

**ESTUDOS COM MAPAS DE PREFERÊNCIA:
ASSOCIAÇÃO COM PROCRUSTES E
CONSTRUÇÃO COM VALORES FALTANTES**

LUCIANA DO NASCIMENTO LANCHOTE

2007

LUCIANA DO NASCIMENTO LANCHOTE

**Estudos com mapas de preferência: associação
com Procrustes e construção com valores faltantes**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Orientador: Prof. Dr. Ruben Delly Veiga

LAVRAS
MINAS GERAIS - BRASIL

2007

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Lanchote, Luciana do Nascimento

Estudos com mapas de preferência: associação com Procrustes e
construção com valores faltantes / Luciana do Nascimento Lanchote.

-- Lavras : UFLA, 2007.

75p. : il.

Dissertação (Mestrado) - Universidade Federal de Lavras, 2007.

Orientador: Ruben Delly Veiga.

Bibliografia.

1. Mapa de Preferência. 2. Procrustes. 3. Análise Sensorial.
4. Componentes Principais. 5. Valores Faltantes. I. Universidade
Federal de Lavras. II. Título.

CDD-519.535

LUCIANA DO NASCIMENTO LANCHOTE

**Estudos com mapas de preferência: associação
com Procrustes e construção com valores faltantes**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de “Mestre”.

APROVADA em 14 de dezembro de 2007.

Profa. Dra. Valéria Paula Rodrigues Minim

UFV

Prof. Dr. Marcelo Silva de Oliveira

UFLA

Prof. Dr. Ruben Delly Veiga
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

AO ERIC

Pela amizade, paciência e disposição em me ajudar sempre!

Este trabalho é tanto meu quanto seu! Ou seria tanto seu quanto meu?!?!

AGRADECIMENTOS

Inúmeras são as pessoas que estiveram presentes neste momento da minha vida... Além de inúmeras também serem responsáveis por cada pedrinha que montou o meu caminho. Falar de todas é impossível, mas acredito que a cada dia eu mostro a elas a importância que elas têm para mim... Nesta página quero contar sobre algumas, cuja a ordem não aumenta ou diminui o mérito de ninguém. . .

Mãe e Pai: pessoas que entregam cada minuto do seu dia para fazer da minha vida algo melhor sempre e a quem eu tenho certeza que retribuo com todo o meu amor!

Madrinha: amiga e companheira, sempre me traz um sorriso ou um consolo, uma oração ou um abraço; sempre está comigo e a quem eu agradeço por ser minha segunda mãe!

Delly: mais que um amigo, um incentivo nas minhas decisões, um apoio no momento de aperto, um superparceiro nos momentos de alegria! Alguém inesquecível...

D: um resumo de muitas coisas e muitos acontecimentos, alguém capaz de amar e ir em busca do que acredita, alguém especial e muito importante neste e em muitos momentos.

Profa. Valéria, Milene, Rita, UFV: que, mesmo sem me conhecerem direito, se empenharam muito para que o experimento desta dissertação acontecesse, principalmente confiando em mim.

D. Vilma, Sr. Ted, Sarah e Ronaldo: que me acolheram com carinho, me deram muita atenção e incentivo, além de casa, comida e muita oração.

Maria, Josi, Joice, Selminha, Edila, Vânia, Thelma, Eric, aos amigos e companheiros do Departamento de Ciências Exatas da UFLA: por todos esse anos de convivência, sorrisos e amizade, por cada Bom Dia! e Boa Tarde!, pelos conselhos e lealdade.

UFLA e CNPq: por todos os recursos fornecidos.

Aos amigos sinceros, aos baladeiros, aos companheiros, aos de longe e de perto, a amizade: família que escolhemos, merecedores de muitas páginas e ocupantes atemporais do meu coração.

As energias positivas, ao meu anjo da guarda e a Deus!

SUMÁRIO

Resumo	i
Abstract	ii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	4
2.1 Análise sensorial	4
2.2 Análise de componentes principais (ACP)	10
2.3 ACP com valores faltantes	19
2.4 Análise de variância	22
2.5 Mapas de preferência	24
2.5.1 Mapa de preferência interno	25
2.5.2 Mapa de preferência externo	27
2.6 Análise generalizada de Procrustes.	35
3 MATERIAL E MÉTODOS	38
3.1 Associação do mapa de preferência com a análise generalizada de Procrustes	38
3.1.1 Experimento 1: avaliação do ensino (à distância)	40
3.1.2 Experimento 2: palmito tratado.	41

3.2	Mapas de preferência com valores faltantes	41
3.2.1	Experimento 3: análise sensorial de doce de leite	42
3.2.2	Experimento 4: banana tratada	42
4	RESULTADOS E DISCUSSÃO	44
4.1	Associação do mapa de preferência com a análise generalizada de Procrustes	44
4.2	Mapas de preferência com valores faltantes	58
5	CONCLUSÕES	71
	Referências	72

RESUMO

LANCHOTE, Luciana do Nascimento. **Estudos com mapas de preferência:** associação com Procrustes e construção com valores faltantes. 2007. 75 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)- Universidade Federal de Lavras, Lavras, MG.*

A análise sensorial é uma ferramenta importante no desenvolvimento de novos produtos. Ao trabalhar com dados de aceitação (hedônicos) fornecidos por consumidores há um menor custo nas pesquisas, porém existem algumas dificuldades como, valores faltantes durante os experimentos por motivos diversos. Para a realização deste trabalho, foram estudadas formas de associação de mapas de preferência com a análise generalizada de Procrustes e a construção de mapas com valores faltantes, por meio do algoritmo NIPALS. Os resultados foram satisfatórios, sendo possível identificar metodologias para a aplicação em diferentes situações práticas.

*Comitê orientador: Ruben Delly Veiga - UFPA (orientador); Valéria Paula Rodrigues Minim (UFV) e Eric Batista Ferreira (Unifal-MG).

ABSTRACT

LANCHOTE, Luciana do Nascimento. **Studies on preference mapping: association with Procrustes and construction with missing values.** 2007. 75 p. Dissertation (Postgraduation Program in Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras, MG.*

Sensorial analysis is an important tool in the development of new products. Working with preference data (hedonics) given by consumers yields smaller costs in sensory researches, however there are some difficulties, as missing values in such experiments for many reasons. In this work ways of associating preference mapping and generalized Procrustes analysis were studied as well as the construction of maps with missing data, through NIPALS algorithm. Results were satisfactory, identifying methodologies to deal with different practical situations.

*Guidance committee: Ruben Delly Veiga - UFLA (supervisor); Valéria Paula Rodrigues Minim (UFV) and Eric Batista Ferreira (Unifal-MG).

1 INTRODUÇÃO

Análise sensorial é uma ciência que utiliza os sentidos humanos para avaliar as características de um produto. Ela tem se mostrado importante ferramenta no processo de identificação e atendimento aos anseios dos consumidores em relação à área alimentícia (Minim, 2006). No entanto, é também utilizada pela indústria de cosméticos (Fava, 2003), ambiental (Ferreira Filho & Alves, 2006), automóveis (Dairou et al., 2004; Galvão, 2007) e outros, tanto no desenvolvimento como na garantia e na otimização da qualidade de seus produtos.

Esse tipo de análise é feito, na maioria das vezes, por meio de questionários que visam obter várias informações de consumidores, diante da aceitação ou da rejeição de um determinado produto. Para isso, utilizam-se de questões que remetem a respostas como notas de 1 a 9, por exemplo ou graus de aceite, que vão desde “desgostei extremamente” a “gostei extremamente”, aplicados a um ou mais atributos. Estas se caracterizam por serem variáveis qualitativas que são discretizadas ao serem associadas com números inteiros.

Com essas características, as pressuposições da análise de variância podem não ser satisfeitas. Com isso, a Estatística dispõe de algumas outras ferramentas, como modelos lineares generalizados (MLG): que admitem modelos estatísticos com erros não normais e/ou heterocedásticos; mapas de preferência: que representam graficamente as preferências e com-

parações entre os produtos, identificadas pelos provadores, com a restrição de que todos os provadores tenham provado todas as amostras e análise generalizada de Procrustes (AGP): a qual realiza análises semelhantes ao mapa, que não exigem qualquer pressuposição.

A análise generalizada de Procrustes é utilizada para reduzir os efeitos das escalas e obter uma configuração consensual. É também aplicada para comparar a proximidade entre os termos que são usados por diferentes provadores ao descrevem os produtos (Ferreira, 2004).

A análise sensorial se depara, ainda, com algumas dificuldades, como: ausência de painel fixo de provadores, quando se refere a provadores treinados. Ao trabalhar com consumidores, há dificuldades para completar o painel (provar todas as amostras), seja por cansaço, fadiga sensorial ou qualquer outro tipo de indisponibilidade, o que acarreta em perda de amostras, retornando desperdícios, dificuldades de análises ou, dependendo do caso, até mesmo a perda do experimento.

Com base nessas afirmações, este trabalho foi realizado com dois propósitos: verificar qual a melhor maneira para realizar as análises estatísticas de dados provenientes de testes de aceitação e, ainda, apresentar uma metodologia capaz de construir os mapas com dados faltantes.

No primeiro propósito, são discutidas três situações diferentes para a geração de mapas de preferência, contrapondo a análise generalizada de Procrustes e o mapa de preferência, identificando-se assim, a que melhor se adapta à criação dos mapas gerados por meio de dados de consumidores. O segundo propósito traz uma proposta de construção de mapas com dados faltantes. Todas as análises estatísticas apresentadas foram realizadas no software R (R Development Core Team, 2007), o que é uma inovação, já que a grande maioria dos estudos realizados com mapa de preferência não o utiliza.

Assim, este trabalho apresenta alternativas interessantes, possibilitando análises em estudos onde ocorra valores faltantes e a apresentação de diferentes métodos de análises para dados de aceitação. Com isso, diminui-se problemas existentes e enfatiza-se a análise sensorial como uma ferramenta ainda mais poderosa para os estudos de mercado.

2 REFERENCIAL TEÓRICO

2.1 Análise sensorial

Segundo Minim (2006), a análise sensorial é utilizada para evocar, medir, analisar e interpretar reações às características de alimentos e outros materiais da forma como são percebidas pelos sentidos da visão, olfato, gosto, tato e audição.

A análise sensorial caracteriza-se por:

- identificar as características ou propriedades de interesse na qualidade sensorial (por exemplo: sabor, cor etc.);
- selecionar o método sensorial mais adequado para quantificar e, ou, qualificar a sensação experimentada pelo homem em resposta ao estímulo provocado pelo alimento, estabelecendo a equipe de julgadores, as condições ambientais do teste;
- selecionar e aplicar o método estatístico mais adequado para avaliar os resultados;

E entre as aplicações na indústria alimentícia e nas instituições de pesquisa, destaca-se a colaboração nas etapas de desenvolvimento de um novo produto, o controle de qualidade e o auxílio na seleção de métodos instrumentais que tenham correlação com atributos sensoriais de alimentos.

O objetivo final é estudar um produto de modo que este envolva o entendimento dos fatores que determinam as percepções do consumidor acerca desse, pois de nada vale um produto possuir excelentes características químicas, físicas ou microbiológicas, que seja considerado de excelente qualidade, se a característica sensorial não preencher os anseios de quem o consumirá.

Por meio da análise sensorial, as características ou propriedade de interesse relativas à qualidade sensorial do objeto de estudo são identificadas e adequadamente estudadas, com base em metodologias sensoriais de coleta de dados e em métodos estatísticos de avaliação e interpretação dos resultados do estudo sensorial desse objeto. É importante ressaltar que os testes sensoriais são dependentes de fatores ambientais e emocionais relacionados ao seu instrumento de medida, o homem, o que pode acarretar resultados tendenciosos para a análise. Deve-se então usar de cautela na preparação dos testes, a fim de evitar algumas das principais desvantagens (falta de acurácia e precisão dos dados, causadas pela avaliação do homem, condicionada aos fatores referidos).

A qualidade sensorial é função tanto dos estímulos procedentes dos alimentos como também das condições fisiológicas e sociológicas dos indivíduos que o avaliam, no contexto ambiental em que se localizam esse indivíduo e o próprio produto, ou seja, não é uma característica própria do alimento, mas o resultado da interação entre ele e o homem.

Com isso, a qualidade sensorial varia de pessoa para pessoa, sendo ditada por diversos fatores. São eles: aparência, sabor, textura, forma, método de preparo, custo e sazonalidade, diretamente ligados aos alimentos; e, idade, sexo, educação, renda, habilidades culinárias, entre outras, referentes aos indivíduos. Além de ser considerado ainda o ambiente urbano e até mesmo a estação do ano em que o alimento é consumido. Considerando portanto tais individualidades, ao avaliar a aceitação de um alimento, deve-

se realizá-lo junto à população de indivíduos a quem o produto de fato se destina, para que se possa obter resultados confiáveis em relação à aceitabilidade do alimento.

Há diversos métodos de análise sensorial e a escolha entre eles baseia-se em três questões fundamentais:

- O produto é aceito/preferido pelos consumidores?
- Existe diferença perceptível entre o produto em estudo e algum produto convencional?
- Quais os principais pontos de diferença e as suas intensidades?

As respostas a estas questões classifica os métodos em testes afetivos (para a primeira), testes discriminatórios ou de diferença (para a segunda) e análise descritiva (para a terceira).

Os métodos discriminatórios são, em geral, de fácil interpretação, requerem pouco tempo, são relativamente baratos e estabelecem a diferença qualitativa e, ou quantitativa entre as amostras.

A análise descritiva caracteriza e descreve as propriedades sensoriais de um produto, empregando linguagem técnica. Estes necessitam de julgadores treinados, implicando em longo tempo de análise e custo mais elevado.

Os métodos afetivos não requerem treinamento dos julgadores e são importantes por expressar a opinião por parte dos consumidores. Estes determinam qual o produto preferido e/ou mais aceito por determinado público-alvo, em função de suas características sensoriais.

Assim, os métodos afetivos são uma importante ferramenta, pois obtêm diretamente a opinião (preferência ou aceitação) do consumidor em relação a idéias, características específicas ou globais de determinado produto.

Estes não substituem a pesquisa de mercado, que identifica os consumidores para os quais os produtos serão direcionados e desenvolvem estratégias para alcançá-los, mas são uma alternativa à pesquisa de mercado, reduzindo, por exemplo, o número de amostras que seriam encaminhadas para a pesquisa mercadológica.

As aplicações dos testes afetivos são:

- Manutenção da qualidade de um determinado produto, avaliando-se a aceitação deste em função de alterações nas formulações;
- Otimização de produtos e, ou processos, buscando-se a melhoria da qualidade e a redução nos custos de produção e de distribuição de determinado produto;
- Desenvolvimento de novos produtos e processos, aplicando-se testes durante o processo de desenvolvimento de uma nova formulação;
- Acesso ao mercado potencial, trabalhando-se em conjunto com o setor de marketing;
- Avaliação e modificação de produtos, visando-se estudar o desempenho de determinado produto em relação aos das empresas concorrentes no mercado.

Os testes afetivos podem ser qualitativos ou quantitativos.

Os testes afetivos qualitativos visam obter respostas subjetivas de um pequeno subgrupo dos consumidores sobre as características sensoriais do produto, por meio da verbalização de suas opiniões sobre o produto, numa entrevista individual ou num pequeno grupo. Os métodos qualitativos são utilizados nas seguintes situações: conhecer o comportamento do consumidor em relação ao uso de um produto; desenvolver a terminologia usada pelos consumidores na descrição dos atributos sensoriais sobre

o conceito ou protótipo de um produto; avaliar as respostas iniciais dos consumidores sobre o conceito e/ou protótipo de um produto.

Os testes afetivos qualitativos podem ser classificados em: grupos de foco, equipes de foco e entrevistas individuais.

Os testes quantitativos são utilizados para avaliar as respostas dos consumidores em relação às suas preferências, gostos e opiniões e aos atributos sensoriais dos produtos. Estes, são aplicados nas seguintes situações: determinar a preferência global ou aceitação de um ou mais produtos por um grupo de consumidores que representem a população-alvo; determinar a preferência ou aceitação do produto em relação ao aroma, sabor, textura e aparência; medir as respostas dos consumidores em relação a atributos sensoriais específicos do produto.

Os testes quantitativos pode ser classificados basicamente em duas categorias:

- Testes de preferência: quando o objetivo é comparar vários produtos quanto à preferência. Embora meçam a preferência dos consumidores, não indicam se eles gostaram ou não dos produtos avaliados, exigindo que o pesquisador tenha conhecimento prévio sobre a avaliação afetiva desses produtos. Nas indústrias alimentícias são usados no desenvolvimento ou melhoria de produtos, alteração em processos ou formulação etc. São exemplos de testes de preferência: testes de comparação pareada, ordenação e comparação múltipla.
- Testes de aceitação: quando o objetivo do teste é avaliar o grau em que os consumidores gostam ou desgostam do produto. Para isso são utilizadas escalas de aceitação, balanceadas ou não. Sendo as mais utilizadas a escala hedônica, a escala de atitude (FACT) e a do ideal.

A Escala hedônica é facilmente compreendida pelos consumidores

e consiste em categorias ou afirmações, das quais as mais utilizadas são separadas em cinco, sete ou nove pontos, podendo ser: verbais (Figura 1), faciais (Figura 2) ou não-estruturadas (Figura 3).

Nome: _____

Sexo: M () F () Idade: _____ anos

Código da amostra: _____

Por favor, avalie a amostra servida e indique o quanto você gostou ou desgostou do produto. Marque a resposta que melhor reflita seu julgamento.

- () Gostei extremamente.
- () Gostei muito.
- () Gostei moderadamente.
- () Gostei ligeiramente.
- () Indiferente.
- () Desgostei ligeiramente.
- () Desgostei moderadamente.
- () Desgostei muito.
- () Desgostei extremamente.

Comentários: _____

Ficha utilizada para a aceitação de marcas comerciais de doce de leite.

FIGURA 1: Modelo de ficha de avaliação para o testes de aceitação utilizando a escala hedônica de nove pontos (verbal).

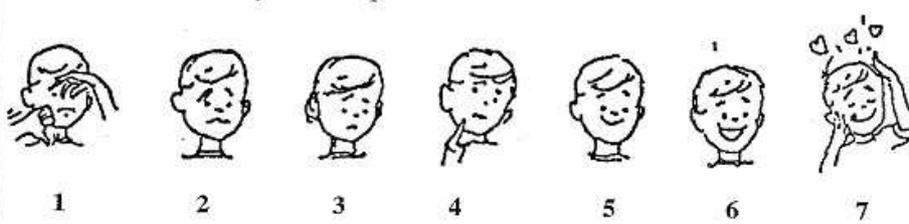
O teste de aceitação ocorre com o julgador recebendo as amostras codificadas, com números de três dígitos aleatórios, e é solicitado a avaliar os seus sentimentos com relação a cada amostra, utilizando uma escala hedônica. Pode-se avaliar a aceitação global (o produto como um todo), ou avaliar a aceitação de atributos do produto. Os julgadores são selecionados ao acaso, dentro de uma população de consumidores do produto.

Os resultados das análises podem ser avaliados: pela análise de distribuição de frequência dos valores hedônicos, por meio de histogramas; pela análise de variância, que considera conjuntamente a avaliação de todos os

Ficha de aplicação do teste de escala Hedônica Estruturada Facial.

Nome:.....Data:.....
Idade:.....

Agora que você acabou de comer a sopa de macarrão com feijão, circule a carinha que expressa o quanto você gostou ou desgostou desse prato.



1 2 3 4 5 6 7

Indique o que você mais gostou na sopa:
Indique o que você menos gostou na sopa:

FIGURA 2: Modelo de ficha de avaliação para o testes de aceitação utilizando a escala hedônica de sete pontos (facial).

consumidores e assume que todos apresentam o mesmo comportamento, desconsiderando suas individualidades; pelo Mapa de Preferência, considerando as individualidades dos consumidores.

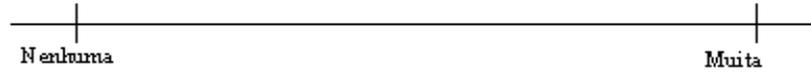
2.2 Análise de componentes principais (ACP)

A análise de componentes principais foi, segundo Ferreira (2007), introduzida por Pearson, em 1901, e é uma técnica multivariada de modelagem da estrutura de covariância.

