**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Modelos para dados categorizados ordinais de ensaios sensoriais com
delineamento em blocos incompletos**

Janaína Marques e Melo

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e Ex-
perimentação Agronômica

**Piracicaba
2021**

Janaína Marques e Melo
Bacharela em Estatística
Licenciada em Matemática

**Modelos para dados categorizados ordinais de ensaios sensoriais com
delineamento em blocos incompletos**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **IDEMAURO ANTONIO RODRIGUES
DE LARA**

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e Ex-
perimentação Agronômica

Piracicaba
2021

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Melo, Janaína Marques e

Modelos para dados categorizados ordinais de ensaios sensoriais com delineamento em blocos incompletos / Janaína Marques e Melo. – – versão revisada de acordo com a resolução CoPGr 6018 de 2011. – – Piracicaba, 2021 .

65 p.

Tese (Doutorado) – – USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Logitos cumulativos 2. Efeitos aleatórios 3. Análise multivariada 4. Análise sensorial. . I. Título.

DEDICATÓRIA

À minha querida família, à qual tanto admiro, dedico o resultado do esforço realizado ao longo deste percurso.

AGRADECIMENTOS

Agradeço a Deus por estar sempre presente na minha caminhada e por proporcionar momentos de alegrias e superações, eu não sou nada sem o Seu amor.

À minha família, por ser o meu alicerce. Minha mãe, Núbia e o meu pai Dionízio, por não deixarem faltar nada, por ajudarem com palavras de conforto e, por fazerem o possível e, às vezes, até um pouco mais... Só cheguei até aqui por eles existirem.

Meus irmãos, João Dionízio e José Marques, por tornarem os meus dias mais felizes e por sempre me darem o suporte para conquistar os meus sonhos.

À minha querida avó Consuelo, por ser uma pessoa forte e determinada, um exemplo de uma mulher guerreira e as minhas tias, tios, primos e sobrinhos, por serem maravilhosos comigo e sempre presentes.

Ao meu orientador Dr. Idemauro, por dar a oportunidade de trabalhar com um tema fantástico que, a cada experiência, deixou mais clara a profissional que eu queria me tornar. Obrigada, professor, por ser tão paciente e passar um pouco do seu conhecimento e, por transmitir a sua satisfação em trabalhar com dados categorizados. Sou muito grata por ter vivenciado esses momentos de sabedoria.

Agradeço também a todos os professores do departamento de ciências exatas do programa de pós-graduação em estatística e experimentação agrônômica da Escola Superior de Agricultura "Luiz de Queiroz", em especial, aos professores Dr. Edwin, Dra. Clarice e Dr. Tadeu, por todos os saberes compartilhados, vocês colaboraram para a minha formação teórica e prática.

À secretária Solange, por ser prestativa, solícita e atenciosa comigo.

À professora Dra. Silvia Maria por estar sempre nessa minha caminhada, desde a graduação, ajudando de alguma forma no meu crescimento intelectual.

Aos companheiros de jornada, minha turma de doutorado, pela ajuda e por compartilharem conhecimentos acadêmicos e pessoais. Obrigada, pelos ensinamentos de vida e profissional Laura, Tatiana, Sidcleide e Danilo.

Aos meus eternos amigos de graduação Kelly, Raquel, Enio, Elaine e Rachel que, mesmo estando longe, se fazem presentes na minha vida.

Agradeço a todos que, de alguma maneira, colaboraram com esse trabalho e com o meu crescimento acadêmico e pessoal.

À coordenação de aperfeiçoamento de pessoal de nível superior (CAPES), pelo auxílio financeiro concedido nessa trajetória.

Posso dizer que chegamos ao fim de mais uma jornada, para dar início a uma nova, sem nunca esquecer o que fez a pessoa que sou hoje, obrigada a todos.

EPÍGRAFE

Não é sobre chegar no topo do mundo, saber que venceu
É sobre escalar e sentir que o caminho te fortaleceu
É sobre ser abrigo e também ter morada em outros corações
E assim ter amigos contigo em todas as situações.

Ana Vilela

LISTA DE FIGURAS

2.1	Estrutura das 13 formulações de bebidas prebióticas de suco de uva, por meio de um fatorial 2^2 de um delineamento composto central rotacional, conforme estudo desenvolvido na Universidade Federal do Ceará, no ano de 2016	28
2.2	Esquema experimental em blocos incompletos balanceados com 130 provadores e 13 formulações de bebidas prebióticas de suco de uva propostas, conforme estudo desenvolvido na Universidade Federal do Ceará, no ano de 2016	29
2.3	Etapas para o modelo dos logitos cumulativos misto	32
2.4	Gráfico entre as categorias atribuídas por cada provador, com relação aos atributos sensoriais e as formulações de bebidas prebióticas de suco de uva, propostas do estudo desenvolvido na Universidade Federal do Ceará, no ano de 2016	33
2.5	Proporções observadas e proporções previstas fornecida pelo o modelo dos logitos cumulativos de chances proporcionais mistos para formulação de suco com relação ao atributo sensorial impressão global do estudo desenvolvido pela Universidade Federal do Ceará, no ano 2016	36
3.1	Representação gráfica bidimensional por meio da análise de correspondência múltipla entre formulações de bebidas prebióticas, atributos sensoriais e as categorias de respostas dos provadores, do estudo desenvolvido na Universidade Federal do Ceará, no ano de 2016 . . .	50
3.2	Probabilidades observadas e previstas para as formulações de bebidas prebióticas com relação aos atributos sensoriais conjuntamente do estudo desenvolvido pela Universidade Federal do Ceará, no ano 2016	53

LISTA DE TABELAS

2.1	Estimativas dos parâmetros, erros padrões, estatística do teste e razão de chances para formulações de bebidas prebiótica de suco de uva com a escala hedônica em relação ao atributo sensorial impressão global do estudo desenvolvido na Universidade Federal do Ceará, no ano 2016	34
2.2	Probabilidades estimadas pelo modelo de chances proporcionais de efeito misto (2.28) de cada categoria de resposta para formulação de bebidas prebiótica de suco de uva com relação ao atributo sensorial impressão global do estudo desenvolvido na UFC, no ano 2016	35
2.3	Probabilidades acumuladas estimadas para as categorias 4 e 5 das formulações de bebidas prebióticas de suco de uva dos atributos sensoriais do estudo desenvolvido na Universidade Federal do Ceará, no ano 2016	36
3.1	Estimativas dos parâmetros, erros padrões, p-valores e razões de chances do modelo dos logitos cumulativos “multi” multivariado de chances não proporcionais mistos para análise conjunta dos atributos sensoriais referente ao estudo desenvolvido na Universidade Federal do Ceará, no ano 2016	51
3.2	Probabilidades das categorias de respostas com relação aos atributos sensoriais e probabilidades acumuladas previstas das categorias 4 e 5, do estudo desenvolvido na Universidade Federal do Ceará, no ano 2016	52

SUMÁRIO

Lista de Figuras	6
Lista de Tabelas	7
Resumo	9
Abstract	10
1 Introdução	11
Referências	12
2 Seleção de formulações de bebidas prebióticas de suco de uva por meio de técnicas para dados multivariados categorizados ordinais	15
Resumo	15
2.1 Introdução	15
2.2 Distribuição de probabilidade multinomial	16
2.3 Modelos para dados categorizados ordinais	17
2.3.1 Modelo dos logitos cumulativos	17
2.3.2 Modelo dos logitos de categorias adjacentes	21
2.3.3 Modelo dos logitos de razão contínua	23
2.4 Modelos lineares generalizados de efeitos mistos	25
2.5 Análise de correspondência	26
2.6 Material	27
2.7 Métodos	30
2.8 Resultados e discussões	32
2.9 Conclusão	37
Referências	37
3 Modelo categorizado ordinal multivariado para análise simultânea de atributos sensoriais	41
Resumo	41
3.1 Introdução	41
3.2 Modelo cumulativos	43
3.3 Modelo dos logitos cumulativos	44
3.4 Versão ampliada do modelo dos logitos cumulativos	44
3.5 Função de ligação e matrizes de delineamento para o modelo dos logitos cumulativos	45
3.6 Material	46
3.7 Método	47
3.7.1 Análise exploratória por meio da análise de correspondência múltipla	47
3.7.2 Proposta do modelo dos logitos cumulativos “multi” multivariado	47
3.8 Resultados e discussões	50
3.9 Conclusão	53
Referências	54
4 Considerações finais	57
Apêndices	59

