



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA DE ALIMENTOS  
DEPARTAMENTO DE PLANEJAMENTO ALIMENTAR  
E NUTRIÇÃO

*Desenvolvimento e Avaliação do Desempenho  
de uma “Escala Hedônica com Referência”  
para Utilização em Teste de Consumidor*

*Selma de Freitas Bergara*

Engenheira de Alimentos

*Profª Drª Maria Aparecida A. P. da Silva*

Orientadora

PARECER

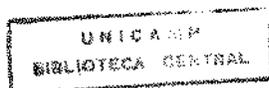
Este exemplar corresponde à redação final da tese defendida por Selma de Freitas Bergara, aprovada pela Comissão Julgadora em 02 de setembro de 1999.

Campinas, 02 de setembro de 1999.

  
Profª Dra Maria Aparecida  
Azevedo Pereira da Silva  
Presidente da Banca

Tese apresentada à Faculdade de Engenharia de Alimentos da Universidade Estadual de Campinas, para obtenção do título de Mestre em Ciência da Nutrição Aplicada a Tecnologia de Alimentos

Campinas – São Paulo  
Julho - 1999



UNIDADE	BC
N.º CHAMADA:	
V.	Ex.
TOMBO BC/	3922A
PROC.	229199
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	23/10/99
N.º CPD	

CM-00136441-1

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DA F.E.A. - UNICAMP

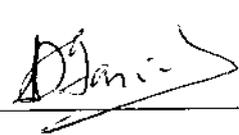
B452d Bergara, Selma de Freitas  
Desenvolvimento e avaliação do desempenho de uma  
"Escala Hedônica com Referência" para utilização em teste  
de consumidor / Selma de Freitas Bergara. - Campinas, SP  
[s.n.], 1999.

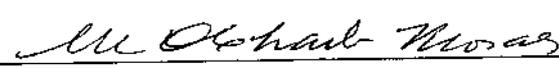
Orientador: Maria Aparecida Azevedo Pereira da Silva  
Dissertação (mestrado) - Universidade Estadual de  
Campinas.Faculdade de Engenharia de Alimentos.

1. Consumidores. 2. Avaliação sensorial. I. Silva, Maria  
Aparecida Azevedo Pereira da. II. Universidade Estadual de  
Campinas.Faculdade de Engenharia de Alimentos. III. Título.

## Membros da Banca Examinadora

  
\_\_\_\_\_  
Profª Drª Maria Aparecida Azevedo Pereira da Silva  
(Orientador)

  
\_\_\_\_\_  
Prof. Dr. Belmer Garcia Negrillo  
(Membro)

  
\_\_\_\_\_  
Profª Drª Maria Amélia Chaib Moraes  
(Membro)

\_\_\_\_\_  
Profª Drª Hilary Castle de Menezes  
(Membro)

Campinas, \_\_\_\_ de \_\_\_\_\_ de 1999

*"Amo ao Senhor, porque Ele ouviu a minha voz e a  
minha súplica, porque inclinou para mim os Seus ouvidos;  
portanto invocá-Lo-ei enquanto viver.  
Que darei eu ao Senhor, por todos os benefícios que  
me tem feito?  
Oferecer-Te-ei sacrifícios de louvor e invocarei o  
nome do Senhor.  
Salmos 116: 1, 12, 17*

*Dedico:  
A Deus, meu Criador e Senhor, único digno de toda  
honra, glória, louvor e poder.*

*Meus especiais agradecimentos:*

*Aos meus pais, Rubens e Damaris, pelo amor,  
carinho, paciência, apoio constante e exemplos de vida;*

*Ao meu irmão Rogério e minha cunhada Eliana, por  
suas companhias, orações e pela contribuição para o meu  
crescimento espiritual;*

*À minha irmã Ana Virginia e meu cunhado Rafael,  
pelo apoio, carinho e pelo empréstimo do computador;*

*Ao meu irmão Márcio, pelo incentivo.*

## AGRADECIMENTOS

À Prof<sup>a</sup> Dr<sup>a</sup> Maria Aparecida (Cida), pela orientação caracterizada por boa dose de objetividade e bom senso, paciência, disponibilidade e dedicação na realização deste trabalho; por saber respeitar o meu ritmo de trabalho, tão diferente ao seu.

Ao CNPq pela concessão da bolsa de estudos.

À minha ex-professora Marta Benassi e à Dr<sup>a</sup> Mércia Betini, pelo incentivo ao ingresso no mestrado.

Aos professores Belmer, Maria Amélia e Hilary pelas sugestões na ocasião da qualificação e/ou correção do "boneco".

À Citrosuco Paulista S.A., na pessoa do Sr. Marcos Valim, pela doação do suco de laranja concentrado congelado.

À Citrovita Agro Industrial Ltda., nas pessoas de Ivan Marie, José Orlando e Mauro Ricardo, pelas informações e suco de laranja cedidos.

À Prof<sup>a</sup> Jane, por permitir o uso do Laboratório de Análise Sensorial do Colégio Técnico da UNICAMP, COTUCA.

À Prof<sup>a</sup> Dr<sup>a</sup> Gabriela Stangenhau, ao José Ramos e à Nilda pela valiosa assistência na área de estatística, que resultou no enriquecimento deste trabalho.

Aos amigos do Laboratório de Análise Sensorial - FEA, Luciane, Regina, Jorge, Lúcio e, mais uma vez, à Nilda, pela troca de experiências e companheirismo.

Ao Chico, pelo auxílio técnico, cafézinhos e pelo carinho.

Aos funcionários do DEPAN, Rose e Soely, pelas sugestões e empréstimos; Bete, Cidinha, Valter e Serginho pelos serviços prestados.

Aos funcionários Margarete e Cosme (Secretaria de Pós-graduação), Toninho, Geraldo, Creusa, Cláudia e Mara (Biblioteca) e Fátima (Microcentro de Informática), pelos serviços prestados.

Ao pessoal (alunos e funcionários) do Departamento de Tecnologia de Alimentos pela convivência nos momentos festivos.

À Nice, pelos serviços prestados, companhia, palavras de conforto e estímulo; pela amizade.

Aos amigos Gabriela, Alessandra (Berta), Alessandra Lopes, Eduardo e, mais uma vez, à Luciane, por muitas vezes terem colocado a mão na massa, ajudando-me nas diversas fases deste

trabalho. Não poderia me esquecer de tantos momentos que passamos juntos durante esses dois anos e meio, nos quais compartilhamos alegrias, conquistas, festas, filosofias de vida ou mesmo as dificuldades de nossas vidas. Vocês foram muito especiais nessa minha caminhada.

Ao Mauro Ricardo, por tantas palavras de estímulo, sobretudo ao ingresso no mestrado, por sempre acreditar na minha capacidade, pela paciência em ouvir meus desabaços, por todo carinho e amizade.

Ao Afranio, por todo carinho, paciência, compreensão e companhia; pelas orações, sem as quais a fase final desse trabalho teria sido muito mais difícil; por estar contribuindo para o meu crescimento espiritual. Agradeço a Deus por tê-lo colocado na minha vida.

Aos meus tios Dionysio e Maria José, pelas orações e companhia quando mais precisei.

Às minhas irmãs em Cristo, do Coral Feminino Laudare, pela amizade e orações, as quais têm me sustentado há alguns meses.

Aos amigos, Carla, Tonhão, Carmem, Sólton, Baby, Fabiana, Joãozinho, Lica, Dilica e, mais uma vez, ao Eduardo e às Alessandras, pelos churrascos, “happy hours” e encontros; por valorizarem nossa amizade, que se iniciou há dez anos; ao Chico Assis pela amizade.

A todos aqueles que participaram como provadores dos testes sensoriais, sem os quais este trabalho não teria sentido.

Rogo a Deus as mais ricas bênçãos a todos que contribuíram direta ou indiretamente na realização deste trabalho.

## ÍNDICE GERAL

<b>ÍNDICE DE FIGURAS</b> .....	III
<b>ÍNDICE DE TABELAS</b> .....	V
<b>RESUMO GERAL</b> .....	VII
<b>GENERAL ABSTRACT</b> .....	X
<b>INTRODUÇÃO GERAL</b> .....	1
<b>CAPÍTULO I. REVISÃO BIBLIOGRÁFICA</b> .....	6
<b>1.1. O USO DE ESCALAS NA ANÁLISE SENSORIAL</b> .....	7
1.1.1. Escalas nominais .....	7
1.1.2. Escalas ordinais .....	8
1.1.3. Escalas de intervalo .....	10
1.1.4. Escalas de razão .....	11
<b>1.2. VANTAGENS, DESVANTAGENS E LIMITAÇÕES DAS PRINCIPAIS ESCALAS UTILIZADAS NA ANÁLISE SENSORIAL</b> .....	13
<b>1.3. ESCALAS QUE UTILIZAM REFERÊNCIA</b> .....	14
<b>1.4. TESTES AFETIVOS E ESCALAS UTILIZADAS EM TESTES DE ACEITAÇÃO</b> .....	15
<b>1.5. ANÁLISES ESTATÍSTICAS</b> .....	21
1.5.1. “Bootstrap” .....	22
1.5.2. Metodologia de Superfície de Resposta (RSM) .....	23
<b>1.6. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	25
<b>CAPÍTULO II - Avaliação do Desempenho de uma “Escala Hedônica com Referência” para Utilização em Teste de Consumidor</b> .....	28
<b>RESUMO</b> .....	29
<b>ABSTRACT</b> .....	31
<b>INTRODUÇÃO</b> .....	33
<b>MATERIAIS E MÉTODOS</b> .....	38
Amostras .....	38
Recrutamento dos consumidores de guaraná .....	38
Condições gerais dos testes .....	39
Consumidores .....	40
Escalas .....	40

Escolha das amostras referências .....	41
Análises estatísticas .....	42
<b>RESULTADOS E DISCUSSÕES</b> .....	46
Poder discriminativo das escalas .....	46
Diagnóstico de normalidade das distribuições das respostas .....	51
Diagnóstico de homogeneidade da variância das respostas .....	58
<b>CONCLUSÃO</b> .....	62
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	64
<b>CAPÍTULO III - Desempenho de uma “Escala Hedônica com Referência” para Obtenção de Modelos Preditivos de Aceitação a Partir de Respostas Obtidas com Consumidores</b> .....	66
<b>RESUMO</b> .....	67
<b>ABSTRACT</b> .....	68
<b>INTRODUÇÃO</b> .....	69
<b>MATERIAIS E MÉTODOS</b> .....	72
Delineamento experimental .....	72
Recrutamento dos consumidores .....	74
Teste de consumidor .....	76
Análise dos resultados .....	78
<b>RESULTADOS E DISCUSSÕES</b> .....	82
<b>CONCLUSÕES</b> .....	93
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	94
<b>CAPÍTULO IV. CONCLUSÕES GERAIS</b> .....	96
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	99

## ÍNDICE DE FIGURAS

### CAPÍTULO I - Revisão Bibliográfica

<b>Figura 1.</b> Exemplos de escala nominal .....	8
<b>Figura 2.</b> Exemplo de escala ordinal verbal .....	10
<b>Figura 3.</b> Exemplo de escala ordinal mista .....	10
<b>Figura 4.</b> Exemplo de escala de intervalo, escala não estruturada .....	11
<b>Figura 5.</b> Exemplo de escala de razão .....	12
<b>Figura 6.</b> Escala hedônica estruturada verbal de 9 pontos .....	18

### CAPÍTULO II - Avaliação do Desempenho de uma “Escala Hedônica com Referência” para Utilização em Teste de Consumidor

<b>Figura 1.</b> Questionário de recrutamento dos consumidores de refrigerante do tipo guaraná .....	39
<b>Figura 2.</b> Ficha de aplicação do teste de aceitação com escala hedônica com referência .....	41
<b>Figura 3.</b> Distribuição dos valores dos logaritmos neperianos dos níveis de significância de $F_{amostra}$ , obtidos através da simulação “bootstrap” para cada escala .....	50
<b>Figura 4.</b> Gráfico do tipo box-whisker a partir das respostas obtidas com a escala hedônica tradicional .....	52
<b>Figura 5.</b> Gráfico do tipo box-whisker a partir das respostas obtidas com a escala hedônica com referência (referência: amostra B) .....	53
<b>Figura 6.</b> Gráfico do tipo box-whisker a partir das respostas da escala hedônica com referência (referência: amostra C) .....	53
<b>Figura 7.</b> Gráfico do tipo box-whisker a partir das respostas da escala hedônica com referência (referência: amostra D) .....	54
<b>Figura 8.</b> Gráfico do tipo box-whisker a partir dos dados normalizados da escala de estimativa da magnitude .....	54
<b>Figura 9.</b> Gráfico da probabilidade normal dos resíduos referentes aos dados obtidos através da escala hedônica tradicional .....	55
<b>Figura 10.</b> Gráfico da probabilidade normal dos resíduos referentes aos dados obtidos através da escala hedônica com referência (referência: amostra B) .....	55
<b>Figura 11.</b> Gráfico da probabilidade normal dos resíduos referentes aos dados obtidos através da escala hedônica com referência (referência: amostra C) .....	56

<b>Figura 12.</b> Gráfico da probabilidade normal dos resíduos referentes aos dados obtidos através da escala hedônica com referência (referência: amostra D) .....	56
<b>Figura 13.</b> Gráfico da probabilidade normal dos resíduos referentes aos dados transformados obtidos através da escala de estimativa da magnitude .....	57
<b>CAPÍTULO III - Desempenho de uma “Escala Hedônica com Referência” para Obtenção de Modelos Preditivos de Aceitação a partir de Respostas obtidas com Consumidores</b>	
<b>Figura 1.</b> Questionário de recrutamento dos consumidores de suco de laranja .....	75
<b>Figura 2.</b> Ficha de aplicação do teste de aceitação com escala hedônica com referência .....	78
<b>Figura 3.</b> Gráfico de probabilidade normal dos resíduos, considerando o modelo ajustado obtido com a escala hedônica .....	87
<b>Figura 4.</b> Gráfico de probabilidade normal dos resíduos, considerando o modelo ajustado obtido com a escala hedônica com referência .....	88
<b>Figura 5.</b> Gráfico dos resíduos em função da variável independente Brix a partir dos dados obtidos com a escala hedônica .....	88
<b>Figura 6.</b> Gráfico dos resíduos em função da variável independente Brix a partir dos dados obtidos com a escala hedônica com referência .....	89
<b>Figura 7.</b> Gráfico dos resíduos em função da variável independente acidez a partir dos dados obtidos com a escala hedônica .....	89
<b>Figura 8.</b> Gráfico dos resíduos em função da variável independente acidez a partir dos dados obtidos com a escala hedônica com referência .....	90
<b>Figura 9.</b> Superfície de resposta gerada do modelo ajustado decodificado a partir dos dados obtidos com a escala hedônica .....	92
<b>Figura 10.</b> Superfície de resposta gerada do modelo ajustado decodificado a partir dos dados obtidos com a escala hedônica com referência .....	92

## ÍNDICE DE TABELAS

### **CAPÍTULO II - Avaliação do Desempenho de uma “Escala Hedônica com Referência” para Utilização em Teste de Consumidor**

<b>Tabela 1.</b> Ordem de apresentação das escalas aos consumidores ao longo dos testes ..	42
<b>Tabela 2.</b> Médias de aceitação das amostras de guaraná obtidas através de cinco escalas .....	46
<b>Tabela 3.</b> Distribuição dos níveis de significância de $F_{amostra}$ obtidos através da simulação “bootstrap” .....	49
<b>Tabela 4.</b> Médias dos logaritmos neperianos dos níveis de significância de $F_{amostra}$ para cada escala, obtidos em simulação “bootstrap” .....	51
<b>Tabela 5.</b> Níveis de significância (p) obtidos nos testes de normalidade de Kormogorov-Smirnov para todas as escalas estudadas .....	58
<b>Tabela 6.</b> Coeficientes de variação médios referentes aos valores hedônicos coletados através de cada escala .....	60
<b>Tabela 7.</b> Níveis de significância (p) obtidos no teste de homogeneidade de Levene, coeficientes de correlação (r) entre os desvios padrões e médias de aceitação das amostras de guaraná, bem como seus níveis de significância ( $p_r$ ), obtidos para cada escala .....	60

### **CAPÍTULO III - Desempenho de uma “Escala Hedônica com Referência” para Obtenção de Modelos Preditivos de Aceitação a partir de Respostas Obtidas com Consumidores**

<b>Tabela 1.</b> Tratamentos correspondentes ao delineamento experimental mostrando valores reais e codificados da variáveis independentes estudadas e valor do “ratio” do suco formulado em cada tratamento .....	73
<b>Tabela 2.</b> Ordem de apresentação das amostras para cada consumidor em cada uma das sessões de teste .....	79
<b>Tabela 3.</b> Médias de aceitação dos tratamentos correspondentes ao delineamento experimental, obtidas com cada escala .....	82
<b>Tabela 4.</b> Coeficientes de regressão, valores t, níveis de significância dos parâmetros, coeficientes de determinação dos modelos completo e ajustado obtidos a partir da escala hedônica tradicional .....	84

<b>Tabela 5.</b> Coeficientes de regressão, valores t, níveis de significância dos parâmetros, coeficientes de determinação dos modelos completo e ajustado obtidos a partir da escala hedônica com referência .....	85
<b>Tabela 6.</b> Análise da variância (ANOVA, $p \leq 0.05$ ) do modelo ajustado a partir dos dados da escala hedônica (Equação V) .....	86
<b>Tabela 7.</b> Análise da variância (ANOVA, $p \leq 0.05$ ) do modelo ajustado a partir dos dados da escala hedônica com referência (Equação VII) .....	86
<b>Tabela 8.</b> Níveis de significância (p) obtidos nos testes de normalidade de Kolmogorov-Smirnov para as duas escalas estudadas .....	87

## RESUMO GERAL

Na última década, a globalização da economia mundial promoveu uma alta competitividade entre produtos que disputam um mesmo mercado consumidor e trouxe como consequência um aumento da necessidade das indústrias alimentícias em comparar, através de testes sensoriais com consumidores, várias formulações / processamentos, relativamente a um líder de mercado.

A escala hedônica estruturada de 9 pontos, a mais utilizada em testes de aceitação, avalia o grau com que as amostras são aceitas pelos consumidores, no entanto, não possibilita a comparação direta entre duas amostras, não permitindo que se coloque os consumidores numa real situação de escolha entre duas ou mais amostras.

O presente trabalho concentrou-se na adaptação da tradicional escala hedônica para utilização em testes com consumidor nos quais o uso de uma amostra referência é desejável. Esta escala adaptada foi por nós intitulada “escala hedônica com referência” e seu desempenho foi avaliado através de dois experimentos.

Num primeiro experimento, o desempenho da escala hedônica com referência foi comparado aos desempenhos apresentados pelas escalas hedônica estruturada de 9 pontos e de estimativa da magnitude, com relação ao poder discriminativo entre amostras, com relação à variabilidade das respostas e à adequação das respostas aos pressupostos necessários à realização da análise da variância (ANOVA). Foram utilizados como critérios para o julgamento do poder discriminativo das escalas o número de discriminações significativas ( $p \leq 0.05$ ) obtidas entre as amostras e a amostra referência, verificadas através dos testes de Tukey e Dunnett, bem como os níveis de significância de  $F_{amostra}$  obtidos na ANOVA e em simulações de reamostragem “bootstrap”. A comparação entre a variabilidade das respostas dos 50 consumidores, obtidas com as escalas estudadas, foi realizada através de teste de Tukey sobre os coeficientes de variação médios das escalas. A normalidade da distribuição das respostas obtidas com as escalas foi avaliada através da construção e análise de gráficos do tipo box-whisker, gráficos da probabilidade normal dos resíduos e teste de normalidade de Kolmogorov-Smirnov. Para se avaliar

a homogeneidade das variâncias das amostras foram realizados os testes de homogeneidade de Levene e de significância da regressão ( $p_r$ ) entre desvios padrões e médias de aceitação. Os resultados revelaram que, embora a escala de estimativa da magnitude tenha apresentado maior poder discriminativo que as demais escalas, suas respostas apresentaram grande desvio da normalidade e variâncias não homogêneas significativamente ( $p \leq 0.05$ ), tornando o uso da ANOVA inadequado para os dados coletados através dessa escala. Quando comparada à escala hedônica tradicional, a escala hedônica com referência apresentou menores valores dos níveis de significância de  $F_{amostra}$ , mostrando portanto maior poder discriminativo que a primeira. Os coeficientes de variação médios dessas duas escalas não diferiram significativamente ( $p \leq 0.05$ ), sugerindo que a variabilidade das respostas dos consumidores foi similar nas duas escalas. As respostas da escala hedônica com referência apresentaram-se mais próximas à distribuição normal e mostraram maior homogeneidade das variâncias do que as respostas obtidas com a escala hedônica tradicional. Assim, esses resultados permitiram concluir que a escala hedônica com referência, além de oferecer maior poder discriminativo ao teste, forneceu dados que melhor atenderam aos pressupostos da ANOVA, conferindo assim maior confiabilidade aos resultados obtidos.

Num segundo experimento, o desempenho da escala hedônica com referência foi comparado ao da tradicional escala hedônica estruturada de 9 pontos, na obtenção de modelos preditivos da aceitação obtida através de respostas com consumidores. Foram elaboradas onze formulações de suco de laranja reconstituído, utilizando-se um delineamento fatorial  $2^2$ , intitulado delineamento composto central. As variáveis independentes estudadas foram: °Brix (9.0, 9.7, 11.5, 13.2, 14.0) e acidez (0.68, 0.74, 0.90, 1.05, 1.12 g/100 mL). A aceitação global das amostras foi avaliada por 44 consumidores, os quais utilizaram duas escalas: a escala hedônica estruturada de 9 pontos e uma escala hedônica com referência. Os dados foram analisados através da Metodologia de Superfície de Resposta, análise dos resíduos dos modelos e teste de média Dunnett. Ambas as escalas geraram modelos preditivos de aceitação que apresentaram altos valores de coeficiente de determinação ( $R^2$ ), quais sejam: 0.89 ( $p = 0.001$ ) para a escala hedônica tradicional e 0.91 ( $p = 0.0005$ ) para a escala hedônica com referência. Os modelos gerados através das duas escalas foram similares com relação à não significância da falta de ajuste a  $p \leq 0.05$ , parâmetros que compuseram o modelo final ajustado, bem como a adequação dos resíduos aos pressupostos estatísticos inerentes à análise de regressão. As regiões

de otimização obtidas foram similares para as duas escalas, bem como os efeitos das variáveis independentes verificados sobre a aceitação do suco. O teste de média Dunnett ( $p \leq 0.05$ ) indicou que a escala hedônica com referência proporcionou maior discriminação das amostras com relação à amostra referência. O desempenho das duas escalas para gerar modelos preditivos de aceitação foi equivalente.

## GENERAL ABSTRACT

In the last decade, the global economy promoted a high competitiveness amongst products that dispute the same consumer market, and as a consequence, brought an increase in the necessity for food industries to compare several formulations or processes with respect to a market leader through sensorial tests with consumers.

The 9-point structured hedonic scale, the most commonly used scale in acceptability tests, evaluates the degree with which the samples are accepted by the consumers. However, it does not allow for a direct comparison between two samples, not putting the consumers in a real situation of choice between two or more samples.

The present work concentrated on the adaptation of the traditional hedonic scale for use in consumer tests when the use of a reference sample is desirable. This adapted scale was called by us an "hedonic scale with reference" and its performance was evaluated through two experiments.

In the first experiment, the performance of the hedonic scale with reference was compared to that of the 9-point structured hedonic scale and that of the magnitude estimation scale, with respect to the discriminative power amongst samples, with respect to the variability of the answers and to the adequacy of the answers to the necessary assumptions for the accomplishment of the analysis of variance (ANOVA). The number of significant discriminations ( $p \leq 0.05$ ) obtained between the samples and the reference sample was used as a criteria to judge the discriminative power of the scales and they were verified using the Tukey and Dunnett tests, as well as the significance levels of  $F_{\text{sample}}$  obtained in ANOVA and in "bootstrap" resampling simulations. The comparison amongst the variability of the consumer responses obtained with the scales studied was accomplished by applying the Tukey test to the mean variation coefficients of the scales. The normality of the response distribution obtained with all scales was evaluated through the normal probability and box-whisker graphic constructions and analyses, and the Kolmogorov-Smirnov normality test. To evaluate the homogeneity of the sample variances, the correlation ( $r$ ) significance between standard deviations and acceptance means was tested and the Levene homogeneity test used. Although the magnitude estimation scale has presented greater discriminative power than the other scales, the results showed that its answers presented a greater

deviation from normality and non significant homogeneous variances ( $p \leq 0.05$ ), which made the use of ANOVA inadequate to analyse the data collected with that scale. When the hedonic scale with reference were compared to the scale traditional hedonic scale, it presented smaller p values, showing larger discriminative power than the traditional hedonic scale. The mean variation coefficients of these two scales did not differ significantly ( $p \leq 0.05$ ), suggesting that the the consumer response variability was similar in both scales. The responses from the hedonic scale with reference were closer to the normal distribution and showed greater variance homogeneity than those obtained from the traditional hedonic scale. Thus, besides offering greater discriminative power, these results allow for the conclusion that the hedonic scale with reference supplied data which better fitted the assumptions of ANOVA, giving greater reliability to the results obtained.

In the second experiment, the performance of the hedonic scale with reference was compared to that of the 9-point traditional structured hedonic scale in obtaining acceptability predictive models from consumer responses. Eleven formulations of reconstituted orange juice were prepared using a  $2^2$  factorial design called the central composite design. The independent variables studied were: °Brix (9.0, 9.7, 11.5, 13.2, 14.0) and acidity (0.68, 0.74, 0.90, 1.05, 1.12 g/100 mL). The global acceptance of the samples was evaluated by 44 consumers using the following two scales: a 9-point structured hedonic scale and a hedonic scale with reference. The data were analysed using Response Surface Methodology, analysis of the model residues and Dunnett's means test. Both scales generated predictive models of acceptance with high coefficients of determination ( $R^2$ ), these being: 0.89 ( $p = 0.001$ ) for the traditional hedonic scale and 0.91 ( $p = 0.0005$ ) for the hedonic scale with reference. The models generated by the two scales were similar with respect to: the non-significance of the lack of fit at  $p \leq 0.05$ , the parameters comprising the final adjusted model and the adequacy of the residues to the statistical assumptions inherent in the regression analysis. The optimized regions were similar for the two scales. Dunnett's means test indicated that the hedonic scale with reference provided a greater discrimination of the samples with respect to the standard sample. The performance of the two scales in generating predictive models of acceptance was equivalent.

## INTRODUÇÃO GERAL

A globalização da economia mundial tem nitidamente aumentado a competitividade entre as indústrias. Como consequência, as diferenças entre as várias marcas que disputam um mesmo mercado consumidor passaram a ser bastante sutis, porém suficientes para dividirem a preferência entre os consumidores. Desta forma, tornou-se maior o interesse das indústrias, sobretudo das alimentícias, em comparar, com o máximo poder discriminativo possível, várias opções de formulações ou processamentos, relativamente a uma amostra alvo, via de regra um líder de mercado. Com esse objetivo, recorre-se aos testes de consumidores, os quais podem ser classificados em testes de preferência ou de aceitação.

Testes de preferência consistem na comparação direta de duas ou mais amostras, forçando o consumidor a escolher uma amostra sobre à(s) demais. Entretanto, muitas vezes esses testes não são utilizados por não informarem o grau com que as amostras foram aceitas, utilizando-se alternativamente testes de aceitação (STONE & SIDEL, 1985; MEILGAARD *et al*, 1988).

Testes de aceitação avaliam o estado afetivo do consumidor em relação ao produto, indicando o grau com o qual o consumidor gostou ou não do mesmo (STONE & SIDEL, 1985; MEILGAARD *et al*, 1988). Testes de aceitação utilizam diferentes tipos de escalas: estruturadas, não estruturadas ou de estimativa da magnitude (MEILGAARD *et al*, 1988); porém, a escala de uso mais freqüente é a escala hedônica estruturada de nove pontos, desenvolvida e utilizada pela primeira vez por PERYAM & PILGRIM (1957). À essa escala atribui-se as vantagens de fácil compreensão pelo consumidor (McDANIEL & SAWYER, 1981; STONE & SIDEL, 1985); razoável reprodutibilidade das diferenças entre produtos em vários grupos consumidores; assim como confiabilidade e validade dos resultados (STONE & SIDEL, 1985).

Quando escalas hedônicas são utilizadas com o propósito de se comparar um ou mais produtos com o correspondente líder do mercado, a preferência do mesmo sobre os demais é inferida através das médias de aceitação: produtos que apresentam médias de aceitação significativamente maiores que os demais ( $p \leq 0,05$ ) são considerados como preferidos.

Entretanto, nesses casos a comparação direta com o objetivo de identificar a preferência / escolha do consumidor não é de fato realizada.

A escala de estimativa da magnitude para respostas hedônicas possibilita a comparação direta entre uma amostra referência com várias amostras, permitindo aos consumidores expressarem suas respostas hedônicas em termos de desvios positivos ou negativos em relação à amostra referência, por exemplo um líder do mercado (MOSKOWITZ & SIDEL, 1971). A sistemática de avaliação característica à escala hedônica de estimativa da magnitude pode ser vantajosa quando o objetivo é comparar várias amostras com uma amostra líder do mercado. Em complementação, é possível que essa escala diminua a variabilidade das respostas dos indivíduos, uma vez que todos os consumidores realizam suas avaliações baseando-se numa mesma referência, contribuindo assim para um aumento do poder discriminativo do teste. São encontrados na literatura vários estudos sobre a escala de estimativa da magnitude para avaliações hedônicas, comparativamente a outras escalas, os quais consideram vários aspectos: poder discriminativo, repetibilidade e comportamento dos consumidores com relação ao uso das escalas (MOSKOWITZ & SIDEL, 1971; McDANIEL & SAWYER, 1981; GIOVANNI & PANGBORN, 1983; VICKERS, 1983). Os resultados obtidos nesses estudos são controversos: MOSKOWITZ & SIDEL (1971) E VICKERS (1983) relataram que nenhuma das escalas estudadas, quais sejam as escalas de estimativa da magnitude, hedônica de 9 pontos e outra escala hedônica estruturada, mostrou-se superior às demais com relação ao poder discriminativo e repetibilidade; McDANIEL & SAWYER (1981) E GIOVANNI & PANGBORN (1983) encontraram maior poder discriminativo para a escala de estimativa da magnitude do que para as escalas hedônica de 9 pontos e hedônica não estruturada, respectivamente. No entanto, GIOVANNI & PANGBORN (1983) concluíram que, a despeito da escala de estimativa de magnitude ter apresentado maior poder discriminativo, esta mostrou-se inadequada para avaliações hedônicas, uma vez que 50% das respostas obtidas com essa escala concentraram-se em uma faixa muito pequena da mesma. De fato, há controvérsias quanto à utilização numérica plena da escala de estimativa de magnitude, ainda que ela seja uma escala aberta (infinita) (STONE & SIDEL, 1985; LAWLESS & MALONE, 1986a). Ainda tem sido atribuídas à essa escala as seguintes limitações: a mesma requer dos provadores habilidade em calcular frações matemáticas (STONE & SIDEL, 1985) e; apresenta dificuldades de entendimento e uso por parte dos consumidores (STONE & SIDEL, 1985; LAWLESS & MALONE, 1986a).

