *Similaridades dos mapas sensoriais*: não mais que 10% dos sub-grupos com coeficiente RV abaixo de 0,90. O coeficiente RV foi calculado pela comparação realizada entre os mapas sensoriais obtidos por cada sub-grupo com  $k$  julgadores e pelo grupo completo.

### 3. Resultados

#### 3.1. ANOVA e mapa sensorial: painel completo

Os grupos completos com 26 julgadores apresentaram caracterização similar para as estimativas do erro experimental, razão de F para o efeito de formulações e significância deste efeito (Tabela 3.4).

Tabela 3.4 - Resumo da ANOVA para os dados do painel completo.

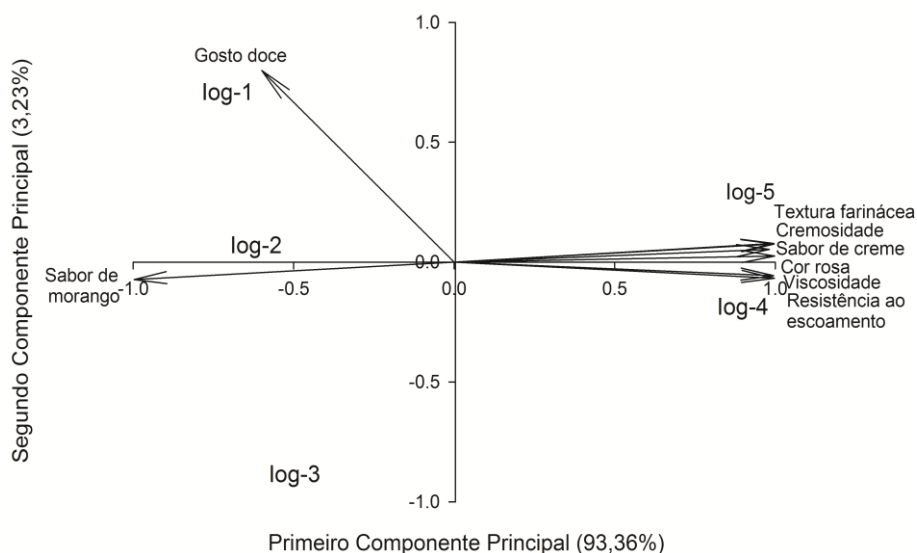
Produtos / Atributos	RMSEL	F <sub>FORMULAÇÃO</sub>	p-valor
<b>Iogurte</b>			
Cor rosa (Cr)	0,107949	346,85	< 0,0001
Gosto doce (Gd-iog)	0,126161	146,60	< 0,0001
Sabor de morango (Sm)	0,134032	371,46	< 0,0001
Sabor de creme (Sc)	0,157952	231,85	< 0,0001
Creiosidade (Cr)	0,123041	509,17	< 0,0001
Viscosidade (Vis)	0,157952	415,73	< 0,0001
Resistência ao escoamento (Re)	0,116549	598,20	< 0,0001
Textura Farinácea (Tf)	0,117703	636,98	< 0,0001
<b>Chocolate</b>			
Cor marrom (Cm)	0,156343	346,68	< 0,0001
Aroma de massa de cacau (Amc)	0,142366	450,21	< 0,0001
Gosto doce (Gd-choc)	0,158141	320,07	< 0,0001
Gosto amargo residual (Gar)	0,148077	348,43	< 0,0001
Sabor de massa de cacau (Smc)	0,109060	849,01	< 0,0001
Dureza (Dur)	0,142868	430,46	< 0,0001
Espalhabilidade (Esp)	0,168534	301,90	< 0,0001

Foi verificada uma amplitude de RMSEL de 0,107949 a 0,168534. O atributo que apresentou a menor variabilidade do erro experimental foi a “cor rosa” dos iogurtes sabor morango e a maior variabilidade foi verificada para a “espalhabilidade” dos chocolates. A estimativa do erro experimental representa a variabilidade nas respostas dos julgadores nas três repetições, mostrando que para o atributo “cor rosa” as notas das

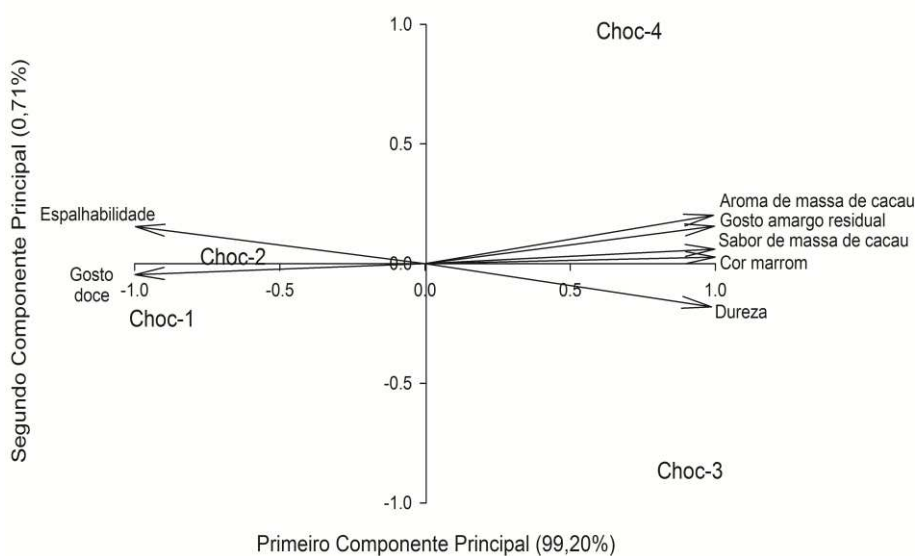
repetições foram mais homogêneas. Entretanto, para o atributo espalhabilidade a variação entre as repetições foi maior. A magnitude desta variação é influenciada pelo tipo de matriz alimentar e do atributo avaliado, pois a complexidade da avaliação sensorial é relacionada ao estímulo provocado pelo alimento, o qual pode ser mais facilmente ou não detectado e quantificado (King et al., 1995). Para todos os atributos, foi verificada a significância da razão de F para o efeito de formulações, a um nível inferior a 0,0001. Portanto, a hipótese de nulidade (igualdade) foi rejeitada, indicando diferença entre as amostras.

Os mapas descritivos (ACPs) obtidos na caracterização sensorial dos produtos (iogurte e chocolate) mostram que, para ambos os produtos, um elevado percentual da variabilidade total dos dados foi explicada pelo primeiro componente principal, sendo 93,36% para os iogurtes e 99,20% para os chocolates. Portanto, somente uma dimensão foi considerada na representação gráfica para propósito de interpretação (Figura 3.1).

Neste tipo de análise, a representação bi-dimensional possibilita uma fácil visualização do mapa sensorial. O mapa representa o perfil sensorial dos produtos, identificando os atributos de maior importância na caracterização de cada uma das formulações. O objetivo nesta pesquisa é garantir que a redução do número de julgadores não altere o perfil sensorial dos produtos obtidos pela equipe completa.



(a)



(b)

Figura 3.1 - Mapas sensoriais obtidas pela Análise de Componentes Principais. (a) iogurtes sabor morango (b) chocolates.

### 3.2. Critério I: estimativa do erro experimental

No critério de comparação da estimativa do erro experimental obtido nos sub-grupos com o valor de referência da literatura (“ponto de corte”), foi verificado que o painel com 16 julgadores apresentou atendimento ao critério estabelecido considerando os dois produtos e os diferentes atributos (Tabela 3.5). Assim, para  $N$  igual a 16 avaliadores, pelo menos 90% dos sub-grupos formados apresentaram estimativa do RMSEL inferior ou igual ao erro obtido pelo método convencional “Perfil Convencional”. O valor do erro estabelecido como “ponto de corte” apresenta 90% de probabilidade de ocorrência em estudos descritivos, representando a variabilidade global de estudos prévios de descrição sensorial.

Verificou-se diferença na necessidade do número de julgadores em relação aos produtos e aos atributos avaliados. Enquanto para o atributo “cor marrom” (Cm) do chocolate foram necessários 16 julgadores para atender ao critério estabelecido, para o atributo “cor rosa” (Cr) somente 2 julgadores foram necessários. A variação existente entre as repetições das avaliações pode apresentar amplitudes diferentes conforme o produto e o atributo, além do que o indivíduo também pode apresentar diferença nesta variabilidade (King et al., 1995; Stone & Sidel, 2012). Na técnica de avaliação do PDO, os julgadores apresentam baixo nível de treinamento, o que pode acarretar em maior

variabilidade interna e, portanto, maior variância aleatória no experimento. Desta forma, como os julgadores podem apresentar níveis diferentes de variabilidade, a inclusão aleatória de um julgador com maior variabilidade residual, acarreta em aumento da variância conjunta do grupo.

Tabela 3.5 - Percentual de atendimento dos sub-grupos ao critério da estimativa do erro experimental.

Produtos / Atributos	Tamanho do painel ( <i>k</i> )										
	2	4	6	8	10	12	14	16	18	20	22
<b>Iogurte</b>											
Cr	<b>94,1</b>	99,2	99,9	100	100	100	100	100	100	100	100
Gd-iog	93,7	98,8	99,8	99,9	100	100	100	100	100	100	100
Sm	89,5	<b>96,5</b>	99,0	99,8	100	100	100	100	100	100	100
Sc	74,4	74,7	80,1	83,3	87,4	89,6	<b>92,6</b>	93,7	96,2	97,3	98,6
Cr	<b>95,2</b>	96,2	99,6	100	100	100	100	100	100	100	100
Vis	87,3	87,8	85,5	86,2	<b>92,4</b>	95,3	98,7	99,9	100,0	100	100
Re	<b>99,1</b>	100	100	100	100	100	100	100	100	100	100
Tf	<b>92,9</b>	98,5	99,8	100	100	100	100	100	100	100	100
<b>Chocolate</b>											
Cm	78,6	77,4	73,5	70,1	74,0	78,0	84,5	<b>90,2</b>	100	100	100
Amc	81,0	87,0	<b>92,9</b>	96,5	98,6	99,9	100	100	100	100	100
Gd-choc	75,8	74,1	74,6	77,8	81,6	86,0	89,5	<b>94,43</b>	98,6	100	100
Gar	85,0	89,8	<b>94,5</b>	98,1	99,7	99,9	100	100	100	100	100
Smc	<b>99,3</b>	100	100	100	100	100	100	100	100	100	100
Dur	86,4	89,6	<b>95,3</b>	98,5	99,8	100	100	100	100	100	100
Esp	89,3	<b>91,5</b>	97,3	99,7	100	100	100	100	100	100	100

Valores em negrito: menor painel com atendimento ao critério I

### 3.3. Critério II: significância do efeito de tratamento

O efeito de tratamento no grupo completo apresentou significância ao nível inferior a 0,0001 de probabilidade para todos os atributos sensoriais, considerando os dois produtos em teste. Portanto, para atendimento ao critério estabelecido, pelo menos 90% dos sub-grupos, considerando um determinado tamanho de painel, também devem apresentar o efeito de formulações significativo ( $p < 0,0001$ ). Foi verificado que, para todos os atributos, os painéis constituídos por somente quatro julgadores atenderam ao critério estabelecido para a significância do efeito de tratamentos (Tabela 3.6).

A significância do efeito de tratamento indica que pelos menos um contraste entre médias difere estatisticamente de zero, ou seja, pelo menos um dos tratamentos difere dos demais, em termos médios. Portanto, esta comparação pode ser realizada entre as duas formulações mais distantes entre si, o que facilitaria a obtenção de efeito significativo. Verifica-se que para 16 julgadores (recomendação obtida pelo critério I), todos os 10.000 sub-grupos simulados apresentaram concordância com o grupo

completo para o critério de significância do efeito de tratamento. Isto ressalta a importância da avaliação conjunta dos critérios.

Tabela 3.6 - Percentual de atendimento dos sub-grupos ao critério de significância do efeito de tratamento.

Produtos / Atributos	Tamanho do painel ( <i>k</i> )										
	2	4	6	8	10	12	14	16	18	20	22
<b>Iogurte</b>											
Cr	<b>92,5</b>	000	100	100	100	100	100	100	100	100	100
Gd-iog	88,5	<b>98,0</b>	99,6	100	100	100	100	100	100	100	100
Sm	<b>95,1</b>	100	100	100	100	100	100	100	100	100	100
Sc	77,8	<b>92,3</b>	98,1	99,5	99,9	99,9	100	100	100	100	100
Cr	<b>96,0</b>	100	100	100	100	100	100	100	100	100	100
Vis	<b>93,0</b>	99,7	100	100	100	100	100	100	100	100	100
Re	<b>99,9</b>	100	100	100	100	100	100	100	100	100	100
Tf	<b>99,3</b>	100	100	100	100	100	100	100	100	100	100
<b>Chocolate</b>											
Cm	88,8	<b>99,9</b>	100	100	100	100	100	100	100	100	100
Amc	100	100	100	100	100	100	100	100	100	100	100
Gd-choc	89,5	<b>99,9</b>	100	100	100	100	100	100	100	100	100
Gar	<b>98,8</b>	100	100	100	100	100	100	100	100	100	100
Smc	68,7	<b>91,0</b>	97,5	99,8	100	100	100	100	100	100	100
Dur	<b>99,7</b>	100	100	100	100	100	100	100	100	100	100
Esp	<b>99,0</b>	100	000	100	100	100	100	100	100	100	100

**Valores em negrito:** menor painel com atendimento ao critério II

### 3.4. Critério III: similaridades dos mapas sensoriais

Na comparação dos mapas sensoriais obtidos para cada um dos sub-grupos compostos por *k* número de julgadores e pelo painel completo, verificou-se que a avaliação de somente dois julgadores na equipe sensorial já permitiu a obtenção de uma configuração espacial muito semelhante à configuração obtida pelo painel completo, apresentando coeficiente RV superior a 0,90 (Tabela 3.7).

Tabela 3.7 - Percentual de atendimento dos sub-grupos ao critério de similaridade do mapas sensoriais.

Produtos	Tamanho do painel ( <i>k</i> )										
	2	4	6	8	10	12	14	16	18	20	22
Iogurte	<b>92,5</b>	95,9	99,2	99,9	100	100	100	100	100	100	100
Chocolate	<b>91,4</b>	93,7	98,2	100	100	100	100	100	100	100	100

**Valores em negrito:** menor painel com atendimento ao critério III

Na GPA, as configurações individuais (painel com *k* número de julgadores e painel completo) são submetidas a três tipos de transformações (escalamento, rotação e translação), na obtenção da configuração de consenso. Após a realização das

transformações necessárias, é medida a distância entre as configurações individuais e a de consenso para o cálculo do coeficiente RV (Dijksterhuis, 1996). Portanto, apenas grandes desvios entre estas configurações são normalmente detectados por meio deste coeficiente. Isto ressalta a importância da avaliação de mais de um critério para determinação do número ideal de julgadores, pois uma única medida pode acarretar na perda de informações importantes sobre a variação dos dados.

#### **4. Discussão**

A recomendação do número de julgadores no PDO variou de acordo com os requisitos considerados, sendo que para o atendimento da estimativa do erro experimental similarmente ao método convencional, dezesseis julgadores foram suficientes. Em contrapartida, para a obtenção do efeito significativo das formulações semelhantemente à significância encontrada no painel completo, somente quatro julgadores na equipe já atenderam ao requisito estabelecido.

A significância do efeito de tratamento indica que pelos menos um dos tratamentos difere dos demais, em média. Portanto, esta comparação pode ser realizada entre as duas formulações mais distantes entre si, o que facilita a obtenção de efeito significativo. Esta característica do teste estatístico faz com que este critério seja mais facilmente atendido do que o critério da variância aleatória (critério I).

Por outro lado, no critério I a regra decisória não foi estabelecida em relação ao grupo completo, mas sim em relação a um ponto de corte global, o qual foi calculado envolvendo diferentes matrizes alimentares e atributos sensoriais. Desta forma, um mesmo “ponto de corte” foi estabelecido para todos os atributos, independente do alimento, o que pode ter ocasionado maior dificuldade no atendimento ao critério. Segundo King et al. (1995), a estimativa da variância aleatória é influenciada pela matriz alimentar em avaliação. Portanto, o estabelecimento de um limite (“ponto de corte”) global para esta variação fez-se interessante e se mostrou mais rigoroso na determinação do número de julgadores na equipe do PDO.

Na verificação de semelhança entre os mapas sensoriais obtidos no painel completo e nos painéis reduzidos, equipes com somente dois julgadores já forneceram medidas de similaridade (RV) superiores a 0,90. As técnicas multivariadas utilizadas para obtenção dos mapas sensoriais medem a similaridade entre as avaliações realizadas

por duas equipes. Neste processo, são avaliadas as configurações das amostras para verificar semelhança entre os mapas sensoriais, não sendo avaliada a magnitude dos escores atribuídos (Dijksterhuis, 1996). Mapas similares indicam que os julgadores estão avaliando as amostras de forma consensual, sendo atribuídos perfis sensoriais semelhantes a cada um dos produtos pelos diferentes julgadores. Portanto, sub-grupos com apenas dois julgadores apresentaram a mesma configuração do que a equipe completa, mostrando que os julgadores apresentaram consenso nas avaliações.

Neste estudo foi observado que os diferentes critérios de avaliação, os tipos de matrizes alimentares e os atributos descritores avaliados apresentaram grande variação na determinação do número ideal de julgadores no PDO. Este comportamento esclarece a discrepância nas recomendações de número de julgadores na literatura para métodos descritivos, sendo recomendados desde dois (Pagès & Perinel, 2003) até vinte julgadores (King et al., 1995).

A diferença na recomendação do número de julgadores verificada quando considerados os diferentes critérios avaliados neste estudo, ressalta a importância da avaliação de mais de um critério para determinação do número ideal de julgadores. A observação de uma única medida, por exemplo, o coeficiente RV, pode acarretar na perda de informação na discriminação das amostras (critério II) e, de maneira mais impactante, à elevada variação residual no experimento (critério I).

Nos estudos realizados com a técnica PDO (Silva et al., 2012: 2013a), foram utilizados quatorze e quinze julgadores, respectivamente. Nesses estudos, foi verificada grande semelhança entre as configurações espaciais obtidas pelo PDO e pelo Perfil Convencional, em relação aos componentes principais da ACP. Este comportamento é coerente com os dados obtidos nesta pesquisa, pois com quatorze julgadores 100% dos sub-grupos simulados apresentaram coeficiente RV superior a 0,90.

Nos estudos prévios do PDO, foi verificado poder de discriminação dos produtos muito semelhante ao Perfil Convencional, quando observadas as razões de F, as significâncias do efeito de formulações e também o poder do teste ( $1 - \beta$ ). O PDO tem se mostrado como uma técnica de descrição sensorial muito semelhante ao Perfil Convencional (PC). Portanto, é coerente que a determinação do número de julgadores neste método se assemelhe às recomendações na literatura para o PC, principalmente em relação aos critérios de semelhança dos mapas sensoriais e de discriminação dos produtos no teste F.

No PDO, o critério da magnitude do erro experimental mostrou-se uma medida mais rigorosa da determinação do número ideal de julgadores. Como o PDO requer baixo nível de treinamento dos julgadores, a avaliação é de extrema importância, pois como os julgadores não são extensivamente treinados, uma variação aleatória residual maior pode ser observada.

Assim, para que um experimento de caracterização sensorial por meio da técnica do PDO atenda ao requisito da magnitude do erro experimental estipulado nesta pesquisa, é recomendada a utilização de equipes com no mínimo dezesseis julgadores. Atendendo este número mínimo de avaliações, a variação aleatória do experimento estará dentro da faixa esperada para estudos descritivos de alimentos, com 90% de probabilidade. O número de julgadores recomendado para o atendimento do critério do erro experimental também possibilitou que 100% dos sub-grupos atendam aos limites estabelecidos para os critérios de significância do efeito de tratamentos e similaridade das configurações espaciais obtidas pelo mapa sensorial.