RESUMO

Modelos para dados categorizados ordinais de ensaios sensoriais com delineamento em blocos incompletos

Dados categorizados e ordinais ocorrem em várias situações experimentais e, em particular, em estudos sensoriais para avaliação de um produto. Nesse contexto, este trabalho descreve uma classe de modelo para variável resposta desta natureza, tendo como motivação um experimento sensorial, desenvolvido na Universidade Federal do Ceará, para avaliação de formulações de bebidas probióticas com suco de uva e amêndoa de castanha de caju. Uma característica particular desse experimento foi o fato do mesmo corresponder a um ensaio com blocos incompletos, uma vez que, cada julgador não avalia todas as formulações de bebidas. Assim, para considerar a estrutura do delineamento é usada a abordagem dos modelos de efeitos mistos. O trabalho está estruturado na forma de artigos. O primeiro artigo traz uma revisão dos modelos para dados categorizados ordinais. A análise dos cinco atributos sensoriais: aroma, corpo, doçura, sabor e impressão global foram feitas por meio da análise de correspondência e do modelo de chances proporcionais com efeito aleatório para provador, sendo ajustado um modelo para cada atributo. Essa análise permitiu identificar as melhores formulações de bebidas, sendo destacada a bebida com 6% de açúcar e 44% de suco de uva. No segundo artigo, apresenta-se a contribuição dessa tese, por meio de um modelo “multi” multivariado de efeito misto, visto que cada modelo em separado já corresponde a uma análise multivariada pela natureza da variável. Nesta abordagem, todos os atributos foram analisados simultaneamente em um modelo único. Os métodos estão centrados na teoria da máxima verossimilhança. Embora o modelo utilizado para a análise única seja o de chances não proporcionais de efeito misto, tal abordagem permitiu concluir que a bebida composta por 6% de açúcar e 43% de suco de uva, obteve uma maior aceitabilidade, representando assim, uma unicidade de análise estatística.

Palavras-chave: Logitos cumulativos, Efeitos aleatórios, Análise multivariada, Análise sensorial, Bebida probiótica.

ABSTRACT

Models for ordinal categorized data from sensory analysis with incomplete block design

Categorized and ordinal data occur in various experimental situations and, in particular, in sensory studies to evaluate a product. In this context, this paper describes a model class for a variable response of this nature, motivated by a sensory experiment, developed at the Federal University of Ceará, to evaluate formulations of probiotic beverages with grape juice and cashew nuts. A particular feature of this experiment was the fact that it corresponds to an assay with incomplete blocks, since, each judge does not evaluate all beverage formulations. Thus, to consider the design structure, the mixed effects models approach is used. The work is structured in the form of articles. The first article provides a review of the models for ordinal categorized data. The analysis of the five sensory attributes: aroma, body, sweetness, flavor and global impression were made through the correspondence analysis and the proportional chance model with random effect for the panelists, adjusting a model for each attribute. This analysis allowed the identification of the best beverage formulations, highlighting the beverage with 6 % sugar and 44 % grape juice. In the second article, the contribution of this thesis is presented, through a mixed effect multivariate “ multi ” model, since each separate model already corresponds to a multivariate analysis by the nature of the variable. In this approach, all attributes were analyzed simultaneously in a single model. The methods are centered on the maximum likelihood theory. Although the model used for the single analysis is that of non-proportional chances of mixed effect, this approach allowed us to conclude that the beverage composed of 6 % sugar and 43 % grape juice, obtained a greater acceptability, thus representing a uniqueness of statistical analysis.

Keywords: Cumulative logit, Random effects, Multivariate analysis, Sensory analysis, Prebiotic beverage

1 INTRODUÇÃO

A partir dos anos 90, tem ocorrido a valorização de produtos com ingredientes prebióticos, em razão de seus benefícios para a saúde, entre os quais se destacam a regulação do sistema digestório e o equilíbrio do organismo como um todo, que provoca a sensação de bem-estar, assim como também possíveis benefícios para a redução de ansiedade, estado de ânimo e sistema imunológico (Zacarchenco et al., 2017). Vale ressaltar que bebidas não alcoólicas estão sendo fortificadas com vitaminas e minerais (cálcio, sódio, ferro e vitamina C), encontrados na amêndoa da castanha de caju, sendo uma alternativa para uma dieta saudável dos consumidores.

Essa procura por alimentos saudáveis, por parte dos consumidores estimulou as indústrias a desenvolverem novos produtos. Antes de um novo produto ser colocado no mercado, faz-se necessária a verificação do mesmo em relação a sua aceitação. Uma ferramenta bastante utilizada pelas indústrias de alimentos, bebidas, cosméticos, perfumes, produtos de limpeza, entre outros, é a avaliação sensorial. Nas indústrias de alimentos, a análise sensorial é o elemento chave para identificar as expectativas dos consumidores (Ledauphin et al., 2008). A metodologia proposta pela análise sensorial é identificar e quantificar, por meio dos sentidos humanos, tais como: visão, olfato, tato, paladar e audição, as características ou atributos de um determinado produto. Essas características sensoriais são classificadas por provadores, segundo uma escala de resposta previamente estabelecida, com base uma classificação de gosto, chamadas de escala hedônica. Acredita-se que um alimento de boa qualidade apresenta características sensoriais agradáveis, próprias do produto, tais como: aroma, corpo, doçura e sabor. É importante ressaltar que a qualidade sensorial não é uma característica própria do alimento e sim um resultado da interação entre o alimento e o homem, isto é, associando as condições fisiológicas, psicológicas, sociológicas e culturais do indivíduo.

A avaliação sensorial é realizada, de acordo com diferentes testes que dependem da sua finalidade (Anzaldúa Morales, 1994). Limitações comuns, nesses testes, referem-se à quantidade de atributos a serem analisados (Rossini et al., 2012), ao número de provadores (Trad, 2009) ou ao número de produtos em teste (Li et al., 2014; Los et al., 2020). No que tange a esse último aspecto, sugere-se um experimento em blocos incompletos, introduzidos por Yates (1936) e generalizados por Bose e Nair (1939), em que os provadores não degustam todas as opções do produto em teste. Tal tipo de experimento reduz o tempo de execução, a fadiga imposta aos provadores e os custos da avaliação. Detalhes adicionais sobre experimento em blocos incompletos podem ser vistos em Cochran e Cox (1992), Hinkelmann e Kempthorne (2005) e Montgomery (2013).

Adicionalmente, nesses delineamentos em blocos incompletos, as variáveis respostas expressas na escala hedônica correspondem, estatisticamente, a uma variável categorizada ordinal. Na área de alimentos, em geral, essas variáveis são, frequentemente, analisadas como variáveis contínuas, como no caso da análise de variância - ANOVA (Li et al., 2014), do método de superfície de resposta (De Souza et al., 2020), da regressão por mínimos quadrados parciais (*Partial least squares regression*) (Belchior et al., 2019), entre outros. No entanto, a utilização da variável categorizada como contínua, não é necessária, pode-se trabalhar com a natureza original da variável resposta. Quando existe uma ordem natural entre as categorias (duas ou mais) de resposta, como no caso da classificação dada aos atributos sensoriais pelos provadores, isso deve ser levado em consideração na especificação do modelo. Uma possível opção para esses tipos de variáveis, os quais permitem calcular as probabilidades das preferências dos provadores, são os modelos dos logitos cumulativos. Esses modelos podem ser de chances proporcionais (McCullagh, 1980), de chances não proporcionais (Williams e Grizzle, 1972) ou de chances parciais (Peterson e Harrell Jr, 1990), todos são extensões multivariadas dos modelos lineares generalizados desenvolvidos por Nelder e Wedderburn (1972). Essa classe de modelos tem sido usada nas áreas de Ciências Agrárias, Ciências Animais, Ciência de Tecnologia de Alimentos, dentre outras, para análise de dados de natureza

ordinal.

Outra problematização para estudos que envolvem análise sensorial é que os atributos são analisados de forma separada. A título de ilustração, Fatores et al. (2018) ajustaram o modelo de chances proporcional misto para estimar as probabilidades para as categorias de respostas considerando duas variedades de tomate, levando em consideração o atributo sensorial sabor. No caso Da Cunha et al. (2019) aplicaram regressão logística ordinal cumulativa com chances proporcional para avaliar a opção de compra de estudantes universitários por alimentos orgânicos com relação aos atributos sensoriais, separadamente, de aspecto visual, sabor, textura, amargura e impressão global.