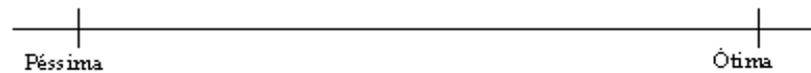
A ACP é associada à idéia de redução da massa de dados e tem como principal objetivo transformar um conjunto original de variáveis em outro conjunto, denominado componentes principais, com a menor perda de informação possível. O objetivo desta técnica é a construção de uma

2- Do Texto acadêmico:

a) Clareza:



b) Redação:

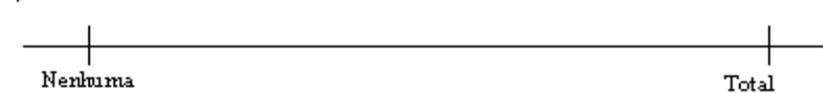


3- Conteúdo:

a) Importância:



b) Utilidade Prática:



e) Participação da turma nessa disciplina

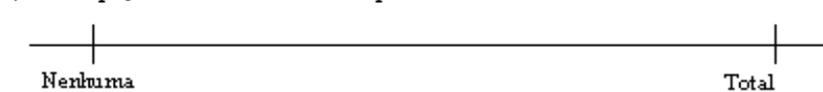


FIGURA 3: Modelo de ficha de avaliação para o testes de aceitação utilizando a escala hedônica não-estruturada de 12 cm entre as âncoras.

combinação linear de medidas que contribuam mais acentuadamente para a variação entre as amostras estudadas (Minim, 2006).

Ferreira (1996) define a ACP como a explicação da estrutura de covariâncias por meio de poucas combinações lineares das variáveis originais e como objetivos a redução da dimensão original e a facilitação da interpretação das análises realizadas. Esse autor enfatiza que a explicação de toda a variabilidade do sistema determinado por p variáveis só pode ser efetuada por p componentes principais. No entanto, grande parte dessa variabilidade

pode ser explicada por um número k menor de componentes $k \leq p$. Esse tipo de análise se presta, fundamentalmente, como um passo intermediário a investigações científicas.

Ferreira (2007) afirma que, para a obtenção dos componentes principais, de forma geral, deve-se definir, inicialmente, um conjunto de p variáveis: (X_1, X_2, \dots, X_p) , com médias $(\mu_1, \mu_2, \dots, \mu_p)$ e variâncias $(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$, respectivamente. Tais variáveis não são independentes e, portanto, possuem covariância entre a i -ésima e k -ésima variável definida por σ_{ik} , para $i \neq k = 1, 2, \dots, p$. Também é possível representar estas p variáveis na forma vetorial por $\mathbf{X} = [X_1, X_2, \dots, X_p]^T$, com vetor de médias $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]^T$ e a matriz de covariâncias $\boldsymbol{\Sigma}$, de dimensão $(p \times p)$. A princípio, nenhuma distribuição é assumida para o vetor \mathbf{X} . Dessa forma, o i -ésimo componente principal Y_i é definido por:

$$Y_i = \mathbf{a}_i^T \mathbf{X} = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p,$$

em que o vetor desconhecido \mathbf{a}_i estabelece a i -ésima combinação linear, para $i = (1, 2, \dots, p)$.

A variável Y_i , componente principal, é uma variável latente, ou seja, não é mensurada a partir do experimento ou do levantamento amostral. O objetivo é determiná-la a partir das p variáveis contidas no vetor \mathbf{X} . Para isso, devem-se projetar os pontos coordenados originais em um plano maximizando a distância entre eles, o que equivale a maximizar a variabilidade da variável latente Y_i . A variância de Y_i é dada por:

$$\begin{aligned} \text{Var}(Y_i) &= \text{Var}(\mathbf{a}_i^T \mathbf{X}) = \mathbf{a}_i^T \text{Var}(\mathbf{X}) \mathbf{a}_i \\ &= \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i \end{aligned}$$

e a covariância entre Y_i e Y_k , ($i \neq k$), por:

$$\begin{aligned} Cov(Y_i, Y_k) &= Cov(\mathbf{a}_i^T \mathbf{X}, \mathbf{a}_k^T \mathbf{X}) = \mathbf{a}_i^T Var(\mathbf{X}) \mathbf{a}_k \\ &= \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_k. \end{aligned}$$

A definição dos componentes principais é baseada na maximização de sua variância. No entanto, o máximo da variância do componente principal não existe, pois, na medida que os componentes do vetor de coeficientes \mathbf{a}_i crescem, a variância crescerá para infinito. Assim, impõe-se como restrição que a soma ao quadrado dos coeficientes do vetor, que estabelece a combinação linear dos componentes principais, seja igual a unidade. Assim, deve-se maximizar a variância $Var(Y_i) = \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i$ com relação ao vetor \mathbf{a}_i , sujeito à restrição de $\mathbf{a}_i^T \mathbf{a}_i = 1$. Para isso, é possível utilizar a técnica de multiplicadores de Lagrange, na qual deve-se maximizar

$$\max_{\mathbf{a}_i} [\mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i - \lambda_i (\mathbf{a}_i^T \mathbf{a}_i - 1)],$$

com relação a \mathbf{a}_i , sendo λ_i o multiplicador de Lagrange.

Uma abordagem mais tradicional impõe a restrição de outra forma, dividindo a função original por $\mathbf{a}_i^T \mathbf{a}_i$, resultando em uma função que deve ser maximizada da seguinte forma:

$$\lambda_i = \max_{\mathbf{a}_i} \frac{\mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i}{\mathbf{a}_i^T \mathbf{a}_i}.$$

Em ambos os casos, para obter-se o máximo, deve-se derivar a função resultante em relação a \mathbf{a}_i e igualar o sistema de equações a zero que, após simplificado, resultará em:

$$(\boldsymbol{\Sigma} - \lambda_i \mathbf{I}) \mathbf{a}_i = 0, \tag{2.1}$$

em que \mathbf{I} é uma matriz identidade de dimensão $(p \times p)$. A demonstração da maximização desta forma quadrática é apresentada pelo teorema da *Maxi-*

mização de formas quadráticas que consiste em que “Se \mathbf{A} é uma matriz simétrica de dimensão $(p \times p)$, então o máximo de $Q(\mathbf{X}) = \mathbf{X}^T \mathbf{A} \mathbf{X}$, restrito a $\mathbf{X}^T \mathbf{X} = 1$ é dado pelo maior autovalor λ_i de \mathbf{A} ($\lambda_i = 1, 2, \dots, p$) e pelo autovetor correspondente X_i , que são soluções de $(\mathbf{A} - \lambda_i \mathbf{I}) \mathbf{X}_i = 0$ ”.

Do sistema de equações (2.1) pode-se concluir que a definição dos componentes principais é, na verdade, um problema de obtenção dos autovalores e autovetores da matriz Σ . Os autovetores \mathbf{a}_i ($i = 1, 2, \dots, p$) representam o sentido da rotação dos eixos coordenados definidos pelas variáveis originais e os autovalores, λ_i ($i = 1, 2, \dots, p$) definem as variâncias desses novos eixos coordenados. Como as covariâncias são nulas entre os diferentes pares de eixos, os novos eixos, que são os componentes principais, são ortogonais e, portanto, representam uma rotação rígida dos eixos originais. De modo geral, opta-se pelo componente principal de maior importância como sendo aquele de maior variância; o segundo mais importante é o que apresentar a segunda maior variância e assim sucessivamente, até o componente principal de menor importância. Ao ordenar $\lambda_1 \geq \lambda_2 \geq \lambda_3 \dots \geq \lambda_p$ é possível definir os componentes principais $Y_1 = \mathbf{a}_1^T \mathbf{X}$, $Y_2 = \mathbf{a}_2^T \mathbf{X}$, \dots , $Y_p = \mathbf{a}_p^T \mathbf{X}$, respectivamente. Assim, é possível obter tantos componentes principais quanto variáveis originais.

Com a decomposição espectral da matriz Σ , dada por $\Sigma = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T$, em que \mathbf{P} é a matriz composta pelos autovetores de Σ em suas colunas e $\mathbf{\Lambda}$, a matriz diagonal de autovalores de Σ , então, pode-se observar que:

$$\begin{aligned} tr(\Sigma) &= tr(\mathbf{P} \mathbf{\Lambda} \mathbf{P}^T) = tr(\mathbf{\Lambda} \mathbf{P}^T \mathbf{P}) = tr(\mathbf{\Lambda} \mathbf{I}) \\ &= tr(\mathbf{\Lambda}) = \sum_{i=1}^p \lambda_i. \end{aligned}$$

Dado que $tr(\Sigma)$ é a soma dos elementos da diagonal, conclui-se que:

$$tr(\Sigma) = \sum_{i=1}^p \sigma_{ii} = \sum_{i=1}^p \lambda_i, \quad (2.2)$$

a variabilidade total contida nas variáveis originais é igual à variabilidade total contida nos componentes principais.

Para gerar um modelo mais parcimonioso para a matriz de covariância, deve-se reter $k < p$ componentes principais e basear-se em algum critério que informe quanto da variabilidade total foi explicada por ele. Para isso, considere inicialmente o vetor de componentes principais $\mathbf{Y} = [Y_1, Y_2, \dots, Y_p]$, que pode ser obtido de forma vetorial por $\mathbf{Y} = \mathbf{P}^T \mathbf{X}$. Como a matriz \mathbf{P} é ortonormal, sua inversa é igual a sua transposta, portanto o vetor \mathbf{X} pode ser recuperado pela transformação não-singular $\mathbf{X} = \mathbf{P} \mathbf{Y}$. Assim, ao reduzir o número de variáveis transformadas, componentes principais, para $k < p$, tem-se o vetor $\mathbf{Y} = [Y_1, Y_2, \dots, Y_p]$, de dimensão $(k \times 1)$ e assume-se que os k primeiros autovetores da matriz \mathbf{P} são utilizados para compor a matriz \mathbf{P}_k de dimensão $(p \times k)$, obtendo-se $\mathbf{Y} = \mathbf{P}_k^T \mathbf{X}$.

A matriz inversa generalizada de $\mathbf{P}_k^T \mathbf{X}$ pode ser obtida por meio da decomposição do valor singular e remete a própria matriz \mathbf{P}_k , devido à ortonormalidade por coluna de \mathbf{P}_k . Assim, é possível recuperar as observações originais com determinado grau de precisão, dependendo de quão adequado for o modelo parcimonioso que se tenha ajustado à matriz de covariâncias. Com isso, as observações das variáveis originais podem ser preditas pelo modelo escolhido por $\tilde{\mathbf{X}} = \mathbf{P}_k \mathbf{Y}$.

Ainda é possível observar que a covariância do vetor \mathbf{Y} , de dimensão $(p \times 1)$, de componentes principais é dada por:

$$\begin{aligned} Cov(\mathbf{Y}) &= Cov(\mathbf{P}^T \mathbf{X}) = \mathbf{P}^T Cov(\mathbf{X}) \mathbf{P} \\ &= \mathbf{P}^T \boldsymbol{\Sigma} \mathbf{P} = \mathbf{P}^T \mathbf{P} \boldsymbol{\Lambda} \mathbf{P}^T \mathbf{P} \\ &= \boldsymbol{\Lambda}. \end{aligned}$$

Assim, a variância total é igual ao traço da matriz de covariâncias e,

portanto, a variância total contida nas variáveis originais é igual ao $tr(\mathbf{\Sigma}) = \sum_{i=1}^p \sigma_{ii}$. Da mesma forma, a variância total de \mathbf{Y} é igual ao $tr(\mathbf{\Lambda}) = \sum_{i=1}^p \lambda_i$.

Verifica-se, da relação estabelecida na equação (2.2), que a variabilidade total contida nas variáveis originais é preservada nas variáveis transformadas, componentes principais. Ao adotar o modelo parcimonioso, com $k < p$ componentes principais, tem-se que a covariância do vetor \mathbf{Y} de dimensão $(k \times 1)$ é:

$$\begin{aligned} Cov(\mathbf{Y}) &= Cov(\mathbf{P}_k^T \mathbf{X}) = \mathbf{P}_k^T Cov(\mathbf{X}) \mathbf{P}_k \\ &= \mathbf{P}_k^T \mathbf{\Sigma} \mathbf{P}_k = \mathbf{P}_k^T \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T \mathbf{P}_k \\ &= \mathbf{\Lambda}_k, \end{aligned}$$

em que

$$\mathbf{\Lambda}_k = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{pmatrix}.$$

Portanto, a variância total do vetor \mathbf{Y} , de dimensão $(k \times 1)$, é $tr(\mathbf{\Lambda}_k) = \sum_{i=1}^k \lambda_i$. É possível, então, estabelecer que a explicação do modelo reduzido em relação ao modelo completo pode ser obtida pela relação das variâncias dos dois modelos, ou seja, quanto da variação total das variáveis originais é explicada pelo modelo de k componentes principais. Essa proporção acumulada da explicação da variação total, expressa em porcentagem, é formalmente dada por:

$$R_k^2 = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \sigma_{ii}} \times 100. \quad (2.3)$$

A explicação individual de cada componente também pode ser considerada e é obtida por meio da proporção de explicação do k -ésimo com-

ponente principal definido por:

$$P_k^2 = \frac{\lambda_k}{\sum_{i=1}^p \sigma_{ii}} \times 100. \quad (2.4)$$

Com base na proporção de explicação da variância total (2.3), que o modelo de k componentes principais é responsável, é possível determinar o número de componentes que se deve reter. Em muitos casos, as evidências empíricas e científicas conduzem a utilizar o critério de reter um número $k < p$ de componentes principais que contemple pelo menos 70% da variação total. Algumas vezes, opta-se por um maior rigor e, assim, deve-se optar por, no mínimo, 80% da variação total. Se o número mínimo de componentes k que explica esta proporção mínima da variação total for muito grande, em muitas aplicações práticas não haverá vantagens em utilizar o método dos componentes principais.

É interessante ressaltar que os componentes principais são mutuamente ortogonais, pois é possível verificar que $\mathbf{P}^T \mathbf{P} = \mathbf{I}$. Do mesmo modo, é possível observar que os componentes principais não são correlacionados, pois $Cov(\mathbf{Y}) = \mathbf{\Lambda}$, que é uma matriz diagonal.

Após a obtenção dos componentes principais, cabe ainda avaliar o grau de associação existente entre eles e as variáveis originais. Para isso, deve-se inicialmente obter a covariância entre o vetor de variáveis originais e os componentes principais por:

$$Cov(\mathbf{Y}, \mathbf{X}) = Cov(\mathbf{P}^T \mathbf{X}, \mathbf{X}) = \mathbf{P}^T Cov(\mathbf{X}) = \mathbf{P}^T \mathbf{\Sigma} = \mathbf{\Lambda} \mathbf{P}^T.$$

De acordo com esta expressão, verifica-se que a covariância entre a k -ésima variável original e o i -ésimo componente principal é dada pelo produto entre o i -ésimo autovalor de $\mathbf{\Sigma}$ e o k -ésimo componente do i -ésimo autovetor, ou seja, por $Cov(Y_i, X_k) = \lambda_i a_{ik}$. Então, é possível determinar a matriz de correlação entre o vetor de componentes principais e o vetor de

variáveis originais por:

$$\begin{aligned}\rho_{\mathbf{Y},\mathbf{X}} &= \mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}\mathbf{P}^T\mathbf{V}^{-1/2} \\ &= \mathbf{\Lambda}^{1/2}\mathbf{P}^T\mathbf{V}^{-1/2},\end{aligned}$$

em que $\mathbf{V} = \text{diag}(\sigma_{ii})$ e para a qual a i -ésima linha e a k -ésima coluna contêm o coeficiente de correlação entre Y_i e X_k (ρ_{Y_i,X_k}), dado de forma escalar por:

$$\rho_{Y_i,X_k} = \frac{\sqrt{\lambda_i}a_{ik}}{\sqrt{\sigma_{kk}}}. \quad (2.5)$$

Vale ressaltar que quando se trabalha com os dados originais, a análise de componentes principais é realizada a partir da matriz de covariância e os escores dos componentes principais são obtidos com as variáveis centradas na média. Quando as medidas são obtidas em escalas e unidades muito diferentes, em geral, a ACP é realizada a partir da matriz de correlação. Neste caso, trabalha-se com variáveis padronizadas com média zero e variância um. Esta técnica permite o agrupamento de indivíduos (produtos, tratamentos, genótipos etc.) similares, mediante exames visuais em dispersões gráficas no espaço bi ou tridimensional de fácil interpretação geométrica (Minim, 2006).

No entanto, Schlich (1995) afirma que a ACP deve ser feita com base na matriz de covariância entre os consumidores para permitir que as preferências dos consumidores sejam leves ou fortes. Se, ao contrário, é assumido que estas diferenças de discrepância individual são artefatos devido a tipos de escalas individuais, então é possível realizar a ACP baseada na matriz de correlação entre consumidores. Esta análise posterior divide as notas dadas por cada consumidor pela divergência de escalas e então posteriormente é possível realizar uma ACP clássica nos resultados. O autor acredita que, em muitos casos, esse procedimento não é sensível.

2.3 ACP com valores faltantes

Ao realizar pesquisas sensoriais com consumidores, em diversos casos não é possível conseguir com que todos os provadores degustem todas as amostras, por limitações diversas: cansaço, fadiga sensorial, indisponibilidade de tempo, entre outros. Quando isso ocorre, na grande maioria das vezes, este provador é descartado e seus dados são eliminados. Grung & Manne (1998) descrevem maneiras distintas de construções da análise de componentes principais com valores faltantes. Entre elas, destaca-se a metodologia usada com auxílio de pacotes computacionais comerciais, ou seja, com o uso do algoritmo Non-Iterative Partial Least Squares - NIPALS.

Assim, para a solução de uma análise de componentes principais com valores faltantes, considere a minimização da função objetiva:

$$F = \sum_{ij} \left(X_{ij} - \sum_k t_{ik} p_{jk} \right)^2, \quad (2.6)$$

em que \mathbf{X} é uma matriz de dados composta por n quantidades escalares obtidas por m diferentes objetos; \mathbf{p}_k^T representa um conjunto de vetores lineares ortonormais e $t_{ik} = (\mathbf{X}_i^T \mathbf{p}_k)$ são escores, assumindo a condição de ortonormalidade $\sum_j p_{jk} p_{jl} = \delta_{kl}$ e a ortogonalidade $\sum_i t_{ik} t_{il} = 0$ para $k \neq l$. Quando há valores faltantes em X_{ij} , o modelo apresentado $\sum_k t_{ik} p_{jk}$ deve ser excluído da função objetiva. Além disso, é necessário distinguir entre a matriz completa, a matriz desconhecida e a parte conhecida. É conveniente denotar tal diferença por: \mathbf{Y} para matriz completa e \mathbf{X} para a parte conhecida de \mathbf{Y} . Na matriz \mathbf{X} os valores faltantes devem ser representados por zeros. Com essa notação, a letra \mathbf{X} é ainda usada para a matriz constituída por valores calculados.

A relação entre \mathbf{X} e \mathbf{Y} é convenientemente expressada com a ajuda da incidência ou presença-ausência da matriz \mathbf{C} de alguma dimensão e com elementos $C_{ij} = 1$ se Y_{ij} é conhecido e $C_{ij} = 0$ se Y_{ij} é desconhecido. Há,

conseqüentemente, as relações $X_{ij} = C_{ij}Y_{ij}$ e $C_{ij}^2 = C_{ij}$, então, a função objetiva deve ser escrita como:

$$F = \sum_{ij} C_{ij} \left(Y_{ij} - \sum_{k=1}^A t_{ik} p_{jk} \right)^2 \quad (2.7)$$

$$= \sum_{ij} \left(X_{ij} - \sum_{k=1}^A t_{ik} C_{ij} p_{jk} \right)^2. \quad (2.8)$$

Christoffersson (1970) considerou a minimização desse problema para um componente principal ($A = 1, t_i = t_{i1}, p_j = P_{j1}$). Igualando-se as derivadas parciais $\delta F/\delta t_i$ e $\delta F/\delta p_j$ a zero, obtém-se:

$$t_i = \frac{\sum_j X_{ij} p_j}{\sum_j C_{ij} p_j^2}, \quad (2.9)$$

$$p_i = \frac{\sum_i t_i X_{ij}}{\sum_i t_i^2 C_{ij}}, \quad (2.10)$$

As equações (2.9) e (2.10) devem ser resolvidas por iteração, com a condição de normalização $\sum_j p_j^2 = 1$. Cabe, então, a aplicação do algoritmo NIPALS padrão.

NIPALS é um algoritmo iterativo e baseado em sucessões de regressões lineares (Tenenhaus, 1998). Estudos iniciais com esse algoritmo foram feitos por Herman Wold (1966), visando trabalhar com quadrados mínimos parciais não iterativos na área da quimiometria, o que explica o nome *Nonlinear iterative partial least squares*. Christofferson (1970) estendeu as aplicações para a geração do primeiro componente principal a partir de uma matriz com dados faltantes. Posteriormente, com maior grau de dificuldade, elaborou melhorias para estimar dois componentes principais, porém, obteve algumas equações que não convergiam a um ponto ótimo, mostrando necessárias adaptações ao mesmo. Assim, o algoritmo NIPALS apresenta algumas modificações, de acordo com a aplicabilidade.