Outra escala utilizada em testes de consumidores é a escala intitulada escala do ideal (just right scale). Essa escala permite ao consumidor avaliar o quanto a intensidade de determinado atributo encontra-se mais forte ou mais fraca em relação à intensidade ideal (MEILGAARD *et al*, 1988). Nesse tipo de avaliação, o consumidor utiliza uma referência virtual própria e por isso não é esperado que o uso desse tipo de referência diminua a variabilidade das respostas entre os consumidores. Estudos comparativos entre o desempenho da escala do ideal e o da escala hedônica mostraram resultados similares. EPLER *et al* (1998) ao avaliarem a doçura ideal de amostras de limonada através de teste de preferência, escala hedônica de 7 pontos e escala do ideal não estruturada e estruturada com 7 categorias, verificaram que os testes de preferência e aceitação apresentaram resultados coincidentes, porém, os mesmos diferiram daqueles obtidos através da escala do ideal. Da mesma forma, SHEPHERD *et al* (1985) compararam também a escala do ideal com a escala hedônica de 9 pontos para avaliação do sabor salgado, encontrando que a concentração ideal de cloreto de sódio indicada pelos consumidores através da escala do ideal, foi significativamente superior à concentração da amostra de maior aceitação obtida através da escala hedônica. A baixa correlação entre os dados das duas escalas sugeriram que a escala do ideal e a escala hedônica não predizem os mesmos valores médios ótimos.

Pelo exposto acima, a identificação de uma escala que permita tanto a comparação direta entre várias amostras com um produto referência, como também a avaliação do grau de aceitação do produto, e ainda seja de fácil uso por consumidores em geral, mostra-se nos dias de hoje como um desafio à área de análise sensorial.

Uma escala com essas características é a escala de diferença do controle, a qual tem sido muito utilizada para avaliações de intensidade, mas não para avaliações hedônicas (MEILGAARD *et al*, 1988). Assim, uma análise exploratória do desempenho dessa escala em teste com consumidores parece ser um campo aberto na área de análise sensorial.

Uma segunda demanda com relação a testes com consumidores diz respeito às análises estatísticas dos dados gerados através dos mesmos. Às respostas hedônicas usualmente são aplicadas análises estatísticas paramétricas, notadamente a análise da variância (ANOVA) (O'MAHONY, 1982; MEILGAARD *et al*, 1988). A ANOVA testa se há diferenças significativas ( $p \leq 0,05$ ) entre as médias de pelo menos duas amostras. Para sua realização, é preciso que os

dados atendam aos pressupostos de normalidade da distribuição das respostas e homogeneidade entre as variâncias das amostras (homoscedasticidade) (O'MAHONY, 1982; O'MAHONY, 1986; DEVORE & PECK, 1986). Se esses requisitos forem violados, o nível de significância ( $p$ ) do  $F_{amostra}$  terá valor apenas aproximado, conferindo baixa confiabilidade aos resultados obtidos e às conclusões delineadas (O'MAHONY, 1982; 1986).

Parte das pesquisas realizadas atualmente na análise sensorial concentram-se no desenvolvimento de modelos preditivos envolvendo como variável dependente a aceitação de produtos por consumidores e como variáveis independentes, várias características de processo ou de formulação do produto, tais como: temperatura de extrusão do produto, níveis de edulcorantes a serem adicionados, etc. Assim, a partir de dados coletados em uma série de experimentos com consumidores, é possível criar uma equação matemática descritiva que resume esses dados, permitindo estabelecer empiricamente o tipo de relação existente entre a variável resposta (crocrância, etc) e os fatores que a influenciam (tempo de fritura, etc) (KHURI & CORNELL, 1987; MOSKOWITZ, 1994). Segundo KHURI & CORNELL (1987), a análise de regressão é uma das ferramentas mais usadas atualmente na investigação da relação causa-e-efeito, com aplicação na física, biologia, ciências sociais e engenharia.

Dado que o interesse pelo desenvolvimento de modelos preditivos envolvendo respostas de consumidores vem aumentando no setor alimentício, torna-se de maior relevância que as escalas utilizadas em testes com consumidores atendam aos requisitos de bom poder discriminativo, baixa variabilidade das respostas entre os consumidores e adequação das mesmas aos pressupostos estatísticos para a obtenção do modelo preditivo.

O'MAHONY (1986) questiona a aplicação de análises paramétricas em dados gerados por escalas estruturada ou de estimativa da magnitude, quer estas tenham sido usadas em avaliações hedônicas, quer em avaliações de intensidade. Segundo este autor, distorções da normalidade da distribuição das respostas podem ser provocadas pelos seguintes efeitos: de compressão da escala, no caso da escala hedônica e; de preferência dos consumidores por determinado(s) valor(es) numérico(s), quer na escala estruturada, quer na escala de estimativa da magnitude. O'MAHONY (1986) afirma que o efeito de compressão provoca também a quebra da homoscedasticidade das respostas sensoriais obtidas através de escalas estruturadas. Segundo

O'MAHONY (1986), uma opção para a avaliação estatística de dados que não atendam aos pressupostos de normalidade da distribuição e homogeneidade das variâncias, são as análises não paramétricas. Contudo, com a utilização de análises não paramétricas pode ocorrer a perda de informações, como por exemplo o grau com que as amostras diferem entre si. Desta forma, a disponibilidade de uma escala que resulte em dados que melhor atendam aos requisitos necessários à realização da ANOVA configura também uma demanda no campo da análise sensorial. Recentes trabalhos, como aqueles realizados por McPHERSON & RANDALL (1985), NAES (1990), GAY & MEAD (1992), BROCKHOFF & SKOVGAARD (1994), WILKINSON & YUKSEL (1997), têm demonstrado a preocupação em solucionar essa questão.

Pelo exposto, apresenta-se hoje, dentro da área de conhecimento intitulada Análise Sensorial, uma enorme demanda em se identificar uma escala para testes de consumidores que: possibilite a avaliação do grau de aceitação de cada amostra com relação a uma referência; seja de fácil entendimento e uso por parte dos consumidores e; produza respostas que atendam aos pressupostos de normalidade da distribuição e homoscedasticidade.

Uma vez que as escalas hedônicas e de diferença do controle apresentam individualmente algumas das várias características anteriormente mencionadas, o presente trabalho concentrou-se na adaptação das mesmas para utilização em testes com consumidor nos quais o uso de uma amostra referência é desejável. Esta escala adaptada foi por nós intitulada “escala hedônica com referência” e seu desempenho foi comparado aos desempenhos apresentados pelas escalas hedônica estruturada de 9 pontos e de estimativa da magnitude, com relação ao poder discriminativo entre amostras, com relação à variabilidade das respostas, à adequação das respostas aos pressupostos necessários à realização da análise da variância (ANOVA) e obtenção de modelos preditivos envolvendo respostas com consumidores.

# **CAPÍTULO I**

## **Revisão Bibliográfica**

## REVISÃO BIBLIOGRÁFICA

### 1.1. O uso de escalas na análise sensorial

Segundo o IFT (Institute of Food Technologists) a análise sensorial é uma disciplina usada para provocar, medir, analisar e interpretar as reações produzidas pelos alimentos e materiais, da forma como são percebidas pelos órgãos da visão, olfato, gosto, tato e audição (STONE & SIDEL, 1985).

Na análise sensorial, para se medir ou comparar estímulos sensoriais, tais como a intensidade da doçura, dureza ou maciez de um alimento, utilizam-se escalas (MEILGAARD *et al*, 1988; STOER & LAWLESS, 1993). Da mesma forma, a reação do indivíduo a um produto, como por exemplo, o grau com que o mesmo gostou do produto, sua intenção de compra, etc, podem também ser avaliados através de escalas (MEILGAARD *et al*, 1988).

Baseados na classificação proposta por STEVENS, 1951, STONE & SIDEL (1985) sugerem que as escalas podem ser classificadas como: escalas nominais, ordinais, de intervalos e de razão.

#### 1.1.1. Escalas nominais

As escalas nominais são ferramentas utilizadas para descrever, codificar ou classificar itens ou respostas de indivíduos com relação a um objeto ou idéia. Dessa forma, são usadas na obtenção de dados demográficos de idade, sexo, renda dos entrevistados, assim como suas atitudes ou opiniões acerca de determinado produto. Cada categoria da escala nominal pode ser composta de valores numéricos que definem a faixa etária do entrevistado ou termos verbais que exemplifiquem sua atitude ou comportamento em relação ao produto. As categorias são perfeitamente intercambiáveis, sem que sejam alteradas a lógica da questão ou o tratamento dos resultados, uma vez que não estão associadas a um valor numérico ou ordem crescente ou decrescente de intensidade. Exemplos da escala nominal são mostrados na Figura 1.

A análise dos dados obtidos com escalas nominais, via de regra, inclui o teste do  $\chi^2$  (qui-quadrado) e comparação da distribuição de frequência dos dados.

Nome:	_____			Data	__ / __ / __
Sexo:	Feminino	_____	Masculino	_____	
Idade:	até 18 anos	_____	de 19 a 30 anos	_____	de 31 a 40 anos
	de 40 a 50 anos	_____	mais que 50 anos	_____	
Por favor, indique se você utiliza alguma dieta especial:					
Para diabéticos	_____	Baixas calorias	_____	Altas calorias	_____
Baixo teor de sal	_____	Outra (Qual?)	_____	Nenhuma	_____
Por favor, indique se você tem alergia ou aversão a algum dos alimentos abaixo:					
Queijo (Que tipo?)	_____	Chocolate	_____	Ovo	_____
Fruta (Qual?)	_____	Leite	_____	Soja	_____
Carne (Qual?)	_____	Frutos do mar (Qual?)		_____	Pimenta
Vegetal (Qual?)	_____				

**FIGURA 1.** Exemplos de escala nominal.

### 1.1.2. Escalas ordinais (“ranking” e “rating scales”)

Escalas ordinais medem a intensidade percebida de um determinado atributo no produto testado. Assim, para estimar a intensidade com que o atributo está presente no produto, essas escalas podem usar tanto números como palavras organizadas de “fraco” a “forte”, de “baixo” a “alto”, de “mais” a “menos” (preferida, doce, salgada, etc). Escalas ordinais podem ainda ser mistas, associando um valor numérico e termo verbal a cada categoria.

Nas escalas ordinais as categorias não são intercambiáveis, uma vez que uma determinada categoria representa uma intensidade maior que a categoria que a antecede e menor que a categoria subsequente.

Infelizmente, não se pode assumir que a distância psicológica entre as categorias seja idêntica ao longo da escala (AMERINE *et al*, 1965; STONE & SIDEL, 1985; O'MAHONY,

1982 e 1986) e essa ocorrência pode vir a ser um fator limitante no uso de algumas análises estatísticas dos dados obtidos através dessas escalas (O'MAHONY, 1982 e 1986).

Quando escalas ordinais são utilizadas para avaliações de intensidade, a necessidade de especificar criteriosamente a característica do produto (atributo) que se deseja avaliar constitui uma limitação da mesma. Nem sempre o atributo a ser avaliado, bem como as variações de intensidade com que o mesmo pode ser encontrado no produto são familiares ao provador, requerendo um treinamento do mesmo anteriormente à realização do teste.

Escalas ordinais encontram aplicação nos testes de comparação pareada e de ordenação tanto em testes de diferença como preferência. São também consideradas escalas ordinais as "rating scales", chamadas por MEILGAARD *et al* (1988) de escalas de categorias ou estruturadas, utilizadas para se avaliar intensidade de atributos ou respostas hedônicas.

As escalas estruturadas podem ter de 5 a 12 categorias, entretanto, grande parte dos pesquisadores utilizam escalas de 8 a 10 categorias. Segundo STONE & SIDEL (1985), há extensa literatura que defende que a escala com nove categorias é a mais eficiente na transmissão de informação. Escalas estruturadas podem utilizar números e/ou palavras ao longo de toda escala para indicar a intensidade do atributo julgado em cada ponto da escala. No entanto, a escala pode também estar ancorada com termos de intensidade somente em seus extremos. A definição do número de categorias e dos termos a ancorarem as escalas constitui-se em grande dificuldade quanto a sua utilização.

A seleção dos termos a ancorarem as escalas, em função do grau de experiência do pesquisador, pode vir a ser bastante arbitrária, gerando termos ambíguos e/ou subjetivos, os quais podem provocar confusão entre provadores durante a avaliação do(s) produto(s), resultando em perda de sensibilidade da escala. Por exemplo, os termos "muito insatisfatório" e "muito satisfatório", ancorando uma escala na avaliação de um produto, podem ter diferentes conotações para diferentes provadores, causando grande variação das respostas obtidas e diminuição do poder discriminativo do teste.

Nas Figuras 2 e 3 são mostrados exemplos de escalas ordinais.

Os resultados obtidos com as escalas ordinais podem ser submetidos a vários tratamentos estatísticos paramétricos, como análise da variância, teste *t*, correlação, média, desvio padrão, etc (STONE & SIDEL, 1985).

Nome: _____		Data ___ / ___ / ___	
Marque um x no quadrado que represente a intensidade de aroma floral presente na amostra de vinho avaliada.			
fraco			forte
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

FIGURA 2. Exemplo de escala ordinal verbal.

Nome: _____		Data ___ / ___ / ___							
Circule o número que corresponda à intensidade de adstringência da amostra de vinho.									
nenhum	fraco	moderado	forte	extremo					
1	2	3	4	5	6	7	8	9	10

FIGURA 3. Exemplo de escala ordinal mista.

### 1.1.3. Escalas de intervalo

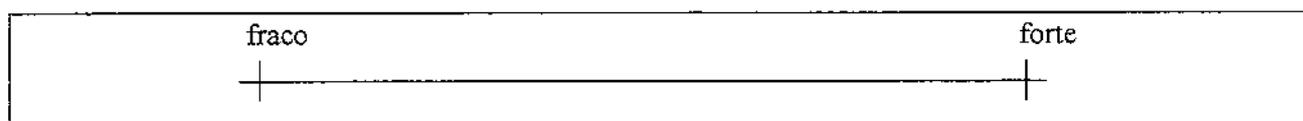
Consideradas escalas quantitativas, as escalas de intervalo consistem em um único intervalo ou numa sequência de intervalos, nos quais as distâncias entre os pontos da escala são iguais, possuindo um ponto zero arbitrário (STONE & SIDEL, 1985).

STONE & SIDEL (1985) assumem que algumas escalas ordinais, como a escala hedônica de nove pontos, são escalas de intervalo, enquanto O'MAHONY (1982 e 1986) considera as escalas estruturadas em geral como escalas de intervalo.

Uma escala de intervalo de grande utilidade na análise descritiva é a escala não estruturada (Figura 4). Ao utilizar essa escala, o provador é solicitado a expressar a intensidade percebida de um estímulo num produto, fazendo uma marca na linha horizontal. A distância entre essa marca e o extremo esquerdo da escala é medida com régua e transformada em valor numérico, o qual pode ser submetido às análises estatísticas como média, desvio padrão, teste t, análise da variância, correlação, análise de fator, regressão, etc (STONE & SIDEL, 1985).

Segundo STONE & SIDEL (1985), as escalas não estruturadas reduzem erros de natureza psicológica devido à ausência de valores numéricos associados à resposta e devido ao uso limitado de palavras, o qual minimiza o efeito da parcialidade (preferência ou aversão) por alguma palavra.

A Figura 4 apresenta um exemplo de escala de intervalo, mais especificamente, uma escala não estruturada.



**FIGURA 4.** Exemplo de escala de intervalo, escala não estruturada (STONE & SIDEL, 1985).

#### 1.1.4. Escalas de razão

Assim como as escalas de intervalo, as de razão também são utilizadas para medir valores / intensidades (STONE & SIDEL, 1985; MEILGAARD *et al*, 1988).

Ao utilizar uma escala de razão (Figura 5), o provador recebe uma amostra referência e é instruído que o atributo a ser julgado no teste sensorial, por exemplo acidez em vinho, apresenta intensidade  $\bar{x}$  em uma determinada amostra referência. Após avaliar a referência, o provador é solicitado a avaliar as demais amostras e indicar quantas vezes a acidez é maior ou menor que  $\bar{x}$  em cada amostra avaliada. Assim, se a acidez da amostra referência for especificada como 10, e a amostra 396 apresenta-se duas vezes mais ácida que a referência, o provador atribuirá a ela o

valor 20 de acidez. Portanto, nas escalas de razão, o provador estabelece a razão entre a sensação provocada por uma referência e aquela provocada pela amostra que está sendo avaliada.

A Figura 5 apresenta um exemplo da escala de estimativa da magnitude, a qual é a escala de razão mais freqüentemente usada.

À escala de estimativa da magnitude são atribuídas as seguintes limitações: a escala requer dos provadores habilidade em calcular frações matemáticas (STONE & SIDEL, 1985); apresenta dificuldades de entendimento por parte dos consumidores sobre como utilizar a escala (STONE & SIDEL, 1985; LAWLESS & MALONE, 1986a) e, finalmente; há controvérsias quanto à utilização numérica plena da escala de estimativa de magnitude, ainda que ela seja uma escala aberta (infinita) (STONE & SIDEL, 1985; LAWLESS & MALONE, 1986a).

Nome: _____		Data ___ / ___ / ___	
Você está recebendo uma amostra referência (R) de suco de laranja e uma amostra codificada. Prove a amostra referência (R), cuja doçura tem intensidade 10. Em seguida, avalie a doçura da amostra codificada atribuindo um valor a sua intensidade que corresponda a quantas vezes é maior ou menor que a doçura da amostra referência. Por exemplo, se a amostra codificada for duas vezes mais doce que a referência, dê valor 20. Se, ao contrário, for duas vezes menos doce que a referência, dê valor 5, e assim por diante.			
Amostra		Doçura	
R		10	
581		_____	

**FIGURA 5.** Exemplo de escala de razão.

Para analisar os dados obtidos através dessa escala, os mesmos devem ser normalizados para que sejam eliminadas as variações de um mesmo provador e entre provadores. Essa normalização pode ocorrer incluindo o cálculo e utilização de médias geométricas (STONE & SIDEL, 1985).

## **1.2. Vantagens, desvantagens e limitações das principais escalas utilizadas na análise sensorial**

Com relação às escalas de intensidade, há trabalhos na literatura que discutem as vantagens, desvantagens, desempenho e limitações de cada uma. Ao se comparar diferentes escalas, muitos pesquisadores avaliaram de forma comparativa, o poder discriminativo e a variabilidade das respostas obtidas em cada uma delas.

A variabilidade das respostas obtidas através de escalas é inerente a todas avaliações sensoriais, e pode ser função das variações entre produtos e também das variações humanas: entre indivíduos ou de um mesmo indivíduo quando as medidas são realizadas com repetições (LAWLESS & MALONE, 1986-a).

A capacidade discriminativa de uma escala pode ser reduzida ou aumentada por influência de alguns fatores. Por exemplo, acredita-se que o uso de escala com menos de cinco categorias torna-a pouco discriminativa (LAWLESS & MALONE, 1986-b). Ao contrário, uma escala pode ter seu poder discriminativo aumentado com a introdução de uma referência (STOER & LAWLESS, 1993) ou com o treinamento dos provadores no uso da escala, principalmente com relação às escalas estimativa da magnitude e não estruturada (LAWLESS & MALONE, 1986-a).

LAWLESS & MALONE (1986-a; 1986-b) avaliaram a intensidade de atributos táteis e olfativos em produtos não alimentícios, utilizando quatro tipos de escala: estruturada numérica, não estruturada, híbrida (com categorias dispostas graficamente numa linha) e estimativa da magnitude. Em um dos estudos, as avaliações foram realizadas apenas por consumidores e no outro, por consumidores e por estudantes universitários. Todas as escalas apresentaram poderes discriminativos equivalentes, com pequena vantagem para a escala estruturada. Em todas as escalas, as faixas de fato utilizadas pelos indivíduos foram equivalentes exceto na escala de estimativa da magnitude que, quando usada pelo grupo de estudantes, apresentou menor faixa utilizada. Em outro estudo, LAWLESS & MALONE (1986-a) verificaram a preferência do usuário quanto a qual escala utilizar, no que se refere a: facilidade de entendimento, rapidez na utilização e restrição quanto ao uso da escala. A escala estruturada mostrou-se de mais fácil

entendimento, maior rapidez de utilização e um tanto restritiva, ou seja, os provadores não se satisfizeram com relação à oportunidade para se expressar; a escala não estruturada mostrou-se de entendimento um pouco difícil, de rapidez de utilização moderada e menos restritiva; finalmente, a escala de estimativa da magnitude mostrou-se pior nos três aspectos.

PANGBORN *et al* (1989) avaliaram o amargor da cafeína em chocolate quente, utilizando as escalas estruturada, não estruturada e de estimativa da magnitude, sendo essa escala com módulos (valores atribuídos à referência) fixo (pré-estabelecido pelo analista sensorial) e livre (atribuído pelo provador). Os autores encontraram que as escalas estruturada e não estruturada foram utilizadas similarmente no que se refere a intensidade estimada e faixa da escala de fato utilizada, a qual ocupou quase toda extensão das escalas, tendo os extremos sido evitados. Não foram observadas diferenças entre os módulos fixo e livre da escala de estimativa da magnitude para avaliações de intensidade.

### **1.3. Escalas que utilizam referência**

Segundo STOER & LAWLESS (1993), há trabalhos na literatura que sugerem que o homem seja pobre como instrumento de medida e que comparações relativas entre estímulos com uma referência especificada são mais consistentes com as habilidades humanas. Esses autores citam PARDUCCI (1974), o qual afirmou que o relativismo é uma característica básica do julgamento humano. PARDUCCI (1974) E RISKEY *et al* (1979), citados por STOER & LAWLESS (1993), colocam que nas avaliações sensoriais, há forte influência dos critérios/referências que o provador utiliza para realizar seu julgamento.

LARSON-POWERS & PANGBORN (1978) compararam escalas de intensidade com e sem referência quando estudaram o efeito de diferentes edulcorantes em bebidas e gelatinas. As autoras observaram melhora na discriminação entre amostras quando utilizada a escala com referência, concluindo que ancorando a descrição a uma referência e expressando os resultados em termos dos desvios positivos e negativos em relação à referência, a precisão e a exatidão da escala são melhoradas. Contudo, LARSON-POWERS & PANGBORN (1993) observaram que o delineamento experimental utilizado nesse trabalho não foi balanceado quanto à ordem de

apresentação das escalas e ao número de vezes que cada escala foi avaliada (uma escala foi apresentada mais vezes que a outra), gerando variações que não foram controladas e que podem ter influenciado os resultados finais.

STOER & LAWLESS (1993) por sua vez compararam o poder discriminativo de escalas estruturadas ancoradas nos extremos com os termos “fraco” e “forte”, com e sem a utilização de amostra referência. Os provadores avaliaram a intensidade de atributos sensoriais de produtos lácteos - leite achocolatado e iogurte de baunilha. Os autores encontraram mais diferenças significativas entre as amostras quando a escala com referência foi utilizada, porém, sendo este total não significativo em relação ao obtido com a escala sem referência. Em complementação, ambas as escalas mostraram-se equivalentes quanto à reprodutibilidade. No entanto, os autores mantiveram as suposições de PARDUCCI (1974) e RISKEY *et al* (1979), sugerindo que a escala com referência possa ser vantajosa em situações nas quais as discriminações são difíceis ou o estudo envolver uma comparação com uma referência principal ou determinada. Portanto, ficando a decisão sobre qual escala utilizar, em função do objetivo do teste.

Com relação à escolha da amostra-referência, STONE & OLIVER (1969), utilizando a escala de magnitude na avaliação da intensidade da doçura de misturas de edulcorantes, relataram que, quando a concentração da amostra referência encontra-se em um dos extremos da escala, os valores de doçura relativa sofrem mudanças consistentes ao longo da série de concentrações. Selecionando-se a referência com concentração no meio da série, parece produzir valores (razão) mais aceitáveis e a distribuição das respostas torna-se mais adequada à função de potência.

#### **1.4. Testes Afetivos e Escalas utilizadas em testes de aceitação**

Os testes afetivos, também conhecidos como testes de consumidores, são pré-requisitos ao lançamento de um produto no mercado consumidor e ao investimento de capital em equipamentos, propaganda, etc. Conhecendo que um produto não atende plenamente às expectativas dos consumidores com relação às suas características sensoriais, obviamente uma empresa deve evitar investir no lançamento do mesmo (STONE & SIDEL, 1985; MEILGAARD *et al*, 1988).

Os testes afetivos podem determinar tanto a preferência como a aceitação dos consumidores por um ou mais produtos. Testes de preferência forçam a escolha de uma amostra sobre a outra, enquanto testes de aceitação medem o quanto o consumidor gostou ou desgostou de cada amostra (STONE & SIDEL, 1985; MEILGAARD *et al*, 1988).

Uma das classificações dos testes afetivos refere-se ao local em que os mesmos devem ser realizados. Assim, de acordo com o local onde os testes afetivos serão realizados, os mesmos podem ser classificados em três categorias: testes de laboratório, testes de uso doméstico ou testes de localização central. Este último pode ser realizado em shopping centers, escolas, supermercados, etc (STONE & SIDEL, 1985; MEILGAARD *et al*, 1987).

Os provadores utilizados em testes afetivos não são treinados, pois são consumidores e devem, preferivelmente, ser inexperientes em avaliações sensoriais, portanto, sem participação anterior em testes sensoriais descritivos ou discriminativos. Os consumidores devem ser recrutados de uma população e mercado alvos previamente definidos. Assim, os indivíduos que participam de testes afetivos, via de regra, são classificados em função de critérios demográficos tais como: idade, sexo, localização geográfica, nacionalidade, raça, religião, grau de instrução, estado civil, frequência, etc (STONE & SIDEL, 1985; MEILGAARD *et al*, 1988).

Os testes afetivos não são diferentes das demais avaliações sensoriais no que se refere à susceptibilidade às influências de determinados fatores de natureza física, fisiológica ou psicológica sobre os resultados. Dentre eles, os efeitos de contraste e adaptação dos indivíduos com relação às amostras são agrupados nos chamados “first-order and carry-over effects”, ou seja, a possibilidade de que uma amostra, sobretudo a primeira amostra avaliada, exerça influência sobre a avaliação da amostra subsequente. Para minimizar esses efeitos em testes sensoriais, a ordem de apresentação das amostras deve ser considerada e balanceada (AMERINE *et al*, 1965; MACFIE *et al*, 1989). Segundo MACFIE *et al* (1989), uma simples aleatorização das posições das amostras não é suficiente para minimizar os efeitos anteriormente mencionados. Sendo assim, MACFIE *et al* (1989) e WAKELING & MACFIE (1995) recomendam o uso do delineamento baseado em quadrados latinos, primeiramente desenvolvido por Williams, em 1949. Tal delineamento consiste em apresentar todas as amostras a cada consumidor e balancear, dentro do grupo de consumidores, a ordem com que as amostras são servidas aos provadores, de

forma que cada amostra ocorra igual número de vezes em cada posição, sendo precedida e sucedida por cada uma das demais amostras.

A escala mais utilizada em testes de aceitação é a escala hedônica estruturada de nove pontos. Na verdade, a escala hedônica estruturada de nove pontos foi desenvolvida por PERYAM & PILGRIM (1957) e ainda hoje é a mais utilizada para se medir a aceitação de consumidores em relação a um ou mais produtos (AMERINE *et al*, 1965; STONE & SIDEL, 1985; MEILGAARD *et al*, 1988).

A escala hedônica pode ser estruturada ou não estruturada, sendo esta última intitulada por MEILGAARD *et al* (1988) como escala linear e por GIOVANNI & PANGBORN (1983) como escala gráfica.

A escala estruturada pode ser verbal ou mista e segundo STONE & SIDEL (1985), quando verbal a escala hedônica apresenta a vantagem de evitar o efeito da preferência ou aversão numérica pelo consumidor.

A escala hedônica estruturada de nove pontos mais utilizada consiste em uma escala bipolar com ponto central neutro, conforme mostra a Figura 6. Essa escala avalia o grau com que o consumidor gostou ou desgostou da amostra (AMERINE *et al*, 1965; STONE & SIDEL, 1985; MEILGAARD *et al*, 1988). Testes com escala hedônica permitem a avaliação de uma amostra isoladamente, não exigindo que a mesma seja comparada com outra(s), nem forçando uma escolha.

STONE & SIDEL (1985) comentam que a escala hedônica possui as virtudes de, com um mínimo de instrução, ser de fácil entendimento pelo consumidor (McDANIEL & SAWYER, 1981), além de mostrar razoável reprodutibilidade dos resultados entre vários grupos de consumidores, apresentando confiabilidade e validade dos resultados. Entretanto, não há evidências definitivas sobre a igualdade de seus intervalos, ou seja, a distância psicológica entre duas categorias não é necessariamente igual à de outras duas, tal como ocorre com a distância física (PERYAM & PILGRIM, 1957; McDANIEL & SAWYER, 1981; STONE & SIDEL, 1985;

O'MAHONY, 1982 e 1986). Conforme comentado anteriormente, essa ocorrência pode limitar o uso de determinados testes estatísticos na avaliação de respostas hedônicas (O'MAHONY, 1982 e 1986).

- 
- Gostei extremamente
  - Gostei muito
  - Gostei moderadamente
  - Gostei ligeiramente
  - Nem gostei / nem desgostei
  - Desgostei ligeiramente
  - Desgostei moderadamente
  - Desgostei muito
  - Desgostei extremamente

---

FIGURA 6. Escala hedônica estruturada verbal de 9 pontos (STONE & SIDEL, 1985).