Objetivou-se, nesse trabalho é desenvolver um método de análise conjunta para variáveis respostas categorizadas ordinais, em um experimento em blocos incompletos, com o propósito de apresentar uma capacidade inferencial dos dados, conseqüentemente, uma indicação mais apropriada, em termos de qualidade, do produto em teste, por meio de um modelo “multi” multivariado, visto que cada modelo em separado já corresponde a uma análise multivariada pela natureza da variável. Para a aplicação desse método, utilizou-se, como motivação, um conjunto de dados de um estudo realizado na Universidade Federal do Ceará (UFC) no ano de 2016. Conduziu-se, esta pesquisa, com o objetivo de analisar a aceitabilidade de formulações de bebidas prebióticas à base de amêndoa da castanha de caju, em relação a atributos sensoriais (impressão global, aroma, cor, corpo, doçura e sabor) mensurados por uma escala hedônica, por meio de modelos para dados categorizados ordinais com efeitos mistos.

Nesse contexto, no presente trabalho, propõe-se como contribuição para a área da Estatística, um modelo único para análise simultânea de todos os atributos, permitindo trabalhar com múltiplas relações, selecionar covariáveis pertinentes, com inferência estatística simplificada para o conjunto de dados. No âmbito da Sensometria, esse trabalho colabora com a seleção do produto com maior aceitabilidade para o consumidor, necessária para estudos de classificação, no contexto da análise sensorial de produtos em desenvolvimento.

O presente trabalho está dividido em quatro capítulos, a saber: no Capítulo 1 apresenta-se a introdução do tema com uma revisão de literatura, o objetivo e a contribuição da proposta. No Capítulo 2, apresentam-se conceitos sobre análise sensorial, delineamento em blocos incompletos balanceados, análise de correspondência, modelos para análises de dados categorizados ordinais e o desenvolvimento de uma aplicação do modelo logito cumulativo de efeitos mistos para avaliar a aceitabilidade de formulações de bebidas prebióticas. No Capítulo 3, encontra-se uma proposta de um modelo único para a análise simultânea dos atributos sensoriais mensurados por uma escala hedônica, permitindo trabalhar com múltiplas relações. Finalmente, no Capítulo 4 apresentam-se as considerações finais.

Referências

- Anzaldúa Morales, A. (1994). *La Evaluación Sensorial de los Alimentos en teoría y la práctica*. Editorial. Acribia.
- Belchior, V., Botelho, B. G., Oliveira, L. S., e Franca, A. S. (2019). Attenuated total reflectance fourier transform spectroscopy (atr-ftir) and chemometrics for discrimination of espresso coffees with different sensory characteristics. *Food chemistry*, 273:178–185.
- Bose, R. C. e Nair, K. (1939). Partially balanced incomplete block designs. *Sankhyā: The Indian Journal of Statistics*, 3:337–372.
- Cochran, W. G. e Cox, G. M. (1992). *Experimental designs*. Wiley, 2ª edition.

- Da Cunha, D. T., Antunes, A. E. C., Da Rocha, J. G., Dutra, T. G., Manfrinato, C. V., Oliveira, J. M., e Rostagno, M. A. (2019). Differences between organic and conventional leafy green vegetables perceived by university students. *British Food Journal*, 121:1579–1591.
- De Souza, V. R., Aniceto, A., Abreu, J. P., Montenegro, J., Boquimpani, B., de Jesus, V. A., Elias Campos, M. d. B., Marcellini, P. S., Freitas-Silva, O., Cadena, R., et al. (2020). Fruit-based drink sensory, physicochemical, and antioxidant properties in the amazon region: Murici (*byrsonima crassifolia* (l.) kunth and *verbascifolia* (l.) dc) and tapereba (*spondia mombin*). *Food Science & Nutrition*, 8(5):2341–2347.
- Fatoretto, M. B., de Lara, I. A. R., Loro, A. C., e Spoto, M. H. F. (2018). Sensory evaluation of dehydrated tomatoes using the proportional odds mixed model. *Journal of food processing and preservation*, 42(11):e13822.
- Hinkelmann, K. e Kempthorne, O. (2005). *Design and analysis of experiments: Advanced experimental design*, volume Volume 2 of *Wiley series in probability and mathematical statistics. Applied probability and statistics*. Wiley, rev. ed. edition.
- Ledauphin, S., Pommeret, D., e Qannari, E. M. (2008). Application of hidden markov model to products shelf lives. *Food Quality and Preference*, 19(2):156–161.
- Li, B., Hayes, J. E., e Ziegler, G. R. (2014). Interpreting consumer preferences: Physicohedonic and psychohedonic models yield different information in a coffee-flavored dairy beverage. *Food quality and preference*, 36:27–32.
- Los, P. R., Marson, G. V., Dutcosky, S. D., Nogueira, A., Marinho, M. T., e Simões, D. R. S. (2020). Optimization of beef patties produced with vegetable oils: a mixture design approach and sensory evaluation. *Food Science and Technology*, (AHEAD).
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127.
- Montgomery, D. C. (2013). *Design and analysis of experiments*. John wiley & sons.
- Nelder, J. A. e Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384.
- Peterson, B. e Harrell Jr, F. E. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 39(2):205–217.
- Rossini, K., Anzanello, M. J., e Fogliatto, F. S. (2012). Seleção de atributos em avaliações sensoriais descritivas. *Production*, 22(3):380–390.
- Trad, L. A. B. (2009). Grupos focais: conceitos, procedimentos e reflexões baseadas em experiências com o uso da técnica em pesquisas de saúde. *Physis: revista de saúde coletiva*, 19(3):777–796.
- Williams, O. D. e Grizzle, J. E. (1972). Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association*, 67(337):55–63.
- Yates, F. (1936). Incomplete randomized blocks. *Annals of Eugenics*, 7(2):121–140.
- Zacarchenco, P., Van Dender, A., e Rego, R. (2017). *Brasil Dairy Trends 2020: Tendência do Mercado de Produtos Lácteos*. Campinas: ITAL. Instituto de Tecnologia de Alimentos.

2 SELEÇÃO DE FORMULAÇÕES DE BEBIDAS PREBIÓTICAS DE SUCO DE UVA POR MEIO DE TÉCNICAS PARA DADOS MULTIVARIADOS CATEGORIZADOS ORDINAIS

Resumo

Atualmente as pessoas estão cada vez mais preocupadas com a saúde e, por consequência, com a alimentação. Com isso, os consumidores estão buscando produtos que tenham nutrientes adequados e acessíveis para incorporarem a sua rotina. Nesse trabalho, descreve-se um estudo para comparação de formulações de bebidas prebióticas à base de amêndoa da castanha de caju adicionada ao suco de uva, que foi desenvolvido na Universidade Federal do Ceará, para avaliar os atributos sensoriais de impressão global, aroma, corpo, doçura e sabor utilizando uma escala hedônica de 9 pontos. A avaliação da aceitação sensorial, das formulações das bebidas, foi realizada em uma única sessão, com 130 julgadores não treinados. As amostras das bebidas foram servidas de forma monádica sequencial, seguindo um delineamento em blocos incompletos balanceados, em que cada provador avaliou 4 das 13 formulações. Para analisar esses dados é proposta a metodologia do modelo dos logitos cumulativos de chances proporcionais de efeito misto. Portanto, esse trabalho apresenta uma estratégia de análise para um conjunto de dados sensoriais, em que as variáveis respostas são categorizadas e ordinais, estruturados em um delineamento em blocos incompletos balanceados. O procedimento apresentou-se adequado para agrupar formulações de bebidas prebióticas de suco de uva com os critérios de notas estabelecidas pelos provadores, bem como selecionar as formulações que tiveram a melhor aceitação ao paladar dos 130 provadores em relação a cada atributo sensorial. Com base nos atributos estudados, verificou-se que as formulações de bebidas prebióticas de suco de uva com melhores aceitações dos provadores, segundo o modelo dos logitos cumulativos de chances proporcionais de efeito misto foram às formulações com 6% de açúcar e 44% de suco de uva (F_6) e 6% de açúcar e 30% de suco de uva (F_{13}).

Palavras-chave: Análise sensorial; Análise de correspondência; Modelo dos logitos cumulativos misto; Variáveis categóricas ordinais; Blocos incompletos balanceados.