Algebricamente, o algoritmo NIPALS adaptado por Dray et al. (2003), disponível em linguagem R (R, 2007), realiza os seguintes passos:

1. Normalização de $Z(Z_0)$
2. Para $h = 1, 2, \dots, a$ (com $a \leq N$):
 - 2.1. Para $t_h = Z_{h-1}[1]$
 - 2.2. Repita até a convergência de p_h
 - 2.2.1. Para $j = 1, 2, \dots, N$:

$$p_h[j] = \frac{\sum_{i=1}^n (\text{se } Z[i, j] \text{ e } t_h[i] \text{ existe}) Z_{h-1}[i, j] t_h[i]}{\sum_{i=1}^n (\text{se } Z[i, j] \text{ e } t_h[i] \text{ existe}) t_h[i]^2}$$
 - 2.2.2. Normalize p_h
 - 2.2.3. Para $i = 1, 2, \dots, n$:

$$t_h[i] = \frac{\sum_{j=1}^N (\text{se } Z[i, j] \text{ existe}) Z_{h-1}[i, j] p_h[j]}{\sum_{j=1}^N (\text{se } Z[i, j] \text{ existe}) p_h[j]^2}$$
 - 2.3. $Z_h = Z_{h-1} - t_h p_h'$

O método também pode ser entendido por meio do seguinte esquema (Figura 4). Assim, o algoritmo NIPALS permite realizar a análise de componentes principais com valores faltantes sem estimar ou excluir registros vazios. Como na análise de componentes principais clássica, o NIPALS permite calcular linhas (t_h) e colunas (p_h) coordenadas, assim como o autovetores do h -ésimo eixo:

$$\lambda_h = \frac{1}{n-1} t_h' t_h.$$

Além disso, valores faltantes podem ser estimados usando as fórmulas de reconstituição clássica na h -ésima ordem (Good, 1969):

$$\hat{Z}_0[i, j] = \sum_{l=1}^h p_l[j] t_l[i].$$

A ACP realizada em \mathbf{Z} tende a encontrar uma seqüência de escores, comum aos espaços vazios observados para todas as variáveis.

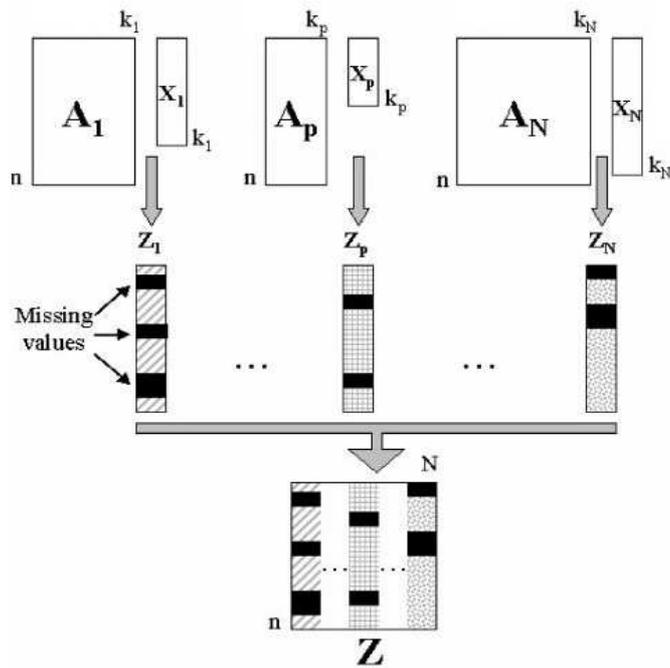


FIGURA 4: Esquemática do método. Painel sensorial com grupo de consumidores ($A_1, \dots, A_p, \dots, A_n$), em que são atribuídas notas para k produtos e n atributos. A média de cada consumidor resulta em um vetor Z_i que constituirá a matriz Z , que será analisada pelo algoritmo NIPALS. **FONTE:** Dray et al. (2003) p. 416.

2.4 Análise de variância

É um método alternativo de análise de dados em que se considera a variação total de uma amostra combinada e a divide em diferentes componentes (Mead & Curnow, 1983). Por exemplo: variação devido a consumidores, variação devido aos produtos, variação puramente aleatória, entre outros. Por meio da análise de variância, pretende-se controlar algumas fontes de variação presentes em um experimento, a partir de um modelo proposto.

Para a análise de variância ser válida é importante conhecer as hi-

póteses formuladas, ou seja, o que deseja-se testar. Além disso alguns pressuposições devem ser atendidas:

- os erros devem: possuir distribuição normal com média zero, ser independentes e possuir variância constante (homocedasticidade);
- o modelo estatístico deve ser aditivo.

O não atendimento de uma ou mais dessas pressuposições pode afetar o nível de significância e a sensibilidade do teste F para a tomada de decisão sobre a hipótese nula (igualdade entre as variâncias). Tais características devem ser verificadas antes da análise.

Teste F é um teste estatístico que está diretamente ligado à análise de variância. Destina-se à comparação de variâncias, ou seja, compara a variação explicada (por ex.: variação entre consumidores) com, na maioria das vezes, a variação não explicada (residual) da variável dependente, tendo então como hipótese de nulidade a igualdade entre as variâncias. Quando esta hipótese é rejeitada, procede-se a comparação múltipla das médias da fonte de variação testada (variáveis qualitativas) ou ajusta-se um modelo de regressão (variáveis quantitativas).

Um teste de comparação múltipla muito utilizado é o **Teste de Tukey**. Este teste serve para verificar qual é a diferença mínima significativa (DMS) entre duas médias para estas serem consideradas diferentes. A DMS deste teste é dada por:

$$\Delta = q_{(I,n)} \times \sqrt{\frac{QME}{J}}.$$

em que: Δ representa a DMS a uma taxa nominal de significância (α), geralmente de 5%; q representa o quantil da amplitude estudentizada, que depende do número de tratamentos I e dos graus de liberdade do resíduo n ; QME é o quadrado médio do resíduo; J é o número de repetições, ou

de unidades experimentais, com que se estimou cada média de tratamento (SISVAR, 2003).

2.5 Mapas de preferência

As primeiras idéias sobre mapa de preferência (MP) surgiram com a Psicometria, nos Estados Unidos, entre 1968 e 1981. Estas então foram introduzidas na indústria alimentícia do Reino Unido em 1989 (Nute et al.; Jones et al.). Em 1995, as técnicas de Mapas de Preferência foram formalizadas por Schlich (1995). Tais técnicas são aplicadas desde então (Oliveira et al., 2004; Guinard et al., 2001; Xiong & Meullenet, 2004).

Schlich (1995) define mapa de preferência como um conjunto de procedimentos estatísticos, baseados em análise de componentes principais, análise de clusters e regressão polinomial múltipla, que auxiliam na verificação da aceitabilidade de um produto. Qualquer tipo de indústria onde a percepção sensorial humana de produtos está envolvida, poderá usar estas técnicas. Assim, MP compara um conjunto de produtos competitivos, cuja aceitação global possa ser medida por um painel de consumidores, e cujas propriedades sensoriais têm sido definidas por um painel de provadores treinados.

Para Minim (2006), a técnica mapa de preferência foi desenvolvida com o objetivo de comparar preferências e relacioná-las com as características de qualidade do produto, auxiliando na segmentação do mercado em grupos definidos de consumidores. Assim, o produto pode ser introduzido no segmento correto do mercado ou ser otimizado a partir das principais características de qualidade e da indicação da direção da preferência do consumidor.

A elaboração do mapa inclui a comparação de n produtos, que podem ser produtos concorrentes, existentes no mercado ou possíveis novos

produtos. Schlich (1995) afirma que, geralmente, esta quantidade varia entre 6 e 12 produtos comparados, mas que o número mínimo de 10 produtos é recomendado.

Do mesmo modo que há uma indicação ao número de produtos, Schlich também ressalta que, para a elaboração de um painel de consumidores não-treinados, é necessário atentar para a representatividade da amostra, utilizando em torno de cem provadores que deverão atribuir notas hedônicas, em escalas pré-definidas, usualmente de 9 pontos, que gerarão a matriz \mathbf{X} , contendo n linhas (produtos) e p colunas (provadores). Caso os provadores sejam treinados, o valor p tende a ser bem menor, visto a capacidade do provador em degustar uma quantidade bem maior de amostras do que um provador não-treinado.

O mapa de preferência pode ser Interno ou Externo.

2.5.1 Mapa de preferência interno

Mapa de preferência interno (MPI) deriva do mapa de preferência dos produtos, segundo Schlich (1995), efetivamente descrito por dimensões de preferências principais básicas, onde são identificados clusters (grupos com preferências similares). É basicamente uma análise de componentes principais com \mathbf{X} matrizes, com p linhas (produtos) e n colunas (consumidores). Assim, as observações estatísticas da ACP são os produtos e as variáveis são os consumidores. Esta análise de componentes principais deve ser baseada na matriz de covariância entre consumidores para permitir as diferenças discrepantes entre as preferências dos consumidores. Por outro lado, se é assumido que essas diferenças devem-se à diferença de escalas, então, é possível executar a ACP baseada na matriz de correlação entre consumidores. O primeiro tipo divide as notas de cada consumidor pelo desvio padrão deles para, posteriormente, executar a ACP clássica nos resultados. Schlich (1995) afirma, ainda que, em muitos casos, essa padronização não

é sensível.

Young et al. (2004), estudando queijo cheddar em diferentes níveis de maturidade, e Thompson et al. (2004), sobre marcas comerciais de achocolatado, afirmam que o mapa de preferência auxilia os pesquisadores a entenderem os atributos sensoriais descritivos que influenciam as preferências dos consumidores. São comumente usados para examinar a relação entre dados sensoriais descritivos e respostas dos consumidores. O mapa de preferência interno é uma análise de componente principal de uma matriz de dados hedônicos que tem, por um lado, os produtos (as observações) e, por outro, o consumidor (a variável), que é extraído para uma matriz de covariância, que permite diferenças nas preferências dos consumidores a serem expressadas.

O MPI é gerado a partir de uma matriz com os dados de aceitação de uma equipe de consumidores, sobre um mesmo conjunto de amostras, não permitindo o uso de delineamento de blocos incompletos (Oliveira et al., 2004). Assim é entendido que, para a construção do MPI, é necessário que todos os consumidores avaliem todas as amostras. No entanto, de acordo com esta afirmação, tem-se uma limitação nos estudos sensoriais, como ocorre freqüentemente visto que, por fatores diversos, como fadiga, indisposição para realizar todas as provas, diferenças entre as habilidades dos provadores, entre outras, os consumidores, muitas vezes, não completam todo o painel sensorial e o pesquisador perde dados. Este trabalho apresenta uma alternativa para solucionar este problema.

Após ter sido gerado, com base somente em dados de aceitação, é possível avaliar o quanto um produto é aceito (ou preferido) no mercado em função de suas características de qualidade.

2.5.2 Mapa de preferência externo

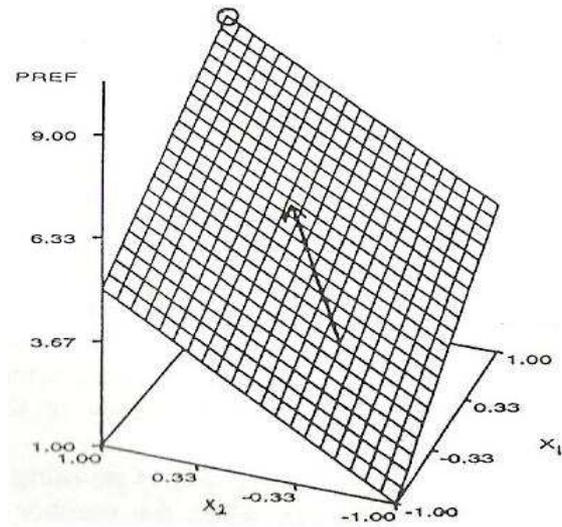
O mapa de preferência externo (MPE) retorna as preferências individuais dos consumidores para o primeiro dos dois componentes principais da matriz de covariância de descrição ou para a avaliação analítica dos produtos (Young et al., 2004). Thompson et al. (2004) ainda acrescentam que o mapa de preferência interno é usado apenas em dados de consumidores que determinam pares de preferências de consumidores e o mapa de preferência externo relaciona dados de preferência dos consumidores com informações da descrição sensorial e ou dados instrumentais. Essas técnicas podem guiar a otimização e o desenvolvimento de novos produtos.

Para Schlich (1995), o MPE localiza o produto ideal no mapa sensorial dos produtos. Esse produto ideal, identificado por diferentes grupos de consumidores, deve se tomar parâmetro para a formulação de novos produtos. Basicamente, é uma regressão polinomial de cada consumidor individual Y_i , $i = 1, \dots, p$ sobre variáveis X_1 e X_2 . A partir do modelo da regressão polinomial, é possível identificar quatro submodelos, aninhados: o vetorial é submodelo do modelo circular, que é submodelo do elíptico, que é, finalmente, um submodelo do quadrático. São eles:

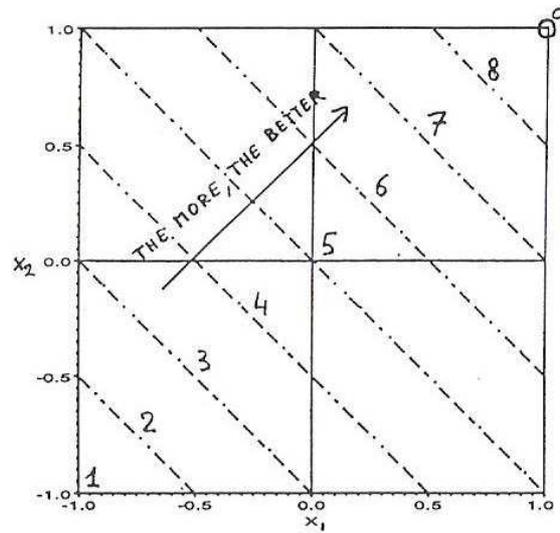
Modelo vetorial é uma regressão linear múltipla de Y_i em X_1 e X_2 :

$$Y_i = a + b_1 X_1 + b_2 X_2.$$

A superfície de resposta (Figura 5a) de cada modelo é uma superfície plana com uma direção mais íngreme (seta) que conduz à preferência ótima igual a 9 para que o ponto sensorial que corresponde a ambos, X_1 e X_2 , se iguale a posição 1 (círculo). O autor acredita ser fácil compreender olhando para o gráfico de curvas de nível (Figura 5b), o qual, neste caso, consiste em linhas diretas paralelas. O modelo vetorial é chamado “quanto



(a)



(b)

FIGURA 5: Modelo vetorial $PREF = 5 + 2X_1 + 2X_2$ (a) superfície de resposta e (b) curvas de nível. **FONTE:** Schlich (1995), p.138

mais, melhor”, indicando que, para qualquer consumidor descrito por este modelo, o produto ideal deve estar localizado até onde for possível para a seta, mostrando, então, que as propriedades sensoriais correlacionadas com essa direção devem ser maximizadas. Alguns consumidores preferem níveis médios de alguns atributos sensoriais, fazendo com que os pontos ideais presentes se localizem dentro do mapa sensorial, em vez de estar fora ou na borda do mapa. Então, o modelo circular permite um melhor ajuste.

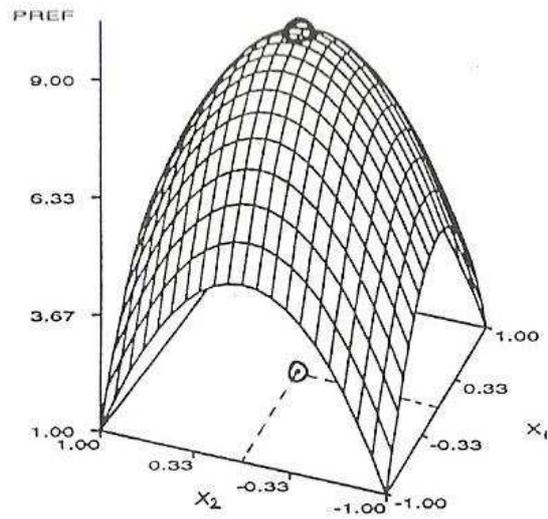
Modelo Circular é o modelo vetorial adicional a um termo quadrático:

$$Y_i = a + b_1X_1 + b_2X_2 + c(X_1^2 + X_2^2)$$

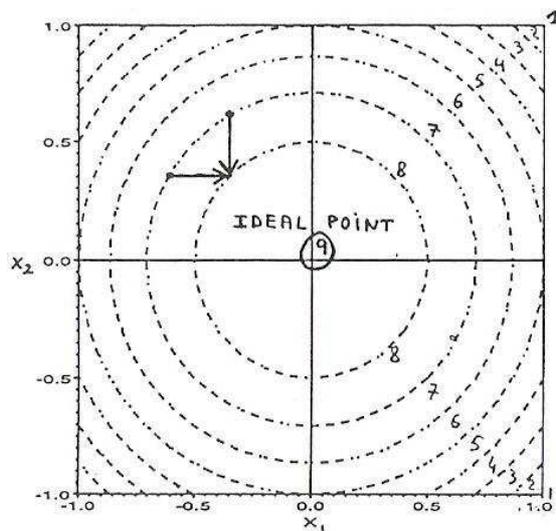
A superfície de resposta (Figura 6a) de cada modelo é uma parábola bi-dimensional, na qual a preferência ótima é marcada pelo número 9, ao centro do espaço sensorial, indicando que esse consumidor está esperando um produto médio em termos dos atributos sensoriais identificados. O gráfico de curvas de nível (Figura 6b) desta superfície, representado por círculos, indica que, no ponto ideal, duas dimensões sensoriais possuem pesos igualmente importantes para preferência. Porém, é provável que esta suposição não esteja clara ao consumidor. Se o sinal da estimação do coeficiente quadrático é positivo, o ponto ideal é negativo, representando que este consumidor deseja algo que tenha características até este ponto no espaço sensorial. Uma das propriedades do modelo circular é que as duas dimensões sensoriais são igualmente importantes para determinar preferências. No entanto, essa propriedade não representa um caso habitual, o que conduz a consideração de um outro modelo, o elíptico.