## **5. Conclusão**

O critério da magnitude da estimativa do erro experimental mostrou-se como uma medida mais robusta na determinação do número de julgadores necessários para a técnica PDO. Como esta técnica descritiva requer baixo nível de treinamento dos julgadores, a avaliação deste critério é de extrema importância, pois uma variação aleatória residual maior pode ser normalmente observada. A utilização de um limite (“ponto de corte”) global para este critério também se mostrou interessante, uma vez que foi verificada que as duas matrizes alimentares avaliadas apresentaram recomendações de número de julgadores diferentes.

Para o atendimento ao critério da magnitude do erro experimental, dezesseis julgadores devem compor a equipe sensorial. Os critérios de significância dos tratamentos e mapeamento sensorial também foram atendidos plenamente quando este número de avaliadores foi utilizado.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

Brandt, M. A., Skinner, E. Z., & Coleman, J. A. (1963). Texture profile method. *Journal of Food Science*, 28, 404–409.

- Cairncross, S. E., & Sjoström, L. B. (1950). Flavour profiles: a new approach to flavor problems. *Food Technology*, 4, 308–311.
- Dijksterhuis, G. (1996). Procrustes analysis in sensory research. In: *Multivariate Analysis of Data in Sensory Science*. Ed. by T. Næs and E. Risvik. Elsevier Science B.V.
- Gacula, M., Rutenbeck, S. (2006). Sample size in consumer test and descriptive analysis. *Journal of Sensory Studies*, 21, 129-145.
- Heymann, H., Machado, B., Torri, L., & Robinson, A.L. (2012). How many judges should one use for sensory descriptive analysis? *Journal of Sensory Studies*, 27, 111-122.
- Hough, G., Wakeling, I., Mucci, A., Chambers, E., Gallardo, I.M., & Alves, L.R. (2006). Number of consumers necessary for sensory acceptability tests. *Food Quality and Preference*, 17, 522-526.
- King, B.M., Arents, P., & Moreau, N. (1995). Cost / Efficiency evaluation of descriptive analysis panels – I. Panel size. *Journal of Sensory Studies*, 6, 245-261.
- Meilgaard, M.C, Civille, G.V., & Carr, B.T. (2006). *Sensory Evaluation Techniques*. (4th ed.). Boca Raton: CRC Press.
- Montgomery, D.C. (2001). *Design and analysis of experiments*. (5th Ed.). New York: John Wiley and Sons. p.699.
- Pagès, J., & Périnel, E. (2003). Panel performance and number of evaluations in a descriptive sensory study. *Journal of Sensory Studies*, 19, 273-291.
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <<http://www.R-project.org>>.
- Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The ‘RV’ coefficient. *Applied Statistics*, 25(3), 257–265.
- Silva, R.C.S.N., Minim, V.P.R., Simiqueli, A.A., Moraes, L.E.S., Gomide, A.I., Minim, L.A. (2012). Optimized Descriptive Profile: a rapid methodology for sensory description. *Food Quality and Preference*, 24, 190-200.
- Silva, R.C.S.N., Minim, V.P.R., Carneiro, J.D.S., Nascimento, M., Della Lucia, S.M., Minim, L.A. (2013a). Quantitative sensory description using the Optimized Descriptive Profile: comparison with conventional and alternative methods for evaluation of chocolate. *Food Quality and Preference*, 30, 169-179.
- Silva, R.C.S.N., Minim, V.P.R., Silva, A.N., Minim, L.A. (2013b). Number of judges necessary for descriptive sensory tests. *Food Quality and Preference*, 31, 22-27.
- Stone, H., & Sidel, J.L. (1985). *Sensory evaluation practices*. (1st ed.) New York: Academic.
- Stone, H., & Sidel, J.L. (2012). *Sensory evaluation practices*. (4th ed.) New York: Academic.



## *Delineamento em Blocos Incompletos*

### *Balanceados: uma alternativa para*

### *coleta de dados no Perfil Descritivo*

## *Otimizado*

#### **RESUMO**

O Perfil Descritivo Otimizado (PDO) é uma nova metodologia descritiva que propõe uma avaliação otimizada das amostras. Para possibilitar que julgadores semi-treinados avaliem as amostras quantitativamente de forma consistente, foi proposta a apresentação dos materiais de referência (“nenhum/fraco” e “forte”) juntamente às amostras, no momento da avaliação sensorial. O protocolo atributo-por-atributo é recomendado por possibilitar a comparação entre as amostras e evitar a fadiga, uma vez que somente um atributo é avaliado por vez. A apresentação simultânea das amostras, juntamente com os materiais de referência, é o “ponto-chave” do PDO e, desta forma, o número de amostras pode ser um fator limitante nesta técnica. A proposta desta pesquisa é avaliar se o Delineamento em Blocos Incompletos Balanceados (DBIB) pode ser uma alternativa para coleta de dados no PDO. No DBIB não é realizada a comparação entre todas as amostras, mas somente entre uma fração delas e, portanto, uma avaliação da qualidade dos dados obtidos por meio deste delineamento se faz necessária para aumentar a aplicabilidade desta técnica. Neste experimento, cinco formulações de suco de uva foram utilizadas, as quais apresentavam variação do teor de suco e açúcar. A avaliação sensorial foi conduzida com três equipes independentes, sendo que cada painel avaliou as amostras segundo uma metodologia e delineamento específico. Duas equipes avaliaram as amostras segundo a técnica do PDO, sendo uma equipe conduzida sob o Delineamento em Blocos Completos Balanceados (DBCB) e outra equipe sob o DBIB. A terceira equipe avaliou as amostras por meio da técnica Perfil Convencional (PC), utilizando equipe treinada. Os dados obtidos pelas três técnicas/delineamentos foram comparados em relação ao mapa sensorial, Análise de Variância e modelos de regressão. Os dados obtidos por meio do PDO conduzido no DBIB apresentaram-se muito semelhantes aos dados do PDO segundo o DBCB e também aos dados obtidos por meio da equipe treinada, no PC. Os escores médios obtidos nas diferentes

técnicas/delineamentos não apresentaram diferença significativa pelo teste F ( $p > 0,05$ ), considerando os diferentes atributos sensoriais. Os mapas sensoriais apresentaram elevada semelhança, com coeficiente RV superior a 0,95. Na ANOVA, as formulações diferiram entre si ( $p < 0,05$ ), em relação a todos os atributos, apresentando poder do teste ( $1 - \beta$ ) superior a 0,95. Os modelos de regressão obtidos nas diferentes técnicas/delineamentos apresentaram *overlapping* dos intervalos de confiança (IC<sub>95%</sub>). Portanto, o DBIB apresenta-se como uma alternativa para coleta de dados no PDO sem perda na qualidade dos resultados.

**Palavras-chave:** Delineamento em Blocos Incompletos Balanceados; Delineamento em Blocos Completos Balanceados; protocolo atributo-por-atributo; *overlapping*.

## **1. Introdução**

O Perfil Descritivo Otimizado (PDO) é uma metodologia descritiva que foi proposta recentemente (Silva et al., 2012) com o objetivo de reduzir o tempo do teste sensorial e, ao mesmo tempo, fornecer informações quantitativas sobre os atributos sensoriais presentes nos alimentos. Nesta técnica um painel de julgadores semi-treinados avalia a intensidade dos estímulos sensoriais em uma escala não estruturada ancorada nos extremos pelos termos “nenhum/fraco” e “forte”, os quais são representados por materiais de referência quantitativos. A avaliação dos produtos pelo painel é realizada por meio do protocolo atributo-por-atributo, o qual foi recomendado por Ishii, Chang e O’Mahony (2007) para julgadores não treinados.

No protocolo atributo-por-atributo, os julgadores recebem todas as amostras de uma só vez e são orientados a compará-las entre si, em relação a um único atributo, e indicar a intensidade do estímulo na escala não estruturada. Nesse protocolo, a re-degustação dos produtos e re-avaliação da pontuação pode ser realizada até o convencimento de que as pontuações representam o "espaçamento" correto entre as intensidades (Ishii, Stampanoni & O’Mahony, 2008).

No PDO, o protocolo atributo-por-atributo é aliado à presença dos materiais de referência no momento da avaliação dos produtos. As referências quantitativas são apresentadas juntamente com as amostras para possibilitar que os julgadores comparem as amostras com as referências, o que facilita a alocação da intensidade dos atributos na escala não estruturada (Silva et al., 2012). Esta estruturação do teste (todas as amostras apresentadas juntas e a presença dos materiais de referência) é o “ponto chave” do PDO, pois possibilita que julgadores com baixo grau de treinamento avaliem a intensidade dos produtos de forma consistente, como foi verificado nos estudos de Silva et al. (2012; 2013a).

A avaliação de um único atributo por sessão é recomendada no PDO com o objetivo de evitar o efeito da fadiga sensorial, uma vez que o julgador deve provar todas as amostras e os materiais de referência (“fraco” e “forte”) antes da alocação da intensidade do estímulo na escala. A re-degustação pode ser necessária em alguns casos, podendo intensificar o efeito de fadiga sensorial se o número de amostras for elevado. Silva et al. (2012) citam o número de amostras como uma possível limitação na metodologia do PDO, uma vez que a apresentação simultânea de um número elevado de

amostras, além das referências, pode provocar a fadiga sensorial e não representar a intensidade real do estímulo sensorial.

Neste sentido, o objetivo desta pesquisa foi avaliar se o Delineamento em Blocos Incompletos Balanceados pode ser uma forma alternativa para a avaliação de um maior número de amostras na metodologia Perfil Descritivo Otimizado.

## 2. Apresentação dos Delineamentos

### 2.1. Delineamento em Blocos Incompletos Balanceados (DBIB)

Os blocos incompletos balanceados são recomendados como delineamento de apresentação das amostras nos métodos descritivos quando o número total de amostras do estudo é superior ao número que pode ser avaliado na sessão antes da fadiga sensorial (Gacula et al., 2008; Best, Rayner, & Allingham, 2011). No DBIB, os julgadores avaliam numa mesma sessão somente uma fração ( $k$ ) do número total de amostras ( $t$ ), sendo  $k < t$ , as quais são apresentadas simultaneamente (Stone & Sidel, 2012). Nesse delineamento todas as amostras são avaliadas um mesmo número de vezes (denotado por  $r$ ) e o número de vezes que um par de amostras avaliado junto é o mesmo para todos os pares e é denotado por  $\lambda$ . O fato de  $r$  e  $\lambda$  serem constantes para todas as amostras no DBIB garante que cada média de tratamento seja estimada com igual precisão e que todas as comparações pareadas entre duas médias de tratamentos sejam igualmente sensíveis (Montgomery, 2001). O número de blocos requeridos para completar uma repetição do delineamento básico é denotado por  $b$  e o número de repetições do delineamento básico é denotado por  $p$ . Uma lista de DBIB é apresentada por Cochran e Cox (1957) e é útil para selecionar um delineamento específico para o estudo.

Segundo Meilgaard, Civille e Carr (2006) duas abordagens do DBIB podem ser adotadas em função do número de blocos, as quais são: (i) *número elevado de blocos*: cada julgador avalia apenas um bloco ( $b$ ) de amostras, sendo, então, necessários  $pb$  julgadores para completar o delineamento básico; (ii) *número pequeno de blocos*: cada julgador retorna diversas vezes até completar toda a repetição do delineamento básico, assim, cada julgador avalia todos os blocos ( $b$ ) de amostras e o número de repetições ( $p$ ) é equivalente ao número de julgadores.

A forma de utilizar a Análise de Variância (ANOVA) para analisar dados provenientes do DBIB depende de como o delineamento é administrado. Assim, se cada julgador avaliar somente um bloco de amostras, caso (i), então o efeito de julgador ficará confundido (“misturado”) com o efeito de blocos. Já no caso (ii), em que o julgador avalia todos os blocos do delineamento básico, o efeito de julgador pode ser particionado da variabilidade total, sendo adotado o modelo na ANOVA descrito da equação (1) (Meilgaard, Civille, & Carr, 2006; Stone & Sidel, 2012).

$$\text{Modelo} = \text{Amostras}_{(\text{ajustada para blocos})} + \text{Julgadores} + \text{Blocos}_{(\text{dentro de julgadores})} + \text{Interação}_{(\text{amostras} \times \text{julgadores})} + \text{Erro.} \quad \text{Eq. (1)}$$

A metodologia do PDO não apresenta nenhum estudo realizado segundo o DBIB e, portanto, existe uma indagação se a técnica apresenta resultados exatos quando os dados são coletados por meio deste delineamento. A proposta do PDO é a comparação entre as amostras e com as referências antes da alocação da intensidade dos estímulos na escala não-estruturada. No DBIB não é possibilitada a comparação entre todas as amostras, mas somente entre uma fração delas e, portanto, a avaliação da viabilidade deste delineamento para coleta de dados no PDO faz-se necessária para aumentar a aplicabilidade desta técnica.

## 2.2. *Delineamento em Blocos Completos Balanceados (DBCB)*

O Delineamento em Blocos Completos Balanceados é o delineamento mais utilizado em testes descritivos e é recomendado quando o número de amostras é pequeno e a fadiga sensorial não é uma preocupação (Meilgaard, Civille, & Carr, 2006). Neste delineamento, os julgadores são os “blocos”, os quais avaliam todas as amostras nas  $r$  repetições. Desta forma, o número de julgadores corresponde ao número de blocos ( $b$ ) e o efeito de julgador é particionado da variância total na Análise de Variância (ANOVA). As amostras são apresentadas aos julgadores de forma aleatória e balanceada e as repetições das avaliações dos julgadores (repetições intra-bloco) permitem a estimação do efeito de interação entre os julgadores e amostras, sendo adotado o modelo descrito na equação (2) (Stone & Sidel, 2012).

$$\text{Modelo} = \text{Amostras} + \text{Julgadores}_{(\text{blocos})} + \text{Interação}_{(\text{amostras} \times \text{julgadores})} + \text{Erro} \quad \text{Eq. (2)}$$

A forma de apresentação das amostras aos julgadores no DBCB pode seguir dois protocolos diferentes, os quais são: (i) *protocolo simultâneo*: todas as amostras são apresentadas simultaneamente aos julgadores em uma mesma sessão, sendo possível a comparação entre elas em relação à lista completa de atributos; (ii) *protocolo monádico*: as amostras são apresentadas monadicamente, ou seja, de forma individualizada. No protocolo monádico, a comparação não é permitida e o julgador avalia uma amostra em relação a todos os atributos listados antes que outra amostra lhe seja apresentada. Em ambos os protocolos a ordem de apresentação das amostras, inclusive das repetições, são aleatorizadas e balanceadas.

Ishii, Stampanoni e O'Mahony (2008) recomendam a avaliação de uma única amostra por vez (protocolo monádico). O agrupamento das amostras em uma única apresentação tem sido observado na literatura; este possibilita ao julgador comparar e ordenar as amostras antes da alocação da intensidade do atributo na escala não estruturada, o que é indesejável para julgadores treinados.

Nos estudos realizados por Silva et al. (2012; 2013a) com a metodologia do PDO (julgadores semi-treinados), as avaliações das amostras foram conduzidas por meio do Delineamento em Blocos Completos Balanceados e segundo o protocolo atributo-por-atributo, o qual se assemelha ao o protocolo simultâneo, porém somente um atributo é avaliado por sessão. Assim, o número de sessões necessárias corresponde ao número de atributos multiplicado pelo número de repetições das avaliações. Nestes mesmos estudos, o PDO foi comparado com a técnica tradicional de descrição sensorial, o Perfil Convencional (PC), o qual utiliza um painel treinado. Na condução desta metodologia (PC), o DBCB foi utilizado aliado ao protocolo monádico, o qual é recomendado para julgadores treinados (Ishii, Stampanoni & O'Mahony, 2008).

### **3. Materiais e Métodos**

#### *3.1. Estímulos sensoriais*

Foram utilizadas cinco formulações de suco de uva, as quais apresentavam diferentes teores de suco concentrado e açúcar. As formulações foram obtidas segundo o esquema fatorial com ponto central (Gacula et al., 2008), sendo dois fatores em dois níveis mais um ponto central ( $2^2 + 1$ ), totalizando cinco formulações (Tabela 4.1).

Tabela 4.1 - Formulações dos sucos de uva.

Formulações	Suco concentrado		Açúcar	
	Codificação	Proporção (suco:água)	Codificação	Concentração (%)
A	+1	1:1	-1	0
B	+1	1:1	+1	2
C	-1	1:3	+1	2
D	-1	1:3	-1	0
E	0	1:1.7	0	1

As formulações foram elaboradas com suco concentrado sabor de uva da marca comercial Bela Ischia® e com açúcar refinado comercial (União®). O suco concentrado foi diluída em água nas proporções desejadas e adicionadas de açúcar nas proporções previamente definidas. Os sucos foram mantidos sob refrigeração e servidos a temperatura de  $4 \pm 1$  °C.

### 3.2. Procedimento da Análise Sensorial

A caracterização sensorial dos sucos de uva foi realizada por meio de três painéis independentes, sendo que cada painel avaliou as amostras segundo uma metodologia e delineamento específico. Duas equipes avaliaram as amostras segundo a técnica do PDO, sendo uma equipe conduzida sob o delineamento em blocos incompletos balanceados (equipe PDO/DBIB) e outra equipe sob o delineamento em blocos completos balanceados (equipe PDO/DBCB). A terceira equipe avaliou as amostras por meio da técnica Perfil Convencional segundo o DBCB (equipe PC/DBCB).

Primeiramente, foram recrutados setenta julgadores, dos quais quarenta e cinco foram pré-selecionados com base no poder discriminatório dos produtos. Todos os participantes pré-selecionados levantaram os termos descritivos a serem avaliados nos sucos de uva e definiram os materiais de referência (“fraco” e “forte”) para cada estímulo sensorial. Após este procedimento, os julgadores foram divididos em três equipes diferentes para avaliação dos produtos. Assim, uma equipe de dezessete julgadores avaliou os sucos por meio da metodologia do PDO segundo o DBCB, outros dezessete julgadores por meio do PDO segundo o DBIB e, por fim, onze julgadores utilizando o PC segundo o DBCB.

Nas três técnicas, os julgadores foram recrutados por meio de questionários estruturados, conforme recomendado por Meilgaard, Civille e Carr (2006). Após o recrutamento, estes julgadores foram pré-selecionados com base na capacidade discriminatória de dois sucos de uva (Suco I e II). Para tanto, foi realizada uma série de quatro testes triangulares e foram selecionados os julgadores que acertaram pelo menos 75% dos testes. As proporções utilizadas na elaboração das formulações foram: Suco I) proporção suco:água igual a 1:1, adicionado de 1% de açúcar; Suco II) proporção suco:água igual a 1:3, sem adição de açúcar. Após a pré-seleção dos participantes com base no poder discriminatório, foram levantados os atributos que descreviam os sucos-teste por meio da técnica de discussão aberta, conforme descrito por Damásio e Costell (1991). Os atributos sensoriais que foram levantados pela equipe foram: cor bordô, aroma de uva, gosto doce, gosto ácido, sabor de uva e sensação adstringente.

Para a elaboração dos materiais de referência que ancoraram os extremos (“fraco” e “forte”) da escala não-estruturada de 9 cm, utilizaram-se formulações com adição de suco e açúcar que permitiram a obtenção de intensidades sensivelmente diferentes e extremas às formulações em avaliação.

### 3.2.1. Perfil Descritivo Otimizado: avaliação por meio do Delineamento em Blocos Incompletos Balanceados

Os dezessete julgadores pré-selecionados destinados à equipe do PDO/BIBD participaram de uma sessão de familiarização com os termos descritivos e seus respectivos materiais de referência. Nesta sessão, o objetivo foi padronizar a forma de avaliação e expor claramente a qual estímulo sensorial estava se referindo cada termo, além de ancorar os extremos da escala não estruturada (“fraco” e “forte”). Nesta sessão de familiarização, foram apresentados aos julgadores, em cabine individual, os termos descritivos e seus respectivos materiais de referência. O julgador foi orientado a ler a definição do atributo sensorial e provar as referências. Após esta etapa, as avaliações dos sucos-teste foram iniciadas.