2.1 Introdução

A crescente tendência dos benefícios funcionais proporcionados pelos frutos é influenciada pelas mudanças nas práticas alimentares e preferências dos consumidores, pelo aumento da idade da população e pela busca de maior qualidade de vida (Lemos et al., 2015). Portanto, produtos alimentícios com prebióticos estão cada vez mais sendo comercializados, por serem mais saudáveis e ajudarem no funcionamento intestinal. Outro aliado à boa saúde é a amêndoa da castanha de caju que é rica em gorduras monoinsaturadas e poli-insaturadas, que auxiliam na diminuição do colesterol ruim e no aumento do bom colesterol. Além disso, o caju é considerado uma boa fonte de vitamina C, minerais (cálcio, sódio e ferro) e substâncias antioxidantes (fenólicos e carotenoides) (Pereira et al., 2013).

A sucessiva busca de alimentos saudáveis, por parte da população, estimulou a indústria a pesquisar e desenvolver novos produtos para atender a esse critério (Abd El-Salam et al., 2011). Uma área que estuda o desenvolvimento de novos produtos é a de Tecnologia dos Alimentos. Pode-se destacar que esse campo de estudo também é responsável pela avaliação dos produtos antes de serem lançados ao mercado para o consumidor. Os produtos são avaliados, por meio de estudos da análise sensorial. Esse método tem como base a avaliação de um determinado produto pelos sentidos humanos, ou seja, visão, olfato, paladar, tato e audição, considerando atributos como aparência, sabor, aroma, textura e som (Sharif et al., 2017). Para analisar a preferência dos consumidores por determinados produtos é feita uma

avaliação que contém uma escala de respostas previamente estabelecida com base em categorias ordinais de gosto, desde “desgosta extremamente” a “gosta extremamente”, conhecida como escala hedônica. Dessa forma, as respostas são associadas a valores numéricos discretos para serem analisadas estatisticamente. Além do mais, várias pesquisas nessa área limitam a quantidade de degustação para cada provador, isso ocorre quando existe uma grande variedade do produto que está sendo proposto para ser analisada, podendo assim comprometer a indicação do melhor produto, por provarem muitas opções, como também, encarecer a pesquisa. Nesse contexto, sugere-se a análise experimental em blocos incompletos, pois os provadores não degustam todas as variações do produto. Por exemplo, Los et al. (2020) usaram o delineamento em blocos incompletos balanceados na análise sensorial, em um ensaio de laboratório, para avaliar 10 formulações de carne de hambúrguer, em que cada participante avaliou 4 carnes de hambúrgueres entre as 10 propostas.

Vale ressaltar que pesquisas que utilizam variáveis categorizadas (escalas hedônicas) para mensurar um determinado atributo sensorial, na área de alimentos são, frequentemente analisadas como variáveis contínuas. Dentre outros, pode-se citar o uso da análise de variância utilizada por Li et al. (2014) para analisar a intensidade de gosto de 20 bebidas lácteas com sabor de café e do método de superfície de resposta aplicada por Rebouças et al. (2018) para selecionar formulações de bebidas com alto apelo nutricional e funcional. Considerar variáveis categorizadas como contínuas podem até resultar em um ajuste razoável de modelos de análise da variância, desde que haja um número de categorias suficientes e poucas observações nas mesmas, mas esse método não retrata a realidade dos dados categorizados (Christensen e Brockhoff, 2013). É difícil quantificar como isso afeta a precisão e a consistência das estimativas dos parâmetros. Em particular, para escalas de classificação com um pequeno número de categorias, os modelos lineares são inadequados. Por outro lado, uma metodologia que leva em consideração a verdadeira natureza da variável de interesse, no caso os atributos sensoriais mensurados por categorias ordinais (escalas hedônicas), são os modelos lineares generalizados para respostas categorizadas, em que o modelo dos logitos cumulativos de chances proporcionais é uma opção. Por meio desse modelo, é possível inferir a aceitação de um atributo sensorial, segundo a percepção dos provadores, em relação à exploração do produto em teste.

Salienta-se que, quando a suposição de chances proporcionais não é atendida, podem-se ajustar os modelos dos logitos cumulativos de chances não proporcionais (Williams e Grizzle, 1972), ou de chances parciais (Peterson e Harrell Jr, 1990), ou modelos dos logitos de categorias adjacentes, ou modelos dos logitos de razão contínua (Agresti, 2010; Giolo, 2017) aos dados. Esses modelos são extensões multivariadas dos modelos lineares generalizados propostos por Nelder e Wedderburn (1972).

Esse trabalho tem como objetivo apresentar a metodologia dos modelos lineares generalizados com efeito misto para dados sensoriais com variáveis respostas categorizadas ordinais em um experimento em blocos incompletos balanceados, tendo como motivação um estudo de formulações de bebidas prebióticas, à base de amêndoa da castanha de caju adicionada ao suco de uva. Para essa análise foi levada em consideração a natureza da variável de interesse (categorizada e ordinal) e se utilizou a combinação de técnicas da análise de correspondência e de modelos de logitos cumulativos, com a inclusão dos efeitos aleatórios para considerar a estrutura do delineamento experimental.

2.2 Distribuição de probabilidade multinomial

Considerando um experimento dividido em m ensaios independentes, no qual cada ensaio resulta em um número finito c de valores possíveis com probabilidades $\pi_1, \pi_2, \dots, \pi_c$ (de modo que $0 \leq \pi_j \leq 1$ para $j = 1, \dots, c$ e $\sum_{j=1}^c \pi_j = 1$). Tomando a variável aleatória Y_j que representa o número de vezes que o índice j foi observado nos m ensaios, o vetor $\mathbf{Y}^T = (Y_1, \dots, Y_c)$ segue uma distribuição multinomial com parâmetros m e $\boldsymbol{\pi}$, em que m é número de ensaios e $\boldsymbol{\pi}^T = (\pi_1, \pi_2, \dots, \pi_c)$. Então

$\mathbf{Y} \sim \text{Multinomial}(m; \pi_1, \pi_2, \dots, \pi_c)$ e a distribuição de probabilidade é definida pela função:

$$\begin{aligned} \mathbb{P}(Y_1 = y_1, \dots, Y_c = y_c; m; \boldsymbol{\pi}) &= \binom{m}{\mathbf{y}} \pi_1^{y_1} \times \pi_2^{y_2} \times \dots \times \pi_c^{y_c} \\ &= \frac{m!}{y_1! \times y_2! \times \dots \times y_c!} \times \pi_1^{y_1} \times \pi_2^{y_2} \times \dots \times \pi_c^{y_c} \\ &= \frac{m!}{\prod_{j=1}^c y_j!} \prod_{j=1}^c \pi_j^{y_j}, \end{aligned} \quad (2.1)$$

com $0 \leq \pi_j \leq 1$, $y_j \in \{0, 1, \dots, m\}$ e $\sum_{j=1}^c y_j = m$.

Sucintamente, um experimento multinomial tem as seguintes características: (i) consiste em m ensaios; (ii) cada ensaio tem um número discreto de resultados possíveis e (iii) os ensaios são independentes entre si, com probabilidade constante de ocorrer um determinado resultado.

Aplicando o logaritmo na distribuição de probabilidade (equação 2.1), tem-se:

$$l(m, \boldsymbol{\pi}) = \log \left(\frac{m!}{\prod_{j=1}^c y_j!} \right) + \sum_{j=1}^c y_j \log \pi_j = \sum_{j=1}^c y_j \log \pi_j + \log m! - \sum_{j=1}^c \log y_j!.$$

Portanto, a distribuição multinomial pertence à família exponencial multiparamétrica.

2.3 Modelos para dados categorizados ordinais

Quando existe uma ordem natural entre as categorias (duas ou mais) de resposta, deve-se levar em consideração na especificação do modelo. As respostas categorizadas ordinais são comuns nas ciências agrárias e, em particular, na sensometria. Nessa seção, serão abordados modelos para análise de dados que apresentam variável resposta politômica ordinal.