Um problema encontrado quando do uso da escala hedônica é a razoável variabilidade dos resultados, devido à diferença entre indivíduos com relação ao grau com que estes gostam ou não do produto. Porém esse fato é inerente às avaliações hedônicas em geral (MOSKOWITZ, 1981). VIE et al (1991) e CHAUCHAN & O'MAHONY (1993) atribuem essa variabilidade à impossibilidade dos provadores serem calibrados tal como instrumentos analíticos, conferindo assim à escala a característica de ser relativa e não absoluta.

Segundo MOSKOWITZ (1981), em avaliações hedônicas os provadores geralmente fazem dois julgamentos: primeiro classificam o estímulo em gostar, não gostar ou neutro; a seguir, avaliam o grau de gostar ou desgostar. Mesmo quando do uso de escala unipolar, o provador formula um ponto neutro na região média da escala dividindo-a entre agradável e desagradável, tarefa que constitui em mais um fator de variabilidade. Portanto, o autor recomenda que seja utilizada uma escala bipolar quando as avaliações sensoriais contemplarem estímulos agradáveis e desagradáveis em única sessão. Entretanto, quando os estímulos avaliados forem todos agradáveis ou todos desagradáveis, é razoável o uso de escala hedônica unipolar. STONE & SIDEL (1985) questionam o uso de escalas bipolares, sugerindo que há uma tendência dos provadores evitarem o ponto neutro da mesma, o que, segundo eles, aumentaria a variabilidade das respostas. Entretanto, esses autores afirmam não haver estudos que evidenciem que esse problema ocorra com a escala hedônica especificamente.

As respostas hedônicas também estão sujeitas à variabilidade provocada pelo contexto em que a avaliação sensorial está inserida. Assim, dependendo da forma com que a amostra é apresentada ao consumidor, este realiza seu julgamento baseando-se em uma referência própria (do seu conhecimento anterior sobre aquele tipo de produto) ou em critérios claros que a amostra forneça. Exemplificando, o aroma de carne pode ser mais facilmente avaliado quando uma amostra de carne é apresentada do que quando este deve ser avaliado em uma solução contendo aroma de carne. Neste caso, pode-se esperar maior variabilidade dos resultados obtidos para a solução contendo aroma de carne (MOSKOWITZ, 1981).

As escalas não estruturadas e de estimativa da magnitude, uni e bipolar, podem também ser utilizadas para medir respostas hedônicas, desde que adaptadas para tal (MEILGAARD et al, 1988).

Na verdade, as escalas estruturada - sobretudo a de 9 pontos -, não estruturada, de estimativa da magnitude e a escala do ideal são freqüentemente alvos de estudos sobre seu desempenho para avaliações hedônicas (MOSKOWITZ & SIDEL, 1971; McDANIEL & SAWYER, 1981; VICKERS, 1983; GIOVANNI & PANGBORN, 1983; McPHERSON & RANDALL, 1985; SHEPHERD et al, 1985; JOHNSON & VICKERS, 1987; PANGBORN et al, 1989; EPLER et al, 1998).

MOSKOWITZ & SIDEL (1971) foram os pesquisadores que primeiramente propuseram o uso da escala de estimativa da magnitude em avaliações hedônicas. Esses autores compararam essa escala com a escala hedônica de 9 pontos com relação à discriminação da aceitação de vários produtos alimentícios e concluíram que ambas as escalas são igualmente sensíveis para diferenciar graus de aceitação entre alimentos.

Entretanto, McDANIEL & SAWYER (1981), comparando também essas mesmas escalas, concluíram que a escala de estimativa da magnitude é mais sensível como ferramenta discriminativa em testes de aceitação.

Em seus estudos, VICKERS (1983) encontrou que as escalas estruturada e de estimativa da magnitude apresentaram resultados consensuais com relação às amostras, mostrando desempenhos similares com relação à variabilidade e reprodutibilidade das respostas dos consumidores.

PANGBORN et al (1989), ao testar escalas de magnitude com módulos fixo e livre, encontraram diferença significativa entre as mesmas em função da referência utilizada. Os autores recomendaram o uso de módulo fixo para tornar a escala de magnitude menos difícil ao provador, bem como para facilitar as análises estatísticas.

GIOVANNI & PANGBORN (1983), ao compararem as escalas não estruturada e de estimativa da magnitude para respostas hedônicas, encontraram maior poder discriminativo para a escala de estimativa da magnitude. No entanto, as autoras concluíram que a escala da magnitude não é apropriada para respostas bipolares como as hedônicas, uma vez que 50% das respostas obtidas com essa escala concentraram-se em uma faixa muito pequena da escala.

McPHERSON & RANDALL (1985) compararam a escala hedônica de 7 pontos com a escala hedônica não estruturada, encontrando alta correlação entre os resultados dessas duas escalas.

Outra escala utilizada em testes de consumidores é a escala intitulada escala do ideal (just right scale). Essa escala permite avaliar o quanto a intensidade de determinado atributo encontra-se mais forte ou mais fraca em relação à intensidade ideal segundo a opinião do consumidor (STONE & SIDEL, 1985; MEILGAARD et al, 1988). Nesse tipo de avaliação, o consumidor utiliza uma referência virtual própria e por isso não é esperado que o uso desse tipo de referência diminua a variabilidade das respostas entre os consumidores.

Segundo EPLER et al (1998) há vários trabalhos na literatura (VICKERS, 1988; JOHNSON & VICKERS, 1987; PANGBORN & GIOVANNI, 1984; STONE & PANGBORN, 1990; McBRIDE, 1985), cujos resultados mostraram que a concentração ideal do atributo avaliado pelos consumidores através da escala do ideal foi inferior àquela do produto que já se

encontrava em comercialização. EPLER et al (1998) compararam as escalas do ideal e hedônica de 9 pontos, realizando paralelamente um teste de preferência para avaliação da opinião do consumidor com relação à doçura de limonada. A amostra mais preferida coincidiu com a amostra mais aceita, ambas de doçura superior àquela considerada ideal pelos consumidores. SHEPHERD et al (1985) compararam também essas duas escalas para avaliação do sabor salgado e encontraram que a concentração ideal, indicada pelos consumidores através da escala do ideal, foi significativamente superior à concentração de maior aceitação obtida através da escala hedônica. Em complementação, esses autores encontraram baixa correlação entre os dados das duas escalas, sugerindo que a escala do ideal e a escala hedônica não predizem os mesmos valores hedônicos médios ótimos.

### 1.5. Análises Estatísticas

Às respostas de avaliações sensoriais, incluindo-se as respostas hedônicas, usualmente são aplicadas análises estatísticas paramétricas, notadamente a análise da variância (ANOVA) (O'MAHONY, 1982; MEILGAARD et al, 1988). A ANOVA testa se há diferença significativa entre as médias de pelo menos duas amostras. Para sua realização, é necessário que os dados atendam aos pressupostos de normalidade da distribuição das respostas e homogeneidade entre as variâncias das amostras (homoscedasticidade) (O'MAHONY, 1982 e 1986; DEVORE & PECK, 1986). Se esses requisitos forem violados, o nível de significância do  $F_{amostra}$  terá valor apenas aproximado, conferindo baixa confiabilidade aos resultados obtidos e às conclusões delineadas (O'MAHONY, 1982; 1986).

O'MAHONY (1986) questiona a aplicação de análises paramétricas em dados gerados por escalas estruturada ou de estimativa da magnitude, quer estas tenham sido usadas em avaliações hedônicas, quer em avaliações de intensidade. Segundo este autor, distorções da normalidade da distribuição das respostas podem ser provocadas pelos seguintes efeitos: de compressão da escala, no caso da escala estruturada e; de preferência dos provadores por determinado(s) valor(es) numérico(s), quer na escala estruturada, quer na escala de estimativa da magnitude. O'MAHONY afirma que o efeito de compressão provoca também a quebra da homoscedasticidade das respostas sensoriais obtidas através de escalas estruturadas. RANDALL & SANJUR (1981) documentaram que o efeito de compressão em um extremo ou outro da escala

hedônica ocorre quanto maior ou menor a média de aceitação da amostra, provocando o truncamento da distribuição das respostas, distanciando-a da normalidade.

Segundo O'MAHONY, uma opção para a avaliação estatística de dados que não atendam aos pressupostos de normalidade da distribuição e homogeneidade das variâncias, são as análises não paramétricas. Contudo, com a utilização de análises não paramétricas pode ocorrer a perda de informações, como por exemplo o grau com que as amostras diferiram entre si. Recentes trabalhos têm demonstrado a preocupação em solucionar essa questão, tais como os trabalhos realizados por McPHERSON & RANDALL (1985), NAES (1990), GAY & MEAD (1992), BROCKHOFF & SKOVGAARD (1994), WILKINSON & YUKSEL (1997).

### **1.5.1. Bootstrap**

Um dos problemas encontrados na realização de um experimento refere-se à validação de seus resultados. Via de regra, a coleta de dados é realizada sob amostragem reduzida, uma vez que é inviável que a mesma ocorra a partir de toda a população estudada porque, geralmente, é alto o custo do planejamento e realização do experimento. Desta forma, usualmente conclusões são extraídas de experimentos isolados e baseados em número reduzido de informações. No caso dos testes de consumidores essa questão constitui-se também uma realidade.

Uma das ferramentas estatísticas disponíveis para validação de resultados de experimentos é a simulação "bootstrap", desenvolvida por EFRON, em 1977. Simulações que utilizam a metodologia "bootstrap" permitem estimar a precisão de resultados gerados a partir de única amostra em relação à população total. Na metodologia "bootstrap", são geradas B amostras "bootstrap" de tamanho n, as quais são geradas através de sorteio dos dados da amostra original de mesmo tamanho n, com reposição. A cada amostra "bootstrap" é aplicado o tratamento estatístico que se deseja estudar. Posteriormente realiza-se a análise da distribuição dos valores dessa estatística, estimando-se, através da mesma, um intervalo de confiança (DIACONIS & EFRON, 1983; EFRON & Tibshirani, 1993). Segundo DIACONIS & EFRON (1983), a amplitude do intervalo de confiança é a medida "bootstrap" da precisão da estatística estudada, sendo que a metade do seu valor pode ser interpretada como o valor médio pelo qual a estatística

de uma amostra aleatória de tamanho  $n$  difere do valor dessa estatística quando considerada a população total. Em outro contexto, a distribuição da estatística que se deseja estudar, obtida a partir de simulações “bootstrap”, pode ser uma forma de avaliar a repetibilidade da mesma, ou de uma metodologia, ao se utilizar condições idênticas àsquelas do experimento que geraram os dados da amostra original. Assim, as conclusões delineadas pelo experimento tornam-se mais robustas porque não são baseadas em apenas um valor da estatística desejada, sem que seja necessário se realizar o experimento repetidas vezes.

### **1.5.2. Metodologia de Superfície de Resposta (RSM)**

Grande parte das pesquisas de natureza exploratória objetivam: i) determinar e quantificar a relação entre valores de uma ou mais características mensuráveis (por exemplo, crocância, gosto doce, etc) e um grupo de fatores experimentais que possam afetá-las (por exemplo, tempo de fritura, porcentagem de sacarose, etc) e; ii) encontrar os fatores experimentais que produzem a melhor resposta (KHURI & CORNELL, 1987).

A partir de dados coletados em uma série de experimentos, a regressão cria uma equação matemática descritiva que resume esses dados, permitindo estabelecer empiricamente o tipo de relação existente entre a variável resposta (crocância, etc) e os fatores que a influenciam (tempo de fritura, etc) (KHURI & CORNELL, 1987; MOSKOWITZ, 1994). Segundo KHURI & CORNELL (1987), a análise de regressão é uma das ferramentas mais usadas na investigação da relação causa-e-efeito, com aplicação na física, biologia, ciências sociais e engenharia.

A Metodologia de Superfície de Resposta (RSM) é uma metodologia adicional empregada antes, durante e após uma análise de regressão sobre alguns dados. Anteriormente à análise de regressão, o experimento deve ser planejado, ou seja, as variáveis devem ser selecionadas e os valores que as mesmas assumirão no decorrer do experimento devem ser definidos. Após efetuar-se a análise de regressão sobre os dados coletados, são utilizados procedimentos para se testar o modelo estatístico, bem como técnicas de otimização (KHURI & CORNELL, 1987). Assim, a RSM é um conjunto de procedimentos, cujos objetivos são: i) planejar uma série de experimentos que irão coletar, de forma adequada e confiável, medidas da resposta de interesse; ii) determinar

um modelo polinomial linear, quadrático ou de ordens superiores que melhor se ajuste aos dados coletados através do planejamento escolhido, realizando testes de hipótese apropriados em torno dos parâmetros do modelo, e; iii) determinar o valor ótimo das variáveis experimentais sob controle do pesquisador, que produzem o valor máximo e/ou mínimo resposta de interesse (KHURI & CORNELL, 1987; MOSKOWITZ, 1994). Modelos gerados pela RSM podem ser úteis na estimativa da resposta sensorial, sobretudo na otimização da resposta do consumidor, permitindo que se encontre a combinação de fatores dentro da faixa estudada que gere a maior aceitação possível pelo consumidor (MOSKOWITZ, 1994).

## 1.6. Referências Bibliográficas

1. AMERINE, M. A.; PANGBORN, R.M.; ROESSLER, E.B. Principles of sensory evaluation of food. In: Food Science and Technology. Academic Press, New York, p.602, 1965.
2. BROCKHOFF, M.; SKOVGAARD, M. Modeling individual differences between assessors in sensory evaluations. Food Quality and Preference, v.5, p.215-224, 1994.
3. CHAUCHAN, J.; O'MAHONY, M. Use of a signal detection ranking analysis to measure preference for commercial and "health modified" cakes. J. Sensory Studies, v.8, p.69-75, 1993.
4. DEVORE, J.L.; PECK, R. Statistics, the Exploration and Analysis of Data. West Publishing Company, St. Paul, p.699, 1986.
5. DIACONIS, P.; EFRON, B. Computer-intensive methods in statistics. Sci. American, v.24 (8), p.116-130, 1983.
6. EFRON, B.; TIBSHIRANI, R. An Introduction to the Bootstrap. Série: Monographs on Statistics and Applied Probability. Chapman & Hall. New York, p.436, 1993.
7. EPLER, S.; CHAMBERS IV, E.; KEMP, K. Hedonic scales are a better predictor than just-about-right scales of optimal sweetness in lemonade. J. Sensory Studies, v.13, p.191-197, 1998.
8. GAY, C.; MEAD, R. A statistical appraisal of the problem of sensory measurement. J. Sensory Studies, v.7, p.205-228, 1992.
9. GIOVANNI, M.E.; PANGBORN, R.M. Measurement of taste intensity and degree of liking of beverages by graphic scales and magnitude estimation. J. Food Science, v.48, p.1175-1181, 1183, 1983.
10. JOHNSON, J.; VICKERS, Z. Avoiding the centering bias or range effect when determining an optimum level of sweetness in lemonade. J. Sensory Studies, v.2, p.283-292, 1987.
11. KHURI, A.I.; CORNELL, J.A. Response Surface: Designs and Analyses. Series: Statistics, Textbooks and Monographs, v.152, D.B. Eds. Owen & W.R. Schucany, Marcel Dekker, Inc, New York, p.510, 1987.
12. LARSON-POWERS, N. & PANGBORN, R.M. Descriptive analysis of the sensory properties of beverages and gelatins containing sucrose or synthetic sweeteners. J. Food Science, v.43, p.47-51, 1978.

13. LAWLESS, H.T.; MALONE, G.J. The discriminative efficiency of common scaling methods. *J. Sensory Studies*, v.1, p.85-98, 1986a.
14. LAWLESS, H.T. & MALONE, G.J. A comparison of rating scales: sensitivity, replicates and relative measurement. *J. Sensory Studies*, v.1, p.155-174, 1986b.
15. MacFIE, H.J.; BRATCHELL, N.; GREENHOFF, K.; VALLIS, L.V. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *J. Sensory Studies*, v.4, p.129-148, 1989.
16. McDANIEL, M.R.; SAWYER, F.M. Preference testing of whiskey sour formulations: magnitude estimation versus the 9-point hedonic. *J. Food Science*, v.46, p.182-185, 1981.
17. McPHERSON, R.S.; RANDALL, E. Line Length measurement as a tool for food preference research. *Ecology of Food and Nutrition*, v.17, p.149-156, 1985.
18. MEILGAARD, M.; CIVILLE, G.V.; CARR, B.T. *Sensory Evaluation Techniques*. CRC Press, Boca Raton, p.281, 1988.
19. MOSKOWITZ, H.R. Sensory intensity versus hedonic functions: classical psychophysical approaches. *J. Food Quality*, v.5: p.109-137, 1981.
20. MOSKOWITZ, H.R. Product testing 2: modeling versus mapping and their integration. *J. of Sensory Studies*, v.9, p.323-336, 1994.
21. MOSKOWITZ, H.R.; SIDEL, J.L. Magnitude and hedonic scales of food acceptability. *J. Food Science*, v.36, p.677-680, 1971.
22. NAES, T. Handling individual differences between assessors in sensory profiling. *Food Quality and Preference*, v.2, p.187-199, 1990.
23. O'MAHONY, M. Some assumptions and difficulties with common statistics for sensory analysis. In: *Overview: Outstanding Symposia in Food Science & Technology*. *Food Technology*, p.75-82, 1982.
24. O'MAHONY, M. *Sensory Evaluation of Food: Statistical Methods and Procedures*. Series: *Food Science and Technology*. Marcel Dekker, Inc. New York, p.487, 1986.
25. PANGBORN, R.M.; GUINARD, J-X.; MEISELMAN, H.L. Evaluation of bitterness of caffeine in hot chocolate drink by category, graphic, and ratio scaling. *J. Sensory Studies*, v.4, p.31-53, 1989.
26. PERYAM, D.R. & PILGRIM, F.J. Hedonic scales of measuring food preferences. *Food Technology*, v.9 (11), p.9, 1957.

27. RANDALL, E.; SANJUR, D. Food preferences - their conceptualization and relationship to consumption. *Ecology Food Nutrition*, v.11, p.151-161, 1981.
28. SHEPHERD, R.; FARLEIGH, C.A.; LAND, D.G.; FRANKLIN, J.G. Validity of a relative-to-ideal rating procedure compared with hedonic rating. In: *Progress in Flavour Research*. Elsevier Science Publishers, Amsterdam, p.103-110, 1984.
29. STOER, N.L.; LAWLESS, H.T. Comparison of single product scaling and relative-to-reference scaling in sensory evaluation of dairy products. *J. Sensory Studies*, v.8, p.257-270, 1993.
30. STONE, H.; OLIVER, S. Measurement of the relative sweetness of selected sweeteners and sweetener mixtures. *J. Food Science*, v.34, p.215, 1969.
31. STONE, H. & SIDEL, J.L. *Sensory Evaluation Practices*. Academic Press, Orlando, p.310, 1985.
32. VICKERS, Z.M. Magnitude estimation vs category scaling of the hedonic quality of food sounds. *J. Food Science*, v.48, p.1183-1186, 1983.
33. VIE, A.; GULLI, D.; O'Mahony, M. Alternative hedonic measures, *J. Food Science*, v.56 (1), p.1-5, 1991.
34. WAKELING, I.N.; MACFIE, H.J. Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of k samples from t may be tested. *Food Quality and Preference*, v.6, p.299-308, 1995.
35. WILKINSON, C.; YUKSEL, D. Modeling differences between panelists in use of measurement scales. *J. Sensory Studies*, v.12, p.55-68, 1997.

## **CAPÍTULO II**

**Avaliação do Desempenho de uma “Escala Hedônica com Referência” para Utilização em Teste de Consumidor**

## RESUMO

Na última década, a globalização da economia mundial promoveu uma alta competitividade entre produtos que disputam um mesmo mercado consumidor e trouxe como consequência um aumento da necessidade das indústrias alimentícias em comparar, através de testes sensoriais com consumidores, várias formulações/processamentos, relativamente a um líder de mercado. Testes de preferência realizam esse tipo de comparação, porém são deficientes porque não informam o grau de aceitação das amostras. A escala hedônica estruturada de 9 pontos, a mais utilizada em testes de aceitação, avalia o grau com que as amostras são aceitas pelos consumidores, no entanto, não possibilita a comparação direta entre duas amostras, não permitindo que se coloque os consumidores numa real situação de escolha entre duas ou mais amostras. A escala de estimativa da magnitude realiza a comparação direta de amostras com uma referência específica e informa o grau de aceitação das amostras com relação à referência, entretanto, essa escala é de difícil entendimento e uso por parte dos consumidores, havendo controvérsias quanto à utilização numérica plena da mesma. Outra escala utilizada em testes de consumidores é a escala do ideal, entretanto, há trabalhos que sugerem que essa escala e a tradicional escala hedônica não predizem os mesmos valores médios ótimos. Uma segunda problemática existente em testes sensoriais com consumidores diz respeito à adequação das respostas coletadas aos pressupostos de normalidade da distribuição e homogeneidade das variâncias das amostras, pois, uma vez violados esses pressupostos, os valores dos níveis de significância de  $F_{amostra}$  obtidos na ANOVA serão apenas aproximados, conferindo baixa confiabilidade aos resultados e às conclusões delineadas. Com o objetivo de se obter uma escala que permitisse tanto a avaliação do estado afetivo do consumidor com relação às amostras como um processo de comparação direta das mesmas com uma referência e que ao mesmo tempo fosse de fácil uso pelos consumidores, esta pesquisa adaptou a escala hedônica tradicional, de forma que a mesma pudesse ser usada em testes nos quais uma amostra referência é desejável. Essa escala foi intitulada escala hedônica com referência e seu desempenho comparado aos das escalas hedônica de 9 pontos e de estimativa da magnitude, com relação a: poder discriminativo, variabilidade das respostas dos consumidores e adequação das mesmas aos requisitos para realização da ANOVA. O número de discriminações significativas ( $p \leq 0.05$ ) obtidas entre as amostras e a amostra referência, verificadas através dos testes de Tukey e Dunnett, bem como os níveis de significância de  $F_{amostra}$  obtidos na ANOVA e nas simulações de reamostragem

“bootstrap”, foram utilizados como critérios para o julgamento do poder discriminativo das escalas. A comparação entre a variabilidade das respostas dos consumidores, obtidas com as escalas estudadas, foi realizada através de teste de Tukey sobre os coeficientes de variação médios das escalas. A normalidade da distribuição das respostas obtidas com as escalas foi avaliada através da construção e análise de gráficos do tipo box-whisker, gráficos da probabilidade normal dos resíduos e teste de normalidade de Kormogorv-Smirnov. Para se avaliar a homogeneidade das variâncias das amostras foram realizados os testes de homogeneidade de Levene e de significância da regressão ( $p_r$ ) entre desvios padrões e médias de aceitação. Os resultados revelaram que, embora a escala de estimativa da magnitude tenha apresentado maior poder discriminativo que as demais escalas, suas respostas apresentaram grande desvio da normalidade e variâncias não homogêneas significativamente, tornando o uso da ANOVA inadequado para os dados coletados através dessa escala. Quando comparada à escala hedônica tradicional, a escala hedônica com referência apresentou menores valores dos níveis de significância de  $F_{amostra}$ , mostrando portanto maior poder discriminativo que a primeira. Os coeficientes de variação médios dessas duas escalas não diferiram significativamente ( $p \leq 0.05$ ), sugerindo que a variabilidade das respostas dos consumidores foi similar nas duas escalas. As respostas da escala hedônica com referência apresentaram-se mais próximas à distribuição normal e mostraram maior homogeneidade das variâncias do que as respostas obtidas com a escala hedônica tradicional. Assim, esses resultados permitiram concluir que a escala hedônica com referência, além de oferecer maior poder discriminativo ao teste, forneceu dados que melhor atenderam aos pressupostos da ANOVA, conferindo assim maior confiabilidade aos resultados obtidos.

## ABSTRACT

In the last decade, the global economy has promoted great competitiveness amongst products that dispute the same consumer market, and as a consequence has resulted in an increase in the necessity of food industries to compare, through sensorial tests with consumers, several formulations or processes with respect to a market leader. Preference tests accomplish this kind of comparison, but they are deficient because they do not provide the degree of sample acceptability. The 9-point structured hedonic scale, which is the most used in acceptability tests, evaluates the degree with which the samples are accepted by the consumers, however, it does not allow for a direct comparison between two samples, not allowing the consumers to be put in a real choice situation between two or more samples. The magnitude estimation scale accomplishes the direct comparison of samples with a specific reference and it provides the degree of sample acceptability with respect to the reference. However, this scale presents difficulty in its understanding and usage by the consumers, and there are controversies with respect to its full numerical usage. Another scale used in consumer tests is the just-right-scale, however some experiments suggest that this and the traditional hedonic scale do not predict the same optimum mean values. A second problem involving consumer tests related to the adequacy of the responses to the assumptions of normal distribution and of the sample variance homogeneity, because when these assumptions are violated, the significance levels of the values of  $F_{\text{sample}}$  provided by ANOVA will just be approximate, giving low reliability to the results and the delineated conclusions. This research concentrated on the adaptation of the traditional hedonic scale for use in consumer tests when a reference sample is desirable, with the objective of obtaining a scale that allows for the evaluation of the affective state of the consumer with respect to the samples, as well as the direct comparison between them and a reference sample, and simultaneously being easy for the consumer to use. This scale was called by us the "hedonic scale with reference" and its performance was compared to those of the 9-point structured hedonic scale and the magnitude estimation scale, with respect to the discriminative power amongst samples, with respect to the variability of the answers and to the adequacy of the answers to the necessary assumptions for the accomplishment of the analysis of variance (ANOVA). The number of significant discriminations ( $p \leq 0.05$ ) obtained between the samples and the reference sample was used as a criteria to judge the discriminative power of the scales and they were verified through the Tukey and Dunnett tests, as well as the significance levels of  $F_{\text{sample}}$  obtained in ANOVA and in the "bootstrap" resampling simulations. The comparison amongst the

variability of the consumer responses that were obtained with the scales studied was accomplished through the Tukey test on the mean variation coefficients of the scales. The normality of the response distribution obtained with all scales was evaluated through the normal probability and box-whisker graphics construction and analysis, and the Kolmogorov-Smirnov normality test. To evaluate homogeneity of the sample variances the correlation ( $p$ .) significance between the standard deviations and acceptance means was tested and the Levene homogeneity test was used. Although the magnitude estimation scale presented greater discriminative power than the other scales, the results showed that its answers presented greater deviation from normality and non significant homogeneous variances ( $p \leq 0.05$ ), which made the use of ANOVA inadequate for the data collected with that scale. When the hedonic scale with reference was compared to the scale traditional hedonic scale, it presented smaller  $p$  values, showing larger discriminative power than the traditional hedonic scale. The mean variation coefficients of these two scales did not differ significantly ( $p \leq 0.05$ ), suggesting that the consumer response variability was similar with both scales. The responses from the hedonic scale with reference were closer to the normal distribution and showed greater variance homogeneity than those obtained from the traditional hedonic scale. Thus, besides offering greater discriminative power to the test, these results allowed for the conclusion that the hedonic scale with reference supplied data that better fitted the assumptions of ANOVA, giving greater reliability to the results obtained.

## INTRODUÇÃO

A globalização da economia mundial tem nitidamente aumentado a competitividade entre as indústrias. Como consequência, as diferenças entre as várias marcas que disputam um mesmo mercado consumidor passaram a ser bastante sutis, porém suficientes para dividirem a preferência entre os consumidores. Desta forma, tornou-se maior o interesse das indústrias, sobretudo das alimentícias, em comparar, com o máximo poder discriminativo possível, várias opções de formulações ou processamentos, relativamente a uma amostra alvo, via de regra um líder de mercado. Com esse objetivo, recorre-se aos testes de consumidores, os quais podem ser classificados em testes de preferência ou de aceitação.

Testes de preferência consistem na comparação direta de duas ou mais amostras, forçando o consumidor a escolher uma amostra sobre à(s) demais. Entretanto, muitas vezes esses testes não são utilizados por não informarem o grau com que as amostras foram aceitas, utilizando-se alternativamente testes de aceitação (STONE & SIDEL, 1985; MEILGAARD *et al*, 1988).

Testes de aceitação avaliam o estado afetivo do consumidor em relação ao produto, indicando o grau com o qual o consumidor gostou ou não do mesmo (STONE & SIDEL, 1985; MEILGAARD *et al*, 1988). Testes de aceitação utilizam diferentes tipos de escalas: estruturadas, não estruturadas ou de estimativa da magnitude (MEILGAARD *et al*, 1988); porém, a escala de uso mais freqüente é a escala hedônica estruturada de nove pontos, desenvolvida e utilizada pela primeira vez por PERYAM & PILGRIM (1957). À essa escala atribui-se as vantagens de fácil compreensão pelo consumidor (Mc Daniel & SAWYER, 1981; STONE & SIDEL, 1985); razoável reprodutibilidade das diferenças entre produtos em vários grupos consumidores; assim como confiabilidade e validade dos resultados (STONE & SIDEL, 1985).

Quando escalas hedônicas são utilizadas com o propósito de se comparar um ou mais produtos com o correspondente líder do mercado, a preferência do mesmo sobre os demais é inferida através das médias de aceitação: produtos que apresentam médias de aceitação significativamente maiores que os demais ( $p \leq 0,05$ ) são considerados como preferidos.

Entretanto, nesses casos a comparação direta com o objetivo de identificar a preferência / escolha do consumidor não é de fato realizada.