As avaliações dos sucos de uva foram conduzidas segundo o Delineamento em Blocos Incompletos Balanceados (DBIB), de forma que os julgadores avaliaram somente uma fração ( $k = 3$ ) das amostras em cada sessão. O delineamento utilizado está representado na Tabela 4.2, o qual foi proposto por Cochran e Cox (1981), considerando os seguintes parâmetros:

- ✓ número de tratamentos ( $t$ ): 5
- ✓ número de amostras avaliadas em cada bloco ( $k$ ): 3
- ✓ número de repetições ( $r$ ): 6
- ✓ número de vezes que um par de amostras é testado junto ( $\lambda$ ): 3
- ✓ número de blocos ( $b$ ): 10

Cada julgador completou as avaliações das amostras nos dez blocos para cada atributo sensorial, sendo apenas um atributo avaliado por sessão, totalizando sessenta sessões. Os blocos foram designados aos julgadores em ordem aleatória.

Tabela 4.2 - Delineamento em Blocos Incompletos Balanceados, considerando  $t=5$ ,  $k=3$ ,  $r=6$ ,  $\lambda=3$ ,  $b=10$ .

Blocos	Amostras				
	A	B	C	D	E
1	x	x	x		
2	x	x			x
3	x			x	x
4		x	x	x	
5			x	x	x
6	x	x		x	
7	x		x	x	
8	x		x		x
9		x	x		x
10		x		x	x

As amostras de cada bloco foram servidas aos julgadores segundo o protocolo atributo-por-atributo, ou seja, foram servidas simultaneamente. Os sucos foram apresentados em sequência aleatória dentro de cada bloco.

No procedimento de avaliação, o julgador era orientado a comparar as amostras entre si e com os materiais de referência e, então, alocar a intensidade do estímulo na escala de intensidade. A ficha de avaliação foi organizada por atributos e continha a escala não estruturada de 9 cm (intervalar) associada a cada formulação.

### 3.2.2. Perfil Descritivo Otimizado: avaliação por meio do Delineamento em Blocos Completos Balanceados

Os dezessete julgadores pré-selecionados destinados à equipe do PDO/DBCB participaram de uma sessão de familiarização com os termos descritivos e seus

respectivos materiais de referência conforme descrito na seção anterior, sendo esta sessão seguida pelas sessões de avaliação das amostras.

As avaliações dos sucos de uva foram conduzidas segundo o Delineamento em Blocos Completos Balanceados (DBCB), de forma que os julgadores avaliaram todas as formulações numa mesma sessão. O protocolo atributo-por-atributo foi utilizado, sendo as amostras servidas de forma simultânea e apenas um atributo avaliado por sessão. Os sucos foram apresentados em sequência aleatória e balanceada. Foram realizadas três repetições da avaliação por julgador. Totalizaram 18 sessões para a avaliação dos produtos em relação aos seis atributos em três repetições.

No procedimento de avaliação, o julgador era orientado a comparar as amostras entre si e com os materiais de referência e, então, alocar a intensidade do estímulo na escala de intensidade. A ficha de avaliação foi organizada por atributos e continha a escala não estruturada de 9 cm (intervalar) associada a cada formulação.

### 3.2.3. Perfil Convencional: avaliação por meio do Delineamento em Blocos Completos Balanceados

Onze julgadores participaram da avaliação dos sucos de uva por meio do Perfil Convencional. Assim, após o levantamento dos atributos sensoriais e da definição das referências, iniciou-se o treinamento. Nessa etapa, eram apresentados os materiais de referência que ancoravam os extremos da escala de intensidade e era solicitado ao julgador que lesse a definição do atributo sensorial e provasse as referências, tentando memorizá-las. Também foram realizadas algumas sessões de exercícios, nos quais os julgadores recebiam as cinco formulações de suco e eram orientados a avaliá-las na escala não estruturada de 9 cm em relação a todos os atributos sensoriais. O treinamento foi realizado até que a equipe não apresentasse dificuldades na avaliação dos sucos. Para tanto, foram realizadas sessenta e cinco sessões de treinamento.

Para avaliar se os julgadores estavam devidamente treinados, foi realizado um teste para verificar o desempenho individual dos voluntários em relação ao poder de discriminação dos produtos e repetibilidade nas avaliações. Para tanto, foram servidos aos julgadores, em cabine individual, dois sucos (Suco I e II) codificados com três dígitos aleatórios. Os julgadores foram solicitados a avaliar os sucos utilizando a ficha de avaliação do PC. Esta ficha foi organizada por amostra e continha a escala não

estruturada de 9 cm associada a cada termo descritivo. Esse teste de desempenho foi realizado com quatro repetições.

Na análise dos resultados individuais dos julgadores, foi considerado o poder de discriminação das formulações de suco e a repetibilidade de resultados. Para tanto, foi realizada uma ANOVA com duas fontes de variação (formulação e repetição) para cada atributo sensorial por julgador. Foi selecionado o julgador que apresentou probabilidade máxima de 30% na discriminação das formulações ( $p.F_{\text{FORMULAÇÃO}} < 0,30$ ) e probabilidade mínima de 20% na repetibilidade dos resultados ( $p.F_{\text{REPETIÇÃO}} > 0,20$ ), conforme critérios adaptados da recomendação de Damásio e Costell (1991). Com base nesses critérios, todos os julgadores apresentaram desempenho satisfatório e participaram da avaliação final dos sucos.

Na avaliação final dos sucos, as formulações foram avaliadas pelos julgadores segundo o Delineamento em Blocos Completos Balanceados, sendo cada julgador considerado um bloco. Cada julgador avaliou as cinco formulações de suco numa mesma sessão, sendo as amostras apresentadas de forma monádica, aleatória e balanceada. Foram realizadas três repetições das avaliações. A avaliação das características sensoriais foi efetuada na ficha de avaliação organizada por amostras, a qual continha a escala não estruturada de 9 cm associada a cada termo descritivo.

### *3.3. Análises estatísticas*

Os dados descritivos foram analisados por meio da Análise de Componentes Principais (ACP), Análise de Variância (ANOVA), e análise de regressão múltipla, as quais foram realizadas no software R (R Core Team, 2012).

A Análise de Componentes Principais foi utilizada para obtenção do mapa sensorial e foi realizada segundo procedimento descrito em Gacula et al. (2008).

Na ANOVA, foram considerados os modelos descritos nas equações (1) e (2), segundo o delineamento apropriado. No caso de interação significativa, o teste F para amostras foi calculado tendo o  $QM_{\text{INTERAÇÃO}}$  como denominador, conforme recomendado por Stone e Sidel (2012). O nível de significância (erro tipo I) da fonte de variação de interesse (amostra), obtido na ANOVA (probabilidade de  $F_{\text{AMOSTRA}}$ ), foi observado como critério para avaliação do poder de discriminação das amostras nas diferentes técnicas/delineamentos. A análise de poder do teste F também foi realizada para complementar a avaliação do poder de discriminação de amostras relativo a cada

técnica/delineamento. O poder do teste foi calculado pela função de probabilidade [1 - P(erro tipo II)] fixando o nível de significância (erro tipo I) a 5%, conforme descrito em O'Brien e Lohr (1984).

A Análise de regressão múltipla foi utilizada para analisar a influência dos fatores quantitativos (concentração de suco e açúcar), variáveis independentes, sobre as variáveis dependentes (atributos sensoriais do suco). O ajuste dos modelos de regressão foi realizado conforme a recomendação de Montgomery (2001), sendo avaliada a falta de ajuste do modelo e a significância dos parâmetros, ao nível de 5% de probabilidade. O coeficiente de determinação também foi obtido.

### 3.3.1. Comparação entre as técnicas/delineamentos

Para comparar os escores médios obtidos pelas diferentes técnicas/delineamentos (PDO/DBIB, PDO/DBCB e PC/DBCB) foi realizada uma ANOVA considerando cada atributo sensorial. A ANOVA foi realizada com duas fontes de variação: técnica/delineamento e amostra, sendo verificado se houve diferença entre os escores médios ao nível de 5% de probabilidade.

#### 3.3.1.1. Comparação entre as configurações espaciais

A comparação dos mapas sensoriais dos sucos de uva obtidos por meio das três técnicas/delineamentos (PDO/DBIB, PDO/DBCB e PC/DBCB) foi realizada por meio do coeficiente RV (Robert & Escoufier, 1976). Este coeficiente é uma generalização multivariada do coeficiente de correlação e, portanto, mede o grau de concordância entre as configurações e varia de 0 (discordância total) a 1 (concordância perfeita) (Dijksterhuis, 1996).

#### 3.3.1.2. Comparação entre os modelos de regressão

A semelhança dos modelos de regressão múltipla ajustados aos dados de descrição sensorial nas diferentes técnicas/delineamentos foi verificada por meio do *overlapping* dos intervalos de confiança dos parâmetros, com 95% de confiança (Van Belle, 2002; Schenker, & Gentleman, 2005).

#### 3.3.4. Estudo dos sub-grupos de amostras no DBIB

Foi realizado um estudo para verificar se houve diferença entre os escores médios dos sucos quando estes foram avaliados nos diferentes blocos (*b*), ou seja, nos diferentes agrupamentos de *k* amostras, semelhante ao procedimento realizado por Villanueva et al. (2005). O objetivo desta avaliação foi verificar se houve alteração significativa nos escores médios atribuídos a uma determinada amostra quando esta se encontrava em comparação com grupos diferentes de amostras, verificando os contrastes entre estímulos.

A comparação dos escores médios das amostras, quando avaliadas nos diferentes blocos, foi realizada por meio da ACP e do teste de médias de *Tukey*, ao nível de 5% de probabilidade. Na ACP, cada formulação de suco teve seis escores médios plotados, sendo cada um referente a uma repetição. Os grupos foram observados pela análise de *cluster*, utilizando o procedimento *k-means clustering*, considerando *k* = 5 grupos (Johnson & Wichern, 2007).

## 4. Resultados

### 4.1. Comparação entre as técnicas/delineamentos (PDO/DBIB, PDO/DBCB e PC/DBCB)

Os escores médios dos atributos sensoriais obtidos por meio das diferentes técnicas/delineamentos não apresentaram diferença significativa ( $p > 0,05$ ) pelo teste F (Tabela 4.3). O PDO conduzido no delineamento em blocos incompletos balanceados apresentou escores médios semelhantes aos dados obtidos quando a mesma metodologia foi conduzida no delineamento em blocos completos balanceados. Houve semelhança também entre os dados do PDO, em ambos os delineamentos, com a técnica tradicional (PC), a qual utiliza julgadores treinados. Este resultado evidencia a capacidade do PDO de produzir avaliações semelhantes à técnica de referência (PC), independente do delineamento de condução da técnica (DBIB ou DBCB).

Tabela 4.3 - ANOVA para comparação das técnicas/ delineamentos

<b>Atributos sensoriais</b>	<b>Fontes de Variação</b>	<b>Graus de Liberdade</b>	<b>Quadrado Médio</b>	<b>F</b>	<b>p-valor</b>
<i>Cor bordô</i>	Técnica/delineamento	2	0,02500	0,02ns	0,9820
	Amostra	4	37,4973		
	Erro	53	1,3775		
<i>Aroma de uva</i>	Técnica/delineamento	2	1,6243	1,05ns	0,3579
	Amostra	4	13,6929		
	Erro	53	1,5502		
<i>Gosto doce</i>	Técnica/delineamento	2	1,8495	1,44ns	0,2460
	Amostra	4	20,2306		
	Erro	53	1,28439		
<i>Gosto ácido</i>	Técnica/delineamento	2	2,0655	1,97ns	0,1492
	Amostra	4	18,0139		
	Erro	53	1,0472		
<i>Sabor de uva</i>	Técnica/delineamento	2	1,1685	1,44ns	0,2467
	Amostra	4	28,7289		
	Erro	53	0,8129		
<i>Adstringente</i>	Técnica/delineamento	2	0,8030	1,02ns	0,3688
	Amostra	4	20,1539		
	Erro	53	0,7899		

ns: não significativo a 0,05 de probabilidade.

#### 4.1.1. Perfil sensorial dos sucos de uva

Na ACP, os dois primeiros componentes principais explicaram juntos mais de 99% da variação total dos dados para as técnicas/delineamentos (PDO/DBIB, PDO/DBCB e PC/ DBCB) e, portanto, somente as duas primeiras dimensões foram consideradas na representação gráfica e interpretação dos resultados. Este tipo de análise fornece uma representação bidimensional de fácil visualização e interpretação do perfil sensorial dos produtos (Figura 4.1). Todos os atributos sensoriais apresentaram correlação com o primeiro componente principal ( $p < 0,05$ ) e somente o atributo gosto doce apresentou correlação com os dois primeiros componentes.

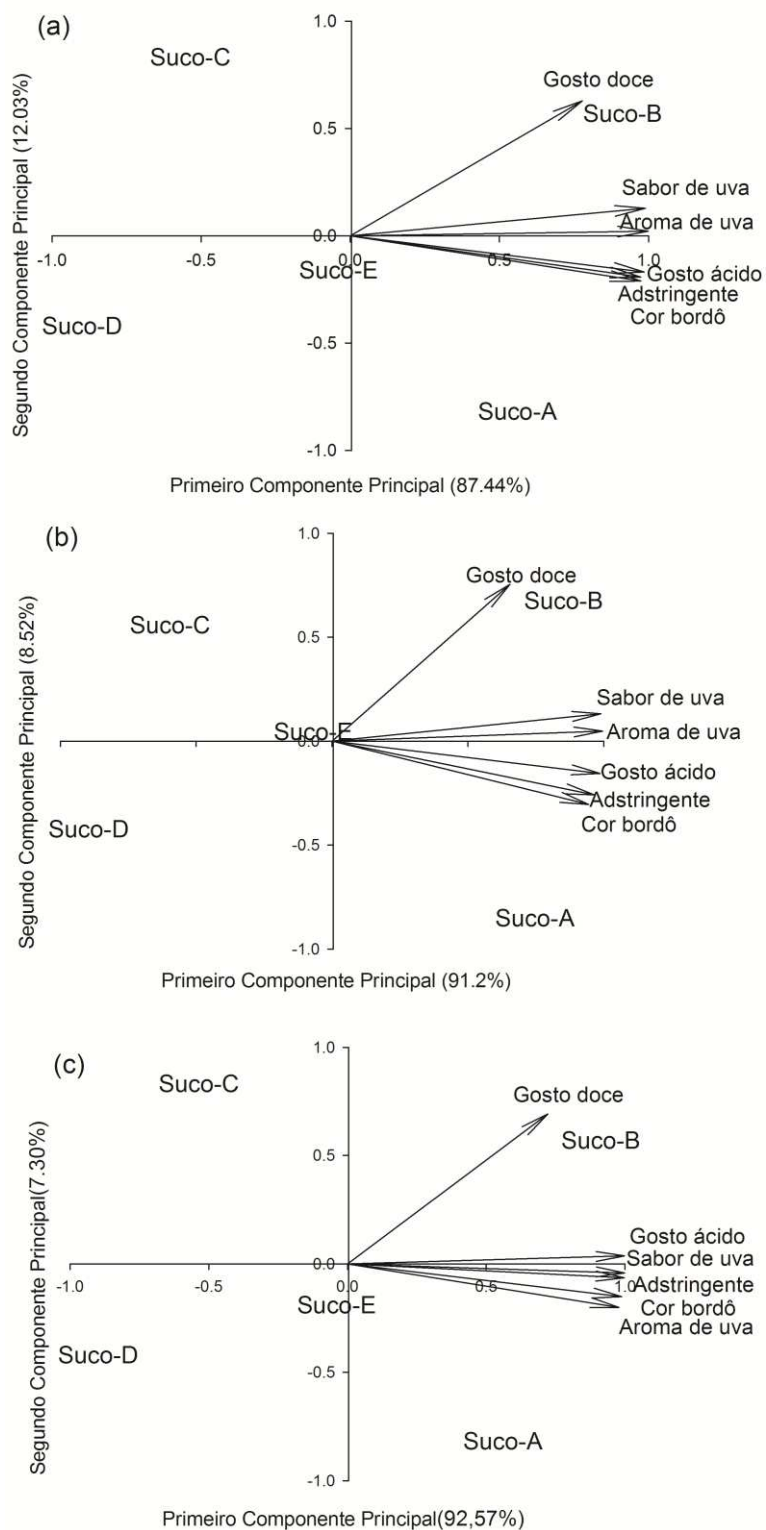


Figura 4.1 - Mapas sensoriais dos sucos de uva para o PDO no BIBD (a) e no BCBD (b) e para o Perfil Convencional no DBC com protocolo de avaliação monádico (c).

Os sucos com maior concentração de suco de uva (A e B) foram caracterizados por apresentar maior intensidade dos atributos cor bordô, aroma e sabor de uva, gosto ácido e sensação adstringente. Já os sucos mais diluídos (C e D) apresentaram perfil sensorial oposto. Os sucos B e C (maior teor de açúcar) apresentaram gosto doce mais intenso. O suco E, com média concentração de suco e de açúcar, apresentou intensidade intermediária dos atributos sensoriais.

As três configurações espaciais apresentaram semelhança na formação dos grupos de amostras e na correlação dos atributos sensoriais com os componentes principais. Tal semelhança foi confirmada pelo coeficiente RV com grau de similaridade superior a 0,95 na comparação das três configurações.

#### 4.1.2. ANOVA: DBIB e DBCB

Um resumo da ANOVA considerando as diferentes técnicas/delineamentos está descrito na Tabela 4.4. No DBCB, para as técnicas PDO e PC, somente quatro fontes de variação foram particionadas da variação total, sendo o “efeito de blocos” representado pela variação entre os julgadores. Já no DBIB, há uma fonte de variação extra, sendo estimada a variabilidade dentro da avaliação de cada julgador devida às repetições, nos diferentes blocos de amostras.

A variação residual ( $QM_{ERRO}$ ) e a variação entre julgadores ( $QM_{JULGADORES}$ ) foram semelhantes nas técnicas/delineamentos, mostrando que a não necessidade de treinamento das equipes do PDO não acarretou aumento na variabilidade entre as avaliações dos diferentes julgadores e nas diferentes repetições. Uma comparação mais efetiva dos componentes de variância é impossibilitada devido ao número diferenciado de julgadores nas técnicas (PDO e PC) e ao número diferente de repetições nos delineamentos (DBIB e DBCB), o que acarreta em diferença nos graus de liberdade das fontes de variação.

#### 4.1.3. Poder de discriminação

A discriminação das amostras e o poder de decisão do teste F foram semelhantes nas três técnicas/delineamentos, sendo as decisões nos testes de hipóteses significativas ao mesmo nível de probabilidade.

Tabela 4.4 - Resumo da ANOVA obtida para cada técnica/delineamento.

Técnica / Delineamento	Fontes de Variação	GL	Quadrado Médio					
			Cor bordô	Aroma de uva	Gosto doce	Gosto ácido	Sabor de uva	Adstringente
PDO / DBIB	Amostras <small>(ajustada para blocos)</small>	4	568,5455	327,7806	371,8679	292,9823	416,4389	314,5753
	Julgadores	16	9,7541	12,6985	10,5825	13,3764	11,7441	17,1937
	Blocos <small>(dentro dos julgadores)</small>	153	3,1057	3,2091	2,6397	3,4986	2,9241	3,5706
	Interação	64	3,2133	6,0704	4,0061	6,5063	5,8122	5,0995
	Erro	272	0,48105	2,0811	1,3557	2,0248	1,0475	2,2097
PDO / DBCB	Amostras	4	299,1208	134,2522	168,5611	151,1814	191,4267	124,3403
	Julgadores (Blocos)	16	5,3929	4,3889	5,3590	14,0418	5,6429	15,9219
	Interação	64	1,9738	4,2534	4,0730	3,9950	4,0688	3,6597
	Erro	170	0,7912	2,9840	1,6660	2,6401	2,0709	2,8607
PC / DBCB	Amostras	4	184,9096	105,7896	96,9279	104,3682	140,2863	109,6030
	Julgadores (Blocos)	10	10,3859	16,1556	10,7409	9,4436	8,0163	6,6339
	Interação	40	2,5389	2,1556	3,8527	5,1522	3,0535	4,7397
	Erro	110	0,6801	1,3101	1,3783	1,3620	1,1567	1,5665

As formulações de suco de uva apresentaram diferença significativa entre si ( $p < 0,05$ ) para todos os atributos sensoriais. O  $F_{AMOSTRA}$  na ANOVA foi dado pela razão entre o  $QM_{AMOSTRA}$  e o  $QM_{INTERAÇÃO}$ , uma vez que o efeito da interação foi significativo ( $p < 0,05$ ) para todos os atributos. O poder de discriminação das amostras foi avaliado pelo nível de significância do efeito de “amostra” (p-valor do  $F_{AMOSTRA}$ ), o qual foi inferior a 0,0001. A rejeição da hipótese de nulidade para o efeito de “amostra” foi realizada com poder de decisão superior a 0,95, apresentando probabilidade de ocorrência do erro tipo II inferior a 0,05.