2.3.1 Modelo dos logitos cumulativos

Um modelo para dados categorizados ordinais é o modelo dos logitos cumulativos que utilizam as probabilidades de resposta acumuladas. Considere $\mathbf{Y}$ um vetor cujos componentes são categorias ordenadas de resposta, com distribuição multinomial como definido em (2.1), e seja $\mathbf{x}^\top = (x_1, x_2, \dots, x_p)$ um vetor de variáveis explicativas, e considerando as quantidades definidas por

$$\begin{aligned} \theta_1 &= \pi_1(\mathbf{x}) = \mathbb{P}(Y \leq 1 \mid \mathbf{x}), \\ \theta_2 &= \pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) = \mathbb{P}(Y \leq 2 \mid \mathbf{x}), \\ \vdots &= \vdots \\ \theta_j &= \pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) + \dots + \pi_{j-1}(\mathbf{x}) + \pi_j(\mathbf{x}) = \mathbb{P}(Y \leq j \mid \mathbf{x}) \\ \vdots &= \vdots \\ \theta_c &= \pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) + \dots + \pi_j(\mathbf{x}) + \dots + \pi_{c-1}(\mathbf{x}) + \pi_c(\mathbf{x}) = \mathbb{P}(Y \leq c \mid \mathbf{x}), \end{aligned}$$

ou seja, $\theta_j = \mathbb{P}(Y \leq j \mid \mathbf{x})$, $j = 1, 2, \dots, c$, corresponde à probabilidade acumulada de ocorrência até a j -ésima categoria para um dado vetor $\mathbf{x}$ de valores de p covariáveis, sujeitas à restrição:

$$\sum_{j=1}^c \pi_j = 1.$$

Como as categorias são ordenadas, as probabilidades acumuladas também seguem uma ordem natural, ou seja, $\mathbb{P}(Y \leq 1 \mid \mathbf{x}) \leq \mathbb{P}(Y \leq 2 \mid \mathbf{x}) \leq \dots \leq \mathbb{P}(Y \leq c-1 \mid \mathbf{x})$, não englobando a última classe, pois $\mathbb{P}(Y \leq c \mid \mathbf{x}) = 1$. Os logitos para as probabilidades acumuladas são definidos como

$$\text{logito}[\mathbb{P}(Y \leq j \mid \mathbf{x})] = \ln \left[\frac{\theta_j(\mathbf{x})}{1 - \theta_j(\mathbf{x})} \right] = \ln \left[\frac{\mathbb{P}(Y \leq j \mid \mathbf{x})}{1 - \mathbb{P}(Y \leq j \mid \mathbf{x})} \right] = \ln \left[\frac{\mathbb{P}(Y \leq j \mid \mathbf{x})}{\mathbb{P}(Y > j \mid \mathbf{x})} \right], \quad (2.2)$$

que corresponde ao logaritmo de chance de ocorrência da resposta em uma das categorias de $\mathbf{Y} \leq j$.

a) Modelo dos logitos cumulativos de chances não proporcionais

Considerando um conjunto de p covariáveis, $\mathbf{x}$ denotando por $\mathbf{x}^\top = (x_1, x_2, \dots, x_p)$, um modelo proposto por Williams e Grizzle (1972) para os logitos cumulativos de chances não proporcionais é dado por:

$$g(\mathbf{x}) = \text{logito}[\mathbb{P}(Y \leq j \mid \mathbf{x})] = \ln \left[\frac{\theta_j(\mathbf{x})}{1 - \theta_j(\mathbf{x})} \right] = \ln \left[\frac{\mathbb{P}(Y \leq j \mid \mathbf{x})}{1 - \mathbb{P}(Y \leq j \mid \mathbf{x})} \right] = \alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}, \quad (2.3)$$

em que, $j = 1, 2, \dots, c-1$, $\boldsymbol{\beta}_j^\top = (\beta_{1j}, \beta_{2j}, \dots, \beta_{pj})$ o vetor de parâmetros de regressão, tal que β_{kj} descreve, para o j -ésimo logito, o efeito da k -ésima covariável, $k = 1, 2, \dots, p$.

Nota-se que o modelo dos logitos cumulativos de chances não proporcionais (equação 2.3) assume que os efeitos das covariáveis diferem entre os $(c-1)$ logitos cumulativos (propriedade de chances não proporcionais), sendo necessários $(c-1)$ vetores de parâmetros de regressão $\boldsymbol{\beta}_j$, $j = 1, 2, \dots, c-1$, para descrever tais efeitos.

Em termos das probabilidades acumuladas $\theta_j(\mathbf{x})$, $j = 1, 2, \dots, c-1$, para o modelo dos logitos cumulativos de chances não proporcionais, é expresso por:

$$\theta_j(\mathbf{x}) = \frac{\exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x})}{1 + \exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x})},$$

de modo que as probabilidades $\pi_j(\mathbf{x})$, para $j = 1, 2, \dots, c$, ficam determinadas por meio da subtração dos $\theta_j(\mathbf{x})$, $j = 1, 2, \dots, c-1$, isto é,

$$\pi_j(\mathbf{x}) = \theta_j(\mathbf{x}) - \theta_{j-1}(\mathbf{x}), \quad \text{com } j = 1, 2, \dots, c.$$

A estimação dos parâmetros do modelo (2.3) é usualmente realizada, por meio do método de máxima verossimilhança (Agresti, 2010). Assim, para uma amostra aleatória de n indivíduos, a função de verossimilhança é dada por:

$$L = \prod_{i=1}^n \left\{ \prod_{j=1}^c [\pi_j(\mathbf{x}_i)]^{y_{ij}} \right\} = \prod_{i=1}^n \left\{ \prod_{j=1}^c [\theta_j(\mathbf{x}_i) - \theta_{j-1}(\mathbf{x}_i)]^{y_{ij}} \right\}, \quad (2.4)$$

em que, $y_{ij} = 1$ se a resposta do i -ésimo indivíduo, $i = 1, 2, \dots, n$, está na j -ésima categoria, $j = 1, 2, \dots, c$ e $y_{ij} = 0$, caso contrário, com $\sum_{j=1}^c y_{ij} = 1$. Nem sempre será possível maximizar tal função de verossimilhança (Giolo, 2017). Superdispersão e/ou observações esparsas podem comprometer a maximização dessa função (Agresti, 2010).

b) Modelo dos logitos cumulativos de chances proporcionais

Um modelo mais simplificado foi proposto por McCullagh (1980), denominado por modelo dos logitos cumulativos de chances proporcionais ou simplesmente modelo de chances proporcionais (MCP) definido por:

$$g(\mathbf{x}) = \text{logito}[\mathbb{P}(Y \leq j \mid \mathbf{x})] = \ln \left[\frac{\theta_j(\mathbf{x})}{1 - \theta_j(\mathbf{x})} \right] = \ln \left[\frac{\mathbb{P}(Y \leq j \mid \mathbf{x})}{\mathbb{P}(Y > j \mid \mathbf{x})} \right] = \alpha_j + \beta^\top \mathbf{x}, \quad (2.5)$$

com, $j = 1, 2, \dots, c-1$, $\beta^\top = (\beta_1, \beta_2, \dots, \beta_p)$ o vetor de parâmetros de regressão comum aos $(c-1)$ logitos. Portanto, o MCP (2.5) assume que os efeitos das covariáveis β não diferem entre os $(c-1)$ logitos (propriedade de chances proporcionais).

O modelo de chances proporcionais tem como característica calcular um valor de intercepto para cada categoria, ou seja, o número de interceptos depende unicamente das categorias e não das variáveis explicativas. Em resumo, pode-se dizer que esse modelo considera $(c-1)$ pontos de cortes das categorias. Por esse motivo, o nome de chances proporcionais, pois esse modelo assume uma idêntica razão de chances nos $(c-1)$ pontos de corte. Esse modelo é considerado o mais simples, nesta classe, pois envolve regressões paralelas, ao supor proporções equivalentes, ou seja, $\beta_j = \beta$ para todo j .

As probabilidades acumuladas são expressas por:

$$\theta_j(\mathbf{x}) = \frac{\exp(\alpha_j + \beta^\top \mathbf{x})}{1 + \exp(\alpha_j + \beta^\top \mathbf{x})}, \quad j = 1, 2, \dots, c-1.$$

Análogo ao modelo dos logitos cumulativos de chances não proporcionais a estimação dos parâmetros do MCP (2.5) pode ser feita, por meio da maximização da função de verossimilhança.

Para discriminar entre esses modelos, usa-se o teste da razão de verossimilhanças (TVR), cujas hipóteses são:

$$\begin{cases} H_0 : \beta_j = \beta, \\ H_1 : \beta_j \neq \beta, \end{cases} \quad \text{para algum } j.$$

A estatística do teste é dada por

$$\Lambda = -2 \log \left(\frac{L_{H_0}}{L_{H_1}} \right), \quad (2.6)$$

em que, L_{H_0} é o valor da função de verossimilhança sob a hipótese H_0 , ou seja, sob o modelo dos logitos cumulativos de chances proporcionais, e L_{H_1} representa o valor da função de verossimilhança sob a hipótese H_1 , ou seja, sob o modelo dos logitos cumulativos de chances não proporcionais. Para a tomada de decisão, se $\Lambda < \chi^2_{(d, 95\%)}$, em que d é o grau de liberdade (diferença do número de parâmetros dos modelos sob H_1 e H_0), H_0 não é rejeitada.