A escala de estimativa da magnitude para respostas hedônicas possibilita a comparação direta entre uma amostra referência com várias amostras, permitindo aos consumidores expressarem suas respostas hedônicas em termos de desvios positivos ou negativos em relação à amostra referência, por exemplo um líder do mercado (MOSKOWITZ & SIDEL, 1971). A sistemática de avaliação característica à escala hedônica de estimativa da magnitude pode ser vantajosa quando o objetivo é comparar várias amostras com uma amostra líder do mercado. Em complementação, é possível que essa escala diminua a variabilidade das respostas dos indivíduos, uma vez que todos os consumidores realizam suas avaliações baseando-se numa mesma referência, contribuindo assim para um aumento do poder discriminativo do teste. São encontrados na literatura vários estudos sobre a escala de estimativa da magnitude para avaliações hedônicas, comparativamente a outras escalas, os quais consideram vários aspectos: poder discriminativo, repetibilidade e comportamento dos consumidores com relação ao uso das escalas (MOSKOWITZ & SIDEL, 1971; McDANIEL & SAWYER, 1981; GIOVANNI & PANGBORN, 1983; VICKERS, 1983). Os resultados obtidos nesses estudos são controversos: enquanto MOSKOWITZ & SIDEL (1971) e VICKERS (1983) relataram que nenhuma das escalas estudadas, quais sejam as escalas de estimativa da magnitude, hedônica estruturada ou não estruturada, mostrou-se superior às demais com relação ao poder discriminativo e repetibilidade; McDANIEL & SAWYER (1981) e GIOVANNI & PANGBORN (1983) encontraram maior poder discriminativo para a escala de estimativa da magnitude do que para outras escalas hedônicas. No entanto, GIOVANNI & PANGBORN (1983) concluíram que, a despeito da escala de estimativa de magnitude ter apresentado maior poder discriminativo, esta mostrou-se inadequada para avaliações hedônicas, uma vez que 50% das respostas obtidas com essa escala concentraram-se em uma faixa muito pequena da mesma. De fato, há controvérsias quanto à utilização numérica plena da escala de estimativa de magnitude, ainda que ela seja uma escala aberta (infinita) (STONE & SIDEL, 1985; LAWLESS & MALONE, 1986a). Ainda tem sido atribuídas à essa escala as seguintes limitações: a mesma requer dos provadores habilidade em calcular frações matemáticas (STONE & SIDEL, 1985) e; apresenta dificuldades de entendimento e uso por parte dos consumidores (STONE & SIDEL, 1985; LAWLESS & MALONE, 1986a).

Outra escala utilizada em testes de consumidores é a escala intitulada escala do ideal (just right scale). Essa escala permite ao consumidor avaliar o quanto a intensidade de determinado atributo encontra-se mais forte ou mais fraca em relação à intensidade ideal (MEILGAARD *et al*, 1988). Nesse tipo de avaliação, o consumidor utiliza uma referência virtual própria e por isso não é esperado que o uso desse tipo de referência diminua a variabilidade das respostas entre os consumidores. Estudos comparativos entre o desempenho da escala do ideal e o da escala hedônica mostraram resultados similares. EPLER *et al* (1998) ao avaliarem a doçura ideal de amostras de limonada através de teste de preferência, escala hedônica de 7 pontos e escala do ideal não estruturada e estruturada com 7 categorias, verificaram que os testes de preferência e aceitação apresentaram resultados coincidentes, porém, os mesmos diferiram daqueles obtidos através da escala do ideal. Da mesma forma, SHEPHERD *et al* (1985) compararam também a escala do ideal com a escala hedônica de 9 pontos para avaliação do sabor salgado, encontrando que a concentração ideal de cloreto de sódio indicada pelos consumidores através da escala do ideal, foi significativamente superior à concentração da amostra de maior aceitação obtida através da escala hedônica. A baixa correlação entre os dados das duas escalas sugeriram que a escala do ideal e a escala hedônica não predizem os mesmos valores médios ótimos.

Pelo exposto, a identificação de uma escala que permita tanto a comparação direta entre várias amostras com um produto referência, como também a avaliação do grau de aceitação do produto, e ainda seja de fácil uso por consumidores em geral, mostra-se nos dias de hoje como um desafio à área de análise sensorial.

Uma escala com essas características é a escala de diferença do controle, a qual tem sido muito utilizada para avaliações de intensidade, mas não para avaliações hedônicas (MEILGAARD *et al*, 1988). Assim, uma análise exploratória do desempenho dessa escala em teste com consumidores parece ser um campo aberto na área de análise sensorial.

Uma segunda demanda com relação a testes com consumidores diz respeito às análises estatísticas dos dados gerados através dos mesmos. Às respostas hedônicas usualmente são aplicadas análises estatísticas paramétricas, notadamente a análise da variância (ANOVA)

(O'MAHONY, 1982; MEILGAARD et al, 1988). A ANOVA testa se há diferenças significativas ( $p \leq 0,05$ ) entre as médias de pelo menos duas amostras. Para sua realização, é preciso que os dados atendam aos pressupostos de normalidade da distribuição das respostas e homogeneidade entre as variâncias das amostras (homoscedasticidade) (O'MAHONY, 1982; O'MAHONY, 1986; DEVORE & PECK, 1986). Se esses requisitos forem violados, o nível de significância do  $F_{amostra}$  terá valor apenas aproximado, conferindo baixa confiabilidade aos resultados obtidos e às conclusões delineadas (O'MAHONY, 1982; 1986).

O'MAHONY (1986) questiona a aplicação de análises paramétricas em dados gerados por escalas estruturada ou de estimativa da magnitude, quer estas tenham sido usadas em avaliações hedônicas, quer em avaliações de intensidade. Segundo este autor, distorções da normalidade da distribuição das respostas podem ser provocadas pelos seguintes efeitos: de compressão da escala, no caso da escala hedônica e; de preferência dos consumidores por determinado(s) valor(es) numérico(s), quer na escala estruturada, quer na escala de estimativa da magnitude. O'MAHONY (1986) afirma que o efeito de compressão provoca também a quebra da homoscedasticidade das respostas sensoriais obtidas através de escalas estruturadas. Segundo O'MAHONY (1982; 1986), uma opção para a avaliação estatística de dados que não atendam aos pressupostos de normalidade da distribuição e homogeneidade das variâncias, são as análises não paramétricas. Contudo, com a utilização de análises não paramétricas, pode ocorrer a perda de informações, como por exemplo o grau com que as amostras diferem entre si. Desta forma, a disponibilidade de uma escala que resulte em dados que melhor atendam aos requisitos necessários à realização da ANOVA configura também uma demanda no campo da análise sensorial. Recentes trabalhos, como aqueles realizados por McPHERSON & RANDALL (1985), NAES (1990), GAY & MEAD (1992), BROCKHOFF & SKOVGAARD (1994), WILKINSON & YUKSEL (1997), têm demonstrado a preocupação em solucionar essa questão.

Pelo exposto, apresenta-se hoje, dentro da área de conhecimento intitulada Análise Sensorial, uma enorme demanda em se identificar uma escala para testes de consumidores que não possibilite apenas a avaliação do grau de aceitação de cada amostra com relação a uma referência, como também seja de fácil entendimento e uso por parte dos consumidores e ainda; produza respostas que atendam aos pressupostos de normalidade da distribuição e homoscedasticidade.

Uma vez que as escalas hedônicas e de diferença do controle apresentam individualmente algumas das várias características anteriormente mencionadas, o presente trabalho concentrou-se na adaptação das mesmas para utilização em testes com consumidor nos quais o uso de uma amostra referência é desejável. Esta escala adaptada foi por nós intitulada “escala hedônica com referência” e seu desempenho foi comparado aos desempenhos apresentados pelas escalas hedônica estruturada de 9 pontos e de estimativa da magnitude, com relação ao poder discriminativo entre amostras, com relação à variabilidade das respostas, e, finalmente, à adequação das respostas aos pressupostos necessários à realização da análise da variância (ANOVA).

## MATERIAIS E MÉTODOS

### **Amostras**

Sete marcas comerciais (A, B, C, D, E, F e G) de bebida refrigerante do tipo guaraná acondicionada em embalagem PET de 2 litros foram utilizadas no presente estudo. Cinco delas referiram-se a refrigerantes líderes nacionais de mercado (A, B, C, E e G), enquanto as demais consistiam de marcas pouco conhecidas, comercializadas apenas na região de Campinas – SP - (D e F).

Inicialmente, as sete amostras foram submetidas a um teste de ordenação-preferência (STONE & SIDEL, 1985) com o objetivo de classificá-las em amostras de baixa, média ou alta preferência e selecionar quais delas seriam avaliadas em posteriores testes de aceitação. No teste de ordenação-preferência as sete amostras foram avaliadas em uma única sessão por 56 consumidores de refrigerante e os resultados foram analisados pelo método de MacFarlane (MEILGAARD *et al.*, 1988).

A partir desses resultados, foram selecionadas as amostras mais e menos preferidas, bem como três amostras de preferências intermediárias, totalizando-se cinco amostras (B, C, D, E e G) que foram avaliadas nos testes de aceitação realizados posteriormente.

### **Recrutamento dos consumidores de guaraná**

Inicialmente, 108 questionários (Figura 1) foram distribuídos a alunos adolescentes do Colégio Técnico da Universidade Estadual de Campinas – Cotuca, com o objetivo de se recrutar consumidores de refrigerante do tipo guaraná que participariam da presente pesquisa.

Com base nas respostas aos questionários, foram recrutados 56 consumidores de guaraná, sendo metade do sexo feminino e metade do sexo masculino, obedecendo-se aos seguintes critérios: i) idade entre 15 e 18 anos; ii) aceitação de guaraná correspondente ou superior ao grau

“gosto moderadamente”; iii) consumir a bebida na frequência igual ou superior a 1 copo / lata por semana e; iv) manifestar interesse e disponibilidade em participar dos testes.

Questionário	
Nome: _____	Turma: _____ Data ____ / ____ / ____
Idade: _____	Sexo: ( ) F ( ) M
Por favor, marque na escala abaixo o quanto você gosta ou desgosta de refrigerante do tipo guaraná.	
<input type="checkbox"/> Gosto extremamente	
<input type="checkbox"/> Gosto muito	
<input type="checkbox"/> Gosto moderadamente	
<input type="checkbox"/> Gosto ligeiramente	
<input type="checkbox"/> Nem gosto / Nem desgosto	
<input type="checkbox"/> Desgosto ligeiramente	
<input type="checkbox"/> Desgosto moderadamente	
<input type="checkbox"/> Desgosto muito	
<input type="checkbox"/> Desgosto extremamente	
Por favor, marque na escala abaixo a frequência com que você costuma consumir guaraná.	
<input type="checkbox"/> Menos de 1 copo / lata por semana	
<input type="checkbox"/> De 1 a 3 copos / latas por semana	
<input type="checkbox"/> De 4 a 6 copos / latas por semana	
<input type="checkbox"/> Mais de 1 copo / lata por dia	

**FIGURA 1.** Questionário de recrutamento dos consumidores de refrigerante do tipo guaraná.

### Condições Gerais dos testes

Os testes sensoriais foram realizados a nível laboratorial (STONE & SIDEL, 1985) no Laboratório de Análise Sensorial do Colégio Técnico da Universidade Estadual de Campinas – Cotuca, em cabines individuais, iluminadas com luz incandescente branca.

As amostras foram servidas à temperatura de 12 +/- 2°C, em quantidade aproximada de 30 mL, em copos plásticos de 50 mL codificados com números de três dígitos definidos de forma aleatória. Nos testes em que uma referência se fez necessária, foram servidos 40 mL da mesma em recipientes identificados com a letra R.

A ordem com que os consumidores avaliaram as amostras seguiu o delineamento descrito por MACFIE *et al* (1989) para balanceamento dos efeitos intitulados “first-order and CARRY-over effect”, para testes de consumidores.

## **Consumidores**

Cinquenta consumidores, anteriormente selecionados, participaram do teste de aceitação, dos quais metade eram do sexo feminino e metade, do sexo masculino.

## **Escalas**

Todos os consumidores avaliaram todas as amostras utilizando as seguintes escalas: hedônica, estimativa de magnitude e hedônicas com referência.

Assim, cada consumidor participou de cinco sessões de avaliação: uma sessão utilizando a tradicional escala hedônica estruturada mista de 9 pontos (1 = desgostei extremamente; 5 = nem gostei / nem desgostei; 9 = gostei extremamente) (STONE & SIDEL, 1985); outra sessão, a escala de estimativa da magnitude para respostas hedônicas (GIOVANNI & PANGBORN, 1983); e, finalmente, três sessões adicionais utilizando, em cada uma delas, uma escala hedônica com referência (Figura 2). Foram três as escalas hedônicas com referência utilizadas, as quais diferiram entre si quanto à amostra utilizada como referência. Para a escala de estimativa de magnitude à amostra referência atribuiu-se o valor 10.

Como pode ser visualizado na Figura 2, ao se utilizar as escalas hedônicas com referência, cada provador foi solicitado a provar primeiramente a amostra referência e em seguida a amostra codificada, e avaliar o grau com que gostou mais ou menos da amostra codificada comparativamente à amostra referência. Este procedimento difere da escala de estimativa de magnitude no sentido de que o provador não é solicitado a trabalhar com frações matemáticas para expressar seu grau de gostar mais ou menos relativamente a uma referência, sendo, portanto, um processo de avaliação mais simples e direto.

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

Você está recebendo uma amostra referência de guaraná, identificada com a letra R e uma amostra codificada. Prove primeiramente a amostra referência (R). Em seguida, prove a amostra codificada e expresse o quanto você gostou ou desgostou da mesma, de forma comparativa à amostra referência R, usando a escala abaixo. Por exemplo, se você gostou muito mais da amostra codificada do que da referência, marque um x no espaço correspondente à frase “Gostei **muito mais da amostra** que da referência”. Ou, se você gostou ligeiramente menos da amostra codificada do que da referência, marque um x no espaço correspondente à frase “Gostei **ligeiramente menos da amostra** que da referência”.

Amostra: \_\_\_\_\_

- Gostei **extremamente mais da amostra** que da referência
- Gostei **muito mais da amostra** que da referência
- Gostei **moderadamente mais da amostra** que da referência
- Gostei **ligeiramente mais da amostra** que da referência
- Gostei igualmente da amostra e da referência
- Gostei **ligeiramente menos da amostra** que da referência
- Gostei **moderadamente menos da amostra** que da referência
- Gostei **muito menos da amostra** que da referência
- Gostei **extremamente menos da amostra** que da referência

Comentários: \_\_\_\_\_

**FIGURA 2.** Ficha de aplicação do teste de aceitação com escala hedônica com referência.

A ordem com que as escalas foram apresentadas aos consumidores obedeceram a um delineamento mostrado na Tabela 1, no qual se exemplifica a forma com que dez consumidores, cinco de cada sexo, foram alocados de forma aleatória a cada uma das seqüências de apresentação das escalas: 1<sup>a</sup>, 2<sup>a</sup>, 3<sup>a</sup>, 4<sup>a</sup> e 5<sup>a</sup>.

### **Escolha das Amostras Referências**

A definição de quais amostras seriam utilizadas como referência foi baseada nos resultados do teste preliminar de preferência. A amostra que obteve preferência intermediária (C) foi utilizada tanto como referência da escala de estimativa da magnitude como em uma das sessões realizadas com a escala hedônica com referência. As amostras mais preferida (B) e menos preferida (D) foram também utilizadas como amostras referências em outras duas sessões de avaliação com a escala hedônica com referência. Assim, foi possível se avaliar o efeito da escolha

da amostra referência sobre os resultados do teste com a escala hedônica com referência aqui proposta.

**TABELA 1.** Ordem de apresentação das escalas aos consumidores ao longo dos testes.

Grupo	Provedor	Escala*				
		1ª Sessão	2ª Sessão	3ª Sessão	4ª Sessão	5ª Sessão
1	de 1 a 10	1	2	3	4	5
2	de 11 a 20	2	3	4	5	1
3	de 21 a 30	3	4	5	1	2
4	de 31 a 40	4	5	1	2	3
5	de 41 a 50	5	1	2	3	4

\*1: escala hedônica tradicional; 2: escala hedônica com referência (ref. = amostra mais preferida); 3: escala hedônica com referência (ref. = amostra de preferência intermediária); 4: escala hedônica com referência (ref. = amostra menos preferida); 5: escala de estimativa de magnitude (ref. = amostra de preferência intermediária)

### **Análises estatísticas**

As escalas tiveram seus desempenhos comparados conforme os seguintes critérios: i) poder discriminativo com relação às amostras; ii) variabilidade entre as respostas dos consumidores e; iii) adequação aos pressupostos da análise da variância, ANOVA, quais sejam: normalidade da distribuição das respostas e homogeneidade da variância entre as amostras.

Para se avaliar o poder discriminativo de cada escala, com os dados obtidos através de cada uma das escalas foram utilizadas as seguintes ferramentas estatísticas: análise da variância (ANOVA), tendo as variáveis amostra e provedor como fontes de variação; teste de média de Tukey ou Dunnett, em função da escala que gerou os dados e; simulações do tipo “bootstrap”. Assim, as médias de aceitação de cada par de amostras, obtidas através da escala hedônica tradicional, foram comparadas através do teste de Tukey, procedimento já usualmente adotado para essa escala. Para as escalas com referência, comparou-se a média de aceitação de cada amostra com a média da amostra referência utilizando-se teste de Dunnett. O número de

discriminações das amostras com relação às referências (B, C e D) foi utilizado como um dos critérios de comparação do poder discriminativo das escalas.

Para se avaliar o desempenho da escala estimativa de magnitude e compará-lo ao das demais escalas, os dados obtidos com essa escala foram transformados através de cálculo baseado no trabalho de McDANIEL (1974): inicialmente, foi calculada a média geométrica de cada provador, com relação às amostras avaliadas. Essa média foi então utilizada para normalizar os resultados de cada provador para cada amostra, gerando os valores individuais amostra x provador, os quais foram submetidos aos mesmos critérios e tratamentos estatísticos utilizados para avaliação do desempenho das demais escalas.

O poder discriminativo de cada escala foi avaliado através da comparação dos valores do nível de significância de  $F_{amostra}$  obtidos para as cinco escalas estudadas: quanto maior o valor do nível de significância de  $F_{amostra}$ , maior foi considerado o poder discriminativo da escala.

Uma vez que, julgar o poder discriminativo das escalas tendo-se como base os valores dos níveis de significância de  $F_{amostra}$  de apenas um experimento conduziria a conclusões bastante frágeis e de baixa confiabilidade, com o objetivo de se obter resultados mais robustos com relação ao poder discriminativo das escalas, optou-se por realizar uma simulação “bootstrap” (EFRON & Tibshirani, 1993) a partir dos dados coletados com cada escala. Simulações utilizando a metodologia “bootstrap” permitem avaliar a repetibilidade dos resultados obtidos através da análise da distribuição da estatística aplicada, a qual neste experimento foi o nível de significância ( $p$ ) de  $F_{amostra}$  das respostas hedônicas atribuídas pelos consumidores às cinco amostras de guaraná.

Assim, para cada escala, foram obtidas 500 amostras “bootstrap”, cada qual formada por 50 consumidores. De forma similar ao delineamento proposto por MACFIE *et al* (1989), cada amostra “bootstrap” foi composta por 5 grupos de 10 consumidores cada. Os grupos que compuseram cada amostra “bootstrap” foram sorteados, com reposição, dentre os grupos previamente definidos no delineamento de apresentação das amostras (MACFIE *et al*, 1989), os quais são os mesmos grupos de consumidores da Tabela 1. Desta forma, um mesmo grupo

poderia estar presente nenhuma ou várias vezes em cada amostragem realizada. Esse procedimento permitiu preservar o balanceamento dos efeitos “first-order” e “carry-over”.

Com as respostas dos consumidores de cada amostra “bootstrap” foi realizada uma análise de variância (ANOVA), computando-se os valores dos níveis de significância de  $F_{amostra}$  para posterior elaboração de uma distribuição desses valores para cada escala. As distribuições dos valores dos níveis de significância de  $F_{amostra}$ , obtidas a partir das 500 amostragens, foram comparadas entre as escalas, tendo-se como critério de avaliação o valor de  $p$  abaixo do qual 90% dos valores dos níveis de significância de  $F_{amostra}$  localizaram-se: quanto menor esse valor, maior foi considerado o poder discriminativo da escala. Os valores dos níveis de significância obtidos foram também transformados em logaritmos neperianos e suas distribuições foram representadas em gráficos do tipo box-whisker, para melhor evidenciar as diferenças entre as escalas. As médias de  $\ln(p)$  foram submetidas ao teste de comparações múltiplas de Ryan-Einot-Gabriel-Welsch (EINOT & GABRIEL, 1975), com o objetivo de se detectar diferenças entre as distribuições dos valores de  $\ln(p)$ .

Uma vez que a análise de variância é baseada nos pressupostos de normalidade da distribuição das respostas e homogeneidade das variâncias das mesmas (DEVORE & PECK, 1986), para cada escala, o diagnóstico da adequação das respostas a tais pressupostos foi realizado através da análise dos resíduos. Para verificar a normalidade das distribuições das respostas foram realizados a análise visual do gráfico da probabilidade normal dos resíduos, bem como a aplicação do teste de normalidade de Kolmogorov-Smirnov (CONOVER, 1980). Uma vez que em amostragens de tamanho reduzido não é possível avaliar-se a normalidade das respostas analisando-se os gráficos com os dados de cada amostra separadamente, para a construção dos gráficos da probabilidade normal deste experimento agrupou-se os dados de todas as amostras, conforme sugestão de DEVORE & PECK (1986).

Para se avaliar a homogeneidade das variâncias em cada escala, foi realizada a análise da correlação ( $r$ ) entre os desvios padrões e as médias das amostras, computando-se o nível de significância de  $r$ . Em complementação, realizou-se também o teste de homogeneidade de Levene (MILLIKEN & JOHNSON, 1984).

A variabilidade entre as respostas dos consumidores em cada escala foi também avaliada, através da comparação dos coeficientes de variação de cada amostra em cada escala, bem como de seus coeficientes de variação médios, os quais foram submetidos a análise da variância (ANOVA) e teste de média de Tukey.

As respostas dos consumidores coletadas através de cada uma das escalas foram representadas em gráficos do tipo box-whisker. A partir desses gráficos as escalas hedônica tradicional e hedônica com referência foram comparadas, de forma exploratória, com relação às seguintes características: simetria da distribuição dos valores hedônicos, presença de valores extremos - "outliers", extensão da faixa de fato utilizada pelos consumidores em cada escala e amplitude interquartis (tamanho da caixa).

A construção dos gráficos do tipo box-whisker e de probabilidade normal dos resíduos, o teste de Levene, e a análise de correlação entre desvios padrões e médias de aceitação foram realizados utilizando-se o pacote estatístico STATISTICA, versão 5.0. As demais análises foram realizadas através do programa estatístico SAS - Statistical Analysis System (SAS Institute Inc., 1985).

## RESULTADOS E DISCUSSÕES

### Poder discriminativo das escalas

As médias de aceitação das amostras, obtidas através das cinco escalas, encontram-se na Tabela 2.

**TABELA 2.** Médias de aceitação das amostras de guaraná obtidas através de cinco escalas.

Amostra	Médias de Aceitação				
	Escala Hedônica <sup>1</sup>	Esc. Hedônica c/referência (refer. = B)	Esc. Hedônica c/referência (refer. = C)	Esc. Hedônica c/referência (refer. = D)	Esc. Est. da Magnitude <sup>2</sup> (refer. = C)
B	7.08 <sup>a</sup>	5.44	5.58	6.56 *	2.22 *
G	6.20 <sup>a</sup>	3.92 #	4.60	5.16	1.22
C	6.08 <sup>ab</sup>	4.20 #	5.16	5.50	1.66
E	5.00 <sup>bc</sup>	3.90 #	3.74 #	4.64	0.85 #
D	4.82 <sup>c</sup>	3.28 #	3.88 #	4.98	0.88 #
<sup>3</sup> p	5.00 x 10 <sup>-8</sup>	9.3 x 10 <sup>-9</sup>	3.42 x 10 <sup>-8</sup>	1.48 x 10 <sup>-6</sup>	1.00 x 10 <sup>-10</sup>

<sup>1</sup>Médias com letras em comum não diferiram significativamente ao nível de 5% quando realizado teste de Tukey; (1=desgostei extremamente; 5 = nem gostei / nem desgostei; 9 = gostei extremamente); #amostras que obtiveram médias de aceitação significativamente menores que a amostra referência a  $p \leq 0.05$  quando realizado teste de Dunnett; \*amostras que obtiveram médias de aceitação significativamente maiores que a amostra referência a  $p \leq 0.05$  quando realizado teste de Dunnett; B = amostra mais preferida; C = amostra de preferência intermediária; D=amostra menos preferida; <sup>2</sup>médias dos dados normalizados; <sup>3</sup>p: nível de significância de  $F_{amostra}$  obtido na análise da variância (ANOVA).

Conforme pode ser observado na Tabela 2, de um modo geral, as cinco escalas apresentaram a mesma tendência de preferência entre os consumidores: em todas as escalas utilizadas, enquanto a amostra B obteve maiores médias de aceitação, as amostras D e E obtiveram consistentemente as menores médias de aceitação.

Através da escala hedônica tradicional (Tabela 2) as amostras foram segmentadas em três grupos distintos: um primeiro grupo formado pelas amostras B e G, as quais apresentaram as maiores médias de aceitação; um segundo grupo, formado pelas amostras C e E, as quais apresentaram valores intermediários de aceitação e, finalmente, a amostra D, que apresentou menor aceitação entre os consumidores. Discriminação significativa a  $p \leq 0.05$  foi alcançada entre a amostra D, de menor aceitação, e as amostras B, G e C, bem como entre a amostra E e as amostras B e G.

Entre as três escalas hedônicas com referência, aquela que utilizou a amostra mais preferida como referência (amostra B) foi a que permitiu maior discriminação das amostras com relação à referência. De fato, observa-se na Tabela 2 que enquanto esta escala discriminou significativamente ( $p \leq 0.05$ ) todas as amostras com relação à referência, a escala que utilizou como referência a amostra de preferência intermediária (amostra C) discriminou significativamente somente duas amostras (D e E), e a escala que utilizou a amostra menos preferida como referência (amostra D) discriminou significativamente apenas uma amostra da referência (B).

Considerando-se o número de amostras que foram discriminadas das amostras referências (B, C e D), os resultados comentados acima sugerem que a escala hedônica tradicional apresenta menor capacidade discriminativa que as escalas hedônicas com referência, tendo como referência as amostras mais preferidas e de preferência intermediária.

Utilizando-se o mesmo critério, verifica-se a escala de estimativa da magnitude apresentou maior poder discriminativo que as demais escalas, pois enquanto a escala de estimativa de magnitude discriminou a amostra C ( $p \leq 0.05$ ) de três outras amostras (B, D e E), a escala hedônica com referência, que utilizou a amostra C como referência, discriminou apenas duas amostras (D e E) e, a escala hedônica tradicional discriminou somente uma amostra (D) da amostra C.

Utilizando-se os valores do nível de significância calculado através da ANOVA para o  $F_{amostra}$  como critério para se julgar o poder discriminativo de cada escala, verifica-se na Tabela 2

que, também sob esse critério, a escala de estimativa da magnitude apresentou maior poder discriminativo, uma vez que mostrou o menor nível de significância ( $p$ ) em relação às demais escalas ( $p = 1.00 \times 10^{-10}$ ). Através desse critério, a escala hedônica com referência, tendo como referência a amostra mais preferida (amostra B), foi a segunda mais discriminativa ( $p = 9.3 \times 10^{-9}$ ), seguida pelas escalas hedônica tradicional e hedônica com referência, tendo como referência a amostra de preferência intermediária (amostra C). A escala hedônica com referência que utilizou a amostra menos preferida como referência (amostra D), foi a escala menos discriminativa.

Embora todas as escalas tenham-se mostrado altamente discriminativas segundo seus valores dos níveis de significância de  $F_{amostra}$ , julgar o poder discriminativo das escalas tendo-se como base os resultados de apenas um experimento poderia conduzir a conclusões bastante frágeis, de baixa confiabilidade. Assim, para se obter resultados mais seguros com relação ao poder discriminativo das escalas, optou-se por realizar uma simulação “bootstrap” (EFRON & Tibshirani, 1993) a partir dos dados coletados com cada escala. Simulações utilizando-se a metodologia “bootstrap” permitem avaliar a repetibilidade dos resultados obtidos através da análise da distribuição da estatística aplicada, a qual neste experimento foi o nível de significância de  $F_{amostra}$  das respostas hedônicas atribuídas pelos consumidores às amostras de guaraná ( $p$ ). A Tabela 3 mostra a distribuição de  $p$  para as cinco escalas estudadas, obtidas através de 500 simulações do tipo “bootstrap” e expressas na forma de percentis. Os resultados apresentados na Tabela 3 confirmam que todas as escalas foram altamente discriminativas, uma vez que os valores dos níveis de significância de  $F_{amostra}$  obtidos em todas elas foram muito baixos e portanto, altamente significativos. Ainda assim, é possível verificar-se que as escalas de estimativa da magnitude e as escalas hedônicas com referência, utilizando as amostras mais preferida e de preferência intermediária como referência (amostra B), foram as escalas mais discriminativas, uma vez que 90% de seus valores de  $p$  situaram-se abaixo de  $10^{-6}$ , enquanto que a escala hedônica tradicional apresentou menor poder discriminativo, pois 90% dos valores de  $p$  localizaram-se abaixo de  $10^{-5}$ . A escala hedônica com referência, tendo a amostra menos preferida como referência (amostra D) apresentou menor poder discriminativo que as demais escalas, pois 90% dos valores de  $p$  mostraram-se inferiores a  $10^{-4}$ .

A Figura 3 apresenta, através de gráficos do tipo box-whisker, as distribuições dos logaritmos neperianos dos valores de  $p$  obtidas através da simulação “bootstrap” ( $\ln(p)$ ) para cada

escala. Esse gráfico complementa as informações contidas na Tabela 3: as escalas de estimativa de magnitude e hedônica com referência, utilizando a amostra mais preferida como referência (amostra B), mostraram menores valores de  $\ln(p)$  e portanto, menores níveis de significância de  $F_{amostra}$ ; a escala hedônica com referência que utilizou a amostra menos preferida como referência (amostra D) apresentou os maiores níveis de significância e; a escala hedônica tradicional, bem como a escala hedônica com referência tendo como referência a amostra de preferência intermediária (amostra C), mostraram-se muito similares com relação às distribuições de  $\ln(p)$ , apresentando valores intermediários em relação às demais escalas.