#### 4.1.4. Influência dos teores de suco e açúcar no perfil sensorial: modelos de regressão

Os teores de suco e de açúcar influenciaram as características sensoriais do suco e seus efeitos foram modelados por meio do modelo linear (Tabela 4.5). A significância e as estimativas dos parâmetros da regressão foram semelhantes nas três técnicas/delineamentos (PDO/DBIB, PDO/DBC B e PC/DBC B), conforme ilustrado na Figura 4.2.

Tabela 4.5 - Modelos de regressão múltipla ajustados para cada atributo nas diferentes técnicas/delineamentos.

Atributos sensoriais	PDO / DBIB	PDO / DBC B	PC / DBC B
Cor bordô	$\hat{Y}_{cor} = 4,11 + 2,40S$ $R^2 = 0,9632$	$\hat{Y}_{cor} = 4,18 + 2,43S$ $R^2 = 0,9831$	$\hat{Y}_{cor} = 4,17 + 2,35S$ $R^2 = 0,9704$
Aroma de uva	$\hat{Y}_{aroma} = 4,30 + 1,77S + 0,35A$ $R^2 = 0,9078$	$\hat{Y}_{aroma} = 3,99 + 1,63S + 0,44A$ $R^2 = 0,9575$	$\hat{Y}_{aroma} = 4,07 + 1,76S + 0,09A$ $R^2 = 0,9378$
Gosto doce	$\hat{Y}_{doce} = 3,74 + 1,06S + 1,60A$ $R^2 = 0,9097$	$\hat{Y}_{doce} = 3,75 + 1,11S + 1,51A$ $R^2 = 0,9801$	$\hat{Y}_{doce} = 3,62 + 1,11S + 1,31A$ $R^2 = 0,9529$
Gosto ácido	$\hat{Y}_{ácido} = 3,89 + 1,66S$ $R^2 = 0,8140$	$\hat{Y}_{ácido} = 3,63 + 1,75S$ $R^2 = 0,9498$	$\hat{Y}_{ácido} = 4,20 + 1,75S$ $R^2 = 0,9361$
Sabor de uva	$\hat{Y}_{sabor} = 4,36 + 1,92S + 0,62A$ $R^2 = 0,9437$	$\hat{Y}_{sabor} = 4,09 + 1,85S + 0,70A$ $R^2 = 0,9755$	$\hat{Y}_{sabor} = 4,47 + 2,03S + 0,26A$ $R^2 = 0,9634$
Adstringente	$\hat{Y}_{adstringente} = 4,15 + 1,73S$ $R^2 = 0,8572$	$\hat{Y}_{adstringente} = 3,77 + 1,57S$ $R^2 = 0,9431$	$\hat{Y}_{adstringente} = 4,33 + 1,78S$ $R^2 = 0,9273$

S: teor de suco concentrado; A: teor de açúcar.

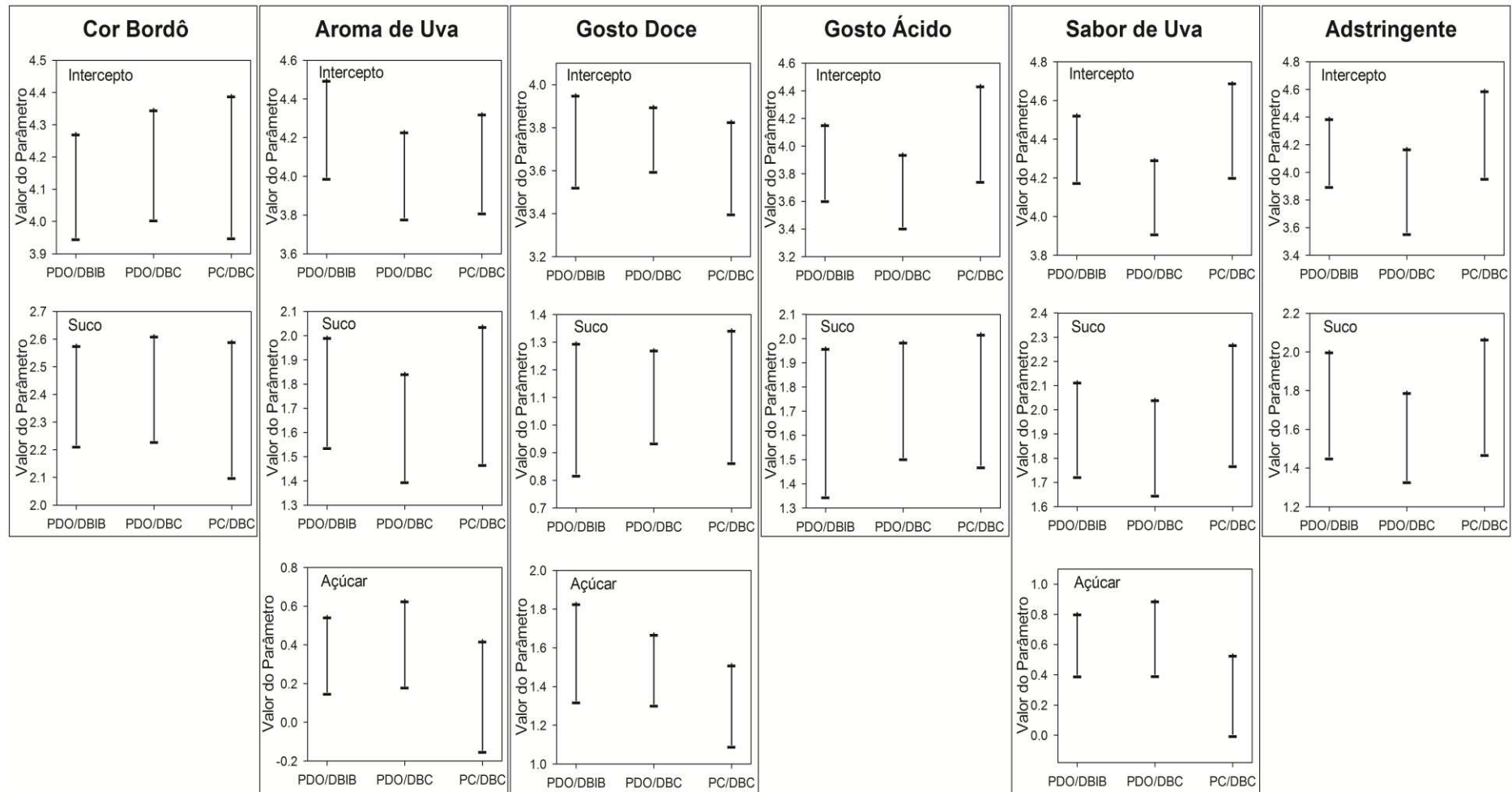


Figura 4.2. Representação dos intervalos de confiança (IC<sub>95%</sub>) dos parâmetros dos modelos lineares para as três técnicas/delineamentos.

O efeito da concentração de suco foi significativa ( $p \leq 0,05$ ) para todos os atributos sensoriais; já o teor de açúcar apresentou influência somente nos atributos gosto doce, aroma e sabor de uva. Todos os modelos de regressão ajustados apresentaram coeficientes de determinação elevados (superiores a 0,85), sendo verificados maiores valores para os dados provenientes do PDO no delineamento em blocos completos, seguido do Perfil Convencional e, por fim, do PDO segundo os blocos incompletos.

#### 4.2. DBIB: comparação entre os sub-grupos de amostras

Na ACP, Figura 4.3, observou-se que as avaliações das amostras de suco foram muito semelhantes quando estas foram avaliadas nos diferentes blocos. Os agrupamentos sugeridos pela ACP foram confirmados pela análise de *cluster*, utilizando o procedimento *k-means clustering* com 1.000.000 de iterações, sendo as repetições de cada amostra agrupadas. Isto indica que o escore médio atribuído às amostras em uma primeira avaliação não foi alterado quando esta mesma amostra foi apresentada pela segunda, terceira ou nas demais vezes. Independente do número de vezes que a amostra já tinha sido apresentada, a intensidade atribuída à amostra foi semelhante em todas as sessões.

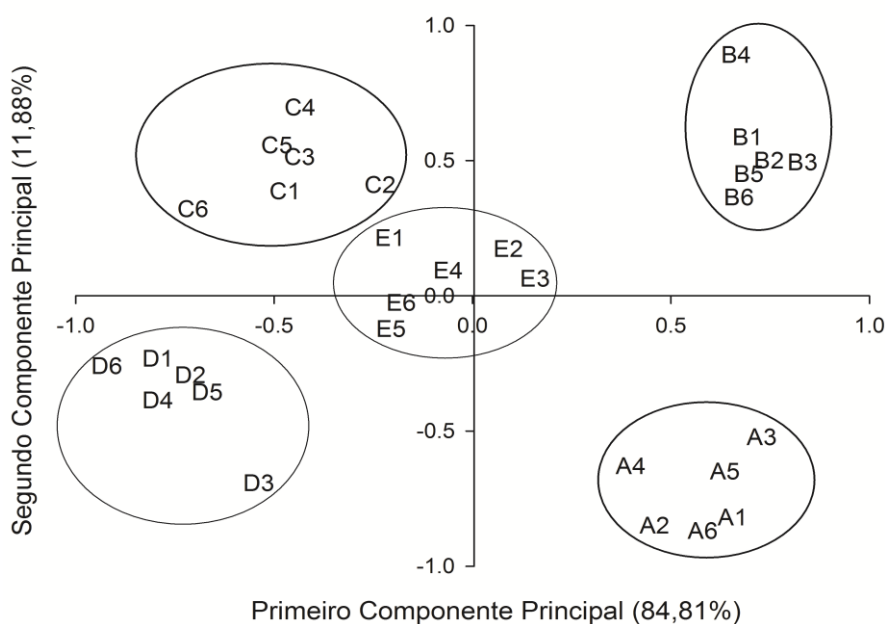


Figura 4.3 - ACP considerando os diferentes blocos do DBIB.

Embora a distribuição espacial das amostras na ACP tenha apresentando semelhança nas repetições das avaliações, quando os escores médios das repetições foram comparados pelo teste de *Tukey* (Tabela 4.6), detectou-se algumas diferenças significativas ( $p \leq 0,05$ ) entre as repetições devido aos contrastes entre estímulos sensoriais.

Tabela 4.6 - Teste de médias de *Tukey* ( $\alpha = 0,05$ ) para os escores médios obtidos nos diferentes blocos do DBIB.

<b>Suco A</b>		<b>Suco B</b>		<b>Suco C</b>		<b>Suco D</b>		<b>Suco E</b>	
<i>bloco</i>	<i>média</i>	<i>bloco</i>	<i>média</i>	<i>bloco</i>	<i>média</i>	<i>bloco</i>	<i>média</i>	<i>bloco</i>	<i>média</i>
<b>Cor bordô</b>									
ABC	7,3a	BAC	7,0a	CAB	1,3b	DAE	1,4a	EAB	3,1a
ABE	6,3ab	BAE	6,3a	CBD	2,7a	DBC	1,9a	EAD	4,0a
ADE	7,1ab	BCD	6,1a	CDE	2,0ab	DCE	2,3a	ECD	4,6a
ABD	6,0b	BAD	6,3a	CAD	2,0ab	DAB	1,5a	EAC	4,3a
ACD	6,7ab	BCE	6,5a	CAE	2,0ab	DAC	1,8a	EBC	4,1a
ACE	6,5ab	BDE	6,3a	CBE	1,3b	DBE	1,4a	EBD	4,1a
<b>Aroma de uva</b>									
ABC	6,3a	BAC	6,1a	CAB	2,7a	DAE	2,6a	EAB	4,4a
ABE	5,5a	BAE	6,3a	CBD	3,9a	DBC	1,9a	EAD	4,7a
ADE	6,1a	BCD	6,0a	CDE	2,3a	DCE	2,5a	ECD	5,4a
ABD	5,1a	BAD	6,5a	CAD	3,3a	DAB	1,7a	EAC	4,7a
ACD	6,4a	BCE	6,4a	CAE	3,0a	DAC	2,5a	EBC	3,6a
ACE	5,4a	BDE	6,8a	CBE	2,5a	DBE	1,6a	EBD	3,6a
<b>Gosto doce</b>									
ABC	2,9a	BAC	6,6a	CAB	3,9a	DAE	1,4a	EAB	3,5ab
ABE	2,4a	BAE	6,9a	CBD	4,2a	DBC	1,7a	EAD	4,6a
ADE	3,8a	BCD	7,0a	CDE	4,6a	DCE	1,1a	ECD	3,9ab
ABD	3,0a	BAD	7,2a	CAD	4,6a	DAB	1,3a	EAC	3,2ab
ACD	3,1a	BCE	6,2a	CAE	4,2a	DAC	1,5a	EBC	2,8b
ACE	2,8a	BDE	5,9a	CBE	2,8a	DBE	1,2a	EBD	3,2ab
<b>Gosto ácido</b>									
ABC	6,4a	BAC	4,8ab	CAB	2,7ab	DAE	2,0ab	EAB	2,7a
ABE	6,0a	BAE	6,0ab	CBD	2,7ab	DBC	2,5ab	EAD	4,0a
ADE	6,3a	BCD	6,3a	CDE	3,3a	DCE	3,5a	ECD	4,0a
ABD	5,5a	BAD	4,1b	CAD	2,2ab	DAB	2,4ab	EAC	3,3a
ACD	6,2a	BCE	4,8ab	CAE	2,2ab	DAC	2,6ab	EBC	2,7a
ACE	6,5a	BDE	5,0ab	CBE	1,4b	DBE	1,0b	EBD	3,3a
<b>Sabor de uva</b>									
ABC	5,8a	BAC	6,8a	CAB	3,3a	DAE	1,6a	EAB	4,1a
ABE	5,5a	BAE	6,4a	CBD	4,0a	DBC	1,8a	EAD	4,2a
ADE	6,6a	BCD	6,7a	CDE	3,1a	DCE	2,0a	ECD	5,1a
ABD	5,2a	BAD	6,9a	CAD	3,4a	DAB	1,5a	EAC	4,6a
ACD	6,3a	BCE	6,8a	CAE	3,0a	DAC	1,9a	EBC	3,8a
ACE	5,8a	BDE	6,5a	CBE	2,5a	DBE	1,1a	EBD	3,9a
<b>Adstringente</b>									
ABC	6,3a	BAC	5,2a	CAB	3,3a	DAE	2,0ab	EAB	3,3a
ABE	6,0a	BAE	6,0a	CBD	3,5a	DBC	2,6ab	EAD	5,0a
ADE	6,6a	BCD	6,3a	CDE	2,6a	DCE	3,8a	ECD	4,2a
ABD	6,1a	BAD	5,3a	CAD	2,3a	DAB	2,6ab	EAC	3,3a
ACD	6,1a	BCE	5,9a	CAE	2,3a	DAC	2,8ab	EBC	3,7a
ACE	6,9a	BDE	5,7a	CBE	1,6a	DBE	1,3b	EBD	3,2a

Letras iguais na coluna, dentro da mesma sessão, não diferem significativamente entre si ao nível de 5%.

Dos trinta e seis grupos formados (blocagens), somente em sete grupos foi verificada diferença significativa ( $p \leq 0,05$ ) entre as repetições realizadas nos diferentes agrupamentos. A percepção sensorial para o suco A apresentou diferença ( $p < 0,05$ ) nos escores médios atribuídos nas diferentes sessões somente em relação à “cor bordô”. A nota atribuída à amostra no bloco ABC diferiu do escore atribuído quando esta se encontrava em comparação com as amostras B e D. A substituição da formulação C para a amostra D no bloco alterou a percepção do estímulo visual do suco A. O contraste entre A e D, formulações com perfis sensoriais opostos (Figura 4.1), provocou uma supressão na percepção do estímulo, sendo este percebido de forma menos intensa (Lim, 2011; Silva et al., 2013b).

Já o suco B teve sua intensidade alterada para o atributo gosto ácido. Quando a amostra B foi testada em comparação com os sucos C e D, a percepção média de acidez foi mais intensa do que quando o mesmo suco foi avaliado em conjunto com as formulações A e D. A alteração da amostra C, com menor intensidade do gosto ácido, para o suco A com maior intensidade do estímulo, provocou uma diminuição na percepção sensorial.

Os sucos C e D foram as formulações mais afetadas pelas blocagens, apresentando diferença em dois atributos, os quais foram: cor bordô e gosto ácido para o suco C e gosto ácido e sensação adstringente para o suco D. Comportamento semelhante foi verificado para os sucos C e D quando estes foram testados em comparação com as amostras E e B. Quando estes estavam em comparação com o suco E (mais diluído) a percepção do gosto ácido foi mais intensa do que quando testados em conjunto com a amostra mais concentrada (suco B).

A formulação E, sem adição de açúcar, teve sua percepção alterada em relação ao gosto doce. Quando este suco foi testado juntamente às formulações com baixo teor de sacarose (A e D), o gosto doce deste suco foi percebido mais intensamente. Já quando o mesmo suco foi avaliado em conjunto com os sucos com maior concentração de sacarose (B e C), a sua doçura foi mais levemente percebida.

## **5. Discussão**

A técnica PDO conduzida segundo o DBIB apresentou perfil sensorial dos sucos de uva muito semelhante ao perfil obtido quando o mesmo método foi conduzido no DBCB (RV igual a 0,99), o qual também foi muito semelhante à avaliação da equipe treinada (PC), apresentando coeficiente RV igual a 0,97. Nos mapas sensoriais, observou-se semelhança na composição dos grupos e na correlação dos atributos sensoriais com os componentes principais. Nos trabalhos de Silva et al. (2012; 2013a) foi verificada elevada semelhança entre os mapas originados pelo PDO e pela técnica convencional (PC). Na presente pesquisa, observa-se que o mesmo perfil sensorial foi obtido para os sucos quando avaliados pelo PDO e pelo PC (equipe treinada), sendo que a configuração espacial não foi alterada quando o delineamento de avaliação das amostras no PDO foi modificado.

Na ANOVA para comparação das técnicas/delineamentos (PDO/DBIB, PDO/DBCB e PC/DBCB) não houve diferença entre os painéis ( $p > 0,05$ ), indicando que o DBIB apresenta-se como alternativa para coleta de dados na técnica PDO sem que ocorra alteração nos escores médios atribuídos às formulações. Isto mostra que a presença dos materiais de referência (“nenhum/fraco” e “forte”) no momento da avaliação das amostras fornece um suporte essencial para a alocação da intensidade dos estímulos na escala de avaliação. Silva et al. (2013a) também verificaram a importância dos materiais de referência para representarem as âncoras da escala não estruturada. Os autores observaram que no PC (equipe treinada e ausência das referências do momento das avaliações) a escala de avaliação não foi utilizada de forma homogênea, sendo que os julgadores tenderam a “ancorar” as amostras, utilizando com maior frequência os extremos da escala de intensidade. Em contrapartida, quando as amostras foram avaliadas por meio da técnica PDO (presença das âncoras), a escala não estruturada foi utilizada de forma mais homogênea.