A pressuposição de chances proporcionais pode ser verificada de maneira global, contendo todas as covariáveis no modelo e individual, covariável por covariável. Outra maneira é utilizando o teste escore (Rao, 1948), pela maximização do logaritmo da função de verossimilhança sob a hipótese nula de proporcionalidade e tendo distribuição assintótica Qui-quadrado com $a(c-2)$ grau de liberdade, sendo a o número de parâmetros no modelo e c o número de classes da variável resposta (Stiger et al., 1999).

Em síntese, se a hipótese nula for rejeitada para todas as covariáveis, utiliza-se o modelo dos logitos cumulativos de chances não proporcionais. Se, contudo, a hipótese nula for rejeitada para parte das covariáveis, recomenda-se utilizar o modelo dos logitos cumulativos de chances proporcionais parciais.

c) Modelo dos logitos cumulativos de chances proporcionais parciais

Como a suposição de chances proporcionais nem sempre pode ser alcançada na prática, alternativamente, tem-se o modelo dos logitos cumulativos de chances proporcionais parciais, ou simplesmente modelo de chances proporcionais parciais (MCP), em que uma parte das covariáveis apresenta a propriedade de chances proporcionais e outra parte não.

Sendo $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^\top$ o vetor de covariáveis particionado, em que $\mathbf{x}_1$ é o vetor de valores observados para as covariáveis que apresentam chances proporcionais e $\mathbf{x}_2$ o vetor para as que não apresentam, o MCP proposto por Peterson e Harrell Jr (1990) é expresso por:

$$\begin{aligned} g(\mathbf{x}_1, \mathbf{x}_2) = \text{logito}[\mathbb{P}(Y \leq j \mid \mathbf{x}_1, \mathbf{x}_2)] &= \ln \left[\frac{\theta_j(\mathbf{x}_1, \mathbf{x}_2)}{1 - \theta_j(\mathbf{x}_1, \mathbf{x}_2)} \right] \\ &= \ln \left[\frac{\mathbb{P}(Y \leq j \mid \mathbf{x}_1, \mathbf{x}_2)}{\mathbb{P}(Y > j \mid \mathbf{x}_1, \mathbf{x}_2)} \right] \\ &= \alpha_j + \boldsymbol{\beta}^\top \mathbf{x}_1 + \boldsymbol{\gamma}_j^\top \mathbf{x}_2, \end{aligned} \quad (2.7)$$

em que, $\boldsymbol{\beta}^\top = (\beta_1, \beta_2, \dots, \beta_{q-1})$ e $\boldsymbol{\gamma}_j^\top = (\gamma_{qj}, \gamma_{q+1j}, \dots, \gamma_{pj})$, $j = 1, 2, \dots, c-1$, os vetores de parâmetros que descrevem os efeitos das covariáveis associados, respectivamente, aos vetores $\mathbf{x}_1$ e $\mathbf{x}_2$. Além do mais, $\mathbf{x}_1$ é um vetor de dimensão $q \times 1$, em que, $q \leq p$, variáveis preditoras do modelo (2.7) as quais o pressuposto de proporcionalidade é satisfeito e $\boldsymbol{\gamma}_j$ é um vetor com $p - q$ parâmetros. Se o parâmetro $\boldsymbol{\gamma}_j$ é nulo, o modelo se reduz ao modelo dos logitos cumulativos de chances proporcionais.

Em termos das probabilidades acumuladas, o modelo dos logitos cumulativos de chances proporcionais parciais é dado por:

$$\theta_j(\mathbf{x}_1, \mathbf{x}_2) = \frac{\exp(\alpha_j + \boldsymbol{\beta}^\top \mathbf{x}_1 + \boldsymbol{\gamma}_j^\top \mathbf{x}_2)}{1 + \exp(\alpha_j + \boldsymbol{\beta}^\top \mathbf{x}_1 + \boldsymbol{\gamma}_j^\top \mathbf{x}_2)}, \quad j = 1, 2, \dots, c-1.$$

Tal qual nos modelos (2.3) e (2.5), as probabilidades $\pi_j(\mathbf{x}_1, \mathbf{x}_2)$, para $j = 1, 2, \dots, c$, são obtidas a partir de subtrações das probabilidades acumuladas, isto é,

$$\pi_j(\mathbf{x}_1, \mathbf{x}_2) = \theta_j(\mathbf{x}_1, \mathbf{x}_2) - \theta_{j-1}(\mathbf{x}_1, \mathbf{x}_2), \quad j = 1, 2, \dots, c. \quad (2.8)$$

Quanto à estimação dos parâmetros do modelo (2.7), esta pode ser realizada, por meio do método de máxima verossimilhança. Detalhes adicionais sobre o modelo dos logitos cumulativos podem ser vistos em Paulino e Singer (2006), Agresti (2010) e Tutz (2012).

2.3.2 Modelo dos logitos de categorias adjacentes

Outro modelo alternativo em que não se utilizam as probabilidades acumuladas e, sim, consideram-se as razões das probabilidades para categorias sucessivas são os modelos dos logitos de categorias adjacentes (MLCA), ou seja, $\frac{\pi_j}{\pi_{j+1}}$, com $j = 1, 2, \dots, c-1$, sendo expresso por:

$$g(\mathbf{x}) = \text{logito}[\mathbb{P}(Y = j \mid \mathbf{x})] = \ln \left[\frac{\pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x})} \right] = \alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}, \quad j = 1, 2, \dots, c-1 \quad (2.9)$$

em que $\pi_j(\mathbf{x}) = \mathbb{P}(Y = j \mid \mathbf{x})$ denota a probabilidade de ocorrência da j -ésima categoria, tal que, $\sum_{j=1}^c \pi_j(\mathbf{x}) = 1$ e $\boldsymbol{\beta}_j$ o vetor de parâmetros que descreve os efeitos das covariáveis para o j -ésimo logito.

Observa-se que o modelo dos logitos de categorias adjacentes de chances não proporcionais assume que os efeitos das covariáveis diferem entre os $(c-1)$ logitos, com o j -ésimo logito representando a chance de ocorrência da j -ésima categoria de resposta em relação à categoria de resposta $j+1$.

Em termos das probabilidades $\pi_j(\mathbf{x})$, tem-se que:

$$\pi_j(\mathbf{x}) = \begin{cases} \frac{\exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x})}{1 + \sum_{j=1}^{c-1} \exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x})}, & j = 1, 2, \dots, c-1, \\ 1 - \sum_{j=1}^{c-1} \pi_j(\mathbf{x}), & j = c. \end{cases}$$

A estimação dos parâmetros do MLCA de chances não proporcionais pode ser realizada maximizando a função de verossimilhança L definida por:

$$L = \prod_{i=1}^n \left\{ \prod_{j=1}^c [\pi_j(\mathbf{x}_i)]^{y_{ij}} \right\} = \prod_{i=1}^n \left\{ \prod_{j=1}^{c-1} [\pi_j(\mathbf{x}_i)]^{y_{ij}} [\pi_c(\mathbf{x}_i)]^{y_{ic}} \right\}, \quad (2.10)$$

em que, $y_{ij} = 1$ se a resposta do i -ésimo indivíduo está na j -ésima categoria, $j = 1, 2, \dots, c$ e $y_{ij} = 0$, caso contrário, com $\sum_{j=1}^c y_{ij} = 1$ e $y_{ic} = 1 - \sum_{j=1}^{c-1} y_{ij}$ e $\mathbf{x}_i$ o vetor de valores das covariáveis do i -ésimo indivíduo, $i = 1, 2, \dots, n$.