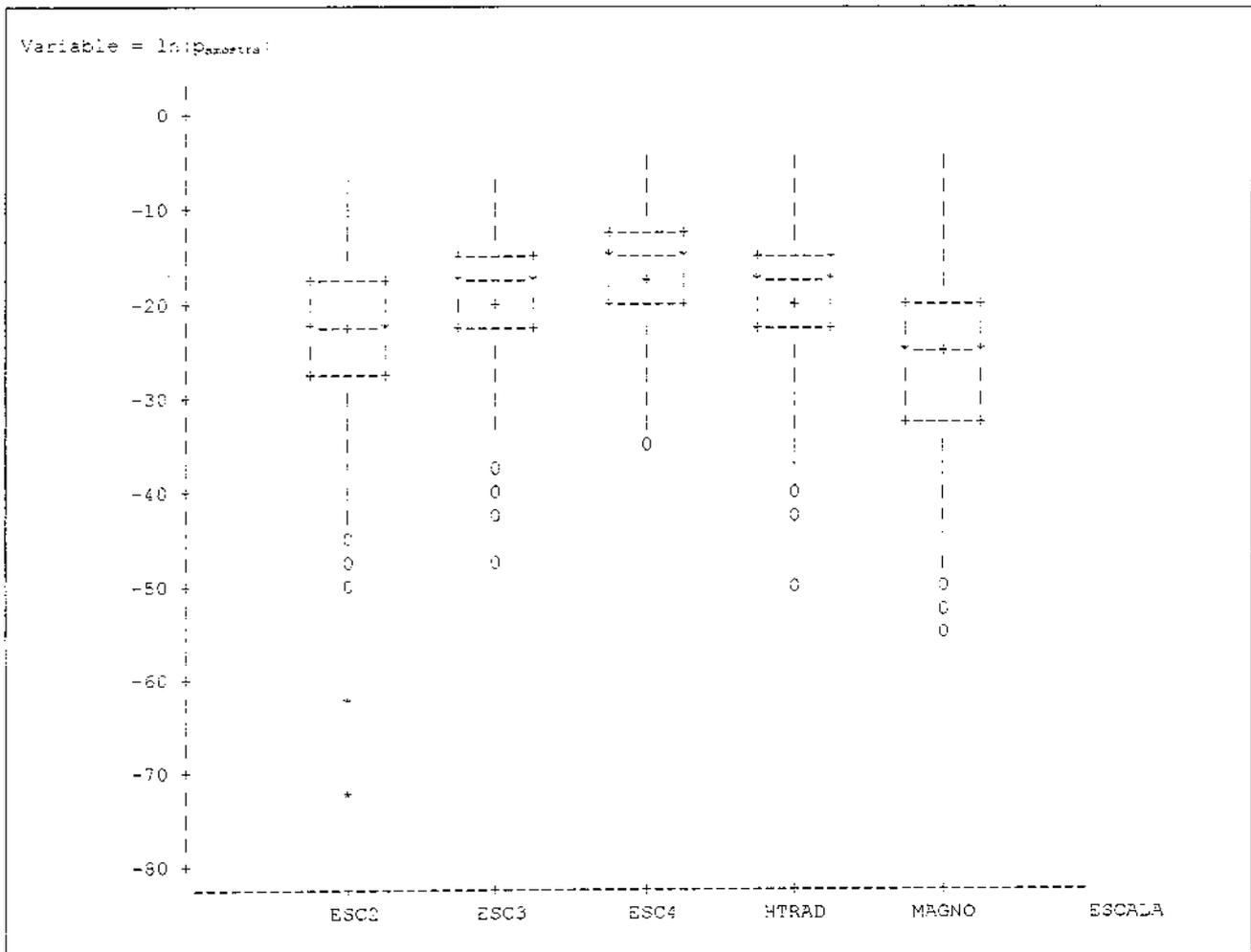
**TABELA 3.** Distribuição dos níveis de significância de  $F_{amostra}$  obtidos através da simulação “bootstrap”.

Escala	Percentuais relativos distribuição dos níveis de significância (p) de $F_{amostra}$ em percentis					
	1 %	5 %	10 %	90 %	95 %	99 %
<b>Hedônica Tradicional</b>	$6.44 \times 10^{-18}$	$2.03 \times 10^{-15}$	$2.66 \times 10^{-14}$	$1.90 \times 10^{-5}$	$6.10 \times 10^{-5}$	$6.17 \times 10^{-4}$
<b>Hedônica com Referência (mais preferida)</b>	$3.98 \times 10^{-22}$	$7.30 \times 10^{-18}$	$5.41 \times 10^{-16}$	$1.88 \times 10^{-6}$	$7.14 \times 10^{-6}$	$9.3 \times 10^{-5}$
<b>Hedônica com Referência (prefer. intermediária)</b>	$7.51 \times 10^{-19}$	$4.15 \times 10^{-15}$	$4.23 \times 10^{-13}$	$8.12 \times 10^{-6}$	$2.60 \times 10^{-5}$	$4.45 \times 10^{-4}$
<b>Hedônica com Referência (menos preferida)</b>	$1.52 \times 10^{-15}$	$4.65 \times 10^{-13}$	$5.85 \times 10^{-12}$	$1.15 \times 10^{-4}$	$3.46 \times 10^{-4}$	$1.27 \times 10^{-3}$
<b>Estimativa da Magnitude <sup>1</sup></b>	$3.55 \times 10^{-23}$	$5.23 \times 10^{-19}$	$1.89 \times 10^{-17}$	$4.46 \times 10^{-7}$	$3.05 \times 10^{-6}$	$1.97 \times 10^{-4}$

<sup>1</sup>análise realizada a partir dos dados normalizados.

Quando comparações múltiplas foram realizadas entre os valores médios de  $\ln(p)$  das escalas, os resultados sugeridos pela Tabela 3 e Figura 3 foram estatisticamente confirmados. A Tabela 4 mostra que a escala de estimativa de magnitude apresentou o maior poder discriminativo, uma vez que obteve a menor média de  $\ln(p)$ , a qual diferiu significativamente ( $p \leq 0.05$ ) das demais escalas. Segundo esse critério, a segunda escala mais discriminativa foi a escala hedônica com referência que utilizou a amostra mais preferida como referência (amostra B), cuja média de  $\ln(p)$  também diferiu significativamente em relação às demais escalas. As médias de  $\ln(p)$  das escalas hedônica tradicional e hedônica com referência, tendo como referência a

amostra de preferência intermediária (amostra C), não diferiram significativamente entre si ( $p \leq 0.05$ ), apresentando valores intermediários em relação às demais escalas. Finalmente, a escala hedônica com referência, cuja referência foi a amostra menos preferida (amostra D), apresentou o menor poder discriminativo, pois o valor da média de  $\ln(p)$  obtido para esta escala foi o maior e diferiu significativamente ( $p \leq 0.05$ ) daqueles encontrados para as demais escalas.



**FIGURA 3.** Distribuição dos valores dos logaritmos neperianos dos níveis de significância de  $F_{amostra}$ , obtidos através da simulação “bootstrap” para cada escala. **ESC2:** escala hedônica com referência (referência = amostra B, mais preferida); **ESC3:** escala hedônica com referência (referência = amostra C, de preferência intermediária); **ESC4:** escala hedônica com referência (referência = amostra D, menos preferida); **HTRAD:** escala hedônica tradicional; **MAGNO:** escala de estimativa da magnitude, utilizando-se os dados normalizados.

Assim, de um modo geral, os resultados da análise da variância (ANOVA) e dos testes de média realizados a partir das respostas hedônicas originalmente obtidas, bem como os resultados da simulação “bootstrap”, mostraram que a escala de estimativa da magnitude apresentou maior

poder discriminativo que as demais escalas, sendo seguida pela escala hedônica com referência, cuja referência foi a amostra mais preferida (amostra B). As escalas hedônica tradicional e hedônica com referência, tendo a amostra de preferência intermediária como referência (amostra C), apresentaram poderes discriminativos similares e intermediários em relação às demais escalas. Finalmente, a escala hedônica com referência que utilizou como referência a amostra menos preferida (amostra D), dentre todas as escalas, foi a que apresentou o menor poder discriminativo.

**Tabela 4.** Médias dos logaritmos neperianos dos níveis de significância de  $F_{amostra}$  para cada escala, obtidos em simulação “bootstrap”.

Escala	<sup>1</sup> Média de ln(p)
Estimativa da Magnitude <sup>2</sup>	-25,99 <sup>a</sup>
Hedônica com Referência (mais preferida)	-23,38 <sup>b</sup>
Hedônica com Referência (preferência intermediária)	-19,69 <sup>c</sup>
Hedônica Tradicional	-19,59 <sup>c</sup>
Hedônica com Referência (menos preferida)	-16,35 <sup>d</sup>

<sup>1</sup>médias com letras em comum não diferiram significativamente a  $p \leq 0.05$ , quando submetidas a teste de comparação múltipla; <sup>2</sup>análise realizada a partir dos dados normalizados.

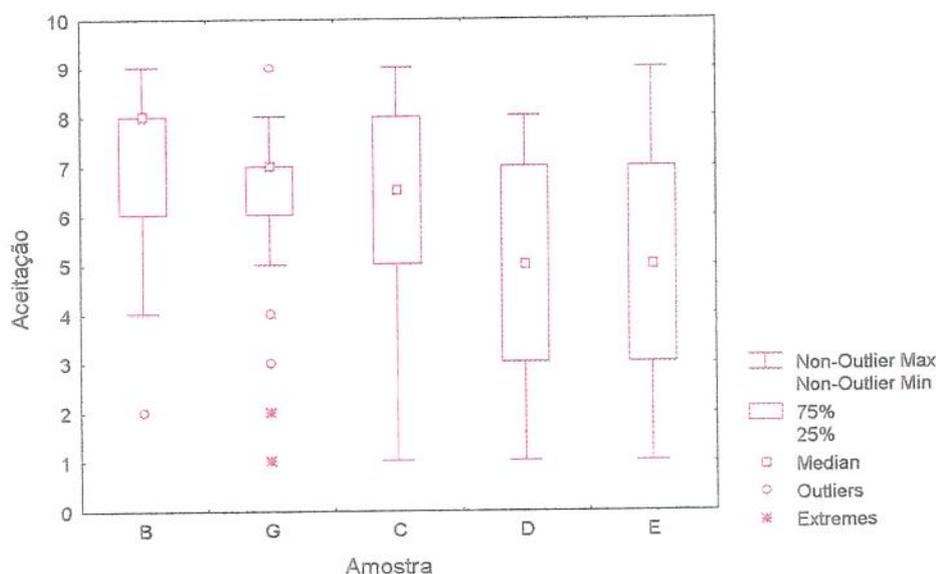
### Diagnóstico de Normalidade das Distribuições das Respostas

As distribuições dos valores hedônicos atribuídos pelos consumidores a cada uma das amostras, encontram-se representadas em gráficos do tipo box-whisker para cada escala (Figuras de 4 a 8).

Em gráficos do tipo box-whisker, a localização da mediana dentro da caixa, bem como o tamanho das caudas, indicam se há ou não simetria da distribuição das respostas e se esta aproxima-se da distribuição normal. Assim, a Figura 4 sugere que a escala hedônica tradicional apresentou forte desvio da normalidade para os valores hedônicos referentes às amostras B e G, uma vez que para essas amostras a mediana está localizada no limite superior da caixa, indicando assimetria da distribuição. O mesmo ocorreu com a escala hedônica com referência que utilizou a

assimetria da distribuição. O mesmo ocorreu com a escala hedônica com referência que utilizou a amostra menos preferida como referência (amostra D), conforme mostra a Figura 7. Nesta escala, a amostra C apresentou evidente assimetria da distribuição das respostas, uma vez que a mediana situou-se também no limite superior da caixa. Com exceção desses casos anteriormente mencionados, de um modo geral, as distribuições das respostas obtidas através das escalas estudadas apresentaram pequenos desvios da normalidade, os quais podem ser visualizados nas Figuras de 4 a 8.

As Figuras de 9 a 13 apresentam os gráficos de probabilidade normal dos resíduos obtidos em cada escala. Na interpretação desse tipo de gráfico, a distribuição das respostas é considerada mais próxima da normal quanto mais os pontos convergirem a uma reta. Desta forma, analisando-se as Figuras de 9 a 13, observa-se claramente que a distribuição dos resíduos referentes à escala de estimativa de magnitude é a única que apresenta um forte desvio da normalidade, enquanto que as distribuições das demais escalas são caracterizadas por apresentarem pequenos desvios em seus extremos, sugerindo que as mesmas não variaram substancialmente de uma distribuição normal (NETER *et al*, 1989).



**FIGURA 4.** Gráfico do tipo box-whisker a partir das respostas obtidas com a escala hedônica tradicional.

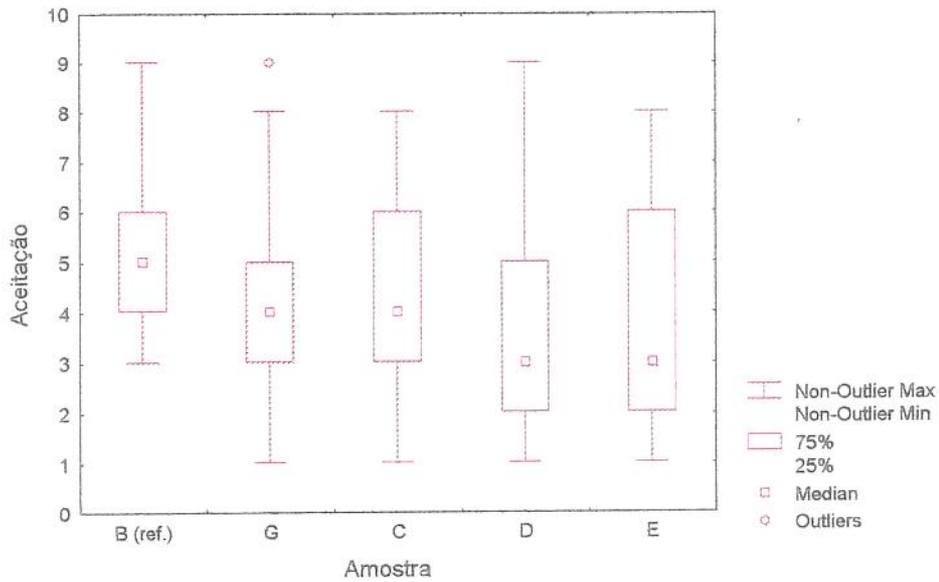


FIGURA 5. Gráfico do tipo box-whisker a partir das respostas obtidas com a escala hedônica com referência (referência: amostra B).

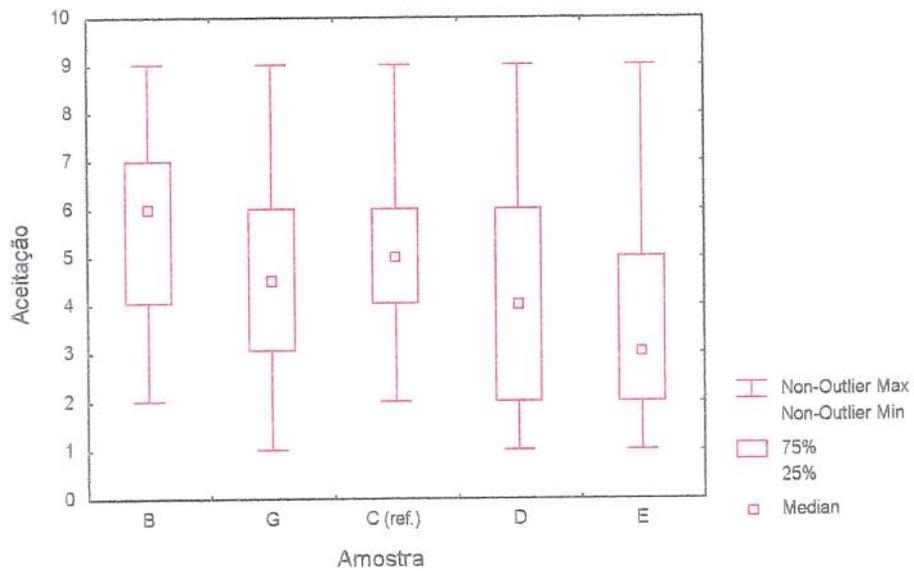


FIGURA 6. Gráfico do tipo box-whisker a partir das respostas da escala hedônica com referência (referência: amostra C).

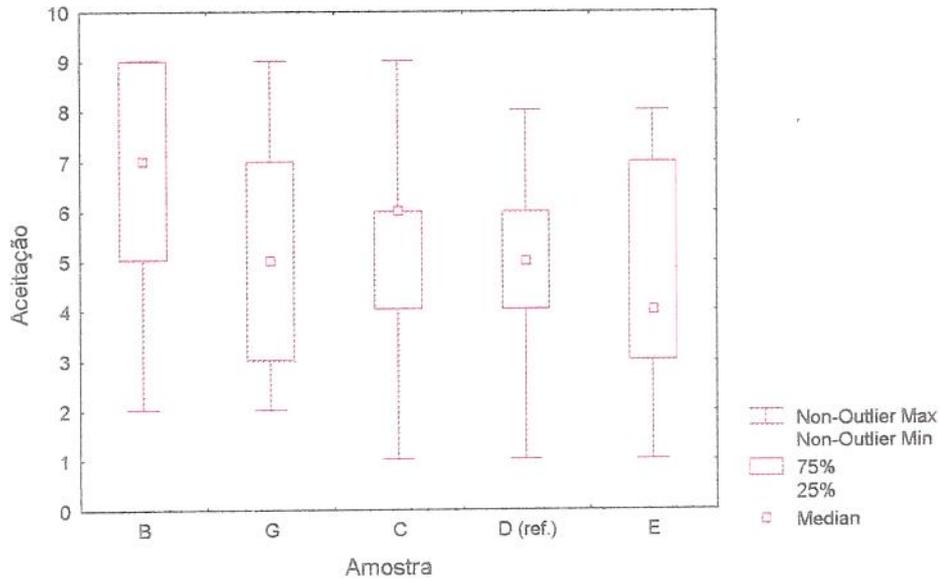


FIGURA 7. Gráfico do tipo box-whisker a partir das respostas da escala hedônica com referência (referência: amostra D).

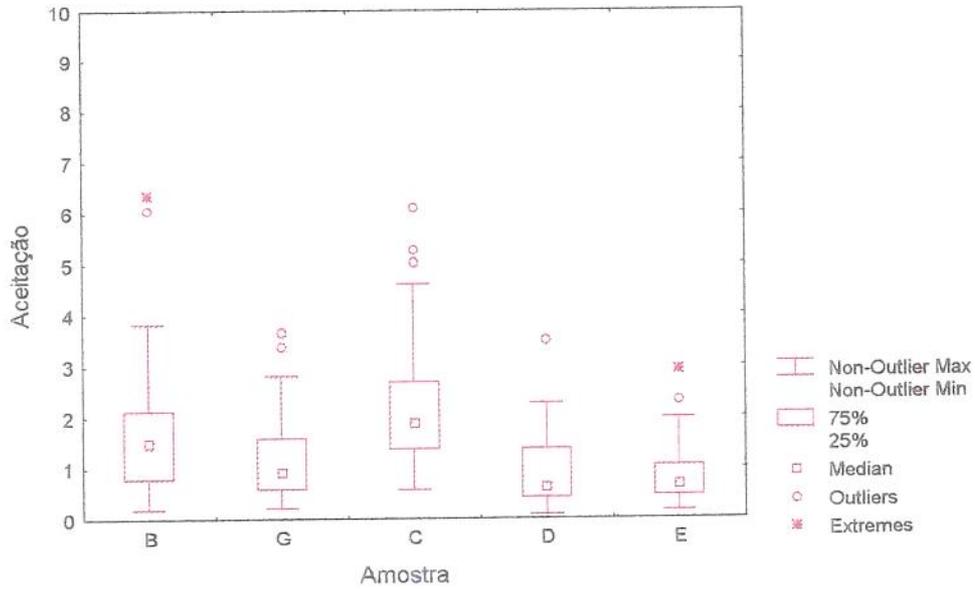
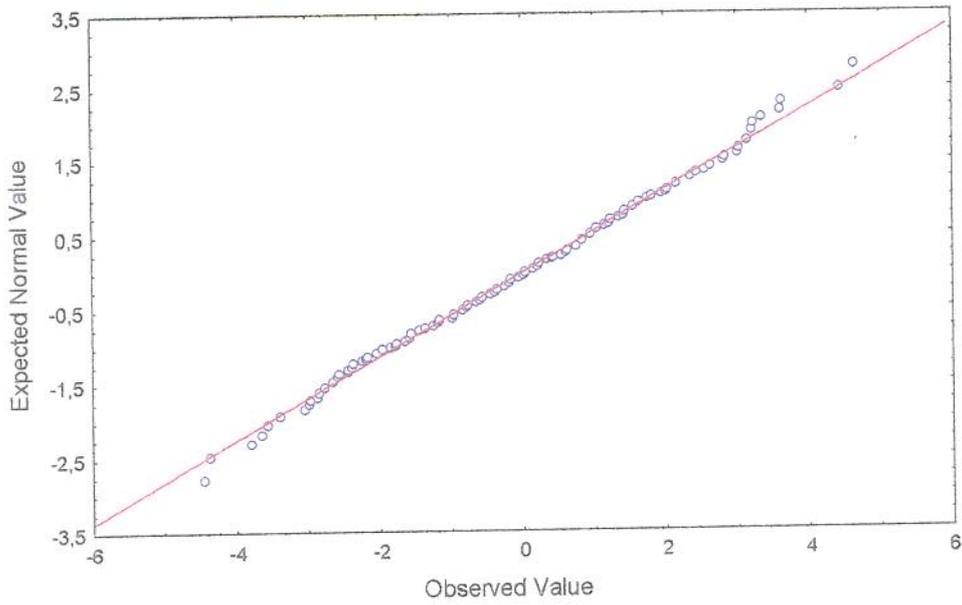
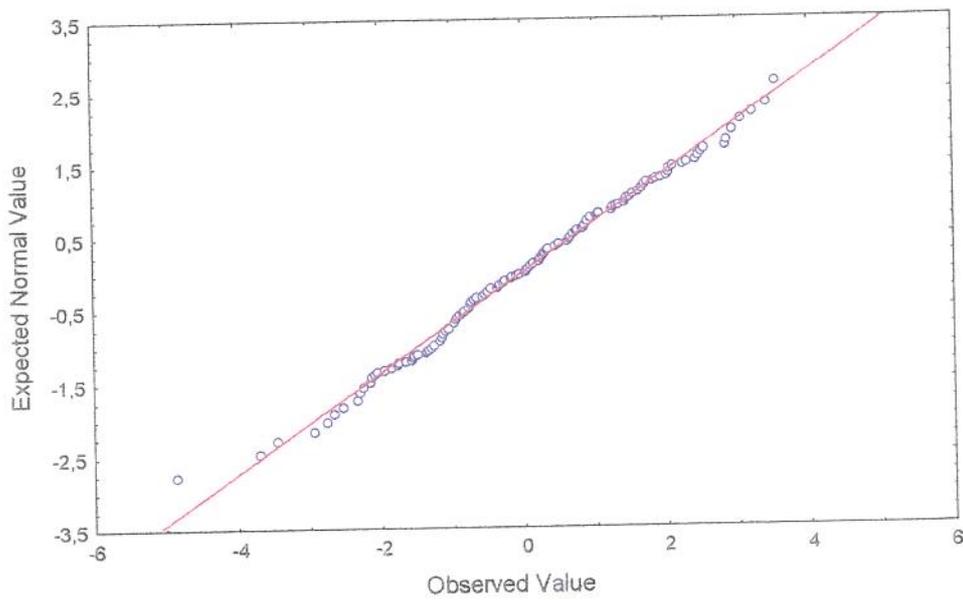


FIGURA 8. Gráfico do tipo box-whisker a partir dos dados normalizados da escala de estimativa de magnitude.



**FIGURA 9.** Gráfico da probabilidade normal dos resíduos referentes aos dados obtidos através da escala hedônica tradicional.



**FIGURA 10.** Gráfico da probabilidade normal dos resíduos referentes aos dados obtidos através da escala hedônica com referência (referência: amostra B).

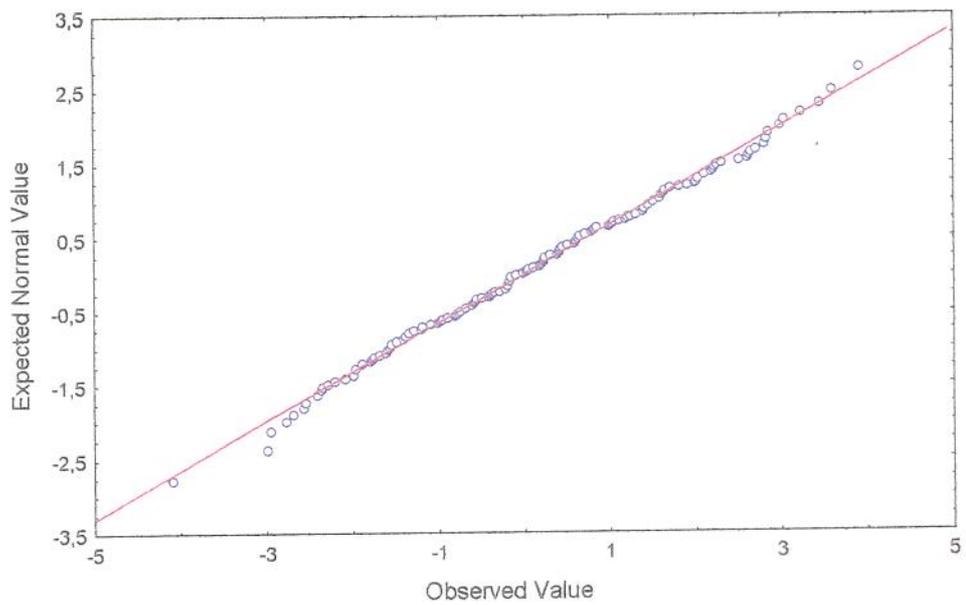


FIGURA 11. Gráfico da probabilidade normal dos resíduos referentes aos dados obtidos através da escala hedônica com referência (referência: amostra C).

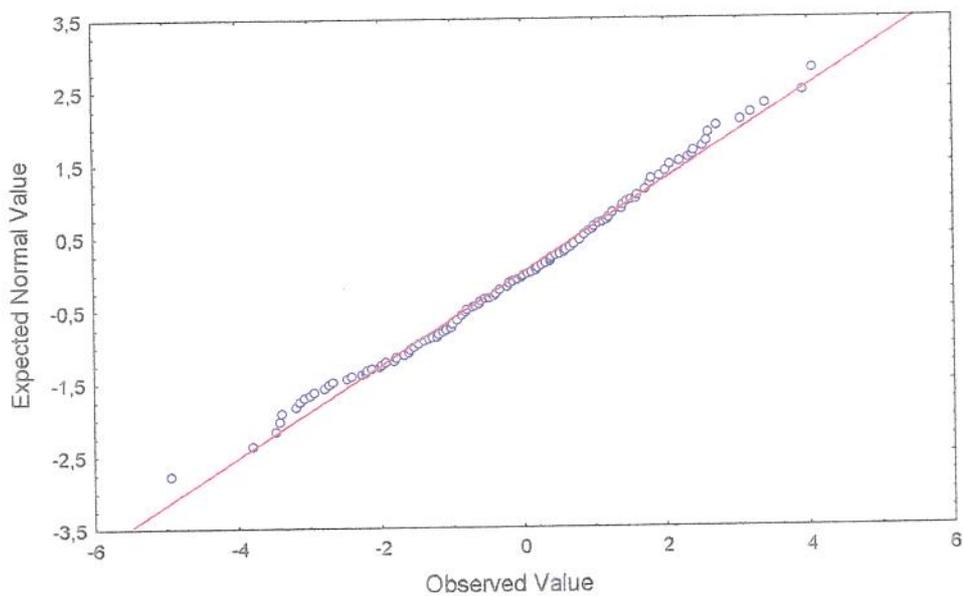


FIGURA 12. Gráfico da probabilidade normal dos resíduos referentes aos dados obtidos através da escala hedônica com referência (referência: amostra D).

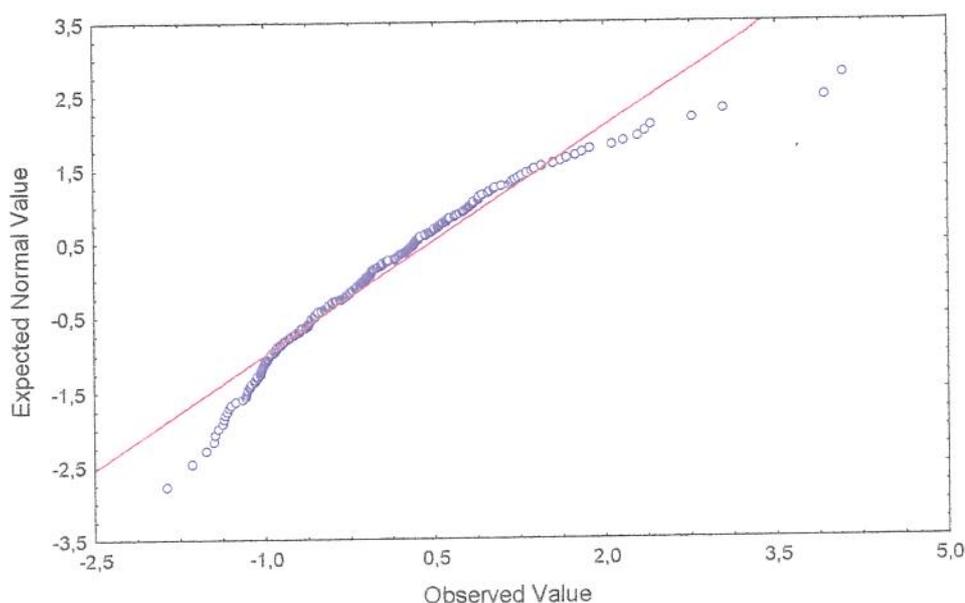


FIGURA 13. Gráfico da probabilidade normal dos resíduos referentes aos dados transformados obtidos através da escala de estimativa da magnitude.

Os resultados da análise da normalidade das distribuições das respostas dos consumidores, obtidos para cada escala através do teste de normalidade de Kolmogorov-Smirnov, encontram-se na Tabela 5. Neste teste, quanto maior for o valor do nível de significância ( $p$ ) obtido, mais próxima da normalidade será a distribuição das respostas hedônicas. Sendo assim, os resultados apresentados na Tabela 5, uma vez mais evidenciam que a distribuição das respostas fornecidas pela escala de estimativa da magnitude foi a que mais se distanciou da distribuição normal, um dado que, dentre todas as escalas, esta apresentou o menor nível de significância no teste de Kolmogorov-Smirnov, mostrando um valor de  $p$  igual a 0.0001. Por outro lado, a escala hedônica com referência que utilizou a amostra mais preferida como referência (amostra B) foi a que apresentou o maior valor de  $p$  neste teste ( $p = 0.5309$ ), sugerindo que a distribuição de suas respostas foi, dentre todas as escalas, a mais próxima da distribuição normal. Observa-se ainda que as outras duas escalas com referência apresentaram também níveis de significância superiores aos obtidos para as escalas de estimativa da magnitude e hedônica tradicional. Assim, neste estudo os resultados do teste de Kolmogorov-Smirnov sugeriram que, de um modo geral, as distribuições das respostas coletadas através das escalas hedônicas com referência apresentaram-se mais próximas da normalidade do que as distribuições das demais escalas.