A discriminação das formulações no PDO, segundo o DBIB e o DBCB, e no PC foram semelhantes, apresentando elevada sensibilidade na diferenciação dos estímulos ( $p < 0,0001$ ). A probabilidade de detectar corretamente diferença entre as amostras também foi elevada (superior a 0,95), possibilitando minimizar o erro tipo II a probabilidades inferiores a 5%. No PDO, a presença dos materiais de referência e o foco na comparação das amostras em relação a apenas um estímulo sensorial possibilitou que

o painel semi-treinado discriminasse as formulações de forma semelhante à equipe com elevado grau de treinamento, mesmo quando não foram apresentadas todas as amostras simultaneamente para a realização das comparações de intensidade.

Na análise da influência dos fatores quantitativos (suco e água) na intensidade dos atributos sensoriais por meio da regressão múltipla, obteve-se a mesma modelagem matemática ( $p > 0,05$ ) para as duas técnicas (PDO e PC), independente do delineamento administrado. Observa-se ainda que o PDO, segundo o delineamento em blocos completos, apresentou um coeficiente de determinação ainda superior à técnica convencional.

É importante ressaltar que no DBIB foram necessárias sessenta sessões para completar a avaliação das amostras e um maior número de repetições das avaliações para obtenção de todas as comparações pareadas. O tempo requerido para totalizar as avaliações pode ter contribuído para a “calibração” da memória sensorial da equipe, constituindo-se como um treinamento do painel. Em contrapartida, quando comparados os escores médios atribuídos às amostras nas diferentes repetições das avaliações, ou seja, ao longo do tempo de execução do experimento, houve semelhança entre as repetições (Figura 4.3). A diferença verificada entre as repetições no teste de médias de *Tukey* foi devida ao efeito de contraste entre estímulos (Lim, 2011; Silva et al., 2013b), o qual ocorreu quando uma determinada amostra era testada em conjunto com outra(s) formulação(ões) com maior(es) intensidade(s) do estímulo, sendo a percepção sensorial suprimida e atribuído um menor escore sensorial. Já quando uma mesma formulação era testada contra formulações com intensidade mais similares à amostra em avaliação, o estímulo era mais perceptível e um maior escore atribuído. Já no DBCB este efeito é minimizado uma vez que todos os estímulos são apresentados de forma simultânea.

Na Tabela 4.7 é mostrada uma análise comparativa entre as técnicas/delineamentos referente ao número de sessões e ao tempo requerido. Verifica-se que o PDO, quando conduzido sob o DBCB, possibilitou uma redução de 68% do número de sessões em relação ao PC, representando uma redução de 58% do tempo requerido. Por outro lado, quando o PDO foi conduzido sob DBIB, a redução do número de etapas foi de apenas 14%, porém em relação ao tempo gasto foi de aproximadamente 25%.

Tabela 4.7 - Comparação entre as técnicas/delineamentos em relação ao número de etapas e tempo requerido.

Critérios	Técnicas/Delineamentos		
	PDO/DBIB	PDO/DBC	PC/DBC
Número de sessões e tempo demandado em cada etapa:			
<i>Pré-seleção (8 min/sessão)</i>	4	4	4
<i>Levantamento dos atributos (30 min/sessão)</i>	2	2	2
<i>Apresentação das referências (5 min/sessão)</i>	1	1	-
<i>Treinamento (5 min/sessão)</i>	-	-	65
<i>Teste de desempenho (15 min/sessão)</i>	-	-	4
<i>Avaliação das amostras</i>	60	18	3
Número total de etapas	67	25	78
Redução do número de sessões em relação ao PC	14%	68%	-
Tempo estimado para completar uma sessão de avaliação das amostras	~ 5 minutos	~ 7 minutos	~ 18 minutos
Redução do tempo de execução em relação ao PC	25%	58%	-

O PDO conduzido no DBIB apresentou menor redução do tempo do teste sensorial quando comparado ao DBC, porém possibilitou a avaliação de um número maior de amostras sem alteração no perfil sensorial dos produtos, na discriminação das amostras e na relação funcional entre as variáveis independentes e dependentes. Desta forma, o número de amostras não é um fator limitante para o PDO, porém quando o delineamento em blocos incompletos é utilizado, a redução de tempo da técnica é menor. Assim, quanto maior for o número de amostras a serem avaliadas, maior será o número de blocos necessários para realizar as comparações entre os tratamentos e, portanto, maior será o tempo requerido para completar as avaliações. Por outro lado, o treinamento do painel para um número elevado de amostras também deve ser mais intenso e o número de sessões para avaliação de todas as amostras e repetições também

aumentará, uma vez que não é recomendada a avaliação superior a seis amostras por sessão, na técnica convencional (Stone & Sidel, 2012). Desta forma, na avaliação de um número mais elevado de amostras, o PDO conduzido segundo DBIB apresenta-se como uma alternativa ao Perfil Convencional, pois a técnica possibilitou a obtenção de resultados semelhantes aos dados provenientes do painel treinado com redução do tempo de execução da técnica.

## **6. Conclusão**

O delineamento em blocos incompletos balanceados apresentou-se como uma alternativa eficiente para a coleta de dados no PDO, sem haver perda de informação no perfil sensorial das amostras. A descrição sensorial por meio do PDO, independente do delineamento, assemelha-se à descrição obtida por meio da equipe treinada utilizando a técnica convencional.

O PDO, em ambos os delineamentos, apresentou elevado poder de discriminação das amostras e possibilitou a obtenção da relação funcional entre as variáveis independentes e dependentes de forma semelhante ao Perfil Convencional.

O DBIB apresentou-se como uma alternativa para avaliação de um número maior de amostras, evidenciando que o número de amostras não é um fator limitante para o PDO. A técnica conduzida sob blocos incompletos possibilitou a obtenção de resultados semelhantes aos dados provenientes do painel treinado com redução do tempo de execução da mesma.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

- Cochran, W., & Cox, G.M. (1981). *Diseños experimentales*. (7th ed.). México: Editorial Trillas.
- Damásio, M. H., & Costell, E. (1991). Análisis sensorial descriptivo: generación de descriptores y selección de catadores. *Revista Agroquímica de Tecnología de alimentos*, 31, 165-178.
- Dijksterhuis, G. (1996). *Procrustes analysis in sensory research*. In: *Multivariate Analysis of Data in Sensory Science*. Ed. by T. Nces and E. Risvik. Elsevier Science B.V.

- Gacula, M.C. Jr., Singh, J., Bi, J., & Altan, S. (2008). *Statistical Methods in Food and Consumer Research* (2nd ed.). Academic Press.
- Ishii, R., Chang, H., & O'Mahony, M. (2007). A comparison of serial monadic and attribute-by-attribute protocols for simple descriptive analysis with untrained judges. *Food Quality and Preference*, 18, 440–449.
- Ishii, R., Stampanoni, C., & O'Mahony, M. (2008). A comparison of serial monadic and attribute-by-attribute descriptive analysis protocols for trained judges. *Food Quality and Preference*, 19, 277-285.
- Johnson, R.A., & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. (6th edition). London: Pearson Education. Inc. p.773.
- Meilgaard, M.C, Civille, G.V., & Carr, B.T. (2006). *Sensory Evaluation Techniques*. (4th ed.). Boca Raton: CRC Press.
- Montgomery, D.C. (2001). *Design and analysis of experiments*. (5th Ed.). New York: John Wiley and Sons. p.699.
- O'Brien, R., & Lohr, V. (1984). *Power analysis for linear models: The time has come*. In Proceedings of the ninth annual SAS User's Group international conference (pp. 840–846).
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <<http://www.R-project.org>>.
- Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The 'RV' coefficient. *Applied Statistics*, 25, 257–265.
- Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by examining overlap between confidence intervals. *The American Statistician*, 55, 182–186.
- Silva, R.C.S.N., Minim, V.P.R., Simiqueli, A.A., Moraes, L.E.S., Gomide, A.I., & Minim, L.A. (2012). Optimized Descriptive Profile: a rapid methodology for sensory description. *Food Quality and Preference*, 24, 190-200.
- Silva, R.C.S.N., Minim, V.P.R., Carneiro, J.D.S., Nascimento, M., Della Lucia, S.M., & Minim, L.A. (2013a). Quantitative sensory description using the Optimized Descriptive Profile: comparison with conventional and alternative methods for evaluation of chocolate. *Food Quality and Preference*, 30, 169-179.
- Silva, A. N., Silva, R.C.S.N., Ferreira, M.A.M., Minim, V.P.R., Costa, T.M.T., & Perez, R. (2013b). Performance of hedonic scales in sensory acceptability of strawberry yogurt. *Food Quality and Preference*, 30, 9-21.
- Stone, H., & Sidel, J.L. (2012). *Sensory evaluation practices*. (4th ed.) New York: Academic.
- Van Belle, G. (2002). *Statistical Rules of Thumb*. New York, NY: John Wiley and Sons.



## *Validação da técnica Perfil Descritivo*

### *Otimizado: exatidão, precisão e robustez*

#### **RESUMO**

O Perfil Descritivo Otimizado é uma metodologia rápida de descrição sensorial, a qual foi proposta recentemente e tem apresentado elevado desempenho. A técnica propõe, além da redução do tempo do teste sensorial, a avaliação quantitativa dos atributos sensoriais, destacando-se entre as técnicas de descrição rápida. O presente estudo teve por objetivo validar esta metodologia, sendo avaliados os critérios exatidão, precisão (repetibilidade e reprodutibilidade) e robustez. A exatidão dos dados gerados pelo PDO foi mensurada por meio da comparação com o método referência, o Perfil Convencional, considerando diferentes matrizes alimentares. A precisão foi mensurada em nível de painel (repetibilidade) e inter-laboratorial (reprodutibilidade). Na robustez foi avaliada a sensibilidade do método quanto à redução do número de julgadores no painel. A metodologia PDO apresentou atendimento aos critérios estabelecidos, apresentando grau de proximidade com a metodologia referência superior a 95%. Na avaliação da repetibilidade de medições, considerando três repetições das avaliações por um mesmo painel, o PDO apresentou igualdade estatística entre as repetições e com grau de proximidade superior a 99%. O método apresentou reprodutibilidade dos dados, quando realizado em diferentes laboratórios, apresentando combinação perfeita entre os perfis sensoriais. O PDO mostrou-se robusto na redução do número de julgadores, sendo que painéis menores apresentaram variação aleatória ainda menor que o painel completo. O PDO foi validado com sucesso e apresentou as medidas de validação com elevado grau de certeza, considerando o conjunto de dados específico.

**Palavras-chave:** técnicas rápidas, PDO, exatidão, repetibilidade, reprodutibilidade, robustez, número de julgadores.

## 1. Introdução

Métodos descritivos rápidos para obtenção do perfil sensorial é um tópico que tem sido amplamente abordado na comunidade científica. Diversas metodologias foram propostas nos últimos anos com o objetivo de reduzir o tempo do teste sensorial, dentre elas: Perfil *Flash* (Dairou & Sieffermann, 2002; Delarue & Sieffermann, 2004), *Napping* (Pagès, 2005); *Free Sorting Task* (Cartier et al., 2006), *Check-all-that-apply* (Adams et al., 2007; Ares et al., 2010), *Pivot Profile* (Thuillier, 2007), *Ultra-Flash* (Perrin et al., 2008), *Polarised Sensory Position* (Teillet et al., 2010), *Ranking Descriptive Analysis* (Richter et al., 2010) e metodologias originadas por variações desses métodos. Essas novas metodologias utilizam o procedimento de classificação e de ordenação para avaliar os produtos em relação aos atributos descritores, obtendo o perfil sensorial dos produtos.

As técnicas supracitadas têm apresentado uma redução no tempo do teste de descrição, dispensando o treinamento da memória sensorial. Porém há limitações nas suas aplicabilidades devido à avaliação qualitativa das amostras. Por meio das técnicas alternativas apresentadas, é possível somente obter o mapa sensorial dos produtos, não sendo possível identificar a magnitude da diferença entre as amostras e, portanto, esses métodos não são recomendados para análises de estabilidade e controle de qualidade.

No sentido de reduzir o tempo do teste sensorial e adicionalmente fornecer medidas quantitativas dos atributos, foi proposto o Perfil Descritivo Otimizado (Silva et al., 2012). Esta técnica apresentou redução de aproximadamente 50% do tempo do teste descritivo quando comparada com o Perfil Convencional (PC), o qual utiliza equipe treinada. Nos estudos realizados para o PDO (Silva et al., 2012; 2013), foi verificada semelhança na obtenção do mapa sensorial por meio da técnica proposta e por meio do PC, além de elevado poder de discriminação das amostras e poder de decisão nos testes estatísticos ( $1 - \beta$ ) superior a 95%. Esta nova técnica sensorial tem mostrado resultados bastante convincentes, embora ainda tenha sido aplicada em poucos estudos.

Para que uma nova metodologia seja aceita na comunidade científica é necessário que esta demonstre capacidade de produzir resultados confiáveis para aplicação ou uso específico pretendido (ISO/IEC, 2005). Em métodos sensoriais descritivos não existe um protocolo específico para validação, sendo encontrados na literatura poucos trabalhos que avaliaram critérios de validação específicos para

algumas técnicas (Monrozier & Danzart, 2001; Pagès & Husson, 2001; Cartier et al., 2006; Blancher et al., 2007; Sinésio et al., 2010; Dehlholm et al., 2012). Na literatura mais recente, encontram-se disponíveis os estudos de validação de duas técnicas descritivas, *Ideal Profile* (Worch et al., 2013) e *Napping Profile* (Louw et al., 2013).

A metodologia *Ideal Profile* (Perfil Ideal) é uma técnica que utiliza os procedimentos descritivos e quantifica a intensidade das características sensoriais, porém esta utiliza um grupo representativo de consumidores. A avaliação dos produtos é realizada para mensurar a intensidade dos atributos sensoriais, não somente para obter o perfil sensorial dos produtos em questão, mas para identificar o grau de aceitabilidade associado às intensidades dos atributos, além de determinar a intensidade ideal de cada atributo no produto em questão. O objetivo desta metodologia não é diminuir o tempo do teste sensorial, mas sim otimizar formulações, identificando um “produto ideal”. Portanto, a validação da técnica consistiu em avaliar dois critérios: (i) consistência sensorial e (ii) consistência hedônica. Na consistência sensorial foi verificado se o perfil sensorial atribuído ao “produto ideal” aproximava-se da amostra que os consumidores mais gostaram entre as amostras avaliadas. A consistência hedônica foi avaliada após o processo de otimização da formulação, sendo verificado se o “produto ideal” era realmente o mais aceito pelos consumidores (Worch et al., 2013).

O *Napping Profile* é uma técnica rápida de descrição sensorial, a qual utiliza julgadores com determinado nível de acuidade sensorial. Nesta técnica, para a descrição dos produtos, os julgadores são orientados a alocar as amostras em uma folha tamanho A3 (60 x 40 cm) de acordo com suas similaridades e dissimilaridades. Os julgadores são instruídos que dois produtos deverão ser alocados próximos se forem identificados como iguais e alocados distantes se forem diferentes. Não há mais instruções de como proceder à alocação das amostras no papel, cada julgador utiliza seu próprio critério (Pagès, 2005). Na técnica *Global Napping* (GP) apenas uma sessão de avaliação é realizada, sendo as amostras alocadas na folha sem nenhuma restrição. Já na técnica *Partial Napping* (PN), são realizadas três avaliações distintas, sendo uma para a aparência dos produtos, outra para o sabor e uma última para a textura. No estudo de validação (Louw et al., 2013), as duas variações do *Napping* (GP e PN) foram avaliadas, sendo determinada a exatidão, a repetibilidade e a reprodutibilidade dos dados de avaliação. Os critérios de validação da técnica foram mensurados por meio da comparação dos perfis sensoriais obtidos em diferentes cenários experimentais,

utilizando o coeficiente RV. Uma avaliação mais robusta dos dados não é possível, devido à avaliação qualitativa dos produtos, não sendo mensurado o grau de similaridade ou dissimilaridade dos produtos.

A metodologia do presente estudo, o Perfil Descritivo Otimizado (PDO), apresenta um protocolo de avaliação das amostras otimizado, dispensando o treinamento da memória sensorial. Além do que as avaliações da intensidade dos atributos sensoriais são mensuradas em uma escala intervalar de intensidade e, portanto, é possível identificar o grau de diferença entre as amostras. A avaliação quantitativa do PDO possibilita um tratamento estatístico dos dados de forma mais robusta e com maior poder de decisão. Neste sentido, o objetivo da atual pesquisa é mensurar a exatidão, a precisão (repetibilidade e reprodutibilidade) e a robustez do Perfil Descritivo Otimizado em conjunto específico de dados, a fim de validar esta técnica como metodologia sensorial descritiva.

## 2. Materiais e Métodos

Foram mensuradas as medidas de validação: (i) exatidão, (ii) precisão (repetibilidade e reprodutibilidade) e (iii) robustez (ISO/IEC, 2005). Estas quais são definidas, de forma adaptada para métodos sensoriais, como:

- i. *Exatidão*: grau de concordância entre o resultado do teste e o valor de referência, ou seja, a habilidade do método em mensurar uma resposta próxima ao valor verdadeiro ou referência. Assim, para estimar a exatidão do PDO, a caracterização sensorial obtida por meio da metodologia em estudo foi comparada com a caracterização fornecida por uma metodologia de referência, o Perfil Convencional (Dairou & Sieffermann, 2002; Delarue & Sieffermann, 2004; Blancher et al., 2007; Sinésio et al., 2010 e Silva et al., 2012; 2013);
- ii. *Precisão*: consiste na repetibilidade e reprodutibilidade de resultados. Repetibilidade é a proximidade entre os resultados de medições sucessivas do mesmo alimento realizadas nas mesmas condições de medição. Reprodutibilidade é o grau de concordância entre os resultados das medições de um mesmo alimento efetuadas sob condições variadas de medições, fornecendo dados de precisão inter-laboratorial (Pagés & Husson, 2001);

- iii. *Robustez*: mede a sensibilidade do método em função de interferentes externos. Um método se diz robusto se revelar praticamente insensível a pequenas variações que possam ocorrer quando esse está sendo executado, como por exemplo, o número de julgadores (Boutrolle et al., 2005).

### 2.1. Estímulos sensoriais

Para avaliação dos três critérios de validação foram utilizados diferentes produtos alimentícios, sendo que cada grupo de alimentos foi utilizado para determinar critérios específicos (Tabela 5.1).

Tabela 5.1 - Tipos de produtos alimentícios utilizados para mensurar os critérios de validação.

<b>Produto</b>	<b>Número de amostras</b>	<b>Variação nas amostras</b>	<b>Critério de validação mensurado</b>
Requeijão	5	conteúdo de gordura e água	exatidão
Suco de uva	5	conteúdo de polpa e açúcar	exatidão e repetibilidade
Chocolate	4	conteúdo de chocolate ao leite e chocolate amargo	exatidão, repetibilidade, reprodutibilidade e robustez
Iogurte de morango	5	conteúdo de leite em pó, açúcar e corante	repetibilidade e robustez

### 2.2. Painéis sensoriais

Na avaliação sensorial descritiva dos produtos alimentícios (requeijão, suco de uva, chocolate e iogurte de morango), utilizando o PDO, foram formados sete painéis diferentes de julgadores semi-treinados. Cada painel trabalhou na avaliação sensorial de um produto específico, com exceção do chocolate, que teve sua avaliação realizada por quatro painéis diferentes, sendo cada painel proveniente de um laboratório situado em uma determinada cidade (Viçosa/MG, Sete Lagoas/MG, Lavras/MG e Alegre/ES). A avaliação de diferentes painéis foi realizada a fim de mensurar a reprodutibilidade do método quando as condições externas foram alteradas (precisão inter-laboratorial).

Os painéis foram compostos por 16 julgadores, conforme o número recomendado para a técnica do PDO. Porém, para a avaliação do critério “robustez”, foi considerado como fator perturbante o número de julgadores e, portanto, um painel

maior foi necessário. Assim, dois painéis de 26 julgadores foram formados, sendo um para a avaliação do chocolate e outro para a avaliação do iogurte (produtos alimentícios utilizados para o critério robustez). É importante ressaltar que os dados de descrição sensorial desses produtos (chocolate e iogurte) também foram utilizados para mensurar os critérios “exatidão” e “precisão” e, portanto, para essas análises foram selecionados aleatoriamente 16 julgadores do painel original contendo 26 julgadores, com o objetivo de padronizar os tamanhos dos painéis e os graus de liberdades de julgadores, possibilitando as comparações pareadas com o mesmo grau de sensibilidade.