Similar ao que foi apresentado para o modelo dos logitos cumulativos, os efeitos das covariáveis podem não diferir entre os logitos (propriedade de chances proporcionais), o que simplifica a expressão do modelo dos logitos de categorias adjacentes de chances não proporcionais para o modelo dos logitos de categorias adjacentes de chances proporcionais (MLCACP):

$$g(\mathbf{x}) = \text{logito}[\mathbb{P}(Y = j \mid \mathbf{x})] = \ln \left[\frac{\pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x})} \right] = \alpha_j + \boldsymbol{\beta}^\top \mathbf{x}, \quad j = 1, 2, \dots, c-1, \quad (2.11)$$

em que, $\boldsymbol{\beta}$ denota o vetor de parâmetros comum aos $(c-1)$ logitos, os quais descrevem os efeitos das covariáveis. Por outro lado, o efeito de algumas covariáveis pode diferir entre os logitos, enquanto o efeito de outras não (propriedade de chances proporcionais parciais). Nesse caso, o modelo fica exposto por:

$$g(\mathbf{x}_1, \mathbf{x}_2) = \text{logito}[\mathbb{P}(Y = j \mid \mathbf{x}_1, \mathbf{x}_2)] = \ln \left[\frac{\pi_j(\mathbf{x}_1, \mathbf{x}_2)}{\pi_{j+1}(\mathbf{x}_1, \mathbf{x}_2)} \right] = \alpha_j + \boldsymbol{\beta}^\top \mathbf{x}_1 + \boldsymbol{\gamma}_j^\top \mathbf{x}_2, \quad (2.12)$$

em que, $j = 1, \dots, c-1$ e $\mathbf{x}_1$ denota o vetor das covariáveis cujos efeitos não diferem entre os logitos e $\mathbf{x}_2$ o vetor das que diferem, com $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}_j$, $j = 1, 2, \dots, c-1$, os vetores de parâmetros que descrevem os efeitos das covariáveis.

A estimação dos parâmetros dos modelos (2.11) e (2.12) também se dá por meio de maximização de suas respectivas funções de verossimilhanças, sendo que para o modelo (2.12) tem-se

$$L = \prod_{i=1}^n \left\{ \prod_{j=1}^c [\pi_j(\mathbf{x}_{1i}, \mathbf{x}_{2i})]^{y_{ij}} \right\} = \prod_{i=1}^n \left\{ \prod_{j=1}^{c-1} [\pi_j(\mathbf{x}_{1i}, \mathbf{x}_{2i})]^{y_{ij}} [\pi_c(\mathbf{x}_{1i}, \mathbf{x}_{2i})]^{y_{ic}} \right\},$$

em que,

$$\pi_j(\mathbf{x}_{1i}, \mathbf{x}_{2i}) = \begin{cases} \frac{\exp(\alpha_j + \boldsymbol{\beta}^\top \mathbf{x}_{1i} + \boldsymbol{\gamma}_j^\top \mathbf{x}_{2i})}{1 + \sum_{j=1}^{c-1} \exp(\alpha_j + \boldsymbol{\beta}^\top \mathbf{x}_{1i} + \boldsymbol{\gamma}_j^\top \mathbf{x}_{2i})}, & j = 1, 2, \dots, c-1, \\ 1 - \sum_{j=1}^{c-1} \pi_j(\mathbf{x}_{1i}, \mathbf{x}_{2i}), & j = c. \end{cases}$$

com $y_{ij} = 1$ se a resposta do indivíduo i pertence à categoria j e 0 caso contrário, $j = 1, 2, \dots, c-1$, $y_{ij} = 1 - \sum_{j=1}^{c-1} y_{ij}$, $\mathbf{x}_{1i}$ e $\mathbf{x}_{2i}$ vetores de covariáveis.

Para selecionar ou discriminar os modelos (2.9), (2.11) e (2.12) recomenda-se o teste de proporcionalidade (Agresti, 2010; Giolo, 2017)

2.3.3 Modelo dos logitos de razão contínua

Outra opção para dados com resposta categorizada ordinal que leva em consideração uma relação entre as proporções de probabilidades $\frac{\pi_1}{\pi_2 + \dots + \pi_c}, \frac{\pi_2}{\pi_3 + \dots + \pi_c}, \dots, \frac{\pi_{j-1}}{\pi_j + \dots + \pi_c}, \dots, \frac{\pi_{c-1}}{\pi_c}$ são os modelos dos logitos de razão contínua (MLRC). Os $(c-1)$ logitos são definidos pelo logaritmo da chance de ocorrência da j -ésima categoria de resposta, $j = 1, 2, \dots, c-1$, relativo às categorias de resposta superiores a j , isto é,

$$g(\mathbf{x}) = \text{logito}[\mathbb{P}(Y = j | \mathbf{x})] = \ln \left[\frac{\mathbb{P}(Y = j | \mathbf{x})}{\mathbb{P}(Y > j | \mathbf{x})} \right] = \ln \left[\frac{\pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x}) + \dots + \pi_c(\mathbf{x})} \right], \quad (2.13)$$

sendo c o número de categorias de Y e $\mathbf{x}$ um vetor de covariáveis.

Esse modelo, portanto, compara a probabilidade de uma resposta igual a uma determinada categoria ($Y = j$) com a probabilidade de resposta no agrupamento de categorias de respostas de ordem mais elevadas ($Y > j$), que podem ser influenciadas pela direção escolhida para o ajuste. Uma de suas vantagens é que eles podem ser ajustados, por meio de j modelos de regressão logística. Alternativamente, os $(c-1)$ logitos podem ser expressos pelo logaritmo da chance de ocorrência da j -ésima categoria de resposta, $j = 1, 2, \dots, c$, relativo às categorias de resposta inferiores a j .

De acordo com Agresti (2010), os logitos de razão contínua (2.13) são particularmente úteis quando a variável resposta Y for caracterizada por processo sequencial no qual cada sujeito tem de ter passado por todas as categorias de respostas anteriores àquela em que se encontra. Por exemplo, o nível de escolaridade de uma pessoa (fundamental, médio, superior e pós-graduação).

Nessa classe de modelos, as probabilidades são expressas na forma condicional:

$$\zeta_j(\mathbf{x}) = \mathbb{P}(Y = j | Y \geq j, \mathbf{x}) = \frac{\pi_j(\mathbf{x})}{\pi_j(\mathbf{x}) + \dots + \pi_c(\mathbf{x})}, \quad j = 1, 2, \dots, c-1,$$

tendo em vista que,

$$\text{logito}[\mathbb{P}(Y = j | \mathbf{x})] = \ln \left[\frac{\zeta_j(\mathbf{x})}{1 - \zeta_j(\mathbf{x})} \right] = \ln \left[\frac{\mathbb{P}(Y = j | Y \geq j, \mathbf{x})}{1 - \mathbb{P}(Y = j | Y \geq j, \mathbf{x})} \right] = \ln \left[\frac{\mathbb{P}(Y = j | \mathbf{x})}{\mathbb{P}(Y > j | \mathbf{x})} \right]. \quad (2.14)$$

Similar ao que foi apresentado para os modelos dos logitos cumulativos e de categorias adjacentes, os efeitos das covariáveis no modelo dos logitos de razão contínua podem ser: a) diferentes entre os logitos (propriedade de chances não proporcionais), b) comuns entre os logitos (propriedade de chances proporcionais) ou, c) diferentes entre os logitos para algumas covariáveis e para outras não (propriedade

de chances proporcionais parciais). Para esses cenários, respectivamente, as expressões para o MLRC são

$$(i) \quad g_1(\mathbf{x}) = \ln \left[\frac{\zeta_j(\mathbf{x})}{1 - \zeta_j(\mathbf{x})} \right] = \ln \left[\frac{\mathbb{P}(Y = j \mid \mathbf{x})}{\mathbb{P}(Y > j \mid \mathbf{x})} \right] = -\alpha_j - \boldsymbol{\beta}_j^\top \mathbf{x}, \quad (2.15)$$

$$(ii) \quad g_2(\mathbf{x}) = \ln \left[\frac{\zeta_j(\mathbf{x})}{1 - \zeta_j(\mathbf{x})} \right] = \ln \left[\frac{\mathbb{P}(Y = j \mid \mathbf{x})}{\mathbb{P}(Y > j \mid \mathbf{x})} \right] = -\alpha_j - \boldsymbol{\beta}^\top \mathbf{x}, \quad (2.16)$$

$$(iii) \quad g_3(\mathbf{x}_1, \mathbf{x}_2) = \ln \left[\frac{\zeta_j(\mathbf{x}_1, \mathbf{x}_2)}{1 - \zeta_j(\mathbf{x}_1, \mathbf{x}_2)} \right] = \ln \left[\frac{\mathbb{P}(Y = j \mid \mathbf{x}_1, \mathbf{x}_2)}{\mathbb{P}(Y > j \mid \mathbf{x}_1, \mathbf{x}_2)} \right] = -\alpha_j - \boldsymbol{\beta}^\top \mathbf{x}_1 - \boldsymbol{\gamma}_j^\top \mathbf{x}_2, \quad (2.17)$$

em que, $\mathbf{x}_1$ e $\mathbf{x}_2$ são vetores de covariáveis e $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}_j$, $j = 1, 2, \dots, c-1$, são vetores de parâmetros que descrevem os efeitos das covariáveis. Em termos da probabilidade de ocorrência da categoria j , o modelo dos logitos de razão contínua com a propriedade de chances proporcionais fica expresso, para um dado vetor $\mathbf{x}$ de covariáveis, por:

$$\pi_j(\mathbf{x}) = \begin{cases} \frac{\exp(-\alpha_j - \boldsymbol{\beta}^\top \mathbf{x}_i)}{\prod_{k=1}^j [1 + \exp(-\alpha_k - \boldsymbol{\beta}^\top \mathbf{x})]}, & j = 1, 2, \dots, c-1, \\ 1 - \sum_{j=1}^{c-1} \pi_j(\mathbf{x}), & j = c. \end{cases}$$