De um modo geral, na presente pesquisa, os diagnósticos de normalidade das distribuições das respostas hedônicas mostraram que a escala de estimativa da magnitude foi aquela que gerou distribuições que mais se distanciaram da distribuição normal. Por outro lado, as distribuições relativas às escalas hedônicas com referência, dentre todas as escalas, foram as que mais se aproximaram de uma distribuição normal, notadamente quanto a referência utilizada foi a amostra mais preferida.

**TABELA 5.** Níveis de significância (p) obtidos nos testes de normalidade de Kolmogorov-Smirnov para todas as escalas estudadas.

Escala	Nível Significância (p)
Hedônica Tradicional	0.1839
Hedônica c/ Referência (+ preferida)	0.5309
Hedônica c/ Referência (preferência intermediária)	0.3233
Hedônica c/ Referência (- preferida)	0.4617
Estimativa da Magnitude <sup>1</sup>	0.0001

<sup>1</sup>análise realizada a partir dos dados normalizados.

### Diagnóstico de homogeneidade da variância das respostas

Gráficos do tipo box-whisker fornecem, de forma exploratória, informações sobre a variabilidade das respostas em cada escala, podendo observar-se a faixa da escala de fato utilizada pelos consumidores e a amplitude interquartis (tamanho da caixa), a qual indica a região da escala em que estão localizadas 50% das respostas. Para cada amostra, quanto maior a faixa utilizada da escala e/ou a amplitude interquartis, maior a variabilidade das respostas. Sendo assim, as Figuras de 4 a 7 sugerem que, de um modo geral, as escalas hedônicas com referência apresentaram maior variabilidade das respostas em relação à escala hedônica tradicional, ainda que somente para algumas amostras. Em contrapartida, essas figuras sugerem também a possibilidade de haver maior homogeneidade entre as variâncias das respostas das amostras nas avaliações realizadas através das escalas hedônicas com referência (Figuras de 5 a 7) em relação à escala hedônica tradicional (Figura 4).

Gráficos do tipo box-whisker (Figuras de 4 a 8) permitem também a detecção de valores extremos - “outliers”, os quais podem provocar distorções na análise dos dados. Através das Figuras de 4 a 8, é possível observar-se que as escalas hedônicas com referência reduziram a ocorrência de “outliers” em relação às demais escalas. Esse resultado sugere uma vantagem das escalas hedônicas com referência sobre as demais.

Comparando-se a variabilidade das respostas hedônicas entre os provadores, utilizando-se como critério os coeficientes de variação das amostras em cada escala (Tabela 6), verificou-se que a escala de estimativa de magnitude apresentou maior variabilidade das respostas, uma vez que seus coeficientes de variação mostraram-se superiores aos observados para as demais escalas, ainda que os mesmos tenham sido obtidos a partir dos dados normalizados, conforme procedimento citado por (McDANIEL, 1974). Ao se realizar análise da variância (ANOVA) e teste de Tukey sobre os coeficientes de variação correspondentes a cada amostra em cada escala, constatou-se que o coeficiente de variação médio da escala de magnitude diferiu significativamente ( $p \leq 0.05$ ) daqueles encontrados para as demais escalas (Tabela 6). As escalas hedônicas com referência mostraram variabilidade das respostas hedônicas similares à escala hedônica tradicional, uma vez que seus coeficientes de variação médios não diferiram significativamente ( $p \leq 0.05$ ) ao desta escala.

Os resultados da análise de homogeneidade das variâncias das amostras, obtidos através do teste de Levene, encontram-se expressos na Tabela 7. O teste de Levene avalia a homogeneidade da variância das respostas, através dos níveis de significância ( $p$ ) relativos à rejeição da hipótese nula  $H_0$ , sendo  $H_0$  uma pressuposição de igualdade da variância entre todas as amostras. Assim, se  $p > 0.05$ , pode-se assumir homogeneidade da variância entre as amostras. Desta forma, a Tabela 7 mostra que as escalas com referência apresentaram maior homogeneidade da variância que a escala hedônica tradicional e, notadamente, que a escala de estimativa da magnitude. Entre as escalas hedônicas com referência, aquelas que utilizaram as amostras mais e menos preferidas como referência (amostras B e D, respectivamente) apresentaram maior homogeneidade da variância, segundo o teste de Levene.

**TABELA 6.** Coeficientes de variação médios referentes aos valores hedônicos coletados através de cada escala.

Amostra	Coeficiente de Variação (%) (CV)				
	Escala Hedônica	Esc. Hedônica c/referência (referência = B)	Esc. Hedônica c/referência (referência = C)	Esc. Hedônica c/referência (referência = D)	Esc. Est. da Magnitude <sup>1</sup> (referência = C)
B	25.01	26.56	30.95	32.47	58.08
G	28.41	48.30	43.70	39.42	74.49
C	32.36	48.58	31.65	36.78	75.85
E	45.89	51.09	49.57	36.62	78.31
D	48.28	60.68	57.12	46.44	69.34
CV Médio <sup>2</sup>	35.99 <sup>b</sup>	47.04 <sup>b</sup>	42.60 <sup>b</sup>	38.35 <sup>b</sup>	71.21 <sup>a</sup>

B = amostra mais preferida; C = amostra de preferência intermediária; D = amostra menos preferida; <sup>1</sup>cálculos realizados a partir dos dados normalizados; <sup>2</sup>Coeficientes de variação médios com letras em comum não diferiram significativamente entre si ( $p \leq 0.05$ ), quando realizado teste de Tukey.

**TABELA 7.** Níveis de significância (p) obtidos no teste de homogeneidade de Levene, coeficientes de correlação (r) entre os desvios padrões e médias de aceitação das amostras de guaraná, bem como seus níveis de significância (p<sub>r</sub>), obtidos para cada escala.

Escala	Teste de Levene	Correlação DP x médias de aceitação	
	Nível de significância (p)	Coefficiente de correlação (r)	Nível de Significância de r (p <sub>r</sub> )
Hedônica Tradicional	0.0015	-0.94	0.019
Hed. c/ Refer. (+ preferida)	0.2174	-0.87	0.054
Hed. c/ Refer.(preferência interm.)	0.1284	ns	0.089
Hed. c/ Refer. (- preferida)	0.2762	ns	0.635
Estimativa da Magnitude <sup>1</sup>	0.000056	0.94	0.018

<sup>1</sup>análise realizada a partir dos dados normalizados; DP: desvio padrão de cada amostra; ns: r não significativo a  $p \leq 0.05$ .

Outro critério de avaliação da homogeneidade da variância é o valor do coeficiente de correlação ( $r$ ) da regressão linear obtida a partir dos desvios padrões em função das médias de aceitação das amostras, bem como o respectivo nível de significância ( $p$ ) de  $r$ . A  $p \leq 0.05$ , quanto mais próximo de 1 ou  $-1$  estiver o valor de  $r$ , maior será a correlação entre os desvios padrões e as médias das amostras e, conseqüentemente, menor homogeneidade haverá entre as variâncias. Assim, os resultados apresentados na Tabela 7 evidenciam que as variâncias das escalas hedônica tradicional e de estimativa da magnitude não mostraram-se homogêneas, uma vez que ambas apresentaram correlação linear significativa ( $p \leq 0.05$ ), negativa e positiva, respectivamente, com valores de  $r$  iguais a 0,94. Esses resultados indicam que para a escala hedônica tradicional, os valores de variância diminuíram com o aumento da aceitação das amostras; o inverso ocorreu com a escala de estimativa da magnitude. A escala hedônica com referência que utilizou a amostra mais preferida como referência (amostra B) pode receber o mesmo diagnóstico das duas escalas mencionadas, uma vez que o nível de significância do seu coeficiente de correlação apresentou-se muito próximo a 5%. A escala hedônica com referência que utilizou a amostra menos preferida como referência (amostra D) merece destaque por ter apresentado o maior nível de significância ( $p = 0.635$ ) em relação às demais escalas, e, conseqüentemente, maior homogeneidade da variância.

Portanto, os diagnósticos de homogeneidade da variância das respostas hedônicas, acima discutidos, revelaram que as escalas de estimativa da magnitude e hedônica tradicional mostraram-se inadequadas quanto à homogeneidade da variância, um pressuposto importante para realização de uma ANOVA. O mesmo diagnóstico foi obtido para a escala hedônica com referência, tendo a amostra mais preferida como referência (amostra B), quando o critério de avaliação utilizado foi a correlação significativa entre os desvios padrões e médias de aceitação. Finalmente, as escalas hedônicas com referência, tendo as amostras menos preferida (amostra D) e de preferência intermediária (amostra C) como referências, foram aquelas que apresentaram maior homogeneidade da variância das respostas hedônicas.

## CONCLUSÃO

Neste estudo, as escalas hedônicas com referência, tendo as amostras mais preferida e de preferência intermediária como referências, permitiram que mais amostras fossem discriminadas significativamente ( $p \leq 0.05$ ) em relação à amostra referência do que permitiu a escala hedônica tradicional. Por sua vez, a escala de estimativa de magnitude apresentou maior poder discriminativo do que as demais escalas, segundo esse critério de avaliação.

Quando observados os níveis de significância de  $F_{amostra}$ , obtidos tanto a partir dos dados originais como através das simulações “bootstrap”, todas as escalas estudadas mostraram-se altamente discriminativas. Entretanto, não se pode garantir que os valores dos níveis de significância de  $F_{amostra}$  relativos às escalas de magnitude e hedônica tradicional (Tabela 2) foram de fato iguais ou inferiores a 5%, uma vez que a estimativa do valor do nível de significância de  $F_{amostra}$  baseia-se em pressupostos de normalidade das distribuições e homogeneidade da variância das respostas, os quais não foram confirmados nos casos das escalas de estimativa de magnitude e hedônica tradicional. Quanto às escalas hedônicas com referência, somente aquela que utilizou como referência a amostra mais preferida (amostra B) apresentou-se ligeiramente inadequada com relação à homogeneidade da variância, uma vez que mostrou correlação significativa ( $p \leq 0.05$ ) das desvios padrões em função das médias de aceitação. As demais escalas hedônicas com referência foram aquelas que melhor atenderam aos pressupostos mencionados, apresentando portanto, alto grau de confiabilidade quanto ao nível de significância relativo aos testes de diferença entre médias de aceitação.

As escalas hedônica tradicional e hedônica com referência apresentaram-se similares com relação à variabilidade das respostas, uma vez que os coeficientes de variação médios dessas escalas não diferiram significativamente entre si a  $p \leq 0.05$ . Por outro lado, a escala de estimativa da magnitude apresentou maior variabilidade que as demais, pois seu coeficiente de variação médio foi superior e diferente significativamente ( $p \leq 0.05$ ) aos das demais escalas.

As conclusões acima sugerem que a escala hedônica com referência pode ser útil em situações nas quais o objetivo é a comparação da aceitação de um ou mais produtos com o

correspondente líder do mercado, uma vez que a mesma pode permitir maior discriminação entre este e as demais amostras, além de oferecer maior confiabilidade com relação à estimativa do nível de significância das diferenças entre médias de aceitação. No entanto, considerando-se a falta de número significativo de estudos que tenham utilizado essa escala em testes de consumidores, acreditamos ser recomendável estudos adicionais que permitam validar os resultados aqui obtidos. A realização de um maior número de pesquisas, que utilizem diferentes produtos e classes de consumidores, poderá confirmar a viabilidade, vantagens e limitações do uso da escala hedônica com referência como uma alternativa à escala hedônica tradicional e à escala de estimativa da magnitude.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. BROCKHOFF, M.; SKOVGAARD, M. Modeling individual differences between assessors in sensory evaluations. **Food Quality and Preference**, v.5, p.215-224, 1994.
2. CONOVER, W.J. **Practical Nonparametric Statistics**. Série: Wiley series in probability and mathematical statistics; John Wiley & Sons, Inc., p.493, 1980.
3. DEVORE, J.L.; PECK, R. **Statistics, the Exploration and Analysis of Data**. West Publishing Company, St. Paul, p.699, 1986.
4. EFRON, B.; TIBSHIRANI, R. **An Introduction to the Bootstrap**. Série: Monographs on Statistics and Applied Probability. Chapman & Hall. New York, p.436, 1993.
5. EINOT, I.; GABRIEL, K.R. A study of the powers of several methods of multiple comparisons. **J. American Statistical Association**, v.70, p.574-583, 1975.
6. EPLER, S.; CHAMBERS IV, E.; KEMP, K. Hedonic scales are a better predictor than just-about-right scales of optimal sweetness in lemonade. **J. Sensory Studies**, v.13, p.191-197, 1998.
7. GAY, C.; MEAD, R. A statistical appraisal of the problem of sensory measurement. **J. Sensory Studies**, v.7, p.205-228, 1992.
8. GIOVANNI, M.E.; PANGBORN, R.M. Measurement of taste intensity and degree of liking of beverages by graphic scales and magnitude estimation. **J. Food Science**, v.48, p.1175-1181, 1183, 1983.
9. LAWLESS, H.T.; MALONE, G.J. The discriminative efficiency of common scaling methods. **J. Sensory Studies**, v.1, p.85-98, 1986a.
10. MACFIE, H.J.; BRATCHELL, N.; GREENHOFF, K.; VALLIS, L.V. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. **J. Sensory Studies**, v.4, p.129-148, 1989.
11. McDANIEL, M.R. Magnitude estimation versus category scaling in food preference testing and food quality rating. Dissertação Ph.D. University of Massachusetts, Amherst, Mass, 1974.
12. McDANIEL, M.R.; SAWYER, F.M. Preference testing of whiskey sour formulations: magnitude estimation versus the 9-point hedonic. **J. Food Science**, v.46, p.182-185, 1981.

13. McPHERSON, R.S.; RANDALL, E. Line Length measurement as a tool for food preference research. **Ecology of Food and Nutrition**, v.17, p.149-156, 1985.
14. MEILGAARD, M.; CIVILLE, G.V.; CARR, B.T. **Sensory Evaluation Techniques**. CRC Press, Boca Raton, p.281, 1988.
15. MILLIKEN, G.A.; JOHNSON, D.E. **Analysis of Messy Data**. v.1: Designed experiments; Mason, R., Scheier, B., Eds. Van Nostrand Reinhold, New York, p. 473, 1984.
16. MOSKOWITZ, H.R.; SIDEL, J.L. Magnitude and hedonic scales of food acceptability. **J. Food Science**, v.36, p.677-680, 1971.
17. NAES, T. Handling individual differences between assessors in sensory profiling. **Food Quality and Preference**, v.2, p.187-199, 1990.
18. NETER, J.; WASSERMAN, W.; KUTNER, M.H. **Applied Linear Regression Models**. Boston, p.667, 1989.
19. O'MAHONY, M. Some assumptions and difficulties with common statistics for sensory analysis. In: **Overview: Outstanding Symposia in Food Science and Technology**. **Food Technology**, 75-82, 1982.
20. O'MAHONY, M. Sensory evaluation of food: statistical methods and procedures. Series: Food Science and Technology. Marcel Dekker, Inc. New York, p.487, 1986.
21. PERYAM, D.R.; PILGRIM, F.J. Hedonic scales of measuring food preferences. **Food Technology**, v.9 (11), p.9, 1957.
22. SAS - Statistical Analysis Systems. SAS user's guide basics. SAS Institute Inc., Cary, 1985.
23. SHEPHERD, R.; FARLEIGH, C.A.; LAND, D.G.; FRANKLIN, J.G. Validity of a relative-to-ideal rating procedure compared with hedonic rating. In: **Progress in Flavour Research**. Elsevier Science Publishers, Amsterdam, p.103-110, 1984.
24. STONE, H.; SIDEL, J.L. **Sensory Evaluation Practices**. Academic Press, Orlando, p.310, 1985.
25. VICKERS, Z.M. Magnitude estimation vs category scaling of the hedonic quality of food sounds. **J. Food Science**, v.48, p.1183-1186, 1983.
26. WILKINSON, C.; YUKSEL, D. Modeling differences between panelists in use of measurement scales. **J. Sensory Studies**, v.12, p.55-68, 1997.

## **CAPÍTULO III**

**Desempenho de uma “Escala Hedônica com Referência”  
para Obtenção de Modelos Preditivos de Aceitação a  
partir de Respostas Obtidas com Consumidores**

## RESUMO

Frequentemente, as indústrias alimentícias necessitam comparar um ou mais produtos com o correspondente líder do mercado para a identificação da preferência dos consumidores. Embora a tradicional escala hedônica estruturada possa ser utilizada nesses casos, o uso de uma amostra referência pode ser vantajoso, uma vez que pode aumentar a precisão e conseqüentemente o poder discriminativo do teste. No passado, a aplicação da escala de magnitude para respostas hedônicas foi exaustivamente estudada, entretanto, o uso da escala hedônica, ajustada a permitir a comparação entre várias amostras e uma amostra referência, não tem sido alvo de um número significativo de estudos. Desta forma, o presente trabalho teve como objetivo avaliar o desempenho de uma escala hedônica com referência para gerar modelos preditivos de aceitação a partir de respostas obtidas com consumidores. Foram elaboradas onze formulações de suco de laranja reconstituído, utilizando-se um delineamento fatorial  $2^2$ , intitulado delineamento composto central. As variáveis independentes estudadas foram: °Brix (9.0, 9.7, 11.5, 13.2, 14.0) e acidez (0.68, 0.74, 0.90, 1.05, 1.12 g/100 mL). A aceitação global das amostras foi avaliada por 44 consumidores, os quais utilizaram duas escalas: a escala hedônica estruturada de 9 pontos e uma escala hedônica com referência. Os dados foram analisados através da Metodologia de Superfície de Resposta, análise dos resíduos dos modelos e teste de média Dunnett. Ambas as escalas geraram modelos preditivos de aceitação que apresentaram altos valores de coeficiente de determinação ( $R^2$ ), quais sejam: 0.89 ( $p = 0.001$ ) para a escala hedônica tradicional e 0.91 ( $p = 0.0005$ ) para a escala hedônica com referência. Os modelos gerados através das duas escalas foram similares com relação à: não significância da falta de ajuste a  $p \leq 0.05$ , parâmetros que compuseram o modelo final ajustado, bem como a adequação dos resíduos aos pressupostos estatísticos inerentes à análise de regressão. As regiões de otimização obtidas foram similares para as duas escalas. O teste de média Dunnett indicou que a escala hedônica com referência proporcionou maior discriminação das amostras com relação à amostra referência. O desempenho das duas escalas para gerar modelos preditivos de aceitação foi equivalente.

## ABSTRACT

Frequently food industries need to compare one or more products with the corresponding market leader to identify consumer preference. Although the traditional structured hedonic scale can be used in these cases, the use of a reference sample can be an advantage, since it increases the precision and consequently the discriminating power of the test. In the past, the use of the magnitude scale for hedonic responses was exhaustively studied, but the use of the hedonic scale, adjusted to allow for the comparison between various samples and the reference sample, has not been the target of a significant number of studies. Thus the objective of this research was to evaluate the performance of a hedonic scale with a reference to generate predictive models of acceptance from the consumer responses obtained. Eleven formulations of reconstituted orange juice were prepared using a  $2^2$  factorial design called the central compound design. The independent variables studied were: °Brix (9.0, 9.7, 11.5, 13.2, 14.0) and acidity (0.68, 0.74, 0.90, 1.05, 1.12 g/100 mL). The global acceptance of the samples was evaluated by 44 consumers using the following two scales: a 9-point structured hedonic scale and a hedonic scale with reference. The data were analysed using Response Surface Methodology, analysis of the model residues and Dunnett's means test. Both scales generated predictive models of acceptance with high coefficients of determination ( $R^2$ ), these being: 0.89 ( $p = 0.001$ ) for the traditional hedonic scale and 0.91 ( $p = 0.0005$ ) for the hedonic scale with reference. The models generated by the two scales were similar with respect to: the non-significance of the lack of adjust at  $p \leq 0.05$ , the parameters comprising the final adjusted model and the adequation of the residues to the presupposed statistical data inherent in the regression analysis. The optimized regions were similar for the two scales. Dunnett's means test indicated that the hedonic scale with reference provided a greater discrimination of the samples with respect to the standard sample. The performance of the two scales in generating predictive models of acceptance was equivalent.

## INTRODUÇÃO

Freqüentemente, as indústrias produtoras de bens de consumo necessitam conhecer a opinião dos consumidores sobre características específicas ou gerais de um produto. Essa necessidade ocorre por razões de desenvolvimento de produtos, avaliação do mercado consumidor potencial, otimização / melhoria de produto ou, simplesmente, manutenção da competitividade do produto no mercado consumidor. Nessas situações, as indústrias recorrem com freqüência aos testes afetivos, sobretudo aos testes de aceitação (MEILGAARD *et al*, 1988). Testes de aceitação avaliam o estado afetivo do consumidor em relação ao produto, indicando o grau com o qual o consumidor gostou ou não do mesmo (STONE & SIDEL, 1985; MEILGAARD *et al*, 1988).

Testes de aceitação utilizam diferentes tipos de escalas: estruturadas, não estruturadas ou de estimativa da magnitude (MEILGAARD *et al*, 1988); porém, a escala de uso mais freqüente é a escala hedônica estruturada de nove pontos, desenvolvida e utilizada pela primeira vez por PERYAM & PILGRIM (1957). À essa escala se atribui as vantagens de fácil compreensão pelo consumidor; razoável reprodutibilidade das diferenças entre produtos em vários grupos consumidores; confiabilidade e validade dos resultados (STONE & SIDEL, 1985). Entretanto, deve-se lembrar que dados gerados através de escalas hedônicas estão sujeitos a variabilidades inerentes às respostas sensoriais em geral: variações entre indivíduos, os quais diferem com relação ao grau com que gostam de cada produto avaliado (MOSKOWITZ, 1981; CHAUCHAN & O'Mahony, 1993); variações devido ao contexto, ou seja, tipo e complexidade do estímulo, forma como o mesmo é apresentado, etc; dentre outras (MOSKOWITZ, 1981).

Freqüentemente, as indústrias alimentícias necessitam comparar um ou mais produtos com o correspondente líder do mercado para a identificação da preferência dos consumidores. Quando escalas hedônicas são utilizadas com esse propósito, a preferência de um produto sobre os demais é inferida através das médias de aceitação: produtos que apresentam médias de aceitação significativamente maiores que os demais ( $p \leq 0,05$ ) são considerados como preferidos. Entretanto, nesses casos a comparação direta com o objetivo de identificar a preferência / escolha do consumidor não é de fato realizada. Por sua vez, os testes de preferência, como o teste pareado, os quais realmente contemplam comparação direta entre produtos, muitas vezes não são

utilizados em testes de consumidor porque não informam o grau de aceitação do produto, apenas a escolha realizada por cada consumidor. Sendo assim, uma escala que permita a comparação direta entre uma ou várias amostras e também possibilite a avaliação do grau de aceitação do (s) produto (s) pode ser muito vantajosa para testes de consumidores. Em complemento, uma escala que permita aos consumidores expressarem suas respostas hedônicas em termos dos desvios positivos e negativos em relação a uma referência, pode inclusive aumentar a precisão e exatidão das respostas dos indivíduos, contribuindo para um aumento do poder discriminativo do teste.

A escala estimativa da magnitude para respostas hedônicas apresenta as características acima citadas, uma vez que permite avaliar uma ou mais amostras com relação a uma referência, por exemplo, um líder do mercado. No entanto, essa escala, foi considerada inadequada para respostas hedônicas (GIOVANNI & PANGBORN, 1983), ainda que VICKERS (1983) tenha verificado desempenho similar da mesma com relação aos das escalas estruturada e não estruturada em testes de consumidores. De fato, a escala estimativa de magnitude apresenta as seguintes limitações: requer dos provadores habilidade em calcular frações matemáticas; apresenta dificuldades de entendimento por parte dos consumidores sobre como utilizar a escala; e, finalmente, há controvérsias quanto à utilização numérica plena da mesma, ainda que ela seja uma escala aberta (infinita) (STONE & SIDEL, 1985; LAWLESS & MALONE, 1986a).

Além da escala de estimativa da magnitude, outra escala utilizada em testes de consumidores é a chamada escala do ideal (*just right scale*), a qual indica o quanto a intensidade de determinado atributo encontra-se mais forte ou mais fraca em relação àquele que o provador considera ideal (MEILGAARD *et al*, 1988); Neste caso o provador utiliza uma referência virtual própria e, assim, a escala do ideal apresenta o inconveniente da referência ser virtual, individual a cada provador e não conhecida.

Há outras escalas disponíveis para avaliações sensoriais que utilizam uma amostra referência. Uma delas é a escala de diferença do controle, que avalia o grau de diferença entre cada amostra e uma amostra controle. No entanto, essa escala está associada a avaliações de intensidade, enquanto que para avaliações hedônicas, estudos que utilizam amostra referência são quase inexistentes na literatura.

Portanto, a identificação de uma escala que permita tanto a comparação direta entre amostras com uma referência, como a avaliação do grau de aceitação do produto e que ainda seja de tão fácil uso que possa ser utilizada por consumidores em geral, mostra-se ainda como um desafio à área de análise sensorial.

Pelo exposto, o presente trabalho concentrou-se na avaliação do desempenho de uma escala caracterizada por: i) permitir a avaliação do grau de aceitação de cada amostra com relação a uma referência; ii) ser estruturada nas tradicionais expressões “gostei ligeiramente”, “moderadamente”, “muito” ou “extremamente” e; iii) apresentar facilidade de uso por parte de consumidores. Esta escala foi por nós intitulada “escala hedônica com referência”.

Uma vez que a obtenção de modelos estatísticos preditivos da aceitação de produtos tem sido objeto de interesse de muitos pesquisadores (CHI & CHEN, 1992; MOSKOWITZ, 1994; GUINARD *et al*, 1996; PASTOR *et al*, 1996; HEINIÖ *et al*, 1997; HOUGH *et al*, 1997) o desempenho da escala hedônica com referência foi avaliado, de forma comparativa ao da escala hedônica estruturada tradicional (PERYAM & PILGRIM, 1957), na obtenção de modelo preditivo de aceitação gerado através da Metodologia de Superfície de Resposta.

## MATERIAIS E MÉTODOS

### Delineamento Experimental

Diferentes amostras de suco de laranja reconstituído foram elaboradas a partir de suco de laranja concentrado congelado e aromatizado pela própria indústria fornecedora do produto (Citrosuco Paulista S.A., Matão, Brasil). O suco concentrado possuía as seguintes características: 65,80 °Brix corrigidos com relação à acidez, acidez 3,37 g/100g e ratio 19,53.

Para os testes sensoriais, o suco concentrado foi reconstituído com água potável a 8,8 °Brix em leitura direta no refratômetro. Onze tratamentos foram formulados adicionando-se diferentes níveis de ácido cítrico (P.A. - Laboratório Synth) e/ou sacarose (P.A. - Laboratório Vetec) de forma a se atingir os níveis de acidez e °Brix especificados no delineamento experimental apresentado na Tabela 1.

Para se determinar os níveis de ácido cítrico e sacarose a serem adicionados a cada tratamento, foram realizados testes preliminares, nos quais metodologias descritas por RED *et al* (1986) foram utilizadas para se verificar °Brix, acidez e ratio nas amostras.

Conforme mostra a Tabela 1, as amostras foram formuladas utilizando-se delineamento fatorial  $2^2$ , com pontos centrais e axiais (Delineamento Composto Central), conforme descrito por KHURI & CORNELL (1987).

Os níveis das variáveis independentes, °Brix e acidez, foram codificados em: i) -1 e +1, representando os níveis do planejamento fatorial  $2^2$ ; ii) 0 (zero), representando o ponto central do planejamento, o qual tornou possível a estimativa da falta de ajuste do modelo estatístico linear obtido, bem como do erro puro dos experimentos e; iii)  $-\alpha$  e  $+\alpha$ , representando os pontos axiais, que viabilizaram o estudo de um modelo estatístico quadrático.

**TABELA 1.** Tratamentos correspondentes ao delineamento experimental, mostrando valores reais e codificados das variáveis independentes estudadas e valor do "ratio" do suco formulado em cada tratamento.

Tratamento	Variáveis Independentes				Ratio
	Codificadas		Decodificadas		
	°Brix	Acidez	°Brix (corrigido)	Acidez (g / 100 mL)	
1	-1	-1	9.7	0.74	13.11
2	-1	1	9.7	1.05	9.24
3	1	-1	13.2	0.74	17.84
4	1	1	13.2	1.05	12.57
5	-1.4142	0	9.0	0.90	10.00
6	1.4142	0	14.0	0.90	15.55
7	0	-1.4142	11.5	0.68	16.91
8	0	1.4142	11.5	1.12	10.27
9	0	0	11.5	0.90	12.78
10	0	0	11.5	0.90	12.78
11	0	0	11.5	0.90	12.78

O número de tratamentos foi determinado pela Equação I,

$$N = F + 2k + n_0 \quad (\text{Equação I})$$

Na Equação I, k corresponde ao número de variáveis independentes ( $k = 2$ ). O valor de F refere-se ao número de tratamentos na porção fatorial que, segundo KHURI & CORNELL (1987), deve ser fixado através da fórmula  $F = 2^k$ . O valor de  $n_0$  refere-se à quantidade de replicatas no ponto central, neste caso  $n_0 = 3$ . Finalmente, N corresponde ao número total de tratamentos do experimento.

Assim, no presente experimento foram 11 (onze) o número total de tratamentos avaliados, dos quais quatro referiram-se à porção fatorial do experimento, quatro referiram-se à porção axial e três replicatas corresponderam ao ponto central, conforme apresentados na Tabela 1.

O valor codificado de  $\alpha$  foi definido em função do número de tratamentos na porção fatorial, segundo a Equação II.

$$\alpha = (F)^{1/4} = (2^k)^{1/4} \quad (\text{Equação II})$$

Portanto, para esse experimento  $\alpha=1,4142$ .

Os limites inferior e superior das faixas de °Brix e acidez estudadas foram selecionados com base em valores reais encontrados em suco natural de laranja proveniente de diferentes safras brasileiras. Desta forma, os valores limites de °Brix no suco pronto para consumo foram 9,0 e 14,0, enquanto que os valores limites de acidez situaram-se entre 0,68 e 1,12 g/100ml.

### **Recrutamento dos consumidores**

Inicialmente, 130 questionários (Figura1) foram distribuídos a estudantes e funcionários da Faculdade de Engenharia de Alimentos – FEA – da UNICAMP, com o objetivo de se recrutar consumidores de suco de laranja que participariam da presente pesquisa.