Modelo Elíptico é um modelo circular, porém, com um coeficiente quadrático diferente para cada dimensão:

$$Y_i = a + b_1X_1 + b_2X_2 + c_1X_1^2 + c_2X_2^2$$

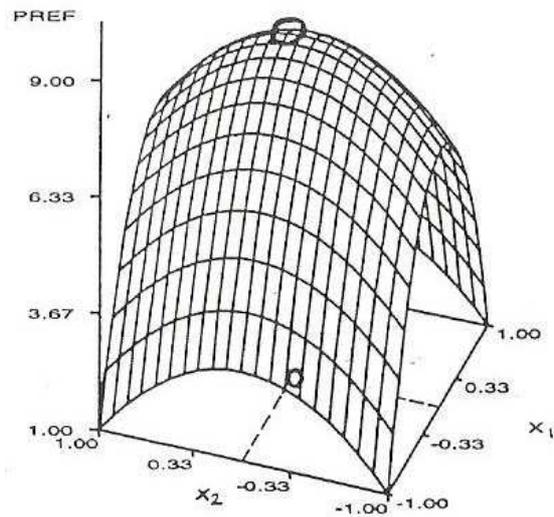


(a)

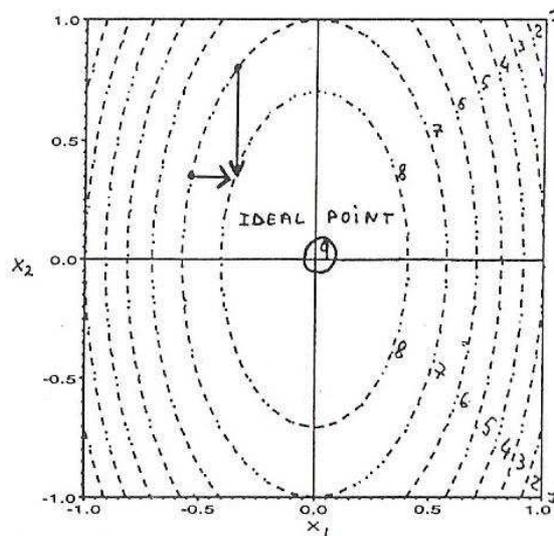


(b)

FIGURA 6: Modelo circular $PREF = 9 - 4X_1^2 - 4X_2^2$ (a) superfície de resposta e (b) curvas de nível. **FONTE:** Schlich (1995), p.138

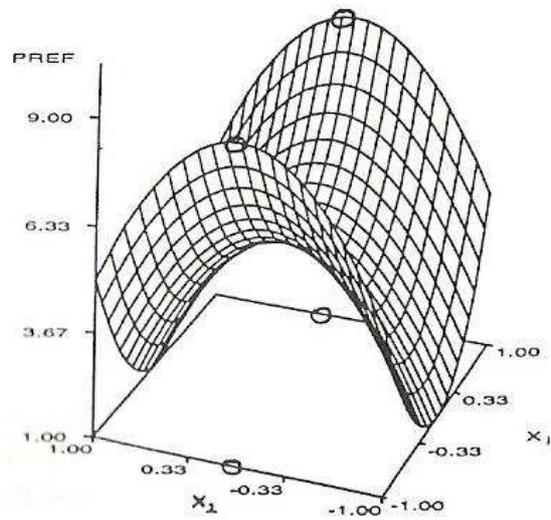


(a)

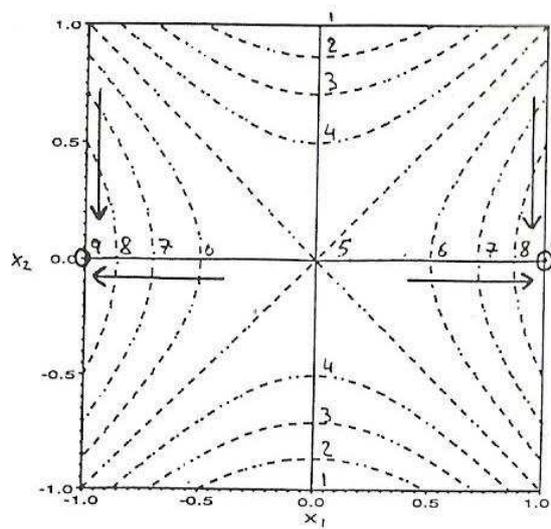


(b)

FIGURA 7: Modelo elíptico $PREF = 9 - 6X_1^2 - 2X_2^2$ (a) superfície de resposta e (b) curvas de nível. **FONTE:** Schlich (1995), p.138



(a)



(b)

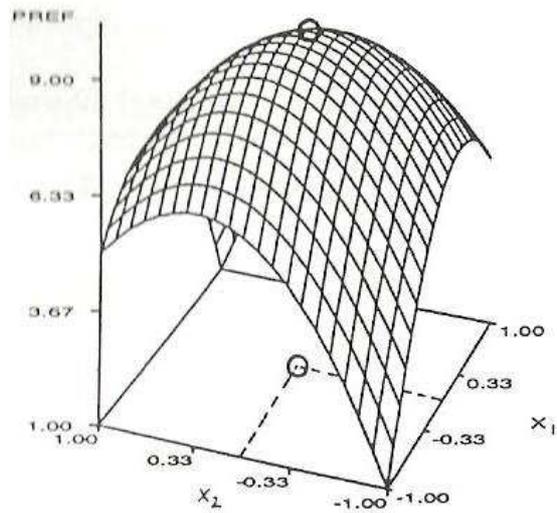
FIGURA 8: Modelo elíptico com ponto de sela $PREF = 5 + 4X_1^2 - 4X_2^2$ (a) superfície de resposta e (b) curvas de nível. **FONTE:** Schlich (1995), p.139

No primeiro exemplo do modelo elíptico, o peso da dimensão sensorial X_1 é 3 vezes o peso de X_2 , conduzindo à superfície de resposta (Figura 7a), na qual há um ponto ideal positivo na origem do espaço sensorial. Porém, observando as curvas de nível, constata-se que está ao redor do ponto ideal (Figura 7b). No entanto, a preferência desse consumidor são mais sensíveis em X_1 do que em X_2 . Assim como no modelo circular, o ponto ideal é chamado negativo quando ambas as estimações dos coeficientes quadráticos são negativos. Neste caso, é mais importante evitar aquele ponto no eixo que tenha o valor absoluto maior do que o coeficiente quadrático em lugar de outro. Mas, permitindo aos coeficientes quadráticos serem diferentes, é possível tê-los com sinais opostos, conduzindo ao chamado “ponto de sela”. No exemplo (Figura 8a), a superfície de resposta parece como a sela de um cavalo com dois ótimos pontos diferentes circulos. Na origem do mapa sensorial há um ponto ideal positivo ao longo de X_2 e um ponto ideal negativo ao longo de X_1 (Figura 8b). Assim, esse consumidor tende a preferir produtos com propriedades sensoriais médias relacionados a X_2 e propriedades sensoriais baixas ou altas relacionadas a X_1 .

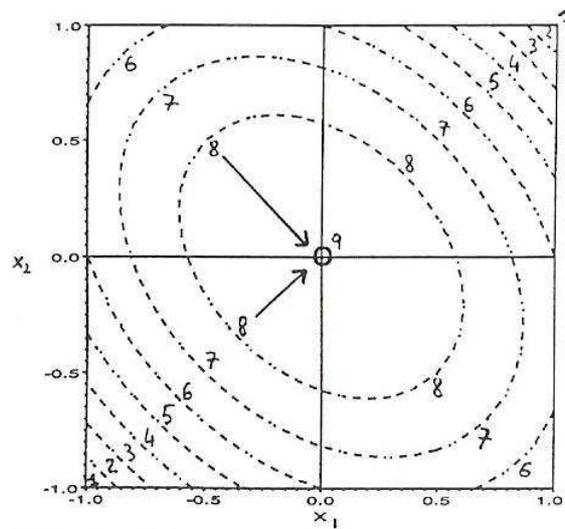
Modelo de superfície quadrático é um modelo elíptico adicionado à interação dos produtos:

$$Y_i = a + b_1X_1 + b_2X_2 + c_1X_1^2 + c_2X_2^2 + dX_1X_2.$$

A superfície de resposta apresentada no exemplo para este modelo (Figura 9a) se parece com uma superfície de resposta elíptica formada por eixos sensoriais não paralelos. A direção sensorial mais importante para a preferência não é nem a de X_1 nem a de X_2 , mas o primeiro bissetor do mapa sensorial. Embora esse consumidor prefira produtos que tenham X_1 e X_2 em níveis médios, se uma das dimensões de X_1 ou X_2 diminui, a outra deve compensar, aumentando na mesma magnitude para persistir a preferência no nível mais alto possível (Figura 9b). Esse tipo de comportamento de preferência é perfeitamente possível na prática.



(a)



(b)

FIGURA 9: Modelo quadrático $PREF = 9 - 3X_1^2 - 3X_2^2 - 2X_1X_2$ (a) superfície de resposta e (b) curvas de nível. **FONTE:** Schlich (1995), p.139

Schlich (1995) menciona não ter conhecimento de publicações usando os modelos elípticos ou quadráticos, certamente devido à dificuldade de interpretação para estes modelos. São mais comuns os modelos vetorial ou circular que possuem ajustes a grupos de consumidores. No entanto, qualquer um dos modelos descritos é, a priori, aceitável e um único modelo deve ser escolhido a posteriori para cada consumidor. Vários testes estatísticos clássicos do modelo linear podem ser feitos, mas, geralmente, são omitidos pelos usuários do mapa de preferência. Tal escolha acaba sendo feita de forma visual.

2.6 Análise generalizada de Procrustes

Procrustes é um personagem da mitologia Grega. Filho de Poseidon, ele possuía uma cama singular que teria o poder de se ajustar a todos que nela se deitassem. Ele costumava oferecer um farto jantar, regado a vinho, aos peregrinos que passavam por sua hospedaria e a eles prometia uma noite em sua cama mágica. Infelizmente, quando se viam sem saída, seus convidados eram ajustados a sua cama. Ele mutilava os muito compridos e esticava os baixinhos. Mas, Procrustes teve um trágico fim. Theseus, quando reclamava o reino de Atenas, matou-o em sua própria cama, em um de seus seis trabalhos.

Segundo Ferreira (2007a), como estavam diante de uma situação que exigia o ajuste de dados, Hurley & Cattell (1962) homenagearam o personagem da mitologia grega formalizando o *problema de Procrustes*, aqui descrito posteriormente.

De acordo com Gower & Dijksterhuis (2004), há três elementos principais na história de Procrustes, o *mal aventurado viajante* (matriz X_1), a *cama de Procrustes* (matriz X_2) e o *esticar* ou *amputar* do infeliz para deixá-lo do mesmo tamanho da cama (matriz T).

Gower (1971) deu os primeiros passo para a *Análise de Procrustes Ortogonal Generalizada* (AGP or APOG). Nesse contexto, X_1 e X_2 são substituídos por k conjuntos X_1, \dots, X_k . Este pode ter sido o primeiro exemplo, na qual as matrizes X_1, \dots, X_k são consideradas como matrizes de configuração em lugar de matrizes coordenadas com fatores de ponderação (Gower & Dijksterhuis, 2004).

Estudando fatores de ponderação, Kristof & Wingersky (1971) consideraram minimização pelo critério de multiconjunto $\sum_{i < i'}^k \|X_i Q_i - X_{i'} Q_{i'}\|$ para determinar matrizes ortogonais e retornar uma “matriz de fator” média referido por $G = \frac{1}{k} \sum_{i=1}^k X_i Q_i$.

Gower (1975) introduziu análise generaliza de Procrustes (AGP) minimizando o mesmo critério de Kristof & Wingersky (1971), mas acrescentando fatores de escala para cada conjunto e vendo todo o problema como uma configuração emparelhada (2.11). Ele também foi o primeiro a criar o termo *Análise Generalizada de Procrustes* e a introduzir fatores de escala s_i ($i = 1, \dots, k$) para cada matriz X_i . O problema é descrito como:

$$\sum_{i < i'}^k \|s_i(X_i - 1u_i')Q_i - s_{i'}(X_{i'} - 1u_{i'}')Q_{i'}\|, \quad (2.11)$$

que é equivalente a minimizar

$$k \sum_{i=1}^k \|s_i(X_i - 1u_i')Q_i - G\|, \quad (2.12)$$

em que G é denominado *configuração média de grupo*, ou seja, representa o consenso entre os juízes sobre o conjunto de produtos analisado (2.13),

$$G = \frac{1}{k} \sum_{i=1}^k s_i(X_i - 1u_i')Q_i. \quad (2.13)$$

Gower (1975), na generalização da análise de Procrustes, apresenta

um algoritmo iterativo que busca os parâmetro de interesse (Q_i , s_i , $1u_i'$ e G) enquanto a distância euclidiana entre os vértices similares das configurações são minimizados. Em 1977, Ten Berge corrigiu o método de Gower (1975) pela determinação dos fatores de escala e melhorou o procedimento com o cálculo das transformações ortogonais. Ten Berge & Knol (1984) estenderam o problema para o caso em que duas ou mais configurações têm número de colunas diferentes (embora apenas para a transformação ortogonal). Mais recentemente, Peay (1988) resolveu o problema de ajuste para mais de duas configurações com diferentes dimensionalidades, incluindo translação e fatores de escala.

A análise generalizada de Procrustes é considerada um método estatístico multivariado descritivo (Ferreira, 2004), um método gráfico que permite a tirada de conclusões relativas em seu plano-resposta. Esse plano é formado por dois eixos principais que representam combinações lineares dos atributos em que os objetos foram julgados.

Estudos recentes realizados por Ferreira (2007a) modificaram o caráter descritivo desse método permitindo que sejam realizadas também inferências estatísticas (testes de hipóteses e estimação de regiões de confiança).

3 MATERIAL E MÉTODOS

3.1 Associação do mapa de preferência com a análise generalizada de Procrustes

Normalmente, na construção de mapas de preferência externos, a nota hedônica dos consumidores é conjuntamente analisada com o resultado da análise descritiva dos produtos feita por provadores treinados ou instrumentos (análises químicas, físicas, físico-químicas, etc). Entretanto, em alguns estudos, os consumidores são chamados a avaliarem atributos sensoriais (Ferreira et al., 2003; Mori et al., 1998). Como consumidores são, na maioria das vezes, pouco precisos e viesados, tais dados são geralmente pouco valorizados. Contudo, consegue-se uma redução considerável no custo da pesquisa com tal procedimento. A seguir, serão apresentadas três maneiras de lidar com dados hedônicos e descritivos de consumidores (provadores não treinados), afim de identificar uma metodologia eficiente, que seja capaz de eliminar o máximo possível do viés existente. Vale notar que, embora as análises já existam, a maneira com que são associadas é contribuição desse trabalho.

Considere uma amostra de n produtos avaliada por k consumidores, que atribuem notas para a aceitação (escala hedônica) e para p atributos sensoriais (contidas nos grupos: aroma, sabor, textura, aparência).

Na SITUAÇÃO I: (Mapa 1) os dados de aceitação são armazenados

em uma matriz \mathbf{H} , de dimensão $n \times k$. As demais variáveis (descritivas) são armazenadas em k matrizes \mathbf{X} , de dimensão $n \times p$. É, então, feita a média dessas k matrizes, resultando em $\mathbf{Y} = \frac{1}{k} \sum_{i=1}^k X_i$. A esta matriz \mathbf{Y} , aplica-se a análise de componentes principais para gerar os eixos do mapa de preferência, calculado com a matriz \mathbf{H} .

A SITUAÇÃO II: (Mapa 2) difere da SITUAÇÃO I apenas quanto à forma de se calcular a matriz \mathbf{Y} . Esta provirá da análise generalizada de Procrustes, ou seja, em vez da média, usa-se o consenso (3.1) entre consumidores,

$$Y = \frac{1}{k} \sum_{i=1}^k X_i^*, \quad (3.1)$$

em que

$$X_i^* = s_i(X_i - \mathbf{1}u_i')Q_i, \quad (3.2)$$

e s_i é o fator de escala do consumidor i ; $\mathbf{1}$ é um vetor com n uns; \mathbf{u}_i é um vetor contendo as médias das colunas de X_i e \mathbf{Q}_i é uma matriz de rotação (matriz ortogonal) própria do consumidor i .

Na SITUAÇÃO III: (Mapa 3) são formadas k matrizes \mathbf{X} , de dimensão $n \times (p + 1)$ que guardam todas as variáveis. A transformação de Procrustes (3.2) é feita nessas matrizes expandidas, com o objetivo de se retirar os vieses de todas as variáveis. Após essa transformação, as colunas referentes à preferência são retiradas e agrupadas numa matriz \mathbf{H}^* . As demais são mantidas intactas e a matriz de consenso \mathbf{Y} é calculada de acordo com (3.1). Por fim, o mapa é feito com \mathbf{H}^* e seus eixos gerados por \mathbf{Y} .

O mapa 1 é construído conforme proposto por Schlich (1995) e os demais mapas são propostas sugeridas neste trabalho. Foram utilizados dados de dois exemplos distintos para a aplicação das metodologias descritas, sendo um aplicado à aceitação de estudantes por disciplinas de um curso a

distância (Experimento 1) e o outro referente à aceitação por palmitos com adição de anti-oxidantes (Experimento 2).

3.1.1 Experimento 1: avaliação do ensino (à distância)

Em um estudo feito por Garcindo (2007), ao final do primeiro encontro do curso de Pós-Graduação *Lato Sensu* em Controladoria e Finanças Empresarias da Universidade Federal de Lavras (UFLA), foram aplicados questionários a 44 estudantes (consumidores), cujo perfil revela que a maioria reside em cidades distantes de Lavras, principalmente em outros estados e que, geralmente, estão empregados, formados em Administração de empresas há muitos anos.

Foram avaliadas as disciplinas (objetos): Metodologia Científica (MC), Teoria e Prática de Administração Financeira (TPAF), Direito e Legislação Tributária (DLT), Comunicação Empresarial (CE) e Matemática Financeira (MF). As disciplinas foram numeradas de 1 a 5, para a apresentação ilustrativa dos resultados. Foram consideradas cinco covariáveis referentes ao docente: domínio do conteúdo (DC), didática (D), clareza (C), utilização de recursos audiovisuais (URA), atividade avaliativa (AAv); duas referentes ao material didático impresso: clareza do material (CMat) e redação do material (RMat); duas sobre o assunto em si: importância (Im) e utilidade prática (Ut) do assunto e uma sobre a turma: participação na disciplina em questão (PT).

Assim, foram avaliados 5 objetos e 10 covariáveis, utilizando-se uma escala hedônica de 9 pontos para avaliar a aceitação aos objetos e escala não estruturada de 12cm para a aceitação as covariáveis.

3.1.2 Experimento 2: palmito tratado

Os dados foram coletados e cedidos por Marisa Carvalho Botelho, Mestranda em Ciência dos Alimentos da Universidade Federal de Lavras.

As amostras de palmito foram tratadas com anti-oxidante em 7 tempos distintos: 0, 2, 6, 8, 10, 12, 14 dias em ação do tratamento, correspondendo, respectivamente, aos tratamentos 1, 2, . . . , 7. Estas foram servidas em pontos de venda do produto, na cidade de Lavras, MG, a um total de 100 consumidores, tendo grande parte deles entre 21 e 50 anos, sendo 48 mulheres e 52 homens, e, principalmente, com o hábito de consumo do produto (70% consomem de 1 a 4 vezes ao mês). As avaliações foram feitas por questionários por meio de escala hedônica de 9 pontos, solicitando respostas para a aceitação e 3 covariáveis: cor, odor e firmeza do produto (aparência).

3.2 Mapas de preferência com valores faltantes

O mapa de preferência foi construído com valores faltantes, tendo, primeiramente, sido obtido os componentes principais por meio do algoritmo NIPALS (Dray et al., 2003), disponibilizado em linguagem R (R, 2007) na página pessoal <http://biomserv.univ-lyon1.fr/dray/software.php>. Em seguida, construiu-se o mapa de preferência por meio da função *carto()*, presente no pacote *SensoMineR* (Husson & Lê, 2006).

Para a exemplificação, foram utilizados dados de dois experimentos, nos quais uma parte dos consumidores não provaram todas as amostras, conforme descrito abaixo. Ressalta-se ainda, que, para a execução desse algoritmo, por restrição do mesmo, foram desconsiderados os provadores que deram nota a apenas um produto ou que repetiram as mesmas notas para todos os produtos provados, resultando em variância igual a zero.

3.2.1 Experimento 3: análise sensorial de doce de leite

Este experimento foi realizado como parte dessa dissertação, na Universidade Federal de Viçosa, sob a supervisão da Profa. Dra. Valéria Paula Rodrigues Minim.

Foram avaliadas seis amostras, codificadas com três dígitos ao acaso, de marcas comerciais de doces de leite, sendo quatro delas escolhidas ao acaso e duas ganhadoras de prêmios de qualidade no estado de Minas Gerais. Dentro de cada marca escolhida, respeitou-se o mesmo lote de fabricação e validade do produto. Os consumidores foram escolhidos por amostragem não-probabilística de auto-escolha.

As amostras foram avaliadas por 147 consumidores, escolhidos de forma aleatória, dos quais 106 provaram todas as amostras, sendo 52 mulheres e 54 homens, com faixa etária entre 16 e 37 anos. O teste de aceitação foi realizado no Laboratório de Análise Sensorial do Departamento de Tecnologia de Alimentos da Universidade Federal de Viçosa, MG, em agosto de 2007. As amostras foram servidas de forma monádica aos consumidores, em cabines individuais. Junto a cada amostra servida, o consumidor recebeu um copo de água em temperatura ambiente para enxaguar a boca entre as avaliações, que foram feitas em duas etapas, sendo 3 amostras aleatórias na primeira etapa e, após um intervalo médio de 15 minutos, as demais amostras. Os consumidores avaliaram a aceitação (impressão global) das seis amostras de doce de leite utilizando a escala hedônica de 9 pontos (Minim, 2006).

3.2.2 Experimento 4: banana tratada

Foram consideradas amostras bananas tratadas com retardadores de amadurecimento (Pinheiro, 2007), servidas no Laboratório de Análise Sensorial do Departamento de Tecnologia de Alimentos da UFLA a consu-

midores escolhidos ao acaso, em sua maioria estudantes da instituição.

As amostras foram servidas em 5 tempos distintos, em 0, 6, 9, 12 e 24 dias, sob ação do retardador de amadurecimento. Inicialmente, participaram 72 consumidores, no entanto, por critérios da metodologia, foram descartados 10 consumidores que haviam provado apenas um produto. O questionário aplicado em escala hedônica de 9 pontos avaliou impressão global e 5 covariáveis: descasque, aroma, sabor, aparência e firmeza do produto.

4 RESULTADOS E DISCUSSÃO

4.1 Associação do mapa de preferência com a análise generalizada de Procrustes

Na análise de componentes principais quanto maior a porcentagem de explicação resultante dos dois primeiros componentes, melhor, pois indica variáveis com grande explicabilidade no modelo. Assim, para os dados do Experimento 1 (educação a distância), o modelo vetorial apresentou visualmente melhor ajuste. Na construção do mapa, pela situação I, o primeiro componente apresentou 87,51% da variabilidade e o segundo componente 8,84%, resultando em 96,35% (Figura 10). Vale ressaltar que, neste método, os eixos são determinados por meio das médias aritméticas das notas atribuídas às covariáveis pelos entrevistados. Além disso, foram identificados 26 grupos distintos (Figura 11), construindo-se um mapa (Figura 12) que apresenta a disciplina 2 como sendo a menos preferida, ficando as demais semelhantes, embora o grupo de disciplinas 3, 4 e 5 agrada consumidores distintos dos que preferem a disciplina 1. Porém, usualmente, os eixos não são construídos com base em nota de consumidores, como foi o caso. Para situações deste tipo, aconselha-se o Mapa 2, discutido a seguir.