Para a mensuração do critério exatidão também foi necessária a formação de painéis treinados para possibilitar a comparação do PDO com os dados provenientes da metodologia referência (Perfil Convencional). Para tanto, foram formados 3 painéis de julgadores treinados, sendo um painel para cada tipo de produto alimentício (requeijão, suco de uva e chocolate). Cada painel foi formado por 10 julgadores.

### *2.3. Procedimento da análise sensorial*

#### *2.3.1. Perfil Descritivo Otimizado*

A avaliação sensorial das amostras-teste (requeijão, suco de uva, chocolate e iogurte de morango) foi realizada por meio do protocolo de avaliação da técnica PDO (Silva et al., 2012; 2013). Esta técnica propõe as seguintes etapas: recrutamento dos voluntários por meio de questionário estruturado, pré-seleção dos julgadores em relação ao grau de acuidade sensorial (75% de acerto em testes triangulares), levantamento dos atributos, familiarização dos julgadores com os atributos e suas referências e, por fim, avaliação das amostras em uma escala de intensidade. É recomendada, neste método, a utilização da escala não estruturada de 9 cm ancorada nos extremos pelos termos “nenhum/fraco” e “forte”, os quais são representados pelos materiais de referência. Assim, foram formados diferentes painéis, sendo que cada painel avaliou a intensidade dos atributos descritores de um tipo de produto alimentício, os quais estão descritos na Tabela 5.2.

Na condução da metodologia, os julgadores selecionados para cada um dos 7 painéis participaram de uma sessão de familiarização com os termos descritivos e os materiais de referência dos atributos específicos dos produtos em avaliação. Assim, cada painel trabalhou especificamente com um tipo de alimento.

Tabela 5.2 - Atributos sensoriais avaliados nos diferentes produtos alimentícios.

<b>Requeijão</b>	<b>Suco de uva</b>	<b>Chocolate</b>	<b>Iogurte de morango</b>
Cor característica	Cor bordô	Cor marrom	Cor rosa
Consistência	Aroma de uva	Aroma de massa de cacau	Gosto doce
Espalhabilidade	Gosto doce	Gosto doce	Sabor de morango
Aroma característico	Gosto ácido	Gosto amargo residual	Sabor de creme
Sabor característico	Sabor de uva	Sabor de massa de cacau	Creiosidade
Viscosidade	Adstringente	Dureza	Viscosidade
Adesividade		Espalhabilidade	Resistência ao escoamento
			Textura Farinácea

Na etapa de familiarização do painel com os atributos sensoriais e seus respectivos materiais de referência, o objetivo foi de padronizar a forma de avaliação e expor claramente a qual estímulo sensorial estava se referindo cada termo, além de ancorar os extremos da escala não estruturada (“fraco” e “forte”). Nesta etapa, foram apresentados aos julgadores, em cabine individual, os termos descritivos e seus respectivos materiais de referência. O julgador foi orientado a ler a definição do atributo sensorial e provar as referências. Após esta etapa, as avaliações dos produtos-teste foram iniciadas.

As avaliações das amostras-teste foram conduzidas segundo o Delineamento em Blocos Completos Balanceados, de forma que os julgadores avaliaram todas as amostras numa mesma sessão. O protocolo atributo-por-atributo foi utilizado, sendo as amostras servidas de forma simultânea e apenas um atributo avaliado por sessão. As amostras-teste foram apresentadas em sequência aleatória e balanceada. No momento da avaliação, os materiais de referência “fraco” e “forte” dos atributos sensoriais estavam disponíveis para consulta, possibilitando os julgadores compararem as amostras com as referências dos extremos da escala de avaliação. Este protocolo de avaliação do PDO possibilita que os julgadores comparem as amostras entre si e com as referências antes da alocação das intensidades na escala não-estruturada, sendo que a re-degustação e a re-avaliação também são permitidas. Foram realizadas três repetições da avaliação por julgador. Portanto, o número total de avaliações ( $N_{TA}$ ) foi dado pela multiplicação do número de atributos ( $n_A$ ) pelo número de repetições da avaliação ( $n_R$ ), assim:  $N_{TA} = n_A \times n_R$ .

No procedimento de avaliação, o julgador era orientado a comparar as amostras entre si e com os materiais de referência e, então, alocar a intensidade do estímulo na escala de intensidade. A ficha de avaliação foi organizada por atributos e continha a escala não estruturada de 9 cm (intervalar) associada a cada amostra-teste.

### 2.3.2. Perfil Convencional

A descrição sensorial das amostras de requeijão, suco de uva e chocolate foram realizadas por meio uma equipe de julgadores treinados por meio do protocolo do Perfil Convencional (PC) descrito em Silva et al. (2012). O procedimento do PC seguiu as etapas: recrutamento dos voluntários por meio de questionário estruturado, pré-seleção dos julgadores em relação ao grau de acuidade sensorial (75% de acerto em testes triangulares), levantamento dos atributos, treinamento da memória sensorial, avaliação do desempenho da equipe e avaliação das amostras em uma escala de intensidade. A avaliação final das amostras no PC foi realizada segundo o Delineamento em Blocos Completos Balanceados e por meio do protocolo amostra-por-amostra. Assim, as amostras foram avaliadas em uma mesma sessão de avaliação de forma monádica, ou seja, era avaliada uma amostra por vez em relação a todos os atributos sensoriais. As amostras foram apresentadas de forma aleatória e balanceada, sendo realizadas três repetições das avaliações. Portanto, o número total de avaliações ( $N_{TA}$ ) foi dado pela multiplicação do número de amostras ( $n_A$ ) pelo número de repetições da avaliação ( $n_R$ ), assim:  $N_{TA} = n_A \times n_R$ .

As etapas iniciais do PC (recrutamento, pré-seleção e levantamento dos atributos sensoriais) são semelhante ao PDO e, portanto, as equipes do PDO e do PC realizaram as etapas iniciais em conjunto, de forma a levantarem os atributos sensoriais de forma consensual (Tabela 5.2). Após o levantamento dos atributos, as equipes seguiram as análises separadamente, cada painel segundo o protocolo específico de cada metodologia.

## 2.4. Mensuração dos critérios de validação

As análises estatísticas utilizadas na mensuração dos critérios de validação da metodologia Perfil Descritivo Otimizado foram realizadas utilizando os *softwares* SAS (licenciado para uso da Universidade Federal de Viçosa) e R (livre).

### 2.4.1. Exatidão

A exatidão mede o grau de concordância entre o resultado do teste e o valor de referência (ISO/IEC, 2005). Assim, para estimar a exatidão do PDO, a caracterização sensorial obtida por meio da metodologia em estudo foi comparada com a

caracterização fornecida por uma metodologia de referência (Perfil Convencional), considerando diferentes matrizes alimentares (requeijão, suco de uva e chocolate).

A comparação entre os perfis sensoriais obtidos pelo PDO e pelo PC foi realizada por meio da Análise de Procrustes Generaliza (APG). Esta técnica estatística multivariada possibilita a comparação de duas ou mais configurações espaciais e utiliza procedimentos de translação, escalonamento e rotação para aproximar as configurações individuais e obter uma configuração de consenso. Para tanto, foram obtidos os mapas sensoriais para cada uma das metodologias individualmente por meio da Análise de Componentes Principais (ACP), utilizando os escores médios de avaliação. As configurações individuais foram aproximadas pela APG e se obteve uma configuração consensual. O coeficiente RV foi utilizado para verificar o consenso entre os mapas individuais. Este coeficiente é uma generalização multivariada do coeficiente de correlação e, portanto, mede o grau de concordância entre as configurações e varia de 0 (discordância total) a 1 (concordância perfeita) (Robert & Escoufier, 1976).

#### 2.4.2. Precisão

A precisão de um método compreende a repetibilidade e a reprodutibilidade de resultados. A repetibilidade representa a proximidade entre os resultados de medições sucessivas do mesmo alimento (amostras-teste); já a reprodutibilidade indica a precisão inter-laboratorial, ou seja, proximidade entre resultados de medições de um mesmo alimento em condições externas variadas (ISO/IEC, 2005).

##### 2.4.2.1. Repetibilidade

O poder de repetibilidade de resultados indica a capacidade do painel em repetir as medidas (escores) nas diferentes repetições das avaliações de uma mesma amostra. A equipe de julgadores que apresenta boa repetibilidade consegue atribuir notas semelhantes nas sucessivas medições do mesmo alimento. Para avaliar o grau de repetibilidade dos painéis sensoriais, quando estes avaliam as amostras por meio da técnica do PDO, foi utilizada a análise de variância (ANOVA) e o coeficiente RV.

O grau de repetibilidade de resultados foi mensurado considerando a variação proveniente das repetições da avaliação de um mesmo painel de julgadores, conforme realizado por León et al. (1999). Esta avaliação foi realizada para cada atributo separadamente, considerando as 3 repetições da avaliação sensorial de uma mesma

amostra para as diferentes matrizes alimentares (suco de uva, chocolate e iogurte de morango). Foi avaliado se houve efeito significativo entre as avaliações sucessivas de um mesmo painel (efeito das repetições -  $e_{ij}$ ) por meio da ANOVA, considerando uma mesma matriz alimentar. Segundo Barbin (1993), o modelo matemático que representa a análise está demonstrado na Equação 5.1. Foi testada a hipótese de nulidade de variabilidade nula entre as repetições das avaliações ( $H_0: \sigma_e^2 = 0$ ).

$$Y_{ijk} = m + T_i + B_j + e_{ij} + \varepsilon_{ijk} \quad \text{Eq. 5.1}$$

Em que:

- $Y_{ijk}$  = representa o escore da amostra  $i$  atribuído pelo julgador  $j$  na repetição  $k$ ;
- $m$  = representa a constante inerente ao modelo ou a média geral;
- $T_i$  = representa o efeito fixo da amostra  $i$ ;
- $B_j$  = representa o efeito aleatório do julgador  $j$ ;
- $e_{ij}$  = efeito aleatório das repetições das avaliações para uma mesma amostra;
- $\varepsilon_{ijk}$  = erro aleatório normal, independente e igualmente distribuído  $(0, \sigma^2)$ .

A capacidade de repetibilidade das avaliações pelos painéis de julgadores utilizando o PDO também foi avaliada por meio da proximidade dos perfis sensoriais nas três repetições das avaliações. Para tanto, foi obtida uma configuração espacial das amostras por meio da Análise de Componentes Principais (ACP) para cada repetição, considerando todos os atributos e julgadores simultaneamente. As três configurações (repetições) de cada atributo foram comparadas por meio do coeficiente RV, sendo mensurado o grau de proximidade das configurações geradas pelas diferentes repetições das avaliações sensoriais pela equipe de julgadores.

#### 2.4.2.2. Reprodutibilidade

A reprodutibilidade indica a precisão inter-laboratorial da técnica sensorial, a qual foi mensurada por meio da comparação dos dados de descrição sensorial de amostras da mesma matriz alimentar (chocolates contendo diferentes proporções de chocolate ao leite e meio amargo) em diferentes laboratórios (UFV-Viçosa/MG, UFSJ-Sete Lagoas/MG, UFLA-Lavras/MG e UFES-Alegre/ES), utilizando a metodologia do PDO. Os resultados obtidos em cada laboratório foram comparados por meio de uma

análise de variância conjunta e por meio da comparação das configurações individuais (mapas sensoriais) por meio do coeficiente RV.

Na ANOVA, o modelo matemático que representa a análise está demonstrado pela Equação 5.2 (Barbin, 1993). Foi testada a hipótese de nulidade de variabilidade nula entre os laboratórios ( $H_0: \sigma_L^2 = 0$ ).

$$Y_{ijk} = m + T_i + L_j + (J/L)_{j/k} + e_{ijk} \quad \text{Eq. 5.2}$$

Em que:

- $Y_{ijk}$  = representa o escore da amostra  $i$  atribuído no laboratório  $j$  pelo julgador  $k$ ;
- $m$  = representa a constante inerente ao modelo ou a média geral;
- $T_i$  = representa o efeito fixo da amostra (tratamento)  $i$ ;
- $L_j$  = representa o efeito aleatório do laboratório  $j$ ;
- $(J/L)_{j/k}$  = representa o efeito aleatório do julgador  $k$  aninhado ao laboratório  $j$ ;
- $e_{ijk}$  = erro aleatório normal, independente e igualmente distribuído  $(0, \sigma^2)$ .

Para os atributos que apresentaram efeito significativo para laboratórios foi realizado um teste de Tukey ( $\alpha = 0,05$ ) para comparar as médias atribuídas às amostras pelos diferentes painéis (laboratórios), com o objetivo de verificar a proximidade quantitativa das avaliações.

As quatro configurações individuais (mapas sensoriais obtidos em cada laboratório) foram comparadas por meio do coeficiente RV, sendo mensurado o grau de proximidade das configurações geradas pelas avaliações dos chocolates por diferentes equipes de julgadores (Robert & Escoufier, 1976; Dijksterhuis, 1996).

#### 2.4.3. Robustez

O critério de validação “robustez” mede a sensibilidade do método em função de interferentes externos e este é considerado robusto se revelar praticamente insensível a pequenas variações que possam ocorrer quando está sendo executado (ISO/IEC, 2005). Neste estudo, foi considerado como fator “perturbante” no PDO o número de julgadores na equipe sensorial, conforme também considerado por Monrozier e Danzart (2001) e Boutrolle et al. (2005) em métodos descritivos convencionais.

Assim, o PDO foi aplicado na caracterização sensorial de duas matrizes alimentares (chocolate e iogurte de morango), utilizando painéis com 26 julgadores. Foi

avaliada a alteração ocorrida na variabilidade do experimento, dada pelo quadrado médio do erro experimental ( $\hat{\sigma}_e^2$ ), quando o número de julgadores nos painéis foi reduzido.

Para avaliar o efeito do tamanho da equipe de julgadores na variabilidade experimental obtida no PDO, foi realizada uma adaptação na metodologia utilizada por Leite et al. (2009). Esta metodologia considera a re-amostragem de sub-painéis com  $k$  julgadores da equipe total, sendo obtidas estimativas da variância experimental para cada sub-painel. Para tanto, foram simulados 10.000 sub-painéis para cada  $k$  número de julgadores, por meio da re-amostragem com reposição,  $k= 2, 4, 6, \dots, 22$  de 2 em 2 julgadores. Para os dados da equipe completa (26 julgadores) foi obtida a estimativa do erro experimental e seu respectivo intervalo de confiança. Assim, os sub-painéis com  $k$  julgadores que apresentarem no mínimo 90% das estimativas da variância experimental dentro do intervalo de confiança proporcionam estimativas do erro ( $QM_{ERRO}$ ) semelhantes, com 95% de confiança, à equipe completa e, desta forma, a metodologia sensorial apresenta-se robusta neste intervalo de número de julgadores na equipe.

As estimativas da variância do erro experimental ( $\hat{\sigma}_e^2$ ) e seu respectivo intervalo de confiança ( $IC_{95\%}$ ) foram obtidos conforme a metodologia descrita em Barbin (1993). A estimativa da variabilidade aleatória do experimento ( $\hat{\sigma}_e^2$ ) foi obtida por meio da estimativa do erro experimental (Quadrado Médio do Erro -  $QM_{ERRO}$ ) na ANOVA com duas fontes de variação (amostras e julgadores, além do efeito de interação entre as amostras e os julgadores). O intervalo de confiança foi obtido considerando a distribuição de qui-quadrado ( $\chi^2$ ), conforme recomendado para componentes de variância. A análise foi realizada considerando cada atributo sensorial individualmente, para as duas matrizes alimentares estudadas.

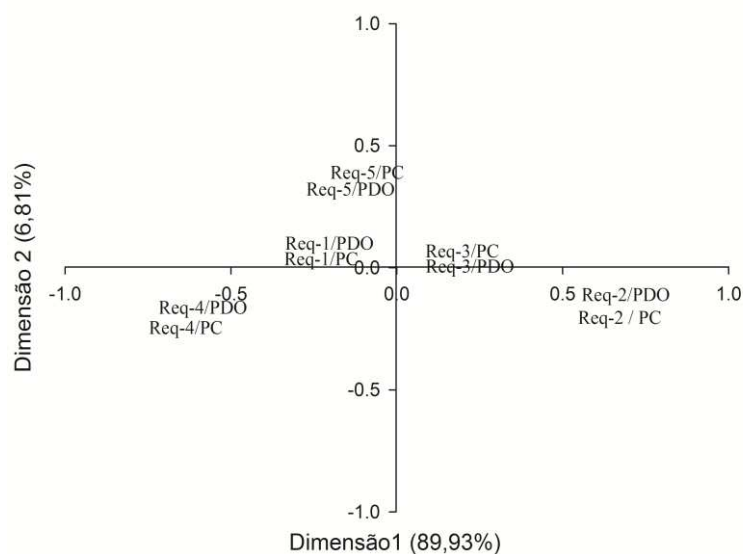
### **3. Resultados**

#### *3.1. Exatidão*

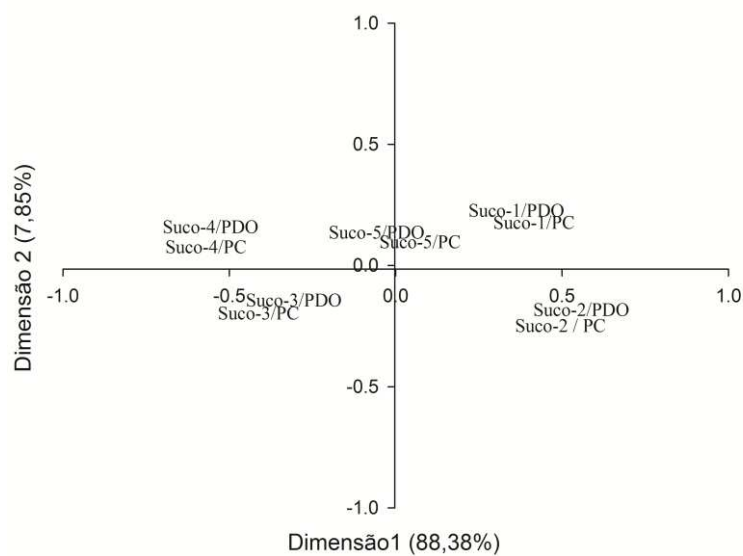
A exatidão dos dados obtidos no PDO foi mensurada por meio da comparação multivariada do mapa sensorial fornecido pela avaliação da equipe semi-treinada, utilizando o PDO, e o mapa referência fornecido pela avaliação da equipe treinada, por meio do protocolo de avaliação do Perfil Convencional. A comparação entre os perfis

sensoriais obtidos pelo PDO e pelo PC foi realizada por meio da Análise de Procrustes Generaliza (APG), obtendo uma configuração de consenso entre os painéis (PDO e PC). O grau de similaridade entre as configurações espaciais foi obtido por meio do coeficiente RV.

Na Figura 5.1 estão representados os mapas consensuais, obtidos na APG, para as matrizes requeijão, suco de uva e chocolate. Na representação gráfica é possível identificar a localização espacial das amostras em cada metodologia (PDO e PC).