Com base nas respostas obtidas, foram recrutados 44 consumidores de suco de laranja, sendo 22 do sexo feminino e 22 do sexo masculino, obedecendo-se aos seguintes critérios: i) idade entre 20 e 45 anos; ii) aceitação do suco de laranja natural ou industrializado correspondente ou superior ao grau “gosto moderadamente”; iii) consumir suco de laranja natural ou industrializado na frequência igual ou superior a 1 copo por semana e; iv) manifestar interesse e disponibilidade em participar dos testes.

### Questionário

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

Contatos: Laboratório: \_\_\_\_\_ ramal: \_\_\_\_\_ Fone (resid.): \_\_\_\_\_  
e-mail: \_\_\_\_\_

Idade: \_\_\_\_\_ Sexo: ( ) F ( ) M

Por favor, marque na escala da esquerda o quanto você gosta ou desgosta de suco de laranja natural e na da direita, a frequência com que você costuma consumi-lo.

- |   |   |
|---|---|
| <input type="checkbox"/> Gosto extremamente       | <input type="checkbox"/> De 1 a 3 copos por mês       |
| <input type="checkbox"/> Gosto muito              | <input type="checkbox"/> Pelo menos 1 copo por semana |
| <input type="checkbox"/> Gosto moderadamente      | <input type="checkbox"/> De 1 a 3 copos por semana    |
| <input type="checkbox"/> Gosto ligeiramente       | <input type="checkbox"/> De 4 a 7 copos por semana    |
| <input type="checkbox"/> Nem gosto / Nem desgosto | <input type="checkbox"/> Mais de um copo por dia      |
| <input type="checkbox"/> Desgosto ligeiramente    |   |
| <input type="checkbox"/> Desgosto moderadamente   |   |
| <input type="checkbox"/> Desgosto muito           |   |
| <input type="checkbox"/> Desgosto extremamente    |   |

Indique na escala abaixo os lugares onde você costuma beber esse produto.

- em casa
- em restaurantes / casas noturnas
- em bares e lanchonetes
- em festas / reuniões sociais

Você já provou suco de laranja concentrado congelado (em lata; ex: Lanjal, Sucky, Max Frut)?

- Sim ( ) Não

Em caso afirmativo, por favor, marque na escala da esquerda o quanto você gosta ou desgosta de suco de laranja concentrado congelado e na da direita, a frequência com que você costuma consumi-lo.

- |   |   |
|---|---|
| <input type="checkbox"/> Gosto extremamente       | <input type="checkbox"/> De 1 a 3 copos por mês       |
| <input type="checkbox"/> Gosto muito              | <input type="checkbox"/> Pelo menos 1 copo por semana |
| <input type="checkbox"/> Gosto moderadamente      | <input type="checkbox"/> De 1 a 3 copos por semana    |
| <input type="checkbox"/> Gosto ligeiramente       | <input type="checkbox"/> De 4 a 7 copos por semana    |
| <input type="checkbox"/> Nem gosto / Nem desgosto | <input type="checkbox"/> Mais de um copo por dia      |
| <input type="checkbox"/> Desgosto ligeiramente    |   |
| <input type="checkbox"/> Desgosto moderadamente   |   |
| <input type="checkbox"/> Desgosto muito           |   |
| <input type="checkbox"/> Desgosto extremamente    |   |

**FIGURA 1.** Questionário de recrutamento dos consumidores de suco de laranja.

Por favor, indique na escala abaixo os lugares onde você costuma beber esse produto.

- em casa
- em restaurantes / casas noturnas
- em bares e lanchonetes
- em festas / reuniões sociais

Você já provou suco de laranja de caixinha, pronto para beber (ex: Izzy, Parmalat, Frully, Danone, Nestle, Melk, Jussy, Leco, etc)

- Sim                       Não

Em caso afirmativo, por favor, marque na escala da esquerda o quanto você gosta ou desgosta de suco de laranja de caixinha, pronto para beber, e na da direita, a frequência com que você costuma consumi-lo.

- |   |   |
|---|---|
| <input type="checkbox"/> Gosto extremamente       | <input type="checkbox"/> De 1 a 3 copos por mês       |
| <input type="checkbox"/> Gosto muito              | <input type="checkbox"/> Pelo menos 1 copo por semana |
| <input type="checkbox"/> Gosto moderadamente      | <input type="checkbox"/> De 1 a 3 copos por semana    |
| <input type="checkbox"/> Gosto ligeiramente       | <input type="checkbox"/> De 4 a 7 copos por semana    |
| <input type="checkbox"/> Nem gosto / Nem desgosto | <input type="checkbox"/> Mais de um copo por dia      |
| <input type="checkbox"/> Desgosto ligeiramente    |   |
| <input type="checkbox"/> Desgosto moderadamente   |   |
| <input type="checkbox"/> Desgosto muito           |   |
| <input type="checkbox"/> Desgosto extremamente    |   |

Por favor, indique na escala abaixo os lugares onde você costuma beber esse produto.

- em casa
- em restaurantes / casas noturnas
- em bares e lanchonetes
- em festas / reuniões sociais

Por motivos de saúde, você tem alguma objeção em participar de um teste de consumidor para avaliar suco de laranja adoçado com açúcar?

- Sim                       Não

Você gostaria de participar desse teste?                       Sim                       Não

FIGURA 1 (continuação).

### Teste de consumidor

Os testes de aceitação foram realizados a nível laboratorial (STONE & SIDEL, 1985) no Laboratório de Análise Sensorial da Faculdade de Engenharia de Alimentos da UNICAMP, em cabines individuais, iluminadas com luz fluorescente branca.

As amostras foram servidas em quantidade aproximada de 30 mL, em copos plásticos de 50 mL, codificados com números de três dígitos definidos de forma aleatória. Nos testes em que uma referência se fez necessária, esta foi apresentada na quantidade de 40 mL de suco.

Todos os provadores avaliaram todas as amostras utilizando uma escala hedônica estruturada mista de 9 pontos (1 = desgostei extremamente; 5 = nem gostei / nem desgostei; 9 = gostei extremamente) (STONE & SIDEL, 1985).

Todas as amostras foram também avaliadas por todos os provadores utilizando uma escala hedônica com referência, conforme mostra a Figura 2.

Como pode ser visualizado na Figura 2, ao utilizar a escala com referência, cada provador foi solicitado a provar primeiramente a amostra referência e em seguida a amostra codificada, e avaliar, o quanto gostou mais ou menos da amostra codificada que da referência.

A amostra utilizada como referência foi idêntica aos tratamentos correspondentes ao ponto central do delineamento apresentado na Tabela 1, tratamentos 9, 10 e 11.

Metade dos provadores avaliou as amostras utilizando primeiramente a escala hedônica tradicional e a outra metade avaliou as amostras utilizando primeiramente a escala hedônica com referência (Figura 2).

Para cada escala, a avaliação das amostras ocorreu através de cinco sessões distintas, servindo-se três amostras em cada sessão. A ordem com que cada provador avaliou cada uma das amostras ao longo das sessões fundamentou-se em delineamento para 11 amostras descrito por WAKELING & MACFIE (1995), o qual balanceia os efeitos intitulados “first-order carry-over effect”.

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

Você está recebendo uma amostra referência de suco de laranja concentrado congeladoreconstituído, identificada com a letra R, e uma amostra codificada. Por favor, prove primeiramente a amostra referência (R). Em seguida, prove a amostra codificada e expresse o quanto você gostou ou desgostou da mesma, de forma comparativa à amostra referência R, usando a escala abaixo. Por exemplo, se você gostou muito mais da amostra codificada do que da referência, faça um círculo no número correspondente à frase “Gostei muito mais da amostra que da referência”. Ou, se você gostou ligeiramente menos da amostra codificada do que da referência, faça um círculo no número correspondente à frase “Gostei ligeiramente menos da amostra que da referência”.

Amostra: \_\_\_\_\_

- +4. Gostei *extremamente mais da amostra* que da referência
- +3. Gostei *muito mais da amostra* que da referência
- +2. Gostei *moderadamente mais da amostra* que da referência
- +1. Gostei *ligeiramente mais da amostra* que da referência
- 0. Gostei igualmente das duas amostras
- 1. Gostei *ligeiramente menos da amostra* que da referência
- 2. Gostei *moderadamente menos da amostra* que da referência
- 3. Gostei *muito menos da amostra* que da referência
- 4. Gostei *extremamente menos da amostra* que da referência

Comentários: \_\_\_\_\_

**FIGURA 2.** Ficha de aplicação do teste de aceitação com escala hedônica com referência.

Uma vez que neste experimento apenas três amostras foram avaliadas por sessão, para se preservar o balanceamento do efeito “carry over”, o delineamento original foi adaptado. Assim, a última amostra testada numa sessão foi avaliada na primeira posição da sessão seguinte, na qual foi considerada como uma “dummy” amostra, sendo seu resultado desconsiderado na análise dos dados. A Tabela 2 exemplifica o delineamento utilizado para 11 provadores.

### **Análise dos resultados**

Para avaliar o desempenho da escala hedônica com referência, foi realizada a comparação dos modelos estatísticos preditivos da aceitação de suco de laranja, gerados pelas duas escalas. Os modelos foram obtidos através da Metodologia de Superfície de Resposta, conforme

BARROS NETO *et al* (1996), utilizando-se o programa STATISTICA, versão 5.0 e o programa SAS - Statistical Analysis System (SAS Institute Inc., 1985).

**TABELA 2.** Ordem de apresentação das amostras para cada consumidor em cada uma das sessões de teste.

Consumidor	Sessão 1			Sessão 2			Sessão3			Sessão4			Sessão 5		
	1	2	3	d	2	3	d	2	3	d	2	3	d	2	3
1	1	2	11	11	3	10	10	4	9	9	5	8	8	6	7
2	2	3	1	1	4	11	11	5	10	10	6	9	9	7	8
3	3	4	2	2	5	1	1	6	11	11	7	10	10	8	9
4	4	5	3	3	6	2	2	7	1	1	8	11	11	9	10
5	5	6	4	4	7	3	3	8	2	2	9	1	1	10	11
6	6	7	5	5	8	4	4	9	3	3	10	2	2	11	1
7	7	8	6	6	9	5	5	10	4	4	11	3	3	1	2
8	8	9	7	7	10	6	6	11	5	5	1	4	4	2	3
9	9	10	8	8	11	7	7	1	6	6	2	5	5	3	4
10	10	11	9	9	1	8	8	2	7	7	3	6	6	4	5
11	11	1	10	10	2	9	9	3	8	8	4	7	7	5	6

d: amostra considerada “dummy”, não utilizada na análise dos dados.

As respostas obtidas com a escala hedônica com referência, as quais consistiam em categorias mistas com números positivos e negativos, foram transformadas em números positivos de 1 a 9 para então serem calculadas as médias de cada tratamento. Foi considerada a seguinte equivalência: em lugar da resposta -4 foi utilizado o número 1; em lugar da resposta -3, o número 2; e assim sucessivamente, passando pelo número 5 ao invés da resposta 0 (zero) e terminando no número 9 em lugar da resposta +4.

As médias de aceitação obtidas através de cada escala, foram submetidas à análise de regressão, sendo modeladas pela equação geral polinomial quadrática expressa pela Equação III:

$$Y = \beta_0 + \beta_1B + \beta_2A + \beta_{11}B^2 + \beta_{22}A^2 + \beta_{12}BA + \varepsilon \quad (\text{Equação III})$$

Na Equação III, Y corresponde à função-resposta aceitação. B e A representam os níveis codificados das variáveis independentes °Brix e acidez, respectivamente.  $\beta_0$  representa o coeficiente relativo à interseção do plano com o eixo da aceitação, que corresponde à média geral das respostas.  $\beta_1$  e  $\beta_2$  são os coeficientes lineares das variáveis independentes.  $\beta_{11}$  e  $\beta_{22}$ , por sua vez, correspondem aos coeficientes das variáveis quadráticas.  $\beta_{12}$  representa o coeficiente de interação das variáveis independentes. Finalmente,  $\varepsilon$  representa o erro experimental.

Para cada modelo procedeu-se ao ajuste por mínimos quadrados, sendo testada a significância de cada parâmetro utilizando-se o teste t. Considerou-se para compor os modelos ajustados somente os parâmetros com nível de significância igual ou inferior a 10% ( $p \leq 0,10$ ). No modelo ajustado, a análise de regressão e teste t foram novamente realizados considerando-se somente esses parâmetros e a significância da falta de ajuste do modelo, foi então avaliada através da análise da variância (ANOVA) usando o teste de distribuição F.

Foi observado em cada modelo o coeficiente de determinação ( $R^2$ ), através do qual é avaliada a adequação do modelo: quanto mais próximo for seu valor de 1 ou 100%, melhor terá sido o ajuste do modelo e sua capacidade preditiva (KHURI & CORNELL, 1987; BARROS NETO *et al*, 1996).

NETER *et al* (1989) afirmam ser importante examinar a adequação do modelo de regressão aos dados estudados. Essa avaliação pode ser realizada de modo indireto através da análise dos resíduos do modelo. Em um modelo de regressão, os resíduos devem possuir as mesmas propriedades do erro verdadeiro ( $\varepsilon$ ), quais sejam: i) ser variável aleatória com distribuição normal; ii) possuir média 0 (zero) e iii) possuir variância constante  $\sigma^2$ . Neste trabalho, a normalidade da distribuição dos resíduos foi verificada através da análise visual do gráfico de probabilidade normal dos resíduos, e também através do nível de significância obtido no teste de normalidade de Kolmogorov-Smirnov (CONOVER, 1980). Também por análise gráfica verificou-se a constância da variância dos resíduos analisando-se o gráfico dos resíduos em função das variáveis independentes, °Brix e acidez.

Assim, os modelos das duas escalas foram então comparados em termos dos critérios acima citados: i) nível de significância ( $p$ ) de cada parâmetro do modelo; ii) observação de quais parâmetros compõem cada modelo; iii) o nível de significância da falta de ajuste do modelo; iv) o coeficiente de determinação de cada modelo ( $R^2$ ) e; v) análise dos resíduos.

A partir das equações de regressão obtidas pelas duas escalas, foi possível elaborar gráficos tridimensionais de superfície de resposta que permitiram visualizar as regiões de otimização da aceitação do suco de laranja.

Escalas cujos modelos apresentaram baixos níveis de significância de seus parâmetros, maior coeficiente de determinação ( $R^2$ ) e bons resultados na análise dos resíduos foram consideradas escalas de bom desempenho para fornecerem modelos preditivos de aceitação.

Para se ter idéia de qual escala proporcionou melhor discriminação entre as amostras, os dados de ambas as escalas foram submetidos à análise da variância (ANOVA) e teste de média Dunnett. Para essa análise foi utilizada a amostra 10 como amostra controle.

## RESULTADOS E DISCUSSÃO

As médias de aceitação das amostras, obtidas através das duas escalas avaliadas neste estudo, são mostradas na Tabela 3.

**TABELA 3.** Médias de aceitação dos tratamentos correspondentes ao delineamento experimental, obtidas com cada escala.

Tratamento	Médias de Aceitação	
	Escala Hedônica <sup>1</sup>	Esc. Hedônica c/ Referência <sup>2</sup>
1	5.98	4.59
2	4.52 <sup>#</sup>	3.43 <sup>#</sup>
3	6.84	6.61 <sup>*</sup>
4	5,82	6.02
5	5.11 <sup>#</sup>	3.89 <sup>#</sup>
6	6.91	6.41 <sup>*</sup>
7	7.02 <sup>*</sup>	6.82 <sup>*</sup>
8	5.52	4.11 <sup>#</sup>
9	6.66	5.95
10	6.11	5.45
11	6.64	6.07
MDS <sup>3</sup>	0.84	0.88

<sup>1</sup>(1 = desgostei extremamente, 5 = nem gostei / nem desgostei, 9 = gostei extremamente); <sup>2</sup>(1 = gostei extremamente menos da amostra que da referência, 5 = gostei igualmente da amostra e da referência, 9 = gostei extremamente mais da amostra que da referência); <sup>\*</sup>amostras que obtiveram média de aceitação significativamente maior que a amostra referência (amostra 10) a  $p \leq 0,05$  quando realizado teste de dunnett; <sup>#</sup>amostras que obtiveram média de aceitação significativamente menor que a amostra referência (amostra 10) a  $p \leq 0,05$  quando realizado teste de dunnett; <sup>3</sup>mínima diferença significativa ( $p \leq 0,05$ ) obtida pelo teste de Dunnett.

Embora as duas escalas utilizem diferentes mecanismos para a avaliação do grau de aceitação das amostras, os resultados sugerem uma mesma tendência de preferência dos

consumidores: as amostras 3, 6 e 7 obtiveram maior aceitação entre os consumidores, enquanto a amostra 2 obteve a menor aceitação. É interessante notar que as amostras 9, 10 e 11, de igual formulação e utilizadas como amostra referência para a escala hedônica com referência, receberam médias de aceitação próximas a 6, correspondendo ao termo hedônico “gostei ligeiramente mais da amostra que da referência” quando o natural seria esperar médias próximas a 5,0 para esses tratamentos, ou seja, correspondente à expressão “gostei igualmente da amostra e da referência”.

As Tabelas 4 e 5 mostram os resultados das análises de regressão que originaram os modelos preditivos completos ( $p \leq 0,05$ ) da aceitação de suco de laranja: a Equação IV para a escala hedônica e a Equação VI para a escala hedônica com referência.

Comparando-se entre as duas escalas os coeficientes de determinação ( $R^2$ ) de cada modelo (Equações IV e VI), os respectivos níveis de significância ( $p$ ), bem como os níveis de significância de cada parâmetro, verifica-se que os resultados obtidos pelas duas escalas são muito similares: ambas produziram modelos completos com  $R^2$  igual a 0,93 e níveis de significância iguais a 0,07 (Equação IV, escala hedônica) e 0,06 (Equação VI, escala hedônica com referência).

Ao comparar um a um os níveis de significância dos parâmetros de cada modelo (Tabelas 4 e 5), verifica-se uma grande similaridade entre as escalas. De fato, em ambas as escalas os parâmetros linear e quadrático da variável Brix e linear da variável acidez foram selecionados para permanecerem no modelo ajustado, enquanto o parâmetro quadrático da acidez, bem como a interação Brix x acidez foram excluídos do modelo final ajustado.

As Equações V e VII (Tabelas 4 e 5) mostram os modelos ajustados referentes às escalas hedônica e hedônica com referência, respectivamente.

Comparando-se os modelos ajustados das duas escalas, observou-se grande proximidade entre os níveis de significância dos parâmetros correspondentes ( $p$ ), bem como dos coeficientes de determinação ( $R^2$ ), sugerindo similaridade de desempenho entre as duas escalas. Conforme

pode ser observado nas Tabelas 4 e 5, para ambas as escalas, o ajuste do modelo provocou a redução dos níveis de significância dos parâmetros mantidos na equação preditiva, com exceção do parâmetro quadrático da variável Brix.

**TABELA 4.** Coeficientes de regressão<sup>1</sup>, valores t, níveis de significância dos parâmetros, coeficientes de determinação dos modelos completo e ajustado obtidos a partir da escala hedônica tradicional.

<b>Modelo Completo</b>			
<b>Parâmetro</b>	<b>Coefficiente de regressão</b>	<b>t</b>	<b><sup>2</sup>p</b>
<b>Média Geral (<math>\beta_0</math>)</b>	6.82	36.41	0.000000
<b>°Brix (B)</b>	0.59	5.39	0.002962
<b>°Brix (<math>B^2</math>)</b>	-0.32	-2.45	0.057892
<b>Acidez (A)</b>	-0.57	-5.28	0.003253
<b>Acidez (<math>A^2</math>)</b>	-0.19	-1.44	ns
<b>Interação (BA)</b>	0.11	0.71	ns
$Y = 6.47 + 0.59B - 0.57A - 0.32B^2 - 0.19A^2 + 0.11BA$ (Equação IV)			
$R^2 = 0.93$		$p = 0.007$	
<b>Modelo Ajustado</b>			
<b>Parâmetro</b>	<b>Coefficiente de regressão</b>	<b>t</b>	<b><sup>2</sup>p</b>
<b>Média Geral (<math>\beta_0</math>)</b>	6.29	46.77	0.000000
<b>°Brix (B)</b>	0.59	5.18	0.001279
<b>°Brix (<math>B^2</math>)</b>	-0.26	-2.04	0.080988
<b>Acidez (A)</b>	-0.57	-5.07	0.001445
$Y = 6.29 + 0.59B - 0.57A - 0.26B^2$ (Equação V)			
$R^2 = 0.89$		$p = 0.001$	

<sup>1</sup>modelo obtido com variáveis independentes codificadas; <sup>2</sup>nível de significância; ns: parâmetro não significativo a  $p \leq 0.10$ .

**TABELA 5.** Coeficientes de regressão<sup>1</sup>, valores t, níveis de significância dos parâmetros, coeficientes de determinação dos modelos completo e ajustado obtidos a partir da escala hedônica com referência.

<b>Modelo Completo</b>			
<b>Parâmetro</b>	<b>Coefficiente de regressão</b>	<b>t</b>	<b><sup>2</sup>p</b>
<b>Média Geral (<math>\beta_0</math>)</b>	5.83	23.42	0.000003
<b>°Brix (B)</b>	1.02	6.71	0.001111
<b>°Brix (<math>B^2</math>)</b>	-0.37	-2.06	0.093778
<b>Acidez (A)</b>	-0.70	-4.57	0.005979
<b>Acidez (<math>A^2</math>)</b>	-0.21	-1.19	ns
<b>Interação (BA)</b>	0.14	0.66	ns
$Y = 5.83 + 1.02B - 0.70A - 0.37B^2 - 0.21A^2 + 0.14BA$ (Equação VI)			
$R^2 = 0.93$		$p = 0.006$	
<b>Modelo Ajustado</b>			
<b>Parâmetro</b>	<b>Coefficiente de regressão</b>	<b>t</b>	<b><sup>2</sup>p</b>
<b>Média Geral (<math>\beta_0</math>)</b>	5.62	31.41	0.000000
<b>°Brix (B)</b>	1.02	6.79	0.000256
<b>°Brix (<math>B^2</math>)</b>	-0.31	-1.81	0.112308
<b>Acidez (A)</b>	-0.70	-4.62	0.002411
$Y = 5.62 + 1.02B - 0.70A - 0.31B^2$ (Equação VII)			
$R^2 = 0.91$		$p = 0.0005$	

<sup>1</sup>modelo obtido com variáveis independentes codificadas; <sup>2</sup>nível de significância; ns: parâmetro não significativo a  $p \leq 0.10$ .

As Tabelas 6 e 7 mostram que os modelos ajustados das duas escalas apresentaram falta de ajuste não significativa ( $p \leq 0,05$ ) e as Tabelas 4 e 5, que os mesmos apresentaram coeficientes de determinação ( $R^2$ ) ligeiramente inferiores aos dos modelos completos. A regressão do modelo ajustado a partir dos dados da escala hedônica explicou 89% da variação da aceitação do suco de laranja, enquanto que a do modelo gerado a partir dos dados da escala

hedônica com referência, 91% (Tabelas 4 e 5). Esses coeficientes de determinação, encontram-se iguais ou superiores àqueles usualmente reportados na literatura (CHI & CHEN, 1992; MOSKOWITZ, 1994; PASTOR *et al*, 1996; HOUGH *et al*, 1997).

**TABELA 6.** Análise da variância (ANOVA,  $p \leq 0,05$ ) do modelo ajustado a partir dos dados da escala hedônica (Equação V).

Fonte de variação	Soma Quadrática	Grau de Liberdade	Média Quadrática	F <sub>calculado</sub>	F <sub>tabelado</sub>
Regressão	5.82	3	1.94	18.91	4.35
Resíduos	0.72	7	0.10		
Falta de ajuste	0.53	5	0.10	1.11	19.30
Erro puro	0.19	2	0.09		
Total	6.54	10			

**TABELA 7.** Análise da variância (ANOVA,  $p \leq 0,05$ ) do modelo ajustado a partir dos dados da escala hedônica com referência (Equação VII).

Fonte de variação	Soma Quadrática	Grau de Liberdade	Média Quadrática	F <sub>calculado</sub>	F <sub>tabelado</sub>
Regressão	12.85	3	4.28	23.59	4.35
Resíduos	1.27	7	0.18		
Falta de ajuste	1.06	5	0.21	1.99	19.30
Erro puro	0.21	2	0.11		
Total	14.12	10			

As Figuras 3 e 4 apresentam, para as duas escalas avaliadas, os gráficos de probabilidade normal dos resíduos. Nos gráficos correspondentes aos dados de ambas as escalas (Figuras 3 e 4), observa-se que os pontos convergem a uma reta, sugerindo normalidade da distribuição dos resíduos. Esses resultados são confirmados através dos níveis de significância obtidos no teste de normalidade de Kolmogorov-Smirnov (Tabela 8), uma vez que seus valores foram superiores a

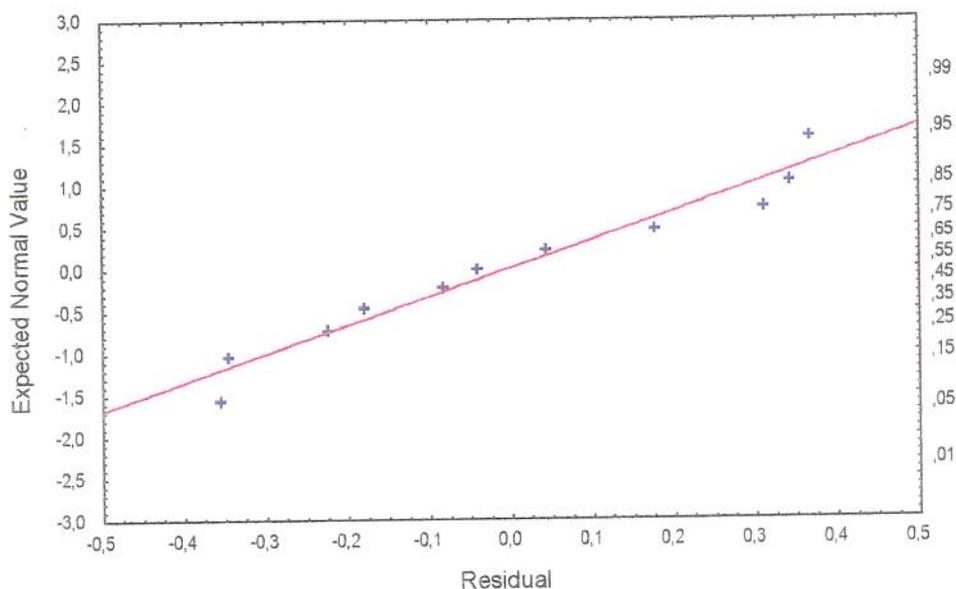
0.05. Assim, os resíduos dos modelos das duas escalas obedecem de forma similar aos preceitos de uma distribuição normal.

**TABELA 8.** Níveis de significância (p) obtidos nos testes de normalidade de Kolmogorov-Smirnov para as duas escalas estudadas.

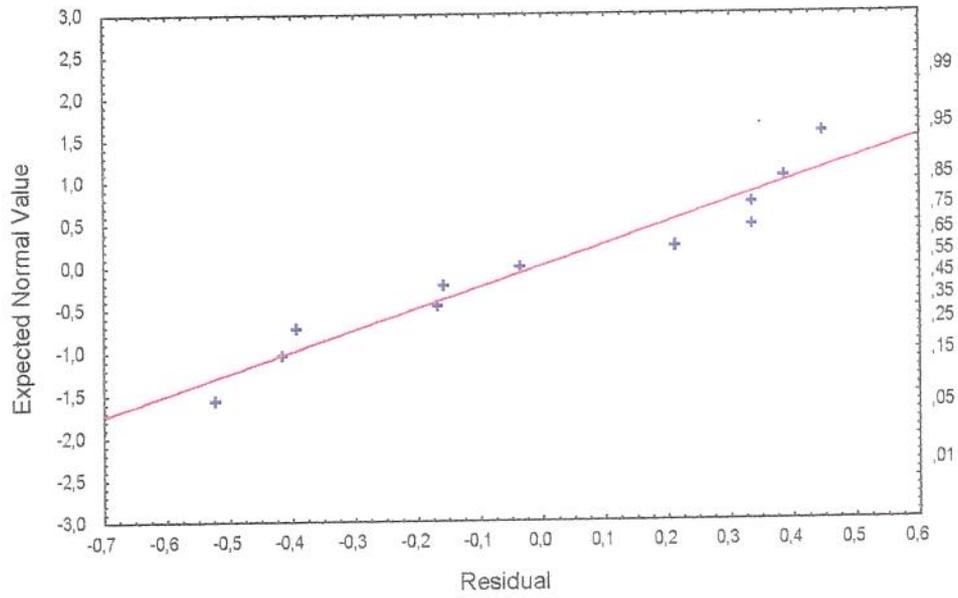
Escala	Nível de Significância (p) <sup>1</sup>
Escala Hedônica Tradicional	0.3479
Escala Hedônica com Referência	0.2179

<sup>1</sup>p > 0.05 indica que a distribuição dos resíduos é normal.

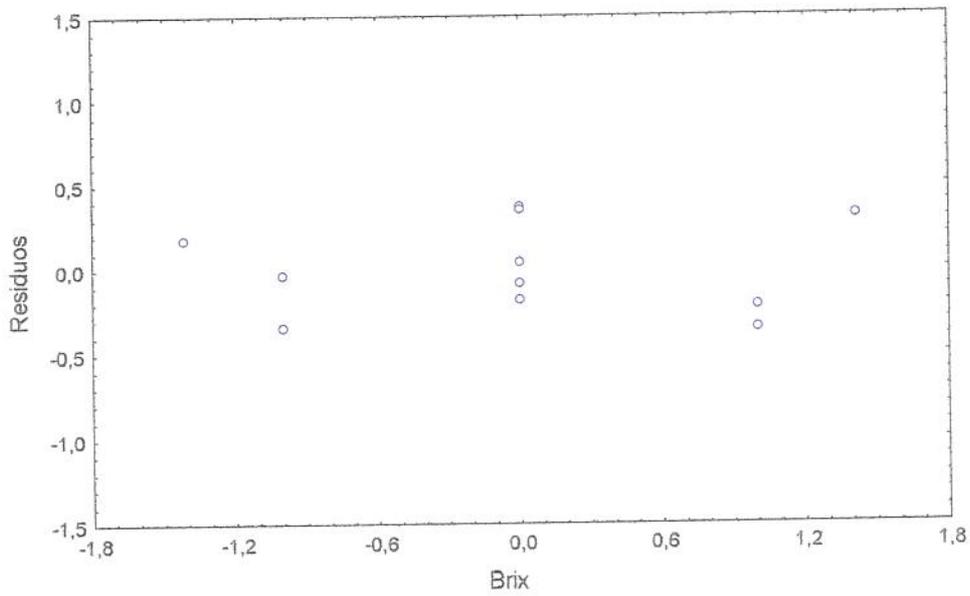
As Figuras 5, 6, 7 e 8 mostram os gráficos dos resíduos dos modelos das duas escalas em função das variáveis independentes °Brix e acidez. Analisando-os visualmente, pode-se concluir que a variância desses resíduos é constante, uma vez que não há evidência de comportamento sistemático ou alguma tendência.