Considerando o segundo método, utilizado para a construção do Mapa 2, ocorreu uma diminuição considerável da variabilidade, visto que, com a análise generalizada de Procrustes, os eixos são determinados por uma “média ponderada”, conhecida como o consenso de Procrustes, das

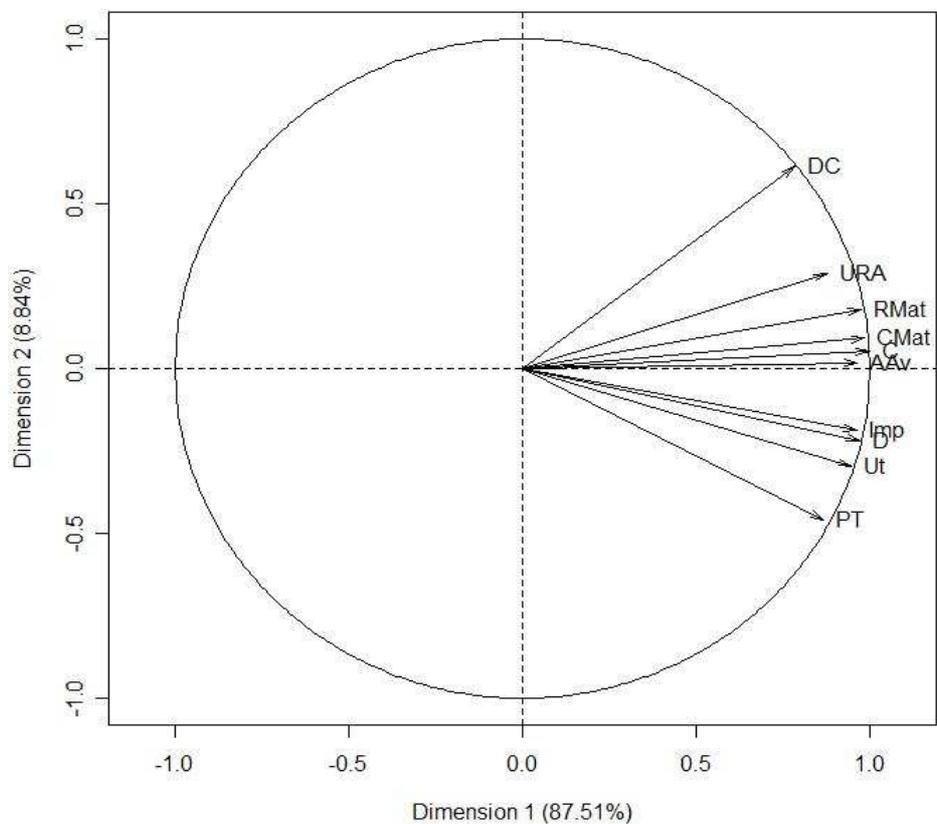


FIGURA 10: Análise de componentes principais - Método 1 aplicado a dados do ensino a distância

covariáveis, o que resulta na diminuição dos vieses presentes nas covariáveis.

A ACP teve 65,77% para o primeiro componente e 20,17% para o segundo componente, no que se refere à variabilidade, resultando em 85,94% (Figura 13). Esta é menor do que no primeiro caso. Entende-se que isso ocorre porque a análise de Procrustes busca o consenso entre os provedores e, portanto retira da variabilidade explicada pelos primeiros componentes aquilo que ela considera ruído (não informação). Por isso, embora os dois primeiros componentes expliquem menos, espera-se que esta explicação seja mais confiável. Além disso, as covariáveis apresentam maior clareza

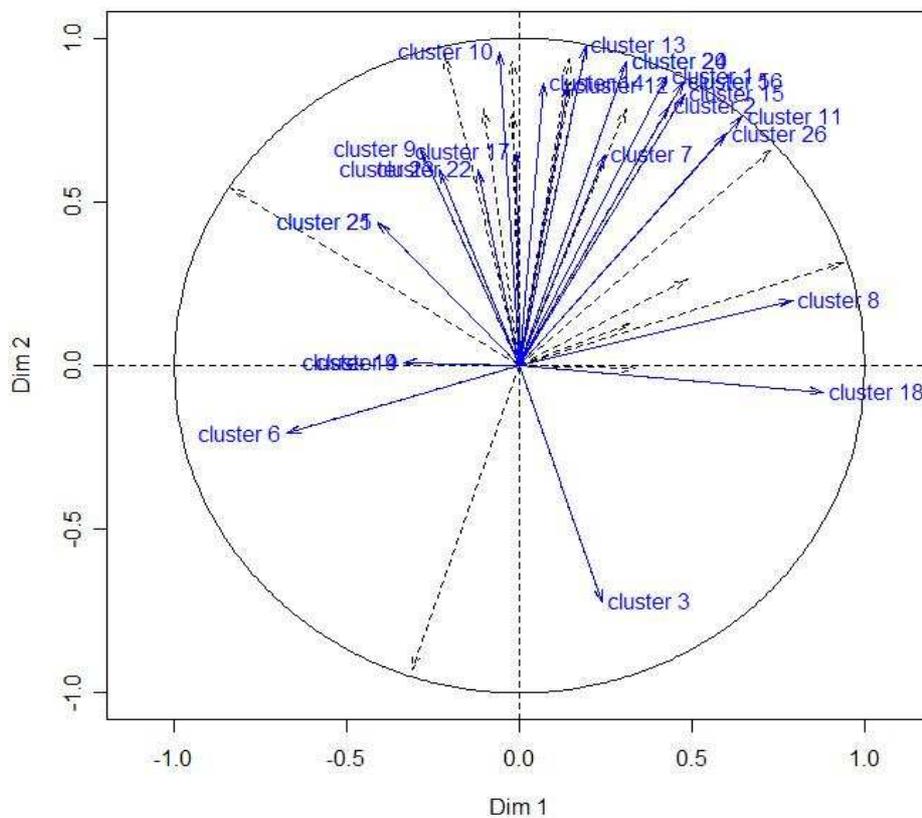


FIGURA 11: Grupos de consumo - Método 1 aplicado a dados do ensino à distância

na explicação, por estarem melhor distribuídas no gráfico, se comparada ao primeiro método (Figura 11).

Na construção do mapa (Figura 15) já é possível observar uma distinção entre as disciplinas 3 e 4 em relação a 5, o que não era possível concluir no mapa anterior. Este segundo método é mais indicado do que o primeiro, haja vista a menor variabilidade e a possibilidade de identificar diferenças pequenas entre produtos que, com determinado rigor, podem ser considerados diferentes, de acordo com o mapa.

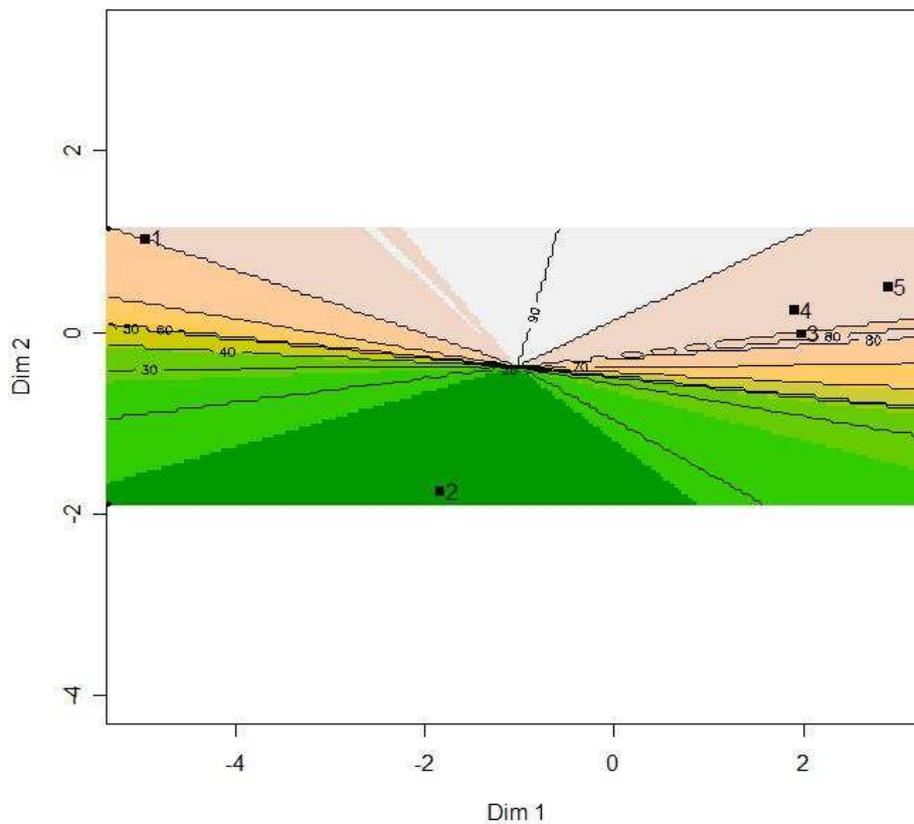


FIGURA 12: Mapa de preferência externo - Método 1 aplicado a dados do ensino a distância

Com a terceira metodologia, a preferência junto às demais covariáveis são analisadas de acordo com Procrustes, visando à minimização de todos os dados, obtendo-se, então, viéses mínimos (Figura 16), implicando em notas parecidas, com a menor variância possível para a primeira dimensão, o que traz um formato mais ortogonal ao mapa (Figura 18).

Neste método o gráfico se formou inverso aos demais, como um "espelhamento" no eixo Y , por questões computacionais, o que não prejudica as conclusões. Nesta construção, fica clara a distinção entre as disciplinas, em que a disciplina 2 foi considerada a menos preferida, ao contrário das

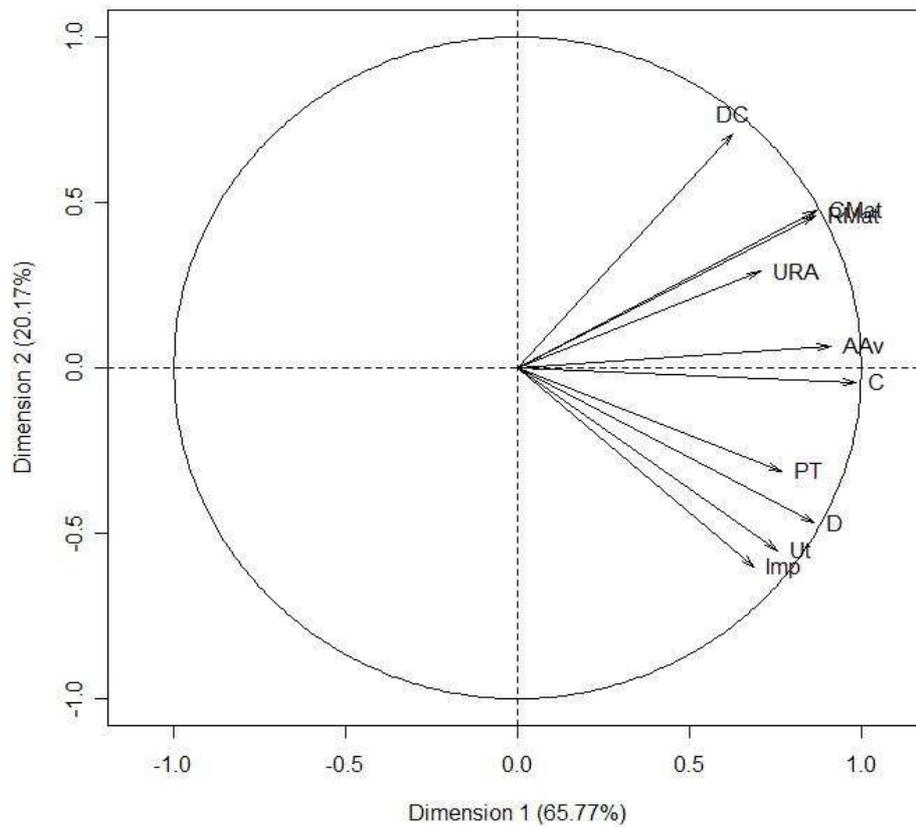


FIGURA 13: Análise de componentes principais - Método 2 aplicado a dados do ensino a distância

disciplinas 3 e 4, que são similarmente as melhores, seguidas pelas disciplinas 1 e 5. Uma outra observação é que os grupos de preferência remetem, a grande maioria, para uma mesma direção, mostrando um maior consenso nas notas (Figura 17).

Com tais resultados, observa-se que o método escolhido dependerá do objetivo da pesquisa realizada. Caso se deseje apenas um mapeamento geral, visando um conhecimento sobre as opiniões de consumidores de determinados produtos, pode-se construir o mapa pela metodologia 1, porém, ressaltando que notas de consumidores para covariáveis não são indicadas

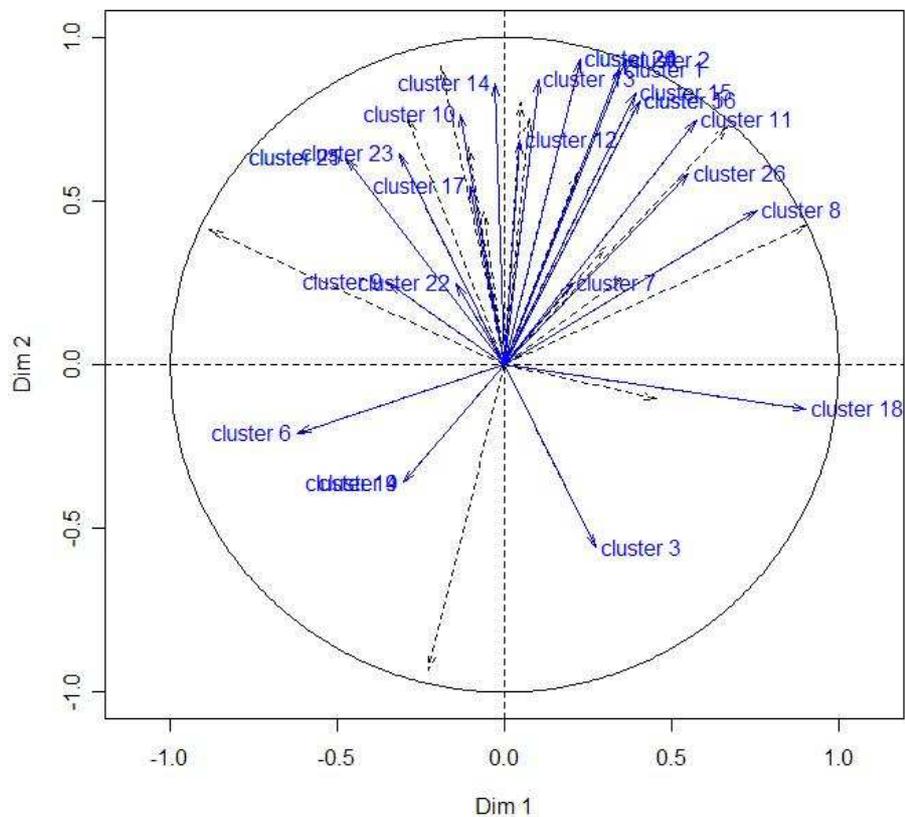


FIGURA 14: Grupos de consumo - Método 2 aplicado a dados do ensino a distância

pela metodologia usual do mapa de preferência externo.

Caso se deseje comparar produtos, verificando-se a preferência entre eles, para fins de acompanhamento de mercado (verificar qual produto deve receber mais investimento), por exemplo, o rigor apresentado pelo método 2 é capaz de fornecer resultados plausíveis para tal análise, apresentando produtos diferentes como aqueles que possuam certo grau real de diferenças. A metodologia 3 apresenta diferenças entre os produtos, mesmo no caso desta ser pequena. Assim, é indicado, caso se deseje, por exemplo, comparar produtos em relação a um concorrente, visando a real necessidade ou não

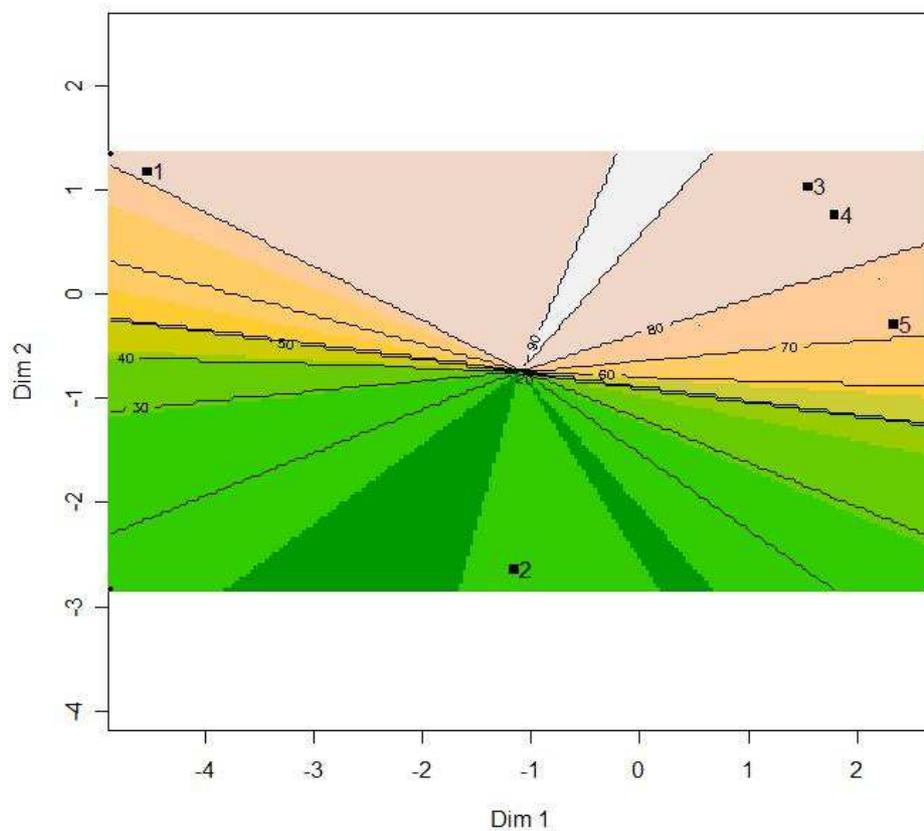


FIGURA 15: Mapa de preferência externo - Método 2 aplicado a dados do ensino a distância

de melhorá-lo para competir.

Inferir conclusões baseadas nos resultados do terceiro método traz maior confiabilidade, pois espera-se que a análise de Procrustes tenha retirado os vieses existentes não apenas nas covariáveis explicativas como também nas notas hedônicas. No terceiro método, os produtos são assumidos como iguais mais dificilmente. Embora mapas de preferência não sejam testes estatísticos, decisões são tomadas com base nos mesmos. Portanto, pode-se fazer uma analogia com testes empíricos e considerar o terceiro método como liberal, ou seja, distingue facilmente os produtos estudados.

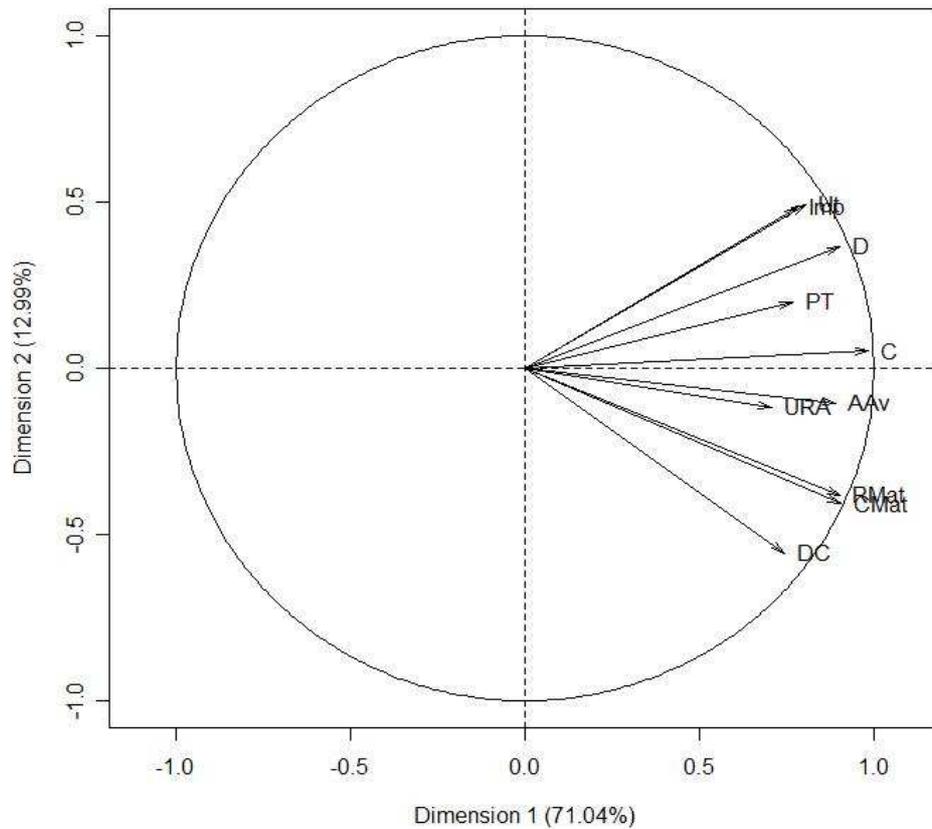


FIGURA 16: Análise de componentes principais - Método 3 aplicado a dados do ensino a distância

Por outro lado, se o objetivo da pesquisa sensorial for afirmar que produtos são distintos quando se tem bastante certeza (o que equivaleria a um teste rigoroso), o segundo método é o mais indicado, pois ele propicia regiões ou faixas de aceitação mais largas.