(a)



(b)

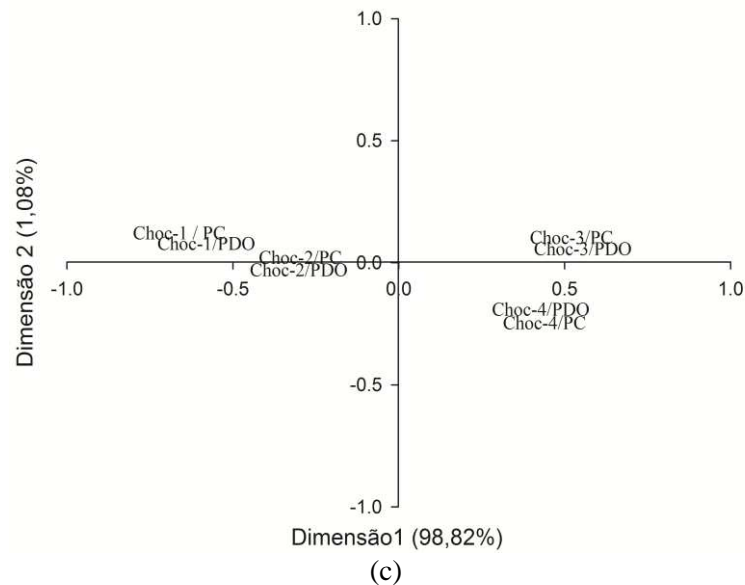


Figura 5.1 - Representação gráfica no mapa sensorial consensual entre as equipes do PDO e do PC, obtidas pela APG, para as matrizes requeijão (a), suco de uva (b) e chocolate (c).

Por meio de uma análise qualitativa das configurações espaciais, observa-se que as configurações individuais estão muito próximas entre si, sendo as amostras alocadas no espaço de forma muito similar nas duas metodologias (PDO e PC). Esta similaridade entre os perfis sensoriais é confirmada pelo coeficiente RV, o qual foi superior a 0,95 nas três comparações (cada matriz alimentar), indicando elevada similaridade entre as configurações. Os perfis dos produtos obtidos pela avaliação sensorial por meio do PDO apresentaram elevada correspondência com a técnica de referência e, portanto, alto nível de exatidão.

### 3.2. Precisão

#### 3.2.1. Repetibilidade

O grau de repetibilidade dos painéis de julgadores quando estes avaliaram as amostras por meio da técnica do PDO foi mensurado por meio da análise de variância (ANOVA) e do coeficiente RV.

Na ANOVA, foi avaliado se houve efeito significativo entre as três avaliações sucessivas de um mesmo julgador (efeito das repetições -  $e_{ij}$ ), considerando uma

mesma matriz alimentar. Na Tabela 5.3 estão apresentados resumos das análises de variâncias realizadas para cada produto e atributo sensorial separadamente.

Tabela 5.3 - Resumo da ANOVA para verificar a significância do efeito das repetições das avaliações ( $e_{ij}$ ).

<b>Produtos</b>	<b>Atributos</b>	<b>GL</b>	<b>F cal</b>	<b>p-valor</b>
Suco de uva	Cor bordô	2	0,46 ns	0,6344
	Aroma de uva	2	0,51 ns	0,6000
	Gosto doce	2	0,71 ns	0,4952
	Gosto ácido	2	1,17 ns	0,3128
	Sabor de uva	2	0,17 ns	0,8473
	Adstringente	2	0,29 ns	0,7522
Chocolate	Cor marrom	2	0,20 ns	0,8229
	Aroma de massa de cacau	2	0,06 ns	0,9444
	Sabor de massa de cacau	2	1,39 ns	0,2527
	Gosto doce	2	0,13 ns	0,8815
	Gosto amargo residual	2	0,49 ns	0,6149
	Dureza	2	0,87 ns	0,4187
	Espalhabilidade	2	0,26 ns	0,7692
Iogurte de morango	Cor rosa	2	1,65 ns	0,1936
	Gosto doce	2	0,37 ns	0,6908
	Sabor de morango	2	0,42 ns	0,6588
	Sabor de creme	2	0,19 ns	0,8267
	Creiosidade	2	1,39 ns	0,2589
	Viscosidade	2	0,98 ns	0,3767
	Resistência ao escoamento	2	3,35*	0,0368
	Textura Farinácea	2	1,48 ns	0,2308

\* significativo a 5% de probabilidade; ns: não significativo a 5% de probabilidade.

O efeito das repetições sucessivas de uma mesma amostra foi não significativo ( $p > 0,05$ ) para os atributos sensoriais. Somente para o atributo “resistência ao escoamento” da matriz iogurte de morango foi observada significância ( $p < 0,05$ ) para o efeito de repetição das avaliações sensoriais, utilizando o PDO.

Na APG, a comparação entre as configurações espaciais geradas nas três repetições das avaliações das amostras, apresentou elevada proximidade entre as medições sucessivas de uma mesma amostra. Na Tabela 5.4 estão apresentados os coeficientes RV obtidos nas comparações pareadas das três repetições das avaliações pelos painéis do PDO, considerando cada produto separadamente.

As comparações pareadas das configurações obtidas em cada repetição das avaliações do painel sensorial, por meio do PDO, apresentaram grau de proximidade superior a 0,99, considerando as três matrizes alimentares. Isto evidencia a elevada repetibilidade das medições fornecidas pelas equipes do PDO.

Tabela 5.4 - Coeficientes RV na comparação das configurações individuais das repetições das avaliações dos painéis de julgadores semi-treinados (PDO).

Repetições	Suco de uva			Chocolate			Iogurte de morango		
	R1	R2	R3	R1	R2	R3	R1	R2	R3
R1	1,0000	-	-	1,0000	-	-	1,0000	-	-
R2	0,9953	1,0000	-	0,9998	1,0000	-	0,9890	1,0000	-
R3	0,9947	0,9958	1,0000	0,9987	0,9988	1,0000	0,9960	0,9952	1,0000

O procedimento da ANOVA para avaliação do critério de validação “repetibilidade” mostrou-se mais rigoroso no estabelecimento da igualdade entre as medições nas diferentes repetições das avaliações. A avaliação quantitativa do PDO possibilitou a utilização de uma técnica mais robusta para análise dos dados provenientes de um painel com baixo grau de treinamento.

### 3.2.2. Reprodutibilidade

A reprodutibilidade (precisão inter-laboratorial) da técnica sensorial Perfil Descritivo Otimizado foi mensurada por meio da comparação dos dados de descrição sensorial de amostras de chocolates contendo diferentes proporções de chocolate ao leite e meio amargo em diferentes laboratórios (Viçosa/MG, Sete Lagoas/MG, Lavras/MG e Alegre/ES). Os resultados obtidos em cada laboratório foram comparados por meio de uma análise de variância conjunta e por meio da comparação das configurações individuais (mapas sensoriais) por meio do coeficiente RV.

Na análise de variância conjunta, Tabela 5.5, foi verificado efeito significativo de laboratórios ( $p \leq 0,05$ ) somente para dois atributos sensoriais (gosto amargo residual e sabor de massa de cacau). Para os demais atributos sensoriais não foi observada variabilidade ( $p > 0,05$ ) entre os locais de avaliação. As equipes sensoriais também apresentaram desempenho nas avaliações bastante similar, uma vez que o efeito de julgadores, aninhado ao efeito de laboratórios, foi significativo ( $p \leq 0,05$ ) somente para dois atributos sensoriais (gosto amargo residual e dureza).

Tabela 5.5 - Resumo da análise de variância conjunta para avaliação da reprodutibilidade de resultados do PDO (precisão inter-laboratorial).

Atributos	FV	GL	QM	F cal	p-valor
Cor marrom	Laboratórios	3	1,8367	2,48 ns	0,0626
	Amostras	3	673,2800	908,26*	<0,0001
	Julgador(laboratório)	60	0,9073	1,22 ns	0,1550
	Erro	189	0,7413		
Aroma de massa de cacau	Laboratórios	3	1,2048	1,06 ns	0,3660
	Amostras	3	583,7138	515,06*	<0,0001
	Julgador(laboratório)	60	1,1336	1,00 ns	0,4850
	Erro	189	1,1333		
Gosto doce	Laboratórios	3	0,3815	0,25 ns	0,8619
	Amostras	3	573,4156	374,42*	<0,0001
	Julgador(laboratório)	60	1,0959	0,72 ns	0,9344
	Erro	189	1,5315		
Gosto amargo residual	Laboratórios	3	10,4724	11,32*	<0,0001
	Amostras	3	588,3597	636,02*	<0,0001
	Julgador(laboratório)	60	2,1785	2,35*	<0,0001
	Erro	189	0,9251		
Sabor de massa de cacau	Laboratórios	3	4,1118	3,89*	0,0100
	Amostras	3	629,9355	595,55*	<0,0001
	Julgador(laboratório)	60	1,2662	1,20 ns	0,1825
	Erro	189	1,0577		
Dureza	Laboratórios	3	0,5722	0,61 ns	0,6102
	Amostras	3	664,5997	706,77*	<0,0001
	Julgador(laboratório)	60	1,3425	1,43*	0,0374
	Erro	189	0,9403		
Espalhabilidade	Laboratórios	3	0,8107	0,52 ns	0,6712
	Amostras	3	542,5378	345,84*	<0,0001
	Julgador(laboratório)	60	1,3581	0,87 ns	0,7394
	Erro	189	1,5687		

\* significativo a 5% de probabilidade; ns: não significativo a 5% de probabilidade.

Embora os laboratórios tenham apresentado efeito significativo na avaliação dos atributos gosto amargo residual e sabor de massa de cacau, as notas atribuídas às amostras pelas diferentes equipes foram bastante semelhantes (Tabela 5.6).

Tabela 5.6 - Médias dos atributos que apresentaram significância no efeito de laboratórios ( $\alpha = 0,05$ ).

Laboratórios	Atributos	
	Sabor de massa de cacau	Gosto amargo residual
Sete Lagoas/MG	4,5a	4,1a
Alegre/ES	4,4ab	3,7ab
Viçosa/MG	4,1ab	3,5bc
Lavras/MG	3,9b	3,2c

Pares de médias seguidas por letras semelhantes na coluna não diferem entre si ao nível de 5% de probabilidade, pelo teste de *Tukey*.

Na APG, os perfis sensoriais das amostras-teste obtidos em cada laboratório apresentaram elevado grau de proximidade das configurações espaciais, com RV superior a 0,99 (Tabela 5.7). Mapas similares indicam que os painéis estão avaliando as amostras de forma consensual, sendo atribuídos perfis sensoriais semelhantes a cada uma das amostras pelos diferentes painéis, provenientes de diferentes laboratórios.

Tabela 5.7 - Coeficientes RV na comparação das configurações individuais do perfil sensorial das amostras obtidas nos diferentes laboratórios, utilizando a técnica PDO.

<b>Laboratórios</b>	Alegre/ES	Viçosa/MG	Lavras/MG	Sete Lagoas/MG
Alegre/ES	1,0000	-	-	
Viçosa/MG	0,9979	1,0000	-	
Lavras/MG	0,9984	0,9962	1,0000	-
Sete Lagoas/MG	0,9944	0,9985	0,9949	1,0000

O coeficiente multivariado de correlação entre configurações espaciais, RV, indicou que os mapas sensoriais dos produtos apresentaram correspondência quase perfeita, com aproximadamente 100% de combinação. Por outro lado, na ANOVA foi verificado efeito significativo de laboratórios, mostrando maior rigor no estabelecimento da igualdade entre as medições nas diferentes condições laboratoriais.

### 3.3. Robustez

A robustez do PDO frente a interferentes externos foi mensurada por meio da alteração ocorrida na variabilidade do experimento, dada pelo quadrado médio do erro experimental ( $QM_{ERRO}$ ), quando o número de julgadores nos painéis foi reduzido.

Foi obtida a estimativa da variância experimental e seu respectivo intervalo de confiança ( $IC_{95\%}$ ) para os dados da equipe completa (26 julgadores), considerando duas matrizes alimentares (chocolate e iogurte). Para os sub-painéis amostrados da equipe completa, com 10.000 iterações para cada  $k$  número de julgadores, também foi obtida a estimativa da variância aleatória, sendo verificado o percentual de sub-painéis que apresentaram a estimativa da variância compreendida no intervalo de confiança do componente de variância aleatória ( $\sigma_e^2$ ). Nas Tabelas 5.8 e 5.9, chocolate e iogurte respectivamente, estão descritos os percentuais de sub-painéis que apresentaram a estimativa do componente de variância dentro do intervalo de confiança ( $IC_{95\%}$ ), abaixo do limite inferior (LI) e acima do limite superior (LS) do intervalo.

Tabela 5.8 - Percentual de sub-painéis que atenderam ao critério ao intervalo de confiança, considerando a matriz alimentar chocolate.

Produtos / Atributos	Tamanho do Painel ( <i>k</i> julgadores)												
	2	4	6	8	10	12	14	16	18	20	22	24	26
<i>Cor marrom</i>													
Acima do LS	16,3	16,4	14,9	8,6	2,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	24,9	33,5	46,5	59,8	71,6	78,5	82,8	86,1	<b>90,2</b>	93,1	96,6	100	100
Abaixo do LI	58,8	50,1	38,6	31,7	25,7	21,5	17,3	13,9	9,8	6,9	3,4	0,0	0,0
<i>Aroma de massa de cacau</i>													
Acima do LS	14,5	5,2	1,9	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	41,0	67,3	79,9	88,1	<b>92,6</b>	95,0	97,4	98,6	99,5	99,9	100	100	100
Abaixo do LI	446	27,5	18,3	11,6	7,4	5,0	2,3	1,4	0,5	0,1	0,0	0,0	0,0
<i>Gosto doce</i>													
Acima do LS	17,6	9,4	4,4	1,5	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	39,3	59,5	70,9	81,7	86,9	<b>90,8</b>	93,9	9,65	98,1	99,3	99,9	100	100
Abaixo do LI	43,1	31,1	24,6	16,8	12,9	9,2	6,1	3,5	1,9	0,7	0,1	0,0	0,0
<i>Gosto amargo residual</i>													
Acima do LS	11,0	4,7	1,4	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	53,8	72,9	84,5	<b>90,6</b>	94,7	96,9	98,5	99,5	99,9	100	100	100	100
Abaixo do LI	35,2	22,4	14,1	9,3	5,3	3,2	1,5	0,5	0,1	0,0	0,0	0,0	0,0
<i>Sabor de massa de cacau</i>													
Acima do LS	10,2	2,6	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	57,8	79,7	88,6	<b>93,8</b>	96,4	98,4	99,3	99,8	100	100	100	100	100
Abaixo do LI	32,1	17,6	11,1	6,2	3,6	1,6	0,7	0,2	0,0	0,0	0,0	0,0	0,0
<i>Dureza</i>													
Acima do LS	8,8	4,0	0,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	60,1	78,2	88,4	<b>93,5</b>	96,2	98,2	99,3	99,7	100	100	100	100	100
Abaixo do LI	31,1	17,8	11,2	6,5	3,8	1,8	0,7	0,3	0,0	0,0	0,0	0,0	0,0
<i>Espalhabilidade</i>													
Acima do LS	8,3	6,3	1,2	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	61,1	71,9	84,4	<b>90,8</b>	94,7	96,8	98,4	99,5	99,9	100	100	100	100
Abaixo do LI	30,7	21,8	14,3	9,1	5,3	3,2	1,6	0,5	0,1	0,0	0,0	0,0	0,0

IC<sub>(95%)</sub>: intervalo com 95% de confiança; LS: limite superior do intervalo de confiança; LI: limite inferior do intervalo de confiança. Valores em negrito indicam o menor tamanho de painel (*k* julgadores) que contém pelo menos 90% dos sub-painéis dentro do IC<sub>95%</sub>.

Tabela 5.9 - Percentual de sub-painéis que atenderam ao critério ao intervalo de confiança, considerando a matriz alimentar iogurte de morango.

Produtos / Atributos	Tamanho do Painel ( <i>k</i> julgadores)												
	2	4	6	8	10	12	14	16	18	20	22	24	26
<i>Cor rosa</i>													
Acima do LS	12,0	8,1	3,5	1,6	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	43,6	59,4	71,4	80,1	85,0	89,0	<b>91,6</b>	93,6	95,8	96,9	98,2	98,6	100
Abaixo do LI	44,4	32,5	25,1	18,3	14,6	11,0	8,4	6,4	4,2	3,2	1,8	1,4	0,0
<i>Gosto doce</i>													
Acima do LS	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	83,9	88,3	<b>91,2</b>	93,9	95,6	96,7	98,1	98,9	99,4	99,6	99,9	99,9	100
Abaixo do LI	16,1	11,8	8,8	6,2	4,4	3,3	2,0	1,1	0,6	0,4	0,1	0,01	0,0
<i>Sabor de morango</i>													
Acima do LS	13,1	6,1	2,5	0,9	0,2	0,0	0,0	0,0	0,0	0,0	0,00	0,0	0,0
Dentro do IC <sub>(95%)</sub>	62,9	83,1	<b>92,2</b>	96,8	98,7	99,5	99,8	100	100	100	100	100	100
Abaixo do LI	23,9	10,8	5,3	2,4	1,2	0,4	0,3	0,1	0,0	0,0	0,0	0,0	0,0
<i>Sabor de creme</i>													
Acima do LS	15,8	6,7	3,0	1,3	0,3	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	50,4	73,6	84,0	<b>91,0</b>	94,7	97,2	97,9	99,1	99,5	99,8	99,9	100	100
Abaixo do LI	33,8	19,7	13,0	7,8	5,0	2,7	2,1	0,9	0,5	0,3	0,1	0,0	0,0
<i>Cremosidade</i>													
Acima do LS	4,7	2,6	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	58,6	69,2	77,7	82,1	86,5	89,0	<b>92,2</b>	94,0	95,7	97,0	98,1	98,8	100
Abaixo do LI	36,8	28,3	22,1	17,9	13,5	11,0	7,8	6,0	4,3	3,0	1,9	1,2	0,0
<i>Viscosidade</i>													
Acima do LS	10,1	10,3	13,8	7,7	4,5	1,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	47,4	57,4	60,3	71,7	79,6	86,1	<b>90,5</b>	93,1	94,7	96,4	97,5	98,1	100
Abaixo do LI	42,5	32,4	25,9	20,6	16,0	12,6	9,4	7,0	5,3	3,6	2,5	1,9	0,0
<i>Resistência ao escoamento</i>													
Acima do LS	2,8	0,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	70,4	82,7	<b>90,0</b>	94,0	96,6	97,8	99,0	99,3	99,7	99,9	100	100	100
Abaixo do LI	26,7	16,8	10,0	6,0	3,4	2,2	1,0	0,7	0,3	0,1	0,1	0,0	0,0
<i>Textura farinácea</i>													
Acima do LS	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Dentro do IC <sub>(95%)</sub>	74,2	78,7	82,4	85,6	87,4	<b>90,2</b>	92,4	94,0	95,7	97,4	98,1	99,0	100
Abaixo do LI	25,8	21,3	17,6	14,5	12,6	9,8	7,7	6,0	4,3	2,6	1,9	1,0	0,0

IC<sub>(95%)</sub>: intervalo com 95% de confiança; LS: limite superior do intervalo de confiança; LI: limite inferior do intervalo de confiança. Valores em negrito indicam o menor tamanho de painel (*k* julgadores) que contém pelo menos 90% dos sub-painéis dentro do IC<sub>95%</sub>.

Os sub-painéis com  $k$  julgadores que apresentam no mínimo 90% das estimativas da variância experimental dentro do intervalo de confiança proporcionam estimativas do erro ( $QM_{ERRO}$ ) semelhantes à equipe completa. Verifica-se que para o chocolate (Tabela 5.8), o PDO apresentou-se robusto até uma redução de 8 julgadores na equipe, de forma que um painel menor de 18 julgadores já apresentou estimativa da variância aleatória significativamente diferente do painel completo (26 julgadores), para pelo menos um atributo sensorial (cor marrom). Para o iogurte, Tabela 5.9, os atributos que apresentaram menor robustez foram: cor rosa, cremosidade e viscosidade, apresentando diferença significativa na estimativa do erro aleatório a partir de 14 julgadores (remoção de 12 julgadores no painel).

#### 4. Discussão

Os critérios de validação exatidão, precisão e robustez do PDO foram mensurados, sendo atendidos aos requisitos estabelecidos. O desempenho do método frente aos critérios de validação foi semelhante quando consideradas diferentes matrizes alimentares e seus respectivos atributos sensoriais.