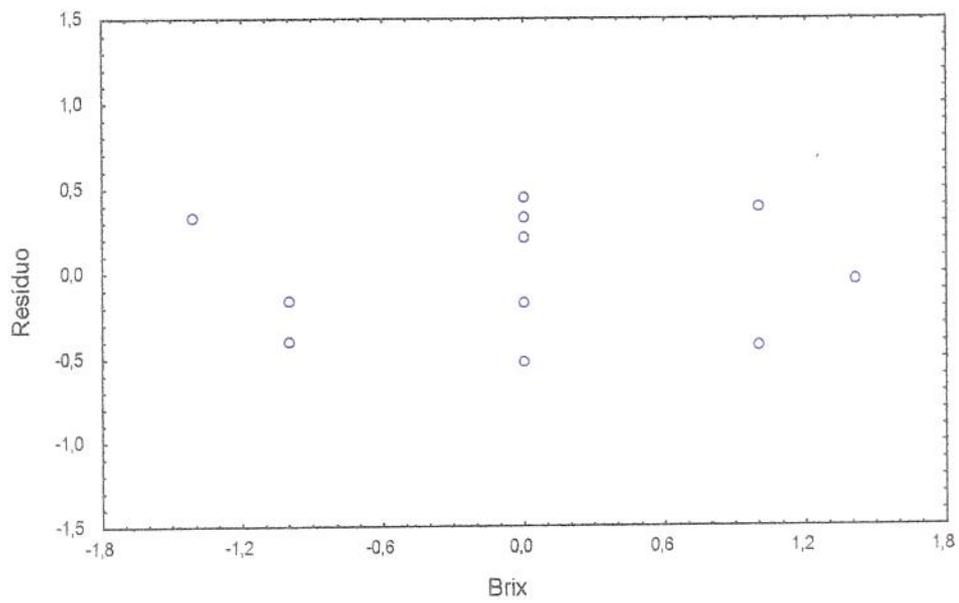
**FIGURA 3.** Gráfico de probabilidade normal dos resíduos, considerando o modelo ajustado obtido com a escala hedônica.



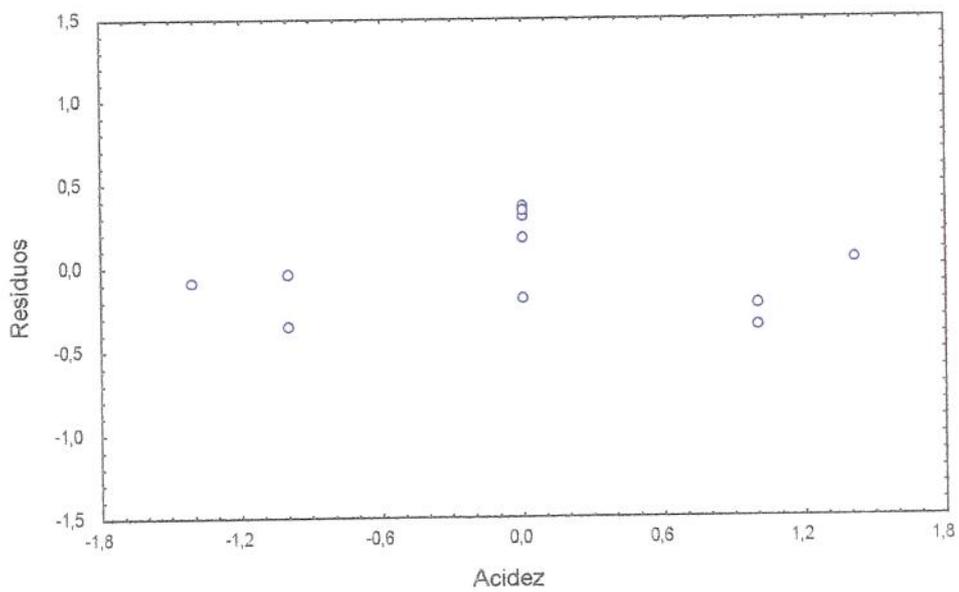
**FIGURA 4.** Gráfico de probabilidade normal dos resíduos, considerando o modelo ajustado obtido com a escala hedônica com referência.



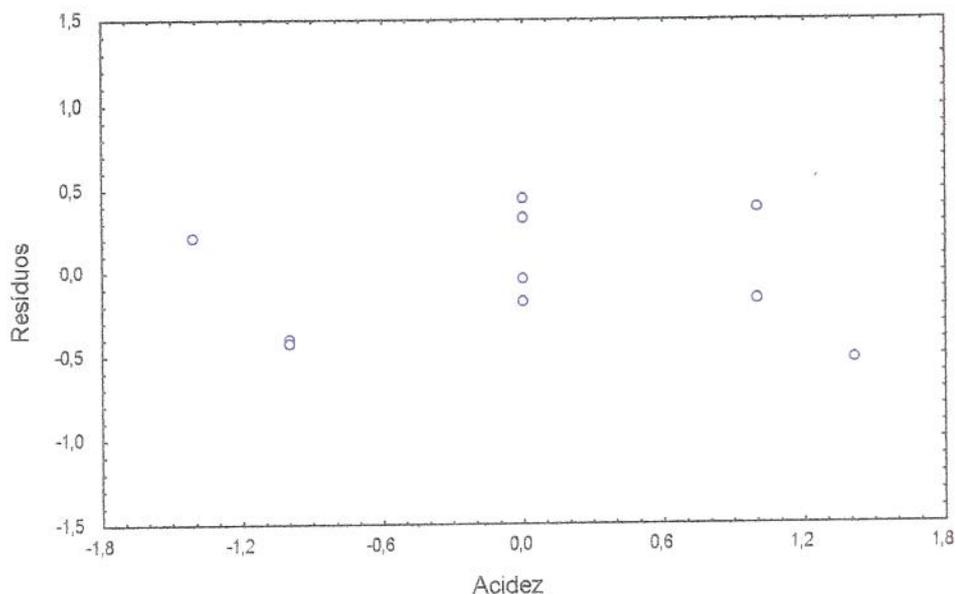
**FIGURA 5.** Gráfico dos resíduos em função da variável independente Brix a partir dos dados obtidos com a escala hedônica.



**FIGURA 6.** Gráfico dos resíduos em função da variável independente Brix a partir dos dados obtidos com a escala hedônica com referência.



**FIGURA 7.** Gráfico dos resíduos em função da variável independente acidez a partir dos dados obtidos com a escala hedônica.



**FIGURA 8.** Gráfico dos resíduos em função da variável independente acidez a partir dos dados obtidos com a escala hedônica com referência.

Considerando-se então, que os resíduos dos modelos das duas escalas apresentam variância constante e distribuição normal, bem como média igual a 0 (zero), é seguro concluir que os modelos preditivos gerados pelas duas escalas adequam-se aos pressupostos estatísticos nos quais se fundamentam.

Assim, os resultados desta pesquisa indicam que as escalas hedônica tradicional e hedônica com referência geraram modelos de igual capacidade preditiva e validade, o que sugere, uma vez mais, equivalência de desempenho das duas escalas na obtenção de modelos preditivos de resposta de consumidores.

As Figuras 9 e 10 referem-se às superfícies de resposta geradas a partir dos modelos ajustados codificados referentes às escalas hedônica e hedônica com referência, respectivamente. Essas figuras mostram os efeitos dos níveis de °Brix e acidez sobre a aceitação do suco de laranja, permitindo a visualização das regiões de otimização das respostas dos consumidores.

Através da Figura 9, observou-se que o modelo preditivo gerado através da escala hedônica indicou que a região de otimização da aceitação do suco de laranja situou-se entre os valores 11,5 e 13,2 °Brix e 0,68 e 0,74 g/100mL de acidez. Nesta região localizaram-se os tratamentos 3 e 7, cujas médias de aceitação foram, respectivamente, 6,84 – entre os termos “gostei ligeiramente” e “gostei moderadamente”- e 7,02- correspondente ao termo “gostei moderadamente”. Os demais tratamentos localizaram-se em regiões de médias de aceitação inferiores a 6,84 – entre as categorias “gostei ligeiramente” e “gostei moderadamente”. A Figura 9 sugere que a aceitação do suco, dentro da faixa estudada de °Brix e acidez, é aumentada quanto maior o nível da variável °Brix e menor o nível de acidez, sendo o efeito causado pela variável acidez bem mais sutil que aquele causado pelo nível de °Brix.

Similarmente, a Figura 10 ilustra que o modelo preditivo gerado através da escala hedônica com referência mostra que a região de otimização situou-se também entre os valores 11,5 e 13,2 °Brix e 0,68 e 0,74 g/100mL de acidez. A Figura 10 sugere que os efeitos das variáveis °Brix e acidez sobre a aceitação do suco, avaliados através da escala hedônica com referência, foram similares aos observados na Figura 9, avaliados através da escala hedônica tradicional.

Ainda que o desempenho das duas escalas tenha sido bastante semelhante, a escala hedônica com referência fornece maior segurança no que diz respeito à discriminação das amostras em relação a amostra utilizada como referência (tratamentos 9, 10 e 11), uma vez que ao utilizá-la, a comparação com essa amostra ocorreu de forma direta pelo consumidor. Em complementação, ao submeter-se as médias de aceitação obtidas com as duas escalas ao teste de Dunnett, a escala hedônica com referência permitiu que seis amostras (2, 3, 5, 6, 7 e 8) se diferenciassem significativamente da amostra 10 ao nível de 5%, enquanto que com a escala hedônica tradicional, apenas três amostras (tratamentos 3, 5 e 7) diferenciaram-se significativamente ao mesmo nível de significância (Tabela 3).

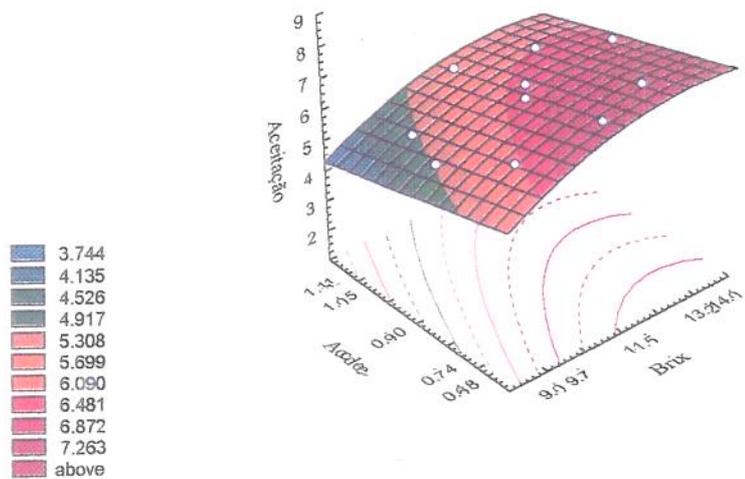


FIGURA 9. Superfície de resposta gerada do modelo ajustado codificado a partir dos dados obtidos com a escala hedônica.

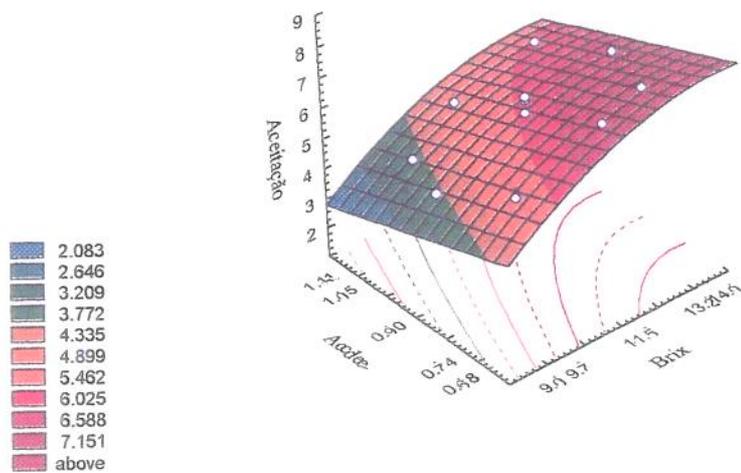


FIGURA 10. Superfície de resposta gerada do modelo ajustado codificado a partir dos dados obtidos com a escala hedônica com referência.

## CONCLUSÕES

Visto que comparativamente os modelos preditivos obtidos pelas duas escalas apresentaram níveis de significância e  $R^2$  muito similares, mesmo nível de significância para parâmetros correspondentes, falta de ajuste não significativa ( $p \leq 0.05$ ) e análise de resíduos igualmente favoráveis, é seguro concluir-se que no presente estudo, o desempenho das duas escalas para se gerar modelos preditivos de aceitação foi equivalente.

Considerando-se os resultados promissores alcançados através da escala com referência e também a falta de número significativo de estudos que tenha utilizado essa escala em testes de consumidores, acreditamos ser recomendável a realização de estudos adicionais que permitam validar os resultados aqui obtidos. A realização de maior número de pesquisas que utilizem diferentes produtos e classes de consumidores poderá confirmar a viabilidade, vantagem e limitações do uso da escala hedônica com referência como uma alternativa à escala hedônica tradicional.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. BARROS NETO, B., SCARMINIO, I.S. and Bruns, R.E. **Planejamento e Otimização de Experimentos**. Ed. Unicamp, Campinas, Brazil, p.297, 1996.
2. CHAUCHAN, J. and O'MAHONY, M. Use of a signal detection ranking analysis to measure preference for commercial and "health modified" cakes. **J. Sensory Studies**, v.8, p.69-75, 1993.
3. CHI, S.P.; CHEN, T.C. Predicting optimum monosodium glutamate and sodium chloride concentrations in chicken broth as affected by spice addition. **J. Food Processing and Preservation**, v.16, 313-326, 1992.
4. CONOVER, W.J. **Practical Nonparametric Statistics**. Series: Wiley series in probability and mathematical statistics; John Wiley & Sons, Inc., p.493, 1980.
5. GIOVANNI, M.E.; PANGBORN, R.M. Measurement of taste intensity and degree of liking of beverages by graphic scales and magnitude estimation. **J. Food Science**, v.48, p.1175-1181, 1183, 1983.
6. GUNARD, J.X.; ZOUMAS-MORSE, C.; MORI, L.; PANYAM, D.; KILARA, A. Effect of sugar and fat on the acceptability of vanilla ice cream. **J. Dairy Science**, v.79, 1922-1927, 1996.
7. HEINIÖ, R-L; URALA, N.; VAINIONPÄÄ, J.; POUTANEN, K.; Tuorila, H. Identity and overall acceptance of two types of sour rye bread. **Int. J. Food Science and Technology**, v.32, p.169-178, 1997.
8. HOUGH, G.; SÁNCHEZ, R.; BARBIERI, T.; MARTINEZ, E. Sensory optimization for a powdered chocolate milk formula. **Food Quality and Preference**, v.8 (3), p. 213-221, 1997.
9. KHURI, A.I.; CORNELL, J.A. **Response Surface: Designs and Analyses**. Series: Statistics, Textbooks and Monographs, v.152, D.B. Eds. Owen & W.R. Schucany, Marcel Dekker, Inc, New York, p.510, 1987.
10. LAWLESS, H.T.; MALONE, G.J. The discriminative efficiency of common scaling methods. **J. Sensory Studies**, v.1, p.85-98, 1986a.
11. MEILGAARD, M.; CIVILLE, G.V.; CARR, B.T. **Sensory Evaluation Techniques**. CRC Press, Boca Raton, p.281, 1988.
12. MOSKOWITZ, H.R. Sensory intensity versus hedonic functions: classical psychophysical approaches. **J. Food Quality**, v.5: p.109-137, 1981.

13. MOSKOWITZ, H.R. Product testing 2: modeling versus mapping and their integration. **J. of Sensory Studies**, v.9, p.323-336, 1994.
14. PASTOR, M.V., COSTELL, E., IZQUIERDO, L. and DURÁN, L. Optimizing acceptability of a high fruit-low sugar peach nectar using aspartame and guar gum. **J. of Food Science**, v.61 (4), p.852-855, 1996.
15. PERYAM, D.R. & PILGRIM, F.J. Hedonic scales of measuring food preferences. **Food Technology**, v.9 (11), p.9, 1957.
16. RED, J.B., HENDRIX Jr., C.M. and HENDRIX, D. L. **Quality control manual for citrus processing plants**. Intercit, Inc, Safety Harbor, p.250, 1986.
17. SAS - Statistical Analysis Systems. SAS user's guide basics. SAS Institute Inc., Cary, 1985.
18. STONE, H. & SIDEL, J.L. **Sensory Evaluation Practices**. Academic Press, Orlando, p.310, 1985.
19. VICKERS, Z.M. Magnitude estimation vs category scaling of the hedonic quality of food sounds. **J. Food Science**, v.48, p.1183-1186, 1983.
20. WAKELING, I.N.; MACFIE, H.J. Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of  $k$  samples from  $t$  may be tested. **Food Quality and Preference**, v.6, p.299-308, 1995.

## **CAPÍTULO IV**

### **Conclusões Gerais**

Neste estudo, as escalas hedônicas com referência que utilizaram como referência as amostras mais preferida e de preferência intermediária, permitiram que mais amostras fossem discriminadas significativamente ( $p \leq 0.05$ ) em relação à amostra referência do que permitiu a escala hedônica tradicional. Por sua vez, a escala de estimativa de magnitude apresentou maior poder discriminativo do que as demais escalas, segundo esse critério de avaliação.

Quando observados os níveis de significância de  $F_{amostra}$ , obtidos tanto a partir dos dados originais como através das simulações “bootstrap”, todas as escalas estudadas mostraram-se altamente discriminativas. Entretanto, não se pode garantir que os valores dos níveis de significância de  $F_{amostra}$  relativos às escalas de magnitude e hedônica tradicional (Tabela 2) fossem de fato iguais aos valores calculados, uma vez que a estimativa do valor do nível de significância de  $F_{amostra}$  baseia-se em pressupostos de normalidade das distribuições e homogeneidade da variância das respostas, os quais não foram confirmados nos casos das escalas de estimativa de magnitude e hedônica tradicional. Os desvios da normalidade dos dados, observados na escala de estimativa da magnitude foram graves, enquanto aqueles observados na escala hedônica tradicional foram ligeiros. Quanto às escalas hedônicas com referência, somente aquela que utilizou como referência a amostra mais preferida (amostra B) apresentou-se ligeiramente inadequada com relação à homogeneidade da variância, uma vez que mostrou correlação significativa ( $p \leq 0.05$ ) dos desvios padrões em função das médias de aceitação. As demais escalas hedônicas com referência foram aquelas que melhor atenderam aos pressupostos mencionados, apresentando portanto, alto grau de confiabilidade quanto ao nível de significância relativo aos testes de diferença entre médias de aceitação.

As escalas hedônica tradicional e hedônica com referência apresentaram-se similares com relação à variabilidade das respostas, uma vez que os coeficientes de variação médios dessas escalas não diferiram significativamente entre si a  $p \leq 0.05$ . Por outro lado, a escala de estimativa da magnitude apresentou maior variabilidade que as demais, pois seu coeficiente de variação médio foi superior e significativamente diferente ( $p \leq 0.05$ ) aos das demais escalas.

O desempenho das escalas hedônica estruturada de 9 pontos e hedônica com referência na obtenção de modelos preditivos de aceitação foi similar, uma vez que os modelos preditivos

obtidos através das duas escalas apresentaram níveis de significância e  $R^2$  muito similares, mesmo nível de significância para parâmetros correspondentes, falta de ajuste não significativa e análise de resíduos igualmente favoráveis. Assim, é seguro concluir-se que no presente estudo, o desempenho das duas escalas para se gerar modelos preditivos de aceitação foi equivalente.

As conclusões acima sugerem que a escala hedônica com referência pode ser vantajosa em situações nas quais o objetivo é a comparação da aceitação de um ou mais produtos com o correspondente líder do mercado, uma vez que a mesma pode permitir maior discriminação entre estes e as demais amostras, além de oferecer maior confiabilidade com relação à estimativa do nível de significância das diferenças entre médias de aceitação. Em complementação, na obtenção de modelos preditivos da aceitação de alimentos, bem como na otimização dos mesmos, a escala hedônica com referência demonstrou igual desempenho à escala hedônica tradicional, com a vantagem de possibilitar a comparação direta de todas as amostras com uma amostra referência.

No entanto, considerando-se a falta de número significativo de estudos que tenham utilizado essa escala em testes de consumidores, acreditamos ser recomendável pesquisas adicionais que permitam validar os resultados aqui obtidos. A realização de um maior número de estudos que utilizem diferentes produtos e classes de consumidores poderá confirmar a viabilidade, vantagens e limitações do uso da escala hedônica com referência como uma alternativa à escala hedônica tradicional e à escala de estimativa da magnitude.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. AMERINE, M. A.; PANGBORN, R.M.; ROESSLER, E.B. Principles of sensory evaluation of food. In: **Food Science and Technology**. Academic Press, New York, p.602, 1965.
2. BARROS NETO, B., SCARMINIO, I.S. and Bruns, R.E. **Planejamento e Otimização de Experimentos**. Ed. Unicamp, Campinas, Brazil, p.297, 1996.
3. BROCKHOFF, M.; SKOVGAARD, M. Modeling individual differences between assessors in sensory evaluations. **Food Quality and Preference**, v.5, p.215-224, 1994.
4. CHAUCHAN, J. and O'MAHONY, M. Use of a signal detection ranking analysis to measure preference for commercial and "health modified" cakes. **J. Sensory Studies**, v.8, p.69-75, 1993.
5. CHI, S.P.; CHEN, T.C. Predicting optimum monosodium glutamate and sodium chloride concentrations in chicken broth as affected by spice addition. **J. Food Processing and Preservation**, v.16, 313-326, 1992.
6. CONOVER, W.J. **Practical Nonparametric Statistics**. Series: Wiley series in probability and mathematical statistics; John Wiley & Sons, Inc., p.493, 1980.
7. DEVORE, J.L.; PECK, R. **Statistics, the Exploration and Analysis of Data**. West Publishing Company, St. Paul, p.699, 1986.
8. DIACONIS, P.; EFRON, B. Computer-intensive methods in statistics. **Sci. American**, v.24 (8), p.116-130, 1983.
9. EFRON, B.; TIBSHIRANI, R. **An Introduction to the Bootstrap**. Série: Monographs on Statistics and Applied Probability. Chapman & Hall. New York, p.436, 1993.
10. EINOT, I.; GABRIEL, K.R. A study of the powers of several methods of multiple comparisons. **J. American Statistical Association**, v.70, 574-583, 1975.
11. EPLER, S.; CHAMBERS IV, E.; KEMP, K. Hedonic scales are a better predictor than just-about-right scales of optimal sweetness in lemonade. **J. Sensory Studies**, v.13, p.191-197, 1998.
12. GAY, C.; MEAD, R. A statistical appraisal of the problem of sensory measurement. **J. Sensory Studies**, v.7, p.205-228, 1992.
13. GIOVANNI, M.E.; PANGBORN, R.M. Measurement of taste intensity and degree of liking of beverages by graphic scales and magnitude estimation. **J. Food Science**, v.48, p.1175-1181, 1183, 1983.

14. GUNARD, J.X.; ZOUMAS-MORSE, C.; MORI, L.; PANYAM, D.; KILARA, A. Effect of sugar and fat on the acceptability of vanilla ice cream. **J. Dairy Science**, v.79, 1922-1927, 1996.
15. HEINIÖ, R-L; URALA, N.; VAINIONPÄÄ, J.; POUTANEN, K.; Tuorila, H. Identity and overall acceptance of two types of sour rye bread. **Int. J. Food Science and Technology**, v.32, p.169-178, 1997.
16. HOUGH, G.; SÁNCHEZ, R.; BARBIERI, T.; MARTINEZ, E. Sensory optimization fo a powdered chocolate milk formula. **Food Quality and Preference**, v.8 (3), p. 213-221, 1997.
17. JOHNSON, J.; VICKERS, Z. Avoiding the centering bias or range effect when determining na optimum level of sweetness in lemonade. **J. Sensory Studies**,v.2, p.283-292, 1987.
18. KHURI, A.I.; CORNELL, J.A. **Response Surface: Designs and Analyses**. Series: Statistics, Textbooks and Monographs, v.152, D.B. Eds. Owen & W.R. Schucany, Marcel Dekker, Inc, New York, p.510, 1987.
19. LARSON-POWERS, N. & PANGBORN, R.M. Descriptive anlysis of the sensory properties of beverages and gelatins containing sucrose or synthetic sweeteners. **J. Food Science**, v.43, p.47-51, 1978.
20. LAWLESS, H.T.; MALONE, G.J. The discriminative efficiency of common scaling methods. **J. Sensory Studies**, v.1, p.85-98, 1986a.
21. LAWLESS, H.T. & MALONE, G.J. A comparison of rating scales: sensitivity, replicates and relative measurement. **J. Sensory Studies**, v.1, p.155-174, 1986-b.
22. MACFIE, H.J.; BRATCHELL, N.; GREENHOFF, K.; VALLIS, L.V. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. **J. Sensory Studies**, v.4, p.129-148, 1989.
24. McDANIEL, M.R. Magnitude estimation versus category scaling in food preference testing and food quality rating. Dissertação Ph.D. University of Massachussetts, Amherst, Mass, 1974.
25. McDANIEL, M.R.; SAWYER, F.M. Preference testing of whiskey sour formulations: magnitude estimation versus the 9-point hedonic. **J. Food Science**, v.46, p.182-185, 1981.
26. McPHERSON, R.S.; RANDALL, E. Line Length measurement as a tool for food preference research. **Ecology of Food and Nutrition**, v.17, p.149-156, 1985.
27. MEILGAARD, M.; CIVILLE, G.V.; CARR, B.T. **Sensory Evaluation Techiques**. CRC Press, Boca Raton, p.281, 1988.

28. MILLIKEN, G.A.; JOHNSON, D.E. **Analysis of Messy Data**. v.1: Designed experiments; Mason, R., Scheier, B., Eds. Van Nostrand Reinhold, New York, p. 473, 1984.
29. MOSKOWITZ, H.R. Sensory intensity versus hedonic functions: classical psychophysical approaches. **J. Food Quality**, v.5: p.109-137, 1981.
30. MOSKOWITZ, H.R. Product testing 2: modeling versus mapping and their integration. **J. of Sensory Studies**, v.9, p.323-336, 1994.
31. MOSKOWITZ, H.R.; SIDEL, J.L. Magnitude and hedonic scales of food acceptability. **J. Food Science**, v.36, p.677-680, 1971.
32. NAES, T. Handling individual differences between assessors in sensory profiling. **Food Quality and Preference**, v.2, p.187-199, 1990.
33. NETER, J.; WASSERMAN, W.; KUTNER, M.H. **Applied Linear Regression Models**. Boston, p.667, 1989.
34. O'MAHONY, M. Some assumptions and difficulties with common statistics for sensory analysis. In: Overview: Outstanding Symposia in Food Science & Technology. **Food Technology**, 75-82, 1982.
35. O'MAHONY, M. **Sensory Evaluation of Food: Statistical Methods and Procedures**. Series: Food Science and Technology. Marcel Dekker, Inc. New York, p.487, 1986.
37. PANGBORN, R.M.; GUINARD, J-X.; MEISELMAN, H.L. Evaluation of bitterness of caffeine in hot chocolate drink by category, graphic, and ratio scaling. **J. Sensory Studies**, v.4, p.31-53, 1989.
38. PASTOR, M.V., COSTELL, E., IZQUIERDO, L. and DURÁN, L. Optimizing acceptability of a high fruit-low sugar peach nectar using aspartame and guar gum. **J. of Food Science**, v.61 (4), p.852-855, 1996.
39. PERYAM, D.R. & PILGRIM, F.J. Hedonic scales of measuring food preferences. **Food Technology**, v.9 (11), p.9, 1957.
40. RANDALL, E.; SANJUR, D. Food preferences - their conceptualization and relationship to consumption. **Ecology Food Nutrition**, v.11, p.151-161, 1981.
41. RED, J.B., HENDRIX Jr., C.M. and HENDRIX, D. L. **Quality control manual for citrus processing plants**. Intercit, Inc, Safety Harbor, p.250, 1986.
42. SAS - Statistical Analysis Systems. SAS user's guide basics. SAS Institute Inc., Cary, 1985.
43. SHEPHERD, R.; FARLEIGH, C.A.; LAND, D.G.; FRANKLIN, J.G. Validity of a relative-to-ideal rating procedure compared with hedonic rating. In: **Progress in Flavour Research**. Elsevier Science Publishers, Amsterdam, p.103-110, 1984.

44. STOER, N.L.; LAWLESS, H.T. Comparison of single product scaling and relative-to-reference scaling in sensory evaluation of dairy products. **J. Sensory Studies**, v.8, p.257-270, 1993.
45. STONE, H.; OLIVER, S. Measurement of the relative sweetness of selected sweeteners and sweetener mixtures. **J. Food Science**, v.34, p.215, 1969.
46. STONE, H. & SIDEL, J.L. **Sensory Evaluation Practices**. Academic Press, Orlando, p.310, 1985.
48. VICKERS, Z.M. Magnitude estimation vs category scaling of the hedonic quality of food sounds. **J. Food Science**, v.48, p.1183-1186, 1983.
50. VIE, A.; GULLI, D.; O'Mahony, M. Alternative hedonic measures, **J. Food Science**, v.56 (1), p.1-5, 1991.
51. WAKELING, I.N.; MACFIE, H.J. Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of  $k$  samples from  $t$  may be tested. **Food Quality and Preference**, v.6, p.299-308, 1995.
52. WILKINSON, C.; YUKSEL, D. Modeling differences between panelists in use of measurement scales. **J. Sensory Studies**, v.12, p.55-68, 1997.