Os mesmos métodos foram aplicados para os dados do Experimento 2 (Palmito tratado) e visualmente ajustados pelo modelo circular. O primeiro método identificou apenas 1 cluster, direcionando preferências bem distintas entre os entrevistados (Figura 19).

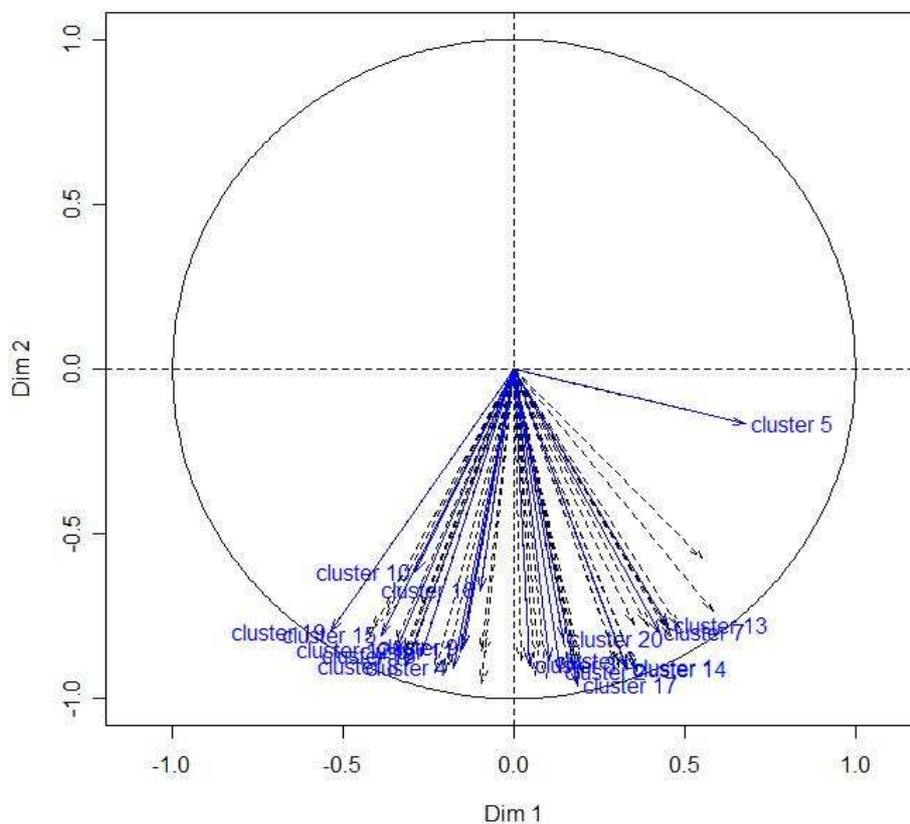


FIGURA 17: Grupos de consumo - Método 3 aplicado a dados do ensino a distância

A análise de componentes principais apresentou 62,57% de variabilidade para o primeiro componente e 27,89% para o segundo, totalizando uma variabilidade de 90,46%, apresentando direções semelhantes para as covariáveis odor e sabor e distinção destas para a aparência. Isso indica que quando sabor e odor são agradáveis a aparência não será (Figura 20).

De acordo com o Teste F da análise de variância não há diferença estatística entre o tratamentos. O mapa apresentado sugere que o palmito no tempo 1 tende a ser melhor do que os demais que aparecem iguais entre si. Esta conclusão é satisfatória para o experimento, visto que o que se

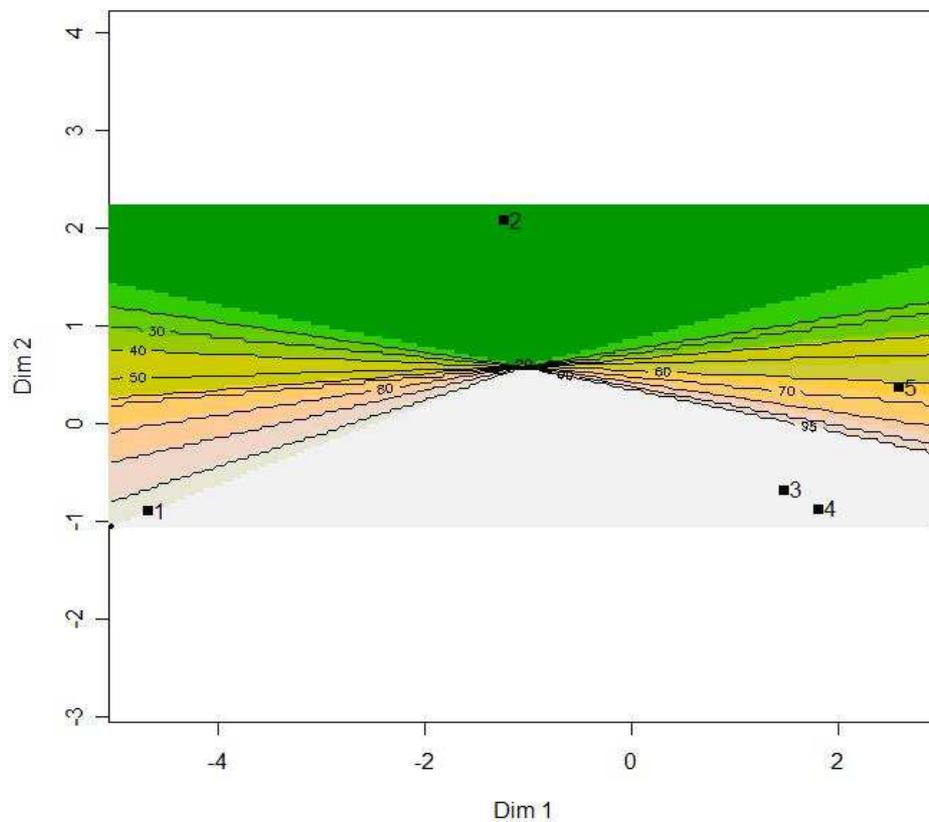


FIGURA 18: Mapa de preferência externo - Método 3 aplicado a dados do ensino a distância

deseja é que o palmito permaneça agradável para consumo por mais tempo (Figura 21).

No segundo método, também identificou apenas 1 cluster (Figura 22) e ainda com preferências bem distintas entre os consumidores. O comportamento das covariáveis apresentou resultado distinto do primeiro método: a aparência, o odor e os sabores estão em direções opostas (Figura 23). A variabilidade explicada foi de 77,21%, sendo 44,34% no primeiro componente e 32,87% no segundo. No mapa gerado, o tempo 1 ainda apresenta-se como o preferido, seguido pelos tempos 2 e 4, 3 e 7, 5 e 6,

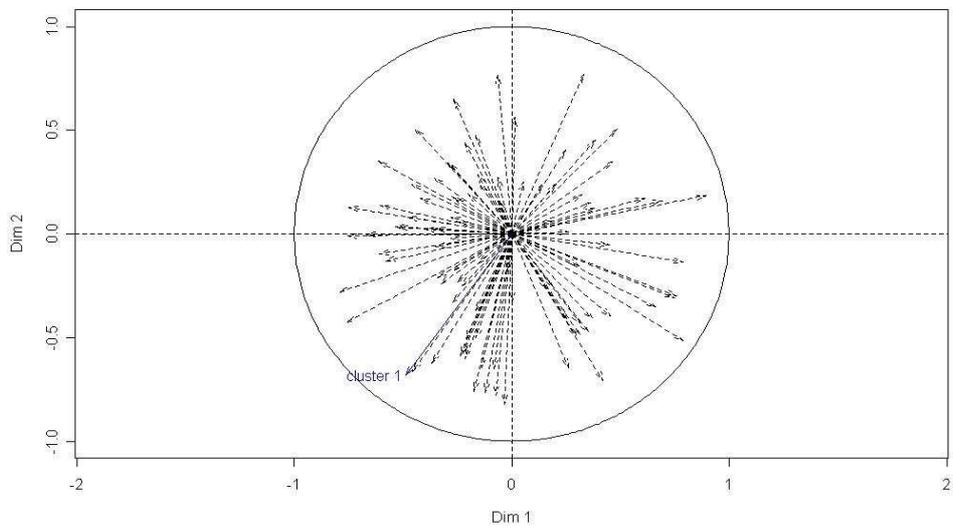


FIGURA 19: Grupos de consumo - Método 1 aplicado a dados de palmitos tratados

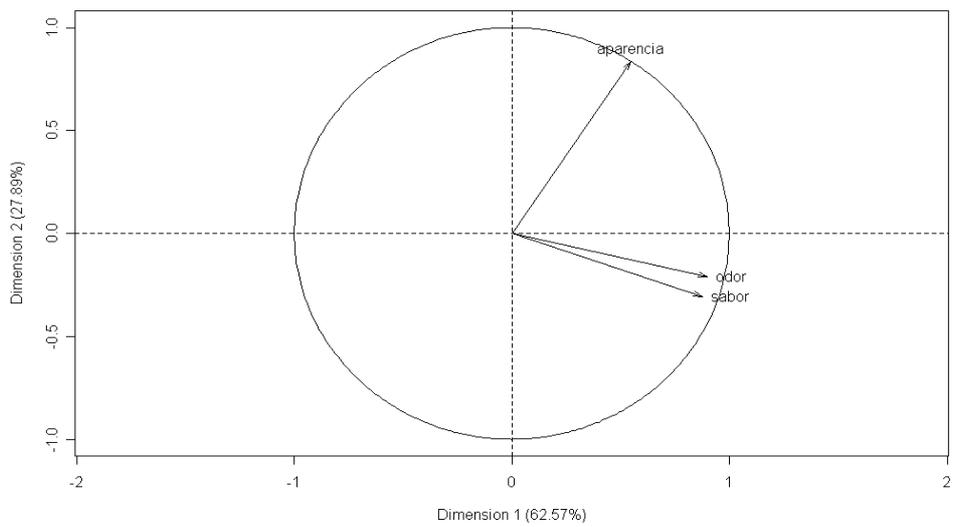


FIGURA 20: Análise de componentes principais - Método 1 aplicado a dados de palmitos tratados

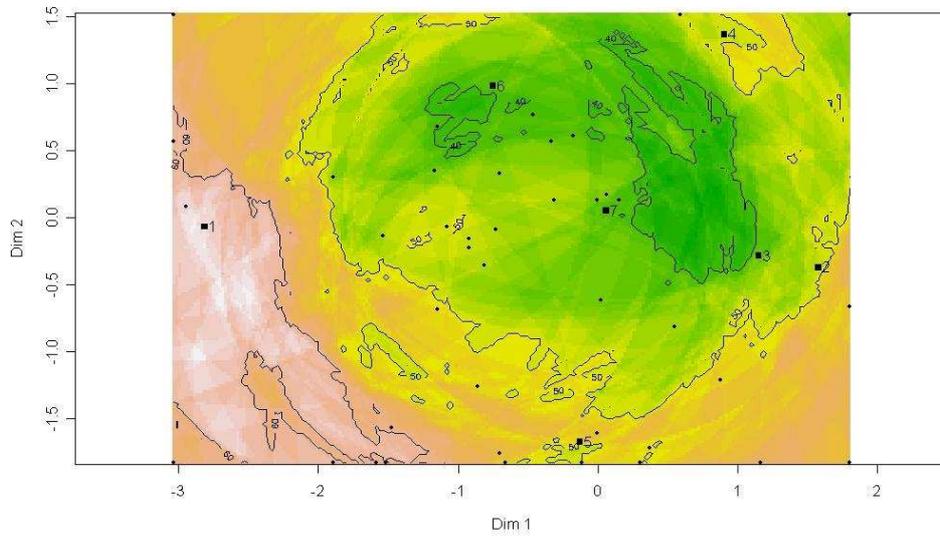


FIGURA 21: Mapa de preferência externo - Método 1 aplicado a dados de palmitos tratados

embora estes estejam bastante próximos entre si (Figura 24).

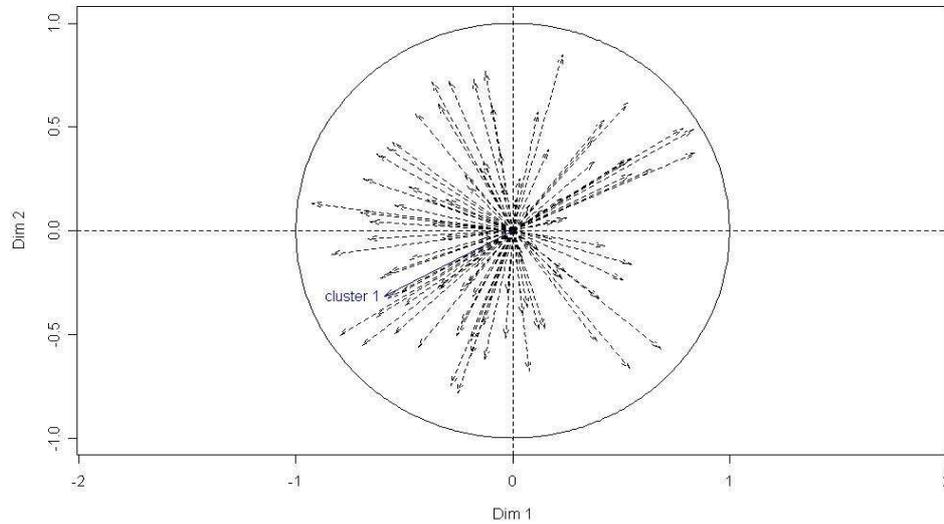


FIGURA 22: Análise de componentes principais - Método 2 aplicado a dados de palmitos tratados

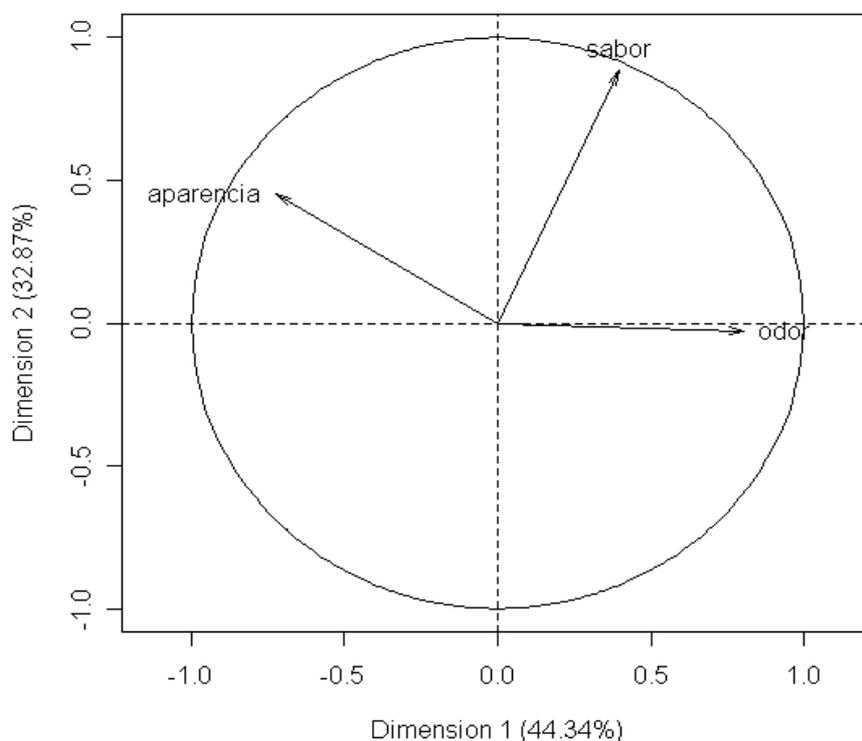


FIGURA 23: Grupos de consumo - Método 2 aplicado a dados de palmitos tratados

A terceira metodologia, quando aplicada aos dados de palmito tratado, permite a identificação de 13 clusters (Figura 25), o que é um ponto a favor desta metodologia, visto a possibilidade de identificação de grupos distintos de preferência, o que não ocorreu nos dois primeiros métodos.

A variabilidade total explicada é de 69,52% (Figura 26). As covariáveis odor, aparência e sabor apresentam-se distintas entre si e indicam que a preferência pelo sabor é inversa à preferência pelo odor ou aparência, remetendo ao fato de que quando o sabor é agradável, provavelmente, o odor e a aparência não serão.

O mapa gerado apresenta uma superfície de resposta com interpre-

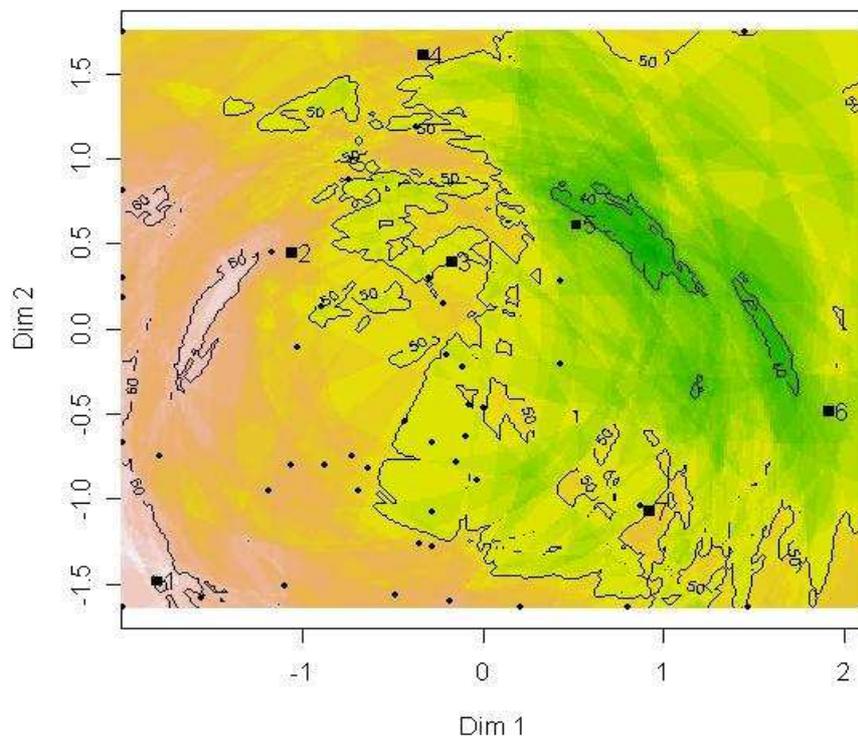


FIGURA 24: Mapa de preferência externo - Método 2 aplicado a dados de palmitos tratados

tação mais fácil, ou seja, melhor ajustada. A interpretação sugere que os tempos preferidos são 1 e 2, seguidos por 4, 5, 7, 3 e 6, cabendo inferir que o antioxidante passa a ter seu efeito diminuído com o passar do tempo (Figura 27).

Enfatiza-se que a opção pelo melhor método dependerá do grau de precisão que se busca, assim como o objetivo e os recursos disponíveis para a pesquisa. Ressalta-se que o método 1 possui maior variabilidade do que o método 2, que possui maior do que o método 3. Portanto, a confiabilidade e a rigosidade do método 3 são maiores do que as do método 2, que é maior do que a do método 1, como ficou demonstrado.