A exatidão do PDO frente ao método referência (PC) foi mensurada com elevado grau de similaridade entre os perfis sensoriais, apresentando coeficiente RV superior para a matriz chocolate (0,9887), seguida pelo requeijão (0,9837) e, por fim, o suco de uva (0,9555). Para todas as matrizes alimentares o grau de proximidade entre as configurações do PDO e do PC foram superiores a 95%, indicando elevada combinação entre os mapas sensoriais. Segundo Faye et al. (2004), um coeficiente RV superior a 0,70 é considerado um nível satisfatório de concordância entre os mapas sensoriais descritivos. Cartier et al. (2006) encontraram grau de proximidade de 0,67 entre as configurações geradas pelos painéis treinados e não treinados, utilizando a metodologia *Free Sorting Task (FST)*. Na comparação da mesma metodologia classificatória (FST) com o método referência (PC), Sinésio et al. (2010) verificaram um valor de 0,80 para o coeficiente RV. Perrin et al. (2008) verificaram grau de proximidade de 0,81 na comparação do mapa sensorial obtido pelo *Flash Profile (FP)* e pelo método convencional (PC). Lown et al. (2013), no estudo de validação da metodologia *Napping*, nas variações *Partial Napping (PN)* e *Global Napping (GN)*, verificaram um

valor do coeficiente RV de 0,88 na comparação do PN com o PC e 0,91 na comparação do GP com o PC. O PDO apresentou neste estudo valores de correlação entre as configurações espaciais superiores aos estudos encontrados na literatura, na comparação de métodos alternativos com a técnica referência (Perfil Convencional).

Na mensuração da precisão do PDO, em quesito de precisão do painel (repetibilidade) e precisão inter-laboratorial (reprodutibilidade), verificou-se nível de precisão bastante satisfatório. As técnicas estatísticas utilizadas na mensuração desses critérios apresentaram diferenças no estabelecimento da igualdade estatística entre as repetições das avaliações, seja em nível de painel ou inter-laboratorial, sendo que a ANOVA se apresentou mais criteriosa que o coeficiente RV.

Na APG, as configurações individuais são submetidas a três tipos de transformações (escalonamento, rotação e translação), na obtenção da configuração de consenso. Após a realização das transformações necessárias, é medida a distância entre as configurações individuais e a de consenso para o cálculo do coeficiente RV (Dijksterhuis, 1996). Neste processo, são avaliadas as configurações das amostras para verificar semelhança entre os mapas sensoriais, não sendo avaliada apenas a magnitude dos escores atribuídos. Portanto, apenas grandes desvios entre estas configurações são normalmente detectados por meio deste coeficiente.

O procedimento da ANOVA para avaliação do critério de validação precisão se mostrou mais rigoroso no estabelecimento da igualdade entre as medições nas diferentes repetições das avaliações e nos diferentes laboratórios. A avaliação quantitativa do PDO possibilitou a utilização de uma técnica mais robusta para análise dos dados provenientes de um painel com baixo grau de treinamento.

Para a repetibilidade foi verificado que apenas um dos vinte um atributos avaliados apresentou efeito significativo ( $p \leq 0,05$ ) das repetições das avaliações, considerando os dados de um mesmo julgador. Os demais atributos, nas diferentes matrizes alimentares, não apresentaram efeito das repetições, sendo o menor valor de significância igual 0,1936. Na comparação dos mapas sensoriais obtidos em cada uma das três repetições das avaliações do painel, foi verificado grau de similaridade (RV) superior a 0,99. Lown et al. (2013), no estudo de validação do *Napping*, verificaram coeficientes RV de 0,923 (PN) e 0,869 (GN), considerando três repetições das avaliações.

Para a reprodutibilidade (precisão inter-laboratorial), foi verificado efeito significativo entre os laboratórios ( $p \leq 0,05$ ) somente para dois atributos sensoriais dos oito atributos avaliados na matriz chocolate. Para os demais atributos sensoriais não foi observada variabilidade ( $p > 0,05$ ) entre os locais de avaliação. Embora os laboratórios tenham apresentado efeito significativo na avaliação dos dois atributos, as notas atribuídas às amostras pelas diferentes equipes foram bastante semelhantes. A escolha aleatória dos laboratórios na avaliação do desempenho do PDO permite a estimação do componente de variância para os laboratórios como um efeito aleatório e, portanto, a decisão pode ser extrapolada para a população. Desta forma, verificou-se que há efeito significativo de laboratórios na medição de alguns atributos sensoriais, porém este efeito é pequeno e, neste caso específico, não acarretou em grande diferença entre os escores médios de avaliação, sendo verificada diferença somente entre dois laboratórios.

Na APG, os perfis sensoriais das amostras-teste obtidos em cada laboratório apresentaram elevado grau de proximidade das configurações espaciais, com RV superior a 0,99. A semelhança nos mapas sensoriais indica que os painéis provenientes de diferentes locais estão avaliando as amostras de forma consensual, gerando perfis sensoriais semelhantes. Blancher et al. (2007) compararam o perfil sensorial de geléias obtido pela técnica *Flash Profile* (FP) e *Free Sorting Task* (FST) em dois países diferentes (França e Vietnã). Os autores verificaram grau de proximidade entre os mapas sensoriais (RV) de 0,73 para o FP e de 0,45 para o FST. Dehlholm et al. (2012) compararam as configurações espaciais de um mesmo grupo de amostras geradas por dois diferentes painéis independentes, considerando as técnicas Perfil Convencional (PC), Partial Napping (PN), Global Napping (GN) e Free Sorting Task (FST). Os autores verificaram semelhança, por meio do coeficiente RV, entre as configurações obtidas pelos dois painéis, de 94% para o PC, 82% para o PN, 71% para o GN e 48% para o FST. No presente estudo, o PDO apresentou semelhança entre os painéis originados de diferentes laboratórios superior ao valor encontrado por Dehlholm e colaboradores para a técnica referência (PC).

Na mensuração do critério de validação robustez para a técnica do PDO, verificou-se que o método apresentou-se robusto até uma redução de 8 julgadores na equipe para a matriz chocolate e de 12 julgadores para o iogurte, considerando o atendimento de todos os atributos ao critério estabelecido (90% dos valores dentro do  $IC_{95\%}$ ). Houve uma grande diferença na robustez do PDO em função dos atributos

sensoriais, sendo que para o atributo cor marrom do chocolate foi tolerada apenas uma redução de 8 julgadores no painel, enquanto para os atributos gosto amargo residual, sabor de massa de cacau, dureza e espalhabilidade foi tolerada uma remoção de 18 julgadores. Para o iogurte também ocorreu diferença na robustez para os diferentes atributos. Os atributos gosto doce, sabor de morango e resistência ao escoamento toleraram uma redução de 20 julgadores na equipe, enquanto os atributos cor rosa, cremosidade e viscosidade foram sensíveis a uma redução de 12 julgadores.

A variação existente entre as repetições das avaliações ( $QM_{ERRO}$ ) podem apresentar amplitudes diferentes conforme o produto e o atributo, além do que o indivíduo também pode apresentar diferença nesta variabilidade (King et al., 1995; Stone & Sidel, 2004). Na técnica de avaliação do PDO, os julgadores apresentam baixo nível de treinamento, o que pode acarretar em maior variabilidade interna e, portanto, maior é a estimativa da variância aleatória no experimento. Desta forma, como os julgadores podem apresentar níveis diferentes de variabilidade, a inclusão aleatória de um julgador com maior variabilidade residual pode acarretar em aumento da variância conjunta do grupo.

Por outro lado, é importante ressaltar que embora a estimativa da variância tenha apresentado diferença em relação à estimativa obtida no grupo completo, se situando fora do intervalo de confiança, esta esteve em maior percentual abaixo do limite inferior (LI) do intervalo do que acima do limite superior (LS). Sub-painéis com valores do  $QM_{ERRO}$  situados abaixo do LI indicam que estes painéis apresentaram menor variabilidade entre as repetições das avaliações e, portanto, apresentam maior consistência nos dados. Assim, estimativas do erro ( $QM_{ERRO}$ ) abaixo do LI são consideradas estatisticamente diferentes da estimativa obtida pelo painel completo e, portanto, o método é sensível a variação no número de julgadores na equipe. Por outro lado, painéis com números reduzidos de julgadores possibilitaram a obtenção de uma menor variação residual. Isto ressalta que um ou dois julgadores com maior variabilidade entre as repetições pode causar um inflação no  $QM_{ERRO}$  do painel como um todo. Se fosse considerado como critério somente o LS do intervalo, admitindo que a sensibilidade do método não varia para valores de  $QM_{ERRO}$  menores do que o LS obtido pelo painel completo, um painel de 6 julgadores satisfaria o critério para o chocolate e 4 julgadores seriam suficientes para o iogurte, considerando todos os atributos sensoriais.

## 5. Conclusão

A metodologia Perfil Descritivo Otimizado apresentou atendimento aos critérios de validação exatidão, precisão e robustez. A exatidão do método frente à metodologia referência foi mensurada com grau de semelhança acima de 95%, considerando diferentes matrizes alimentares. A precisão em nível de painel (repetibilidade) foi satisfatória, apresentando igualdade entre as repetições das avaliações individuais dos julgadores e do painel como um todo. Na medida de precisão inter-laboratorial (reprodutibilidade), os diferentes laboratórios apresentaram desempenho semelhante nas avaliações. Na mensuração da robustez da técnica, quando reduzido o número de julgadores no painel, foi observada sensibilidade na estimativa da variação aleatória a partir de 8 julgadores removidos do painel completo com 26 participantes. Por outro lado, embora o método tenha sido sensível a esta redução, a diferença verificada na variação interna dos painéis com menos de 14 julgadores foi menor do que a variação obtida na equipe completa, mostrando que os painéis menores foram diferentes do painel completo, porém se mostram melhores, apresentando menor variação aleatória.

As técnicas estatísticas utilizadas na mensuração dos critérios de validação apresentaram diferença no estabelecimento da igualdade estatística entre as repetições das avaliações, seja em nível de painel ou inter-laboratorial, sendo que a ANOVA apresentou-se mais criteriosa que o coeficiente RV obtido na Análise de Procrustes Generalizada.

A validação da metodologia PDO apresentou maior nível de proximidade com o método referência, Perfil Convencional, do que outras técnicas rápidas, em estudos prévios encontrados na literatura. Adicionalmente, o PDO apresenta uma avaliação quantitativa das características sensoriais e, portanto, a identificação da magnitude de diferença entre os produtos é possibilitada, além de uma avaliação estatística dos dados mais robusta e com maior poder de decisão.

## REFERÊNCIAS BIBLIOGRÁFICAS

Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of check-all-that-apply response compared to tradicional scaling of attributes for salty snacks. In: *7th Pangborn Sensory Science Symposium*. Minneapolis, USA, 12-16, August, 2007.

- Ares, G., Deliza, R., Barreiro, C., Giménez, A., & Gámbaro, A. (2010). Comparison of two sensory profiling techniques based on consumer perception. *Food Quality and Preference*, 21, 417-426.
- Barbin, D. (1993). *Componentes de Variância: teoria e aplicações*. 2ª ed. revista e ampliada. Piracicaba: FEALQ. 120p.
- Blancher, G., Chollet, S., Kesteloot, R., Nguyen Hoang, D., Cuvelier, G., & Sieffermann, J.-M. (2007). French and Vietnamese: How do they describe texture characteristics of the same food? A case study with jellies. *Food Quality and Preference*, 18, 560–575.
- Boutrolle, I., Arranz, D., Rogeaux, M., & Delarue, J. (2005). Comparing central location test and home use test results: Application of a new criterion. *Food Quality and Preference*, 16, 704 –713.
- Cartier, R., Ritz, A., Lecomte, A., Poblete, F., Krystlik, J., Belin, E., & Martin, N. (2006). Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference*, 17, 562–571.
- Dairou, V., & Sieffermann, J. M. (2002). A comparison of 14 jams characterized by conventional profile and a quick original method, the Flash Profile. *Journal of Food Science*, 67, 826–834.
- Dehlholm, C., Brockhoff, P.B., Meinert, L., Aaslyng, M.D., & Bredie, W.L.P. (2012). Rapid descriptive sensory methods – Comparison of Free Multiple Sorting, Partial Napping, Napping, Flash Profiling and conventional profiling. *Food Quality and Preference*, 26, 267-277.
- Delarue, J., & Sieffermann, J.M. (2004). Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference*, 15, 383–392.
- Dijksterhuis, G. (1996). *Procrustes analysis in sensory research*. In: Multivariate Analysis of Data in Sensory Science. Ed. by T. Nces and E. Risvik. Elsevier Science B.V.
- ISO/IEC 17025 (2005). Associação Brasileira de Normas Técnicas – ABNT. “*Requisitos Gerais para Competência de Laboratórios de Ensaio e Calibração*”, ABNT, Rio de Janeiro, setembro.
- King, B.M., Arents, P., & Moreau, N. (1995). Cost / Efficiency evaluation of descriptive analysis panels – I. Panel size. *Journal of Sensory Studies*, 6, 245-261.
- Leite, M. S. O., Peternelli, L. A., Barbosa, M. H. P., Cecon, P. R., & Cruz, C. D. (2009). Sample size for full-sib family evaluation in sugarcane. *Pesquisa Agropecuária Brasileira*, 44 (12), 1562-1574.
- Léon, F., Couronne, T., Marcuz, M.C., & Koster, E.P. (1999). Measuring food liking in children: a comparison of non verbal methods. *Food Quality and Preference*, 10, 93-100.
- Louw, L., Malherbe, S., Naes, T., Lambrechts, M., Van Rensburg, P., & Nieuwoudt, H. (2013). Validation of two Napping® techniques as rapid sensory screening tools for high alcohol products. *Food Quality and Preference*, 30, 192-201.

- Monrozier, R., & Danzart, M. (2001). A quality measurement for sensory profile analysis The contribution of extended cross-validation and resampling techniques. *Food Quality and Preference*, 12, 393–406.
- Pagès, J., & Husson, F. (2001). Inter-laboratory comparison of sensory profiles: methodology and results. *Food Quality and Preference*, 12, 297–309.
- Pagès, J. (2005). Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley. *Food Quality and Preference*, 16, 642–649.
- Perrin, L., Symoneaux, R., Maître, I., Asselin, C., Jourjon, F., & Pagès, J. (2008). Comparison of three sensory methods for use with the Napping procedure: Case of tem wines from Loire valley. *Food Quality and Preference*, 19, 1–11.
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <<http://www.R-project.org>>.
- Richter, V. B., Almeida, T. C. A., Prudencio, S. H., & Benassi, M.T. (2010). Proposing a ranking descriptive sensory method. *Food Quality and Preference*, 21, 611 - 620.
- Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The 'RV' coefficient. *Applied Statistics*, 25(3), 257–265.
- SAS. *Statistical Analysis System*. Institute Inc., North Carolina, USA, 1999, versão 9.1, licenciado para a Universidade Federal de Viçosa.
- Silva, R. C. S. N., Minim, V. P. R., Simiqueli, A. A., Moraes, L. E. S., Gomide, A. I., & Minim, L. A. (2012). Optimized Descriptive Profile: a rapid methodology for sensory description. *Food Quality and Preference*, 24, 190-200.
- Silva, R.C.S.N., Minim, V.P.R., Carneiro, J.D.S., Nascimento, M., Della Lucia, S.M., & Minim, L.A. (2013). Quantitative sensory description using the Optimized Descriptive Profile: comparison with conventional and alternative methods for evaluation of chocolate. *Food Quality and Preference*, 30, 169-179.
- Sinésio, F., Peparaió, M., Moneta, E., & Comendador F.J. (2010). Perceptive maps of dishes varying in glutamate content with professional and naive subjects. *Food Quality and Preference*, 21, 1034–1041.
- Stone, H., & Sidel, J.L. (2004). *Sensory evaluation practices*. (3th ed.) New York: Academic.
- Teillet, E., Schlich, P., Urbano, C., Cordelle, S. & Guichard, E. (2010). Sensory methodologies and the taste of water. *Food Quality and Preference*, 21, 967–976.
- Thuillier, B. (2007). *Rôle du CO<sub>2</sub> dans l'Appréciation Organoleptique des Champagnes – Expérimentation et Apports Méthodologiques*. Reims, France: Thèse de l'URCA.
- Worch, T., Crine, A., Gruel, A., & Lê, S. Analysis and validation of the Ideal Profile Method: Application to a skin cream study. *Food Quality and Preference* (2013), doi: <http://dx.doi.org/10.1016/j.foodqual.2013.08.005>

# Conclusão Geral

A metodologia Perfil Descritivo Otimizado possibilitou a obtenção do perfil sensorial dos alimentos de forma muito similar Perfil Convencional (PC) com menor tempo demandado na execução do teste sensorial, considerando diferentes matrizes alimentares. A técnica requer a formação de um painel de, no mínimo, 16 julgadores semi-treinados os quais irão avaliar as amostras em relação aos atributos sensoriais descritores, utilizando uma escala de intensidade intervalar. Os materiais de referência que ancoram os extremos da escala de intensidade devem estar presentes no momento da avaliação das amostras para possibilitar que julgadores com baixo nível de treinamento avaliem as amostras consistentemente.

No PDO, uma maior variação aleatória (entre as repetições das avaliações) foi observada frente ao método convencional. Esta variação foi o principal critério para a determinação do número de julgadores, tanto para os métodos descritivos convencionais quanto para o PDO, acarretando na necessidade de um painel maior. O número de julgadores também alterou sensivelmente a magnitude da variação aleatória, quando avaliada a robustez do método frente à redução do tamanho do painel. O PDO se mostrou robusto até uma redução de 8 julgadores na equipe, de forma que um painel menor de 18 julgadores já apresentou estimativa da variância aleatória significativamente diferente do painel completo (26 julgadores), para pelo menos um atributo sensorial. Na determinação do tamanho ideal do painel sensorial, considerando o critério da magnitude do erro experimental, foi observada coerência entre as recomendações.

A avaliação das amostras pelo painel de julgadores semi-treinados deve ser realizada segundo o protocolo atributo-por-atributo, sendo as amostras servidas simultaneamente e avaliadas em relação a um único atributo por vez. A apresentação de todas as amostras de uma só vez, segundo o Delineamento em Blocos Completos Balanceados (DBCB), limita o número de amostras no experimento. Um número elevado de amostras juntamente aos materiais de referência pode ocasionar fadiga nos julgadores. Para contornar esta situação, o Delineamento em Blocos Incompletos Balanceados (DBIB) pode ser utilizado para a coleta de dados no PDO, o qual

apresentou desempenho similar ao PDO conduzido no DBCB. No DBCB, todas as amostras são servidas simultaneamente aos julgadores, sendo possível a comparação entre todas as amostras e também com as referências. Já no DBIB, somente uma fração das amostras é testada por vez, sendo a comparação entre as amostras realizada mais eficazmente com os materiais de referência.

O PDO apresentou atendimento aos critérios de validação exatidão, precisão (repetibilidade e reprodutibilidade) e robustez. O método possibilitou a obtenção do perfil sensorial das diferentes matrizes alimentares de forma muito similar ao método referência, com grau de similaridade superior a 95%. Na avaliação da repetibilidade de medições, considerando três repetições das avaliações por um mesmo painel, o PDO apresentou igualdade estatística entre as repetições e com grau de proximidade superior a 99%. O método apresentou reprodutibilidade dos dados, quando as condições laboratoriais foram alteradas, apresentando combinação perfeita (RV igual a 100%) entre os perfis sensoriais. O PDO se mostrou robusto na redução do número de julgadores, sendo que painéis menores apresentaram variação aleatória ainda menor que o painel completo. O PDO foi validado com sucesso e apresentou as medidas de validação com elevado grau de certeza.

# Anexos

Nome: \_\_\_\_\_ Data: \_\_/\_\_/\_\_

## FICHA DE AVALIAÇÃO

Por favor, prove as amostras e marque com um traço vertical nas escalas abaixo, a posição que identifique melhor a intensidade da característica avaliada.

**Atributo: Gosto doce**

Código \_\_\_\_\_  
Fraco |-----| Forte

Código \_\_\_\_\_  
Fraco |-----| Forte

Código \_\_\_\_\_  
Fraco |-----| Forte

Código \_\_\_\_\_  
Fraco |-----| Forte

Anexo 1. Ficha de avaliação utilizada na metodologia PDO, um exemplo para o atributo gosto doce e quatro amostras.