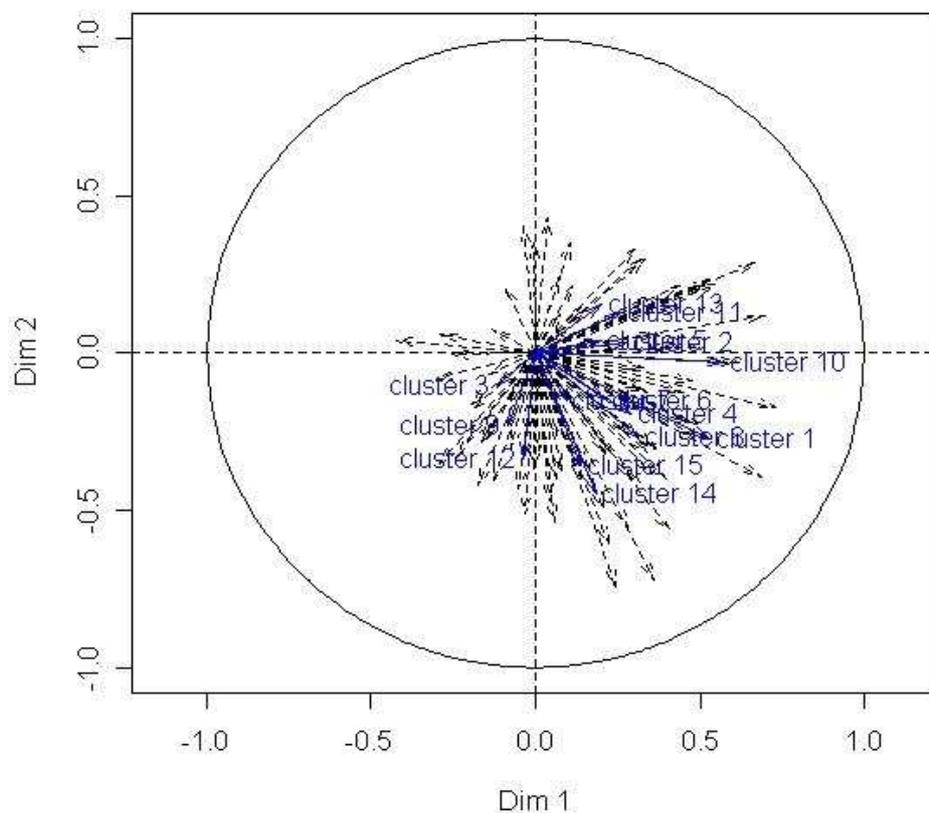


FIGURA 25: Análise de componentes principais - Método 3 aplicado a dados de palmitos tratados

4.2 Mapas de preferência com valores faltantes

No Experimento 3 (doce de leite), considerando-se apenas os dados completos, é possível realizar a análise de variância, em que as amostras 3 e 4 não diferem entre si, pelo teste Tukey e são melhores do que as demais, que são iguais entre si. As amostras 3 e 4 obtiveram médias 8,40 e 7,31, respectivamente, podendo ser classificadas como “gostei moderadamente” e “gostei muito“. As demais amostras: 1, 2, 5 e 6 receberam médias 5,79, 6,01, 5,43 e 5,61, respectivamente, o que classifica este grupo como

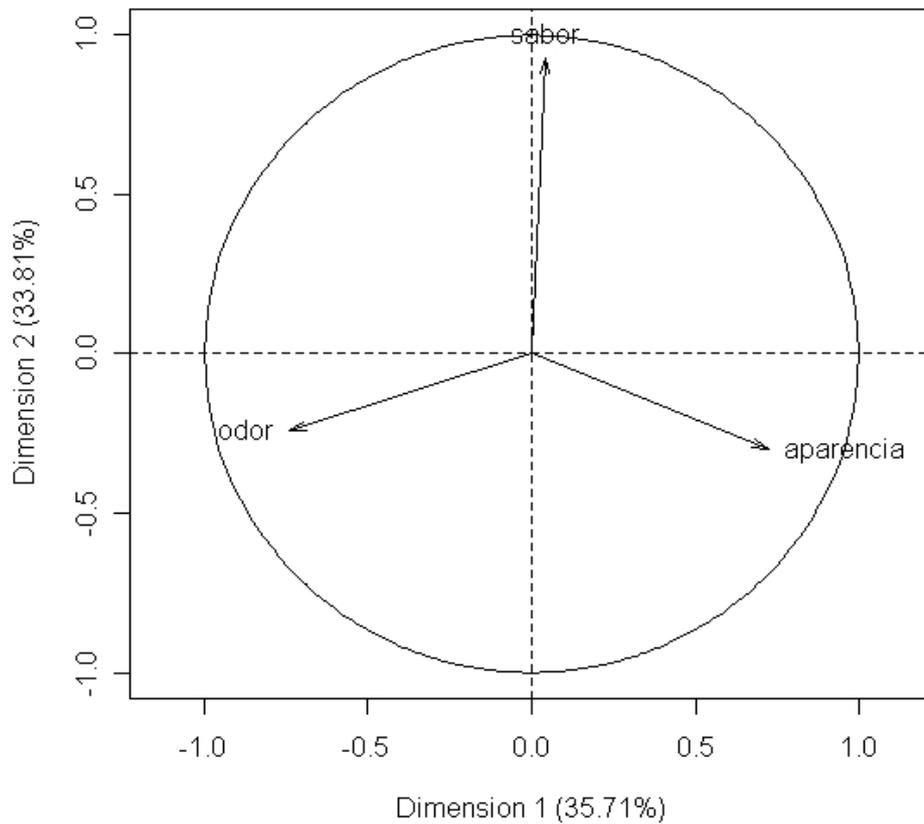


FIGURA 26: Grupos de consumo - Método 3 aplicado a dados de palmitos tratados

”indiferente“ ou “gostei ligeiramente”.

Com os dados completos (106 consumidores), realizou-se a análise de componentes principais (Figura 28), acompanhada da identificação dos clusters (Figura 29) e, conseqüentemente, o mapa de preferência externo (Figura 30). Com estes dados, a variabilidade dos dados foi explicada por 65,73%, sendo 46,33% pelo primeiro componente. O mapa concorda com a análise de variância, apontando os produtos 3 e 4 como os preferidos pelos consumidores e os demais inferiores e iguais entre si.

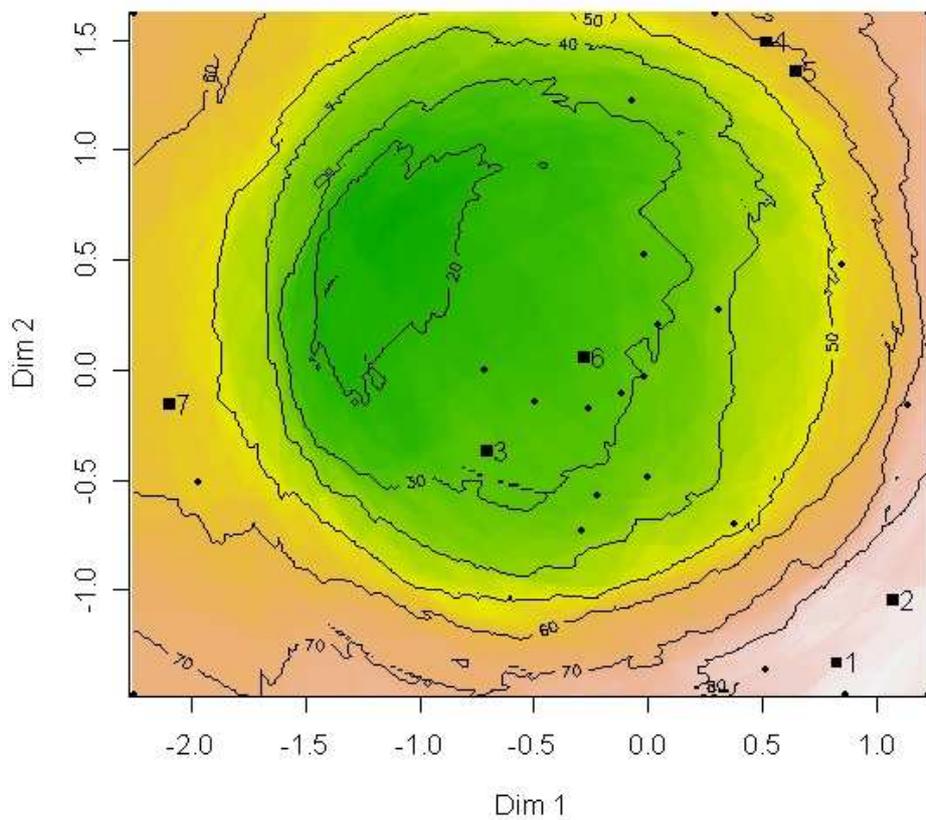


FIGURA 27: Mapa de preferência externo - Método 3 aplicado a dados de palmitos tratados

Ressalta-se que, neste estudo, não há covariáveis, portanto, utilizaram-se os dados hedônicos tanto para a construção dos eixos quanto para a construção dos mapas, representando a direção de aceitação dos consumidores (Figura 31), em que a ponta de cada seta pode ser entendida como a nota atribuída de cada consumidor, representando, assim, a direção de aceitação da maioria.

Ao considerarem-se todos os dados obtidos na pesquisa (147 consumidores, dos quais 41 não provaram todas as amostras), após aplicar o algoritmo NIPALS, tem-se a análise de componentes principais (Figura 32)

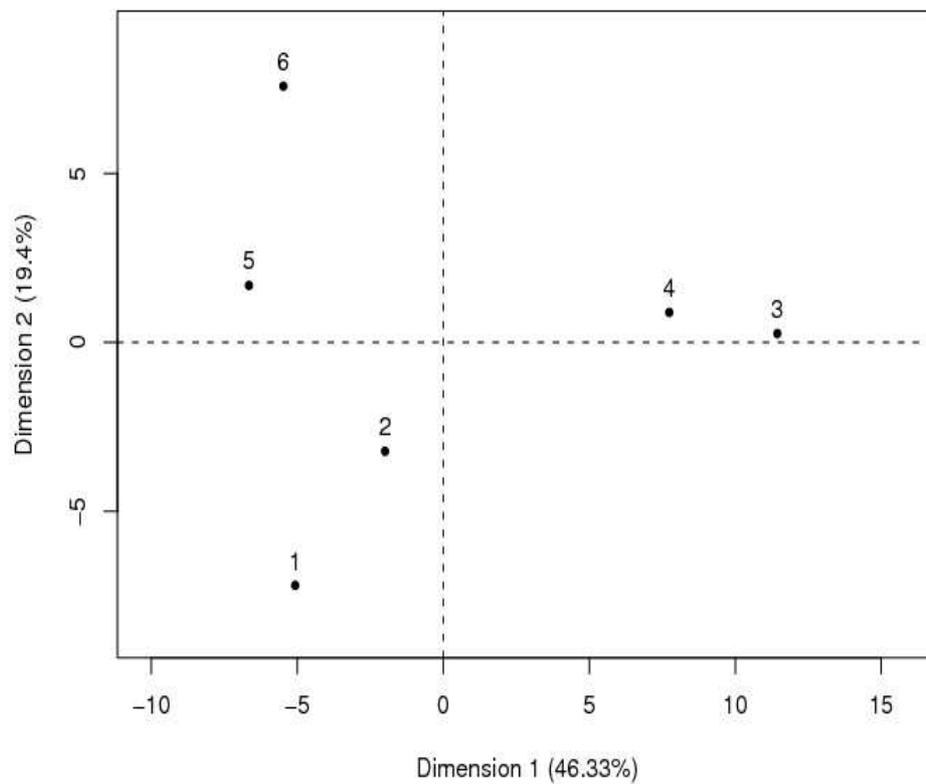


FIGURA 28: Análise de componentes principais com preferências individuais referente à preferência por doce de leite, sem valores faltantes

com 100% da variabilidade, sendo explicada pelos dois primeiros componentes principais. Esta altíssima explicação pode ser atribuída ao fato de que os eixos são formados por dados hedônicos e não por covariáveis explicativas. Há, ainda, a obtenção de clusters (Figura 33) e do mapa de preferência individual dos consumidores (Figura 34).

Por meio dos mapas pode-se confirmar a conclusão obtida anteriormente, no que se refere ao produtos preferidos (Figura 35). No entanto entre as demais destaca-se como a de menor preferência a amostra 5. Assim, há coerência entre os mapas e ainda uma boa explicação obtida pelo

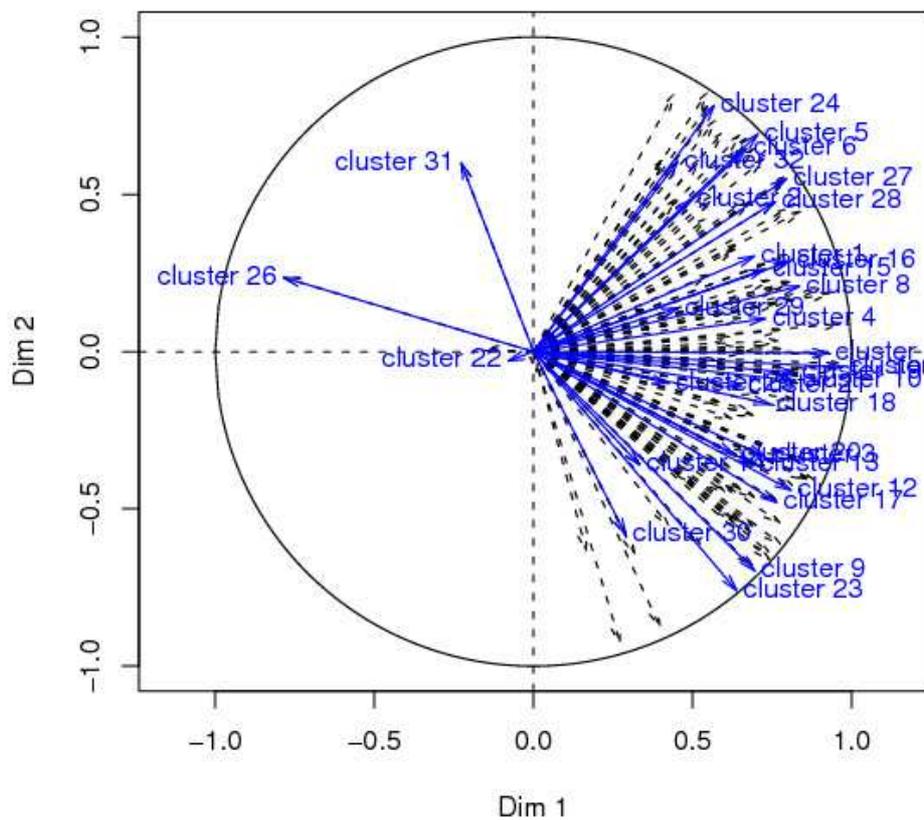


FIGURA 29: Grupos de preferência de doce de leite, sem valores faltantes

mapa gerado com parcelas faltantes.

Estes resultados foram satisfatórios, pois foi possível a construção do mapa, com resultados plausíveis, com valores faltantes. Além disso o consumidor conseguiu diferenciar os produtos com certa coerência, visto que identificaram as amostras que já foram consideradas, em anos distintos, como o “melhor doce de leite de Minas Gerais” em concurso promovido por meio de notas de provadores treinados, que são as amostra 3 e 4. Além disso o mapa apresentou resultados similares aos esperados, considerando as parcelas descartadas, porém, com uma quantidade maior de informação,

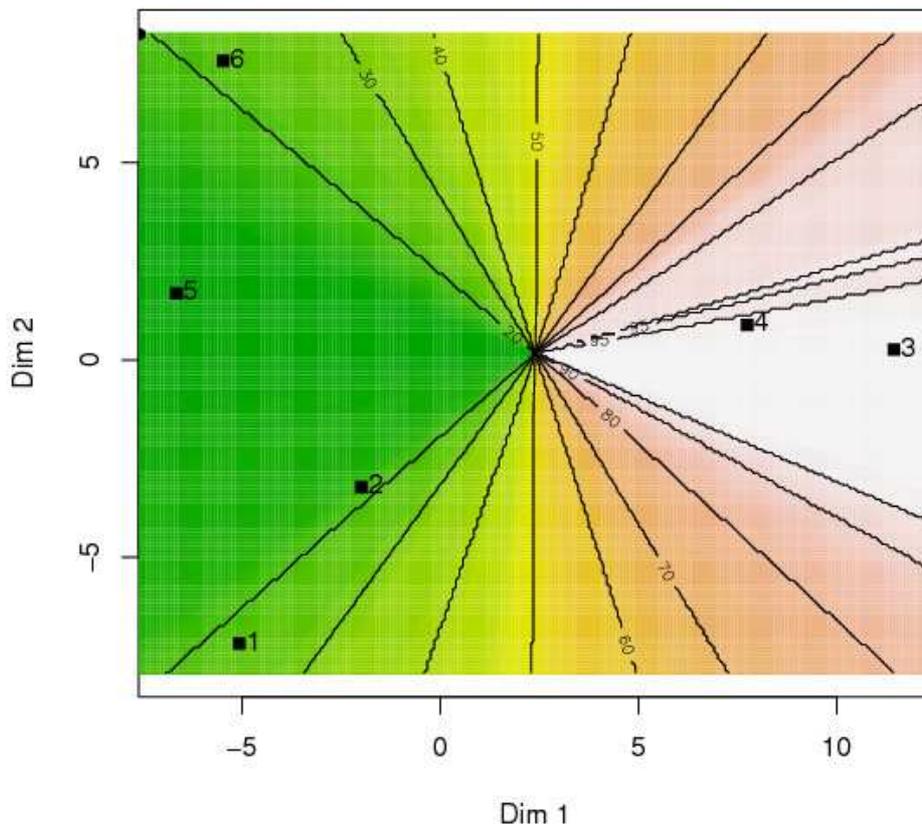


FIGURA 30: Mapa de preferência externo de doce de leite, sem valores faltantes

mais graus de liberdade, aumentando a credibilidade do resultado.

No Experimento 4 (banana tratada), a porcentagem de perda foi muito elevada. Entre os 72 provadores, 10 pontuaram apenas uma amostra e 6 provaram todas as amostras e todos os demais falharam em pelo menos uma amostra. Devido a essa pequena quantidade de dados completos, não foi possível construir o mapa de preferência pelo método usual. Com isso, trabalhou-se com 62 provadores e, por meio do algoritmo NIPALS, foi possível obter a análise de componentes principais (Figura 36), com 93,08% de variabilidade explicada, dos quais 77,69% se encontram no

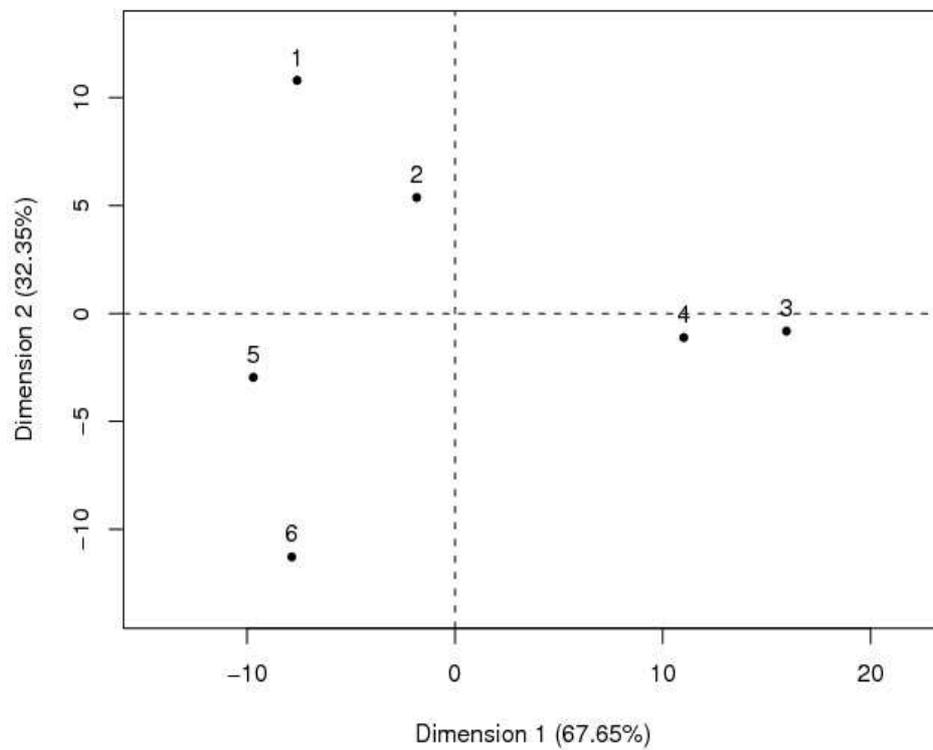


FIGURA 32: Análise de componentes principais referente à preferência por doce de leite, com valores faltantes

A proposta de utilizar o algoritmo NIPALS é, então, validada pelos exemplos apresentados. Cabem ainda alguns estudos, principalmente no que se refere à quantidade máxima de valores faltantes, assim como a influência das covariáveis explicativas.

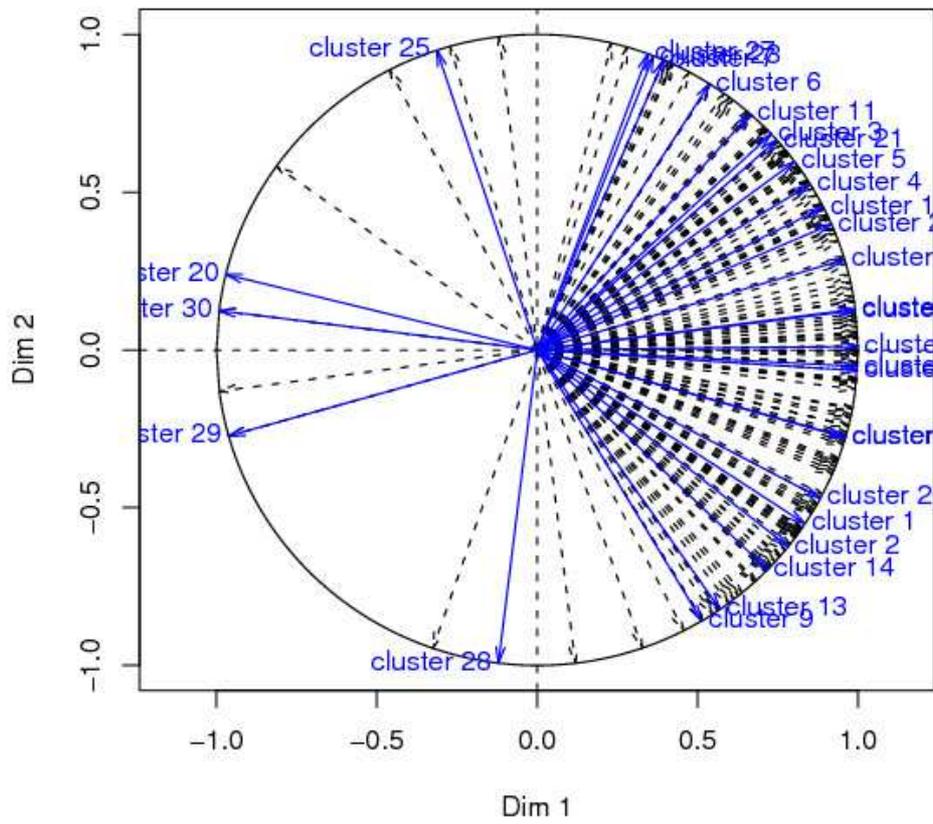


FIGURA 33: Grupos de preferência por doce de leite, com valores faltantes

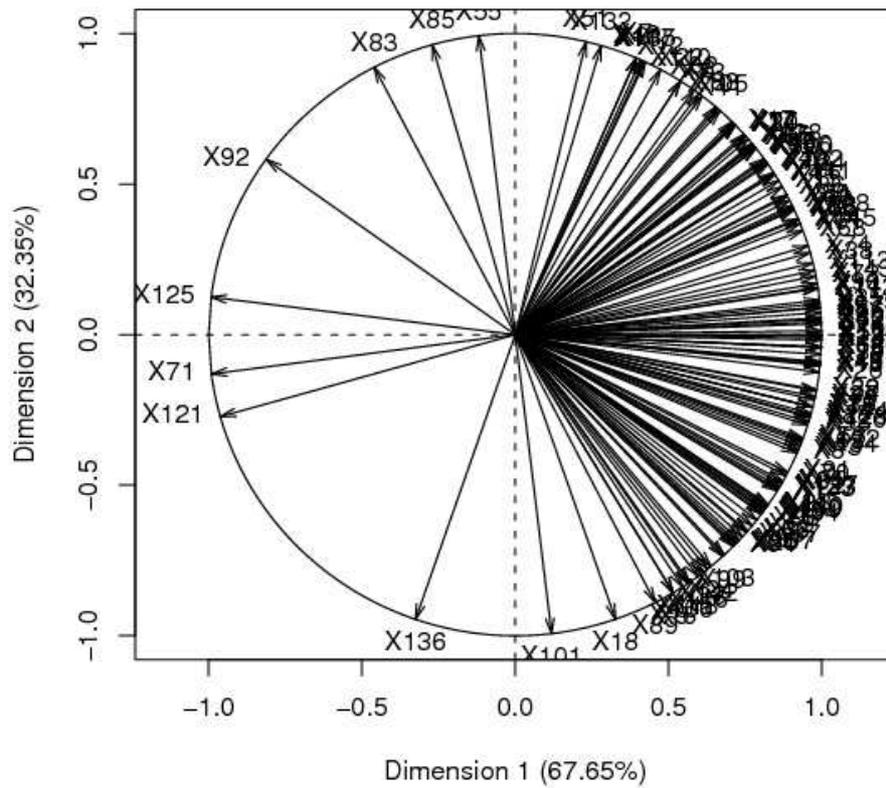


FIGURA 34: Preferência individual dos consumidores de doce de leite, com valores faltantes.

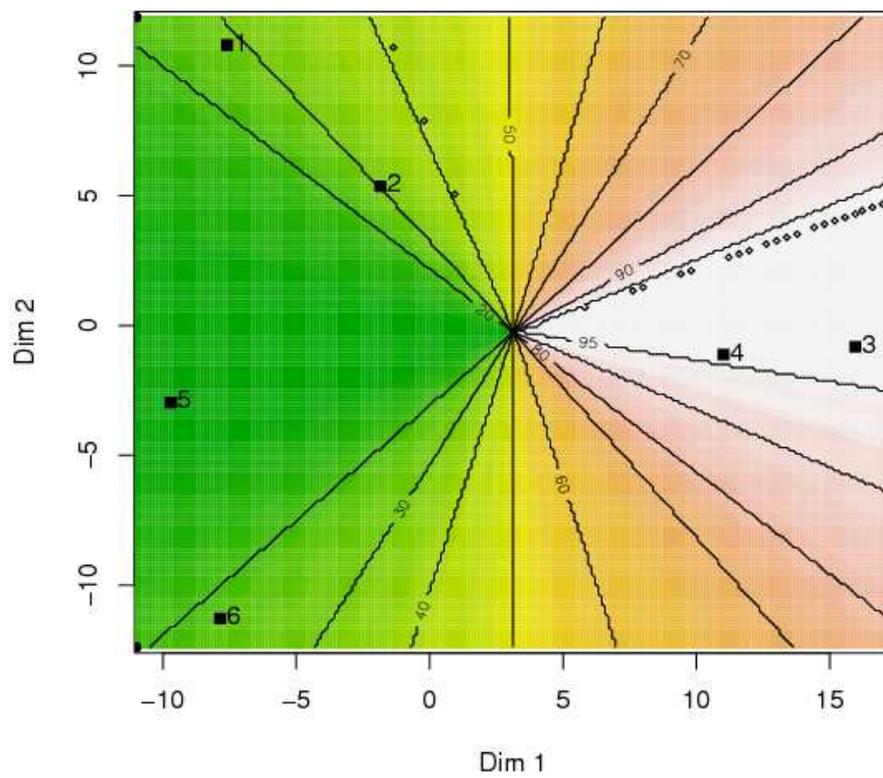


FIGURA 35: Mapa de preferência externo de doce de leite, com valores faltantes

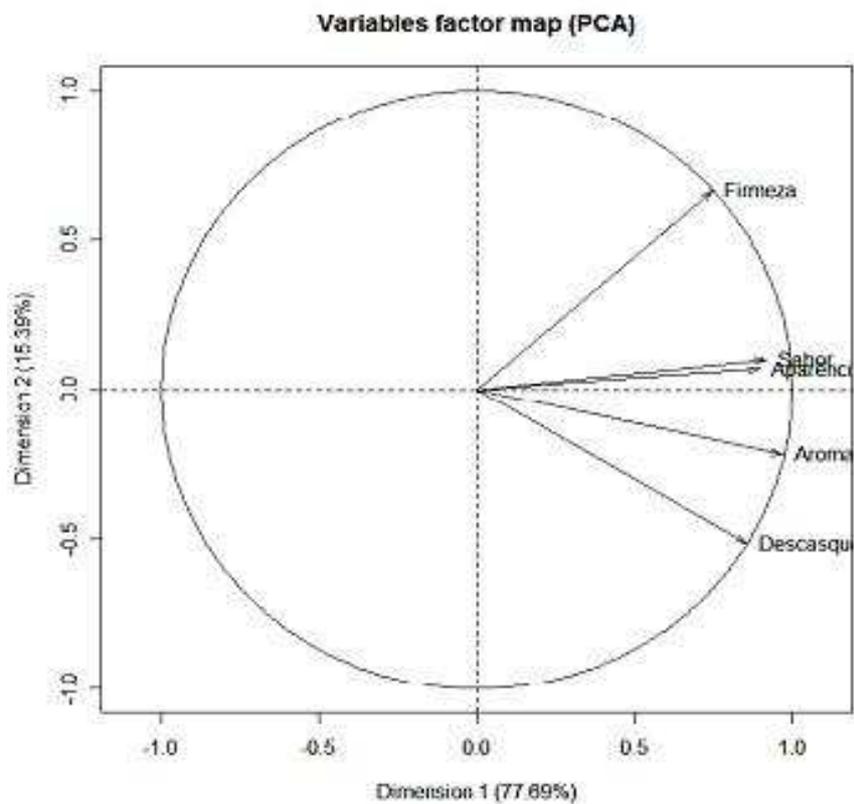


FIGURA 36: Análise de componentes principais de banana tratada, com valores faltantes

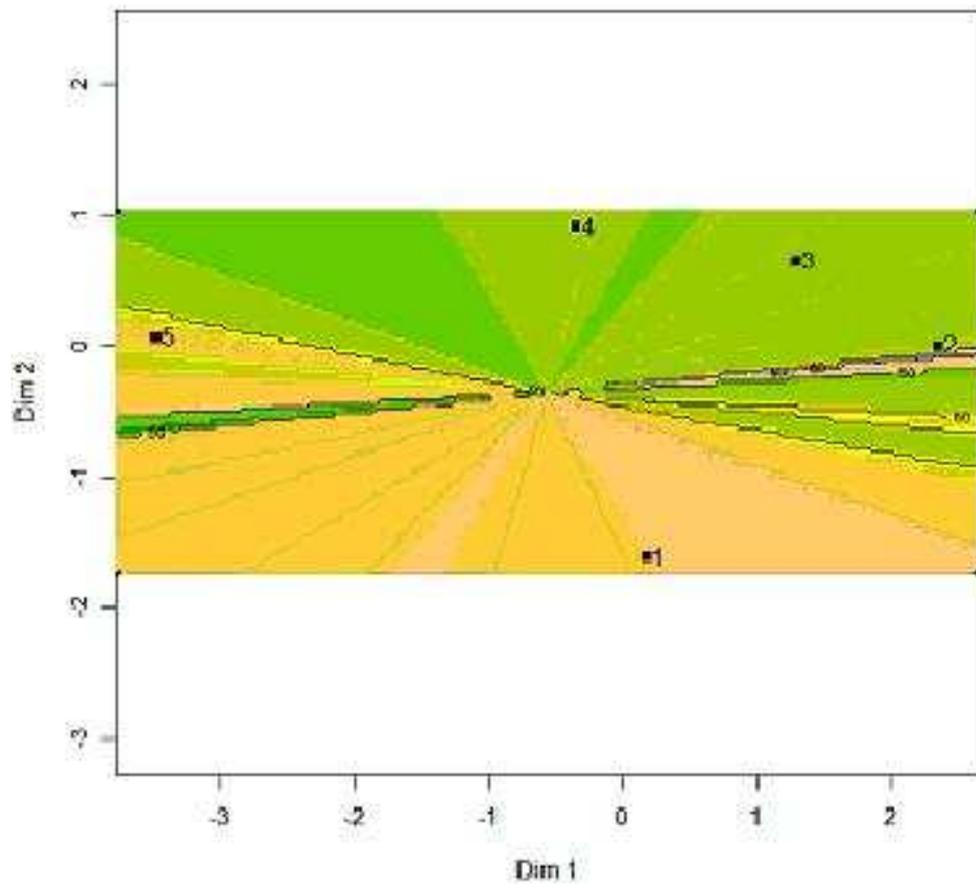


FIGURA 37: Mapa de preferência externo de banana tratada, com valores faltantes

5 CONCLUSÕES

Os propósitos, discutir a construção de mapas de preferência associado com a análise de Procrustes e construir mapas de preferência com valores faltantes, foram cumpridos com conclusões satisfatórias. Para o primeiro caso, conclui-se que a escolha da metodologia para a construção do mapa depende do objetivo do estudo. Para um maior rigor ao afirmar que produtos são diferentes, deve-se aplicar o Mapa 2 e, quando se deseja uma alto rigor ao dizer que produtos são iguais, deve-se optar pelo Mapa 3. O Mapa 1, embora muito usado, não precisar ser recomendado visto que ele é um caso particular do Mapa 2 (quando não há viéses a se retirar) e é muito útil em pesquisas com consumidores (onde esperam-se viéses).

No segundo caso, o NIPALS é um algoritmo capaz de possibilitar a construção de mapas de preferência, mesmo com valores faltantes. Portanto valores faltantes não implicam na não aplicação desta técnica.

REFERÊNCIAS

CARNEIRO, J.D.S.; REIS, R.C.; MINIM V.P.R.; CARNEIRO, J.C.S.; REGAZZI, A.J.; COIMBRA, J.S.R.; MINIM, L.A. Avaliação sensorial e mapa de preferência interno de marcas comerciais de refrigerante sabor guaraná. **Boletim do CEPPA**, v. 21, n. 2, p. 279-292, 2003.

CHRISTOFFERSSON, A. **The one component model with incomplete data**. 1970. Tese (Doutorado em Quimiometria) - Uppsala University, Uppsala, Suécia.

DAIROU, V.; SIEFFERMANN, J.; PRIEZ, A.; DANZART, M. Sensory evaluation of car brake systems - the use of flash profile as a preliminary study before a convention profile. In: **SAE WORLD CONGRESS**. 2003, Detroit, MI. Disponível em: <http://www.perception-sensorielle.com/Congress2003/Dairou_SAE.pdf>. Acesso em: 2 ago. 2004.

DRAY, S.; PETTORELLI, N.; CHESSEL, D. Multivariate analysis of incomplete mapped data. **Transactions in GIS**, v. 7, n. 3, p. 411-422, 2003.

FAVA, A.R. Farmaceutica analisa nuances do perfume. **Jornal da Unicamp**, n. 212, p. 8, maio, 2003.

FERREIRA, D.F. **Análise multivariada**. Lavras: UFLA/DEX, 1996. p. 223-250. Apostila.

FERREIRA, D. F. Análises estatísticas por meio do Sisvar para Windows versão 4. 0.. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 45., 2000, São Carlos. **Programas e Resumos...** São Carlos, SP: UFSCar, 2000. p. 255-258. Software.

FERREIRA, S. M. R.; FREITAS, R. J. S.; BASSLER, T. C. Terminologia descritiva para análise sensorial de tomate de mesa. **Visão Acadêmica**, v. 4, n. 1, p. 7-12, 2003.

FERREIRA, E. B. **Análise generalizada de Procrustes via R: uma aplicação em laticínios.** 2004. 216 p. Dissertação (Mestrado em Agronomia - Área de concentração: Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.

FERREIRA FILHO, S.S.; ALVES, R. Técnicas de avaliação de gosto e odor em águas de abastecimento: método analítico, análise sensorial e percepção dos consumidores. **Engenharia Sanitária Ambiental**, v. 11, n. 4, p. 362-370, 2006.

FERREIRA, E. B. **Some topics in procrustes analysis applied to sensometrics.** 2007a. 131 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.

FERREIRA, D.F. **Estatística multivariada.** Lavras: UFLA, 2007b. 642 p. No prelo.

GALVÃO, G. Tecnologia: os 5 sentidos na construção do automóvel. Adaptado: Renault. **Motor Clube**, 13 out. 2007. Disponível em: <<http://www.motorclube.com.br/materias/tecnologia-os-cinco-sentidos-na-construcao-do-automovel.aspx>>. Acesso em: 19 out. 2007.

GARCINDO, L. A. M. **Uma proposta de gestão de cursos a distância baseada em análise sensorial e um estudo de caso.** 2007. 88 p. Monografia (Especialização em Educação) - Universidade Federal de Lavras, Lavras, MG.

GOOD, I. J. Some applications of the singular decomposition of a matrix. **Technometrics**, v. 11, p. 823-831, 1969.

GOWER, J. C. Statistical methods for comparing different multivariate analysis of the same data. In: HODSON, J.R.; KENDALL, D.G.; TAUTU, P. (Ed.) **Mathematics in the archaeological and historical sciences.** Edingurg: Edinburgh University, 1971.

GOWER, J. C. Generalised procrustes analysis. **Psychometrika**, v. 40, n. 1, p. 33-51, 1975.

GOWER, J. C.; DIJKSTERHUIS, G. B. **Procrustes problems.** Oxford: Oxford University, 2004. 256 p.

GRUNG, B.; MANNE, R. Missing value in principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, n. 42, p. 125-139, 1998.

GUINARD, J.X.; UOTANI,B.; SCHLICH, P. Internal and external mapping of preferences for commercial lager beers: comparison of hedonic ratings by consumers blind versus with knowledge of brand and price. **Food Quality and Preference**, n. 12, p. 243-255, 2001.

HURLEY, J. R.; CATTLE, R. B. The procrustes program: producing direct rotation to test a hypothesized factor structure. **Behavioral Science**, v. 7, p. 258-262, 1962.

HUSSON, F.; LÊ, S. SensomineR: a package for sensory data analysis with r. use r! **Science aliments**. v. 26, n. 4, p. 355-356, 2006.

JONES, P.N., MACFIE, H. J. H., BEILKEN, S.L. **Journal of Science Food Agriculture**, n. 47, p. 113-123, 1989.

KRISTOF, W.; WINGERSKY, B. Generalization of the orthogonal Procrustes rotation procedure for more than two matrices. In: ANNUAL CONVENTION OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION, 79., 1971. **Proceedings...**, AMERICAN PSYCHOLOGICAL ASSOCIATION, 1971. v. 6, p. 89-90.

MEAD, R.; CURNOW, R. N. **Statistical methods in agriculture and experimental biology** London: Chapman and Hall, 1983. 335 p.

MINIM, V. P. R. **Análise Sensorial : estudos com consumidores**. Viçosa: Universidade Federal de Viçosa, 2006. 225 p.

MORI, E. E. M.; YOTSUYANAGI, K.; FERREIRA, V. L. F. Análise sensorial de goiabadas de marcas comerciais. **Ciência e Tecnologia de Alimentos**, Campinas, v. 18, n. 1, 1998. Disponível em: <<http://www.scielo.br/scielo.php>>. Acesso em: 17 jun 2007.

NUTE, G. R.; MACFIE, H. J. H.; GREENHOFF, K. **Practical application in preference mapping**. London: Elsevier Science, 1989. (Food Acceptability).

OLIVEIRA, A.P.V de; FRASSON, K.; ALMEIDA T.C.A; BENASSI M.T. Aceitação de sobremesas lácteas dietéticas e formuladas com açúcar: Teste afetivo e mapa de preferência interno. **Ciência e Tecnologia de Alimentos**, v. 24, n. 4, p. 627-633, 2004.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**, v. 2, n. 6, p. 559-572, 1901.

- PEAY, E. R. Multidimensional rotation and scaling of configurations to optimal agreement. **Psychometrika**, v. 53, p. 199-208, 1988.
- PINHEIRO, A. C. M. **Pós-colheita de bananas maçã submetidas ao 1-MCP**. 2007. Tese (Doutorado em Ciência dos Alimentos) - Universidade Federal de Lavras, Lavras, MG.
- R DEVELOPMENT CORE TEAM. **R**: a language and environment for Statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2007.
- SCHLICH, P. Preference mapping: Relating consumer preferences to sensory or instrumental measurements. **Bioflavour**, v. 95, n. 75, p. 135-150, 1995.
- SILVA, A. F.; MINIM V.P.R.; RIBEIRO, M.M. Análise sensorial de diferentes marcas comerciais de café (*coffea arabica l.*) orgânico. **Ciência e agrotecnologia**, Lavras, v. 29, n. 6, p. 1224-1230, 2005.
- TEN BERGE, J. M. F. Orthogonal procrustes rotation for two or more matrices. **Psychometrika**, v. 42, p. 267-276, 1977.
- TEN BERGE, J. M. F.; KNOL, D. L. Orthogonal rotations to maximal agreement for two or more matrices of different column orders. **Psychometrika**, v. 49, p. 49-55, 1984.
- TENENHAUS, M. **La régression PLS: théorie et pratique**. Paris: Technip, 1998.
- THOMPSON, J.L; DRAKE, M.A.; LOPETCHARAT,K.; YATES, M.D. Preference mapping of commercial chocolate milks. **Journal of Food Science**, v. 9, p. 406-413, 2004.
- XIONG, R.; MEULLENET, J. F. Application of multivariate adaptive regression splines (mars) to the preference mapping of cheese sticks. **Journal of Food Science**, v. 69, n. 4, p. 131-139, 2004.
- YOUNG, N.D; DRAKE, M.A.; LOPETCHARAT,K.; McDANIEL, M.R. Preference mapping of cheddar cheese with varying maturity levels. **Jornal of Dairy Science**, n. 87, p. 11-19, 2004.