Registra-se, também, que a estimação dos parâmetros do modelo dos logitos de razão contínua é realizada mediante ao método de máxima verossimilhança. A avaliação da suposição de chances proporcionais desse modelo também se dá de maneira similar à apresentada para os modelos dos logitos cumulativos e categorias adjacentes.

Os logitos de razão contínua também podem ser definidos como

$$g(\mathbf{x}) = \text{logito}[\mathbb{P}(Y = j \mid \mathbf{x})] = \ln \left[\frac{\mathbb{P}(Y = j \mid \mathbf{x})}{\mathbb{P}(Y > j \mid \mathbf{x})} \right] = \ln \left[\frac{\pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x}) + \dots + \pi_c(\mathbf{x})} \right], \quad (2.18)$$

em que $j = 1, 2, \dots, c-1$, com c o número de categorias de Y e $\mathbf{x}$ um vetor de valores de covariáveis.

De tal forma que o modelo dos logitos de razão contínua de chances proporcionais também pode ser denotado por

$$g(\mathbf{x}) = \text{logito}[\mathbb{P}(Y = j \mid \mathbf{x})] = \ln \left[\frac{1 - \zeta_j(\mathbf{x})}{\zeta_j(\mathbf{x})} \right] = \ln \left[\frac{\mathbb{P}(Y > j \mid \mathbf{x})}{\mathbb{P}(Y = j \mid \mathbf{x})} \right] = \alpha_j + \boldsymbol{\beta}^\top \mathbf{x}, \quad (2.19)$$

com probabilidades expressas na forma

$$\pi_j(\mathbf{x}) = \begin{cases} \frac{\exp(\alpha_j + \boldsymbol{\beta}^\top \mathbf{x}_i)}{\prod_{k=1}^j [1 + \exp(\alpha_k + \boldsymbol{\beta}^\top \mathbf{x})]}, & j = 1, 2, \dots, c-1, \\ 1 - \sum_{j=1}^{c-1} \pi_j(\mathbf{x}), & j = c. \end{cases}$$

2.4 Modelos lineares generalizados de efeitos mistos

A classe dos modelos lineares generalizados mistos estendem a classe dos modelos lineares para que: (i) a resposta esteja linearmente relacionada aos fatores e covariáveis, por meio de uma função de ligação especificada; (ii) a resposta possa ter uma distribuição não normal e (iii) as observações possam ser correlacionadas (Nelder e Wedderburn, 1972; Cordeiro e Demétrio, 2008).

Os modelos lineares generalizados de efeitos mistos abrangem uma ampla variedade de modelos, desde regressão linear simples até modelos multinomiais. Vale ressaltar que os modelos mistos são classificados assim quando em sua estrutura apresentam tanto fatores de efeitos fixos como aleatórios, além do erro experimental. A inclusão de efeitos aleatórios tem por objetivo, entre outros, captar e explicar dependências entre observações que podem existir em função do tipo de estudo ou delineamento realizado. Com a inclusão de efeitos aleatórios em um modelo, pode-se considerar uma matriz de covariâncias das variáveis respostas para descrever correlação entre as observações relacionadas entre si (Pinheiro e Bates, 2000). Os modelos mistos são conhecidos também por outras denominações como modelos hierárquicos, multiníveis ou ainda de componentes de variância, dependendo do contexto em que são aplicados (Diggle et al., 2002; Verbeke et al., 2014).

Além do mais, quando um modelo é considerado de efeito misto, sua análise de variância apresenta algumas peculiaridades, como a composição das esperanças matemáticas dos quadrados médios, cujo conhecimento permite o estabelecimento correto dos testes de hipóteses (Hicks, 1964).

O modelo linear misto pode ser expresso da seguinte forma:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad (2.20)$$

em que, $\mathbf{Y}$ é o vetor ($n \times 1$) de observações, $\boldsymbol{\beta}$ é o vetor ($p \times 1$) de parâmetros desconhecidos, $\mathbf{u}$ é o vetor ($q \times 1$) de variáveis aleatórias não observadas, $\mathbf{X}$ e $\mathbf{Z}$ são as matrizes de delineamento para os efeitos fixos e aleatórios, com dimensões ($n \times p$) e ($n \times q$), respectivamente, e $\boldsymbol{\epsilon}$ é o vetor de erros aleatórios ($n \times 1$). É, usualmente, assumido que $\mathbf{u}$ e $\boldsymbol{\epsilon}$ são independentes e normalmente distribuídos com vetor de médias $\mathbf{0}$, tendo $\mathbf{u} \sim N_q(\mathbf{0}, \mathbf{G})$ e $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$ (Laird e Ware, 1982).

Para os modelos lineares generalizados misto é necessária a inclusão de um efeito aleatório $\mathbf{u}$ no preditor linear, ajustando-se uma função de $\boldsymbol{\eta}$ do vetor de médias condicionais $\boldsymbol{\mu} = \mathbb{E}[\mathbf{Y}|\mathbf{u}]$ (McCulloch et al., 2011) para qualquer distribuição pertencente à família exponencial. Assim, os modelos lineares generalizados mistos podem ser definidos por:

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u},$$

em que

$$\mathbb{E}[Y_i|\mathbf{u}] = \mu_i$$

$$\text{Var}[Y_i|\mathbf{u}] = \phi V(\mu_i),$$

sendo $g(\cdot)$ a função de ligação, $\mathbf{x}_i$ a i -ésima linha da matriz de especificação do modelo associada aos efeitos fixos, $\boldsymbol{\beta}$ o vetor $(p \times 1)$ de parâmetros de efeitos fixos, $\mathbf{z}_i$ a i -ésima linha da matriz associada aos efeitos aleatórios, $\mathbf{u}$ o vetor $(q \times 1)$ de efeitos aleatórios, independentes e assumindo uma particular distribuição, como, por exemplo, Normal multivariada $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, em que $\mathbf{0}$ é o vetor de zeros e $\mathbf{G}$ é a matriz de componentes de variâncias desconhecidas associadas aos efeitos aleatórios presentes no modelo.

Mediante o teste da razão de verossimilhanças também é possível verificar se o efeito de $\mathbf{u}$ é significativo. Para esse teste, considera-se como hipótese nula $H_0 : \sigma_u^2 = 0$, em que, σ_u^2 é a variância de $\mathbf{u}$. Se a hipótese nula não for rejeitada, a inclusão do efeito aleatório não é necessária para o modelo. Esse teste possui distribuição assintótica χ^2 com 1 grau de liberdade, porém sendo conservador ao testar na fronteira do espaço paramétrico, em que, é aconselhável considerar uma mistura de Qui-quadrados $(\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2)$ (Hartzel et al., 2001; McCulloch et al., 2011).

2.5 Análise de correspondência

A análise de correspondência é um método multivariado que tem por objetivo analisar associações, que podem ou não existir, em uma tabela de contingência não havendo necessidade de assumir modelos e nem distribuições fundamentais. A técnica permite a representação simultânea das categorias (linhas e colunas) da tabela de contingência, em dimensão reduzida, que pode ser representada graficamente (Coldebella, 2000). Para a utilização desse método, é necessário verificar se existe associação entre as categorias das variáveis (linha e coluna). Um teste que verifica essa associação é o teste Qui-quadrado.

Nesse contexto, a análise de correspondência parte de uma matriz de dados representados por uma tabela de contingência. O desenvolvimento do algoritmo e a sua geometria fornecem as regras básicas para a interpretação. O algoritmo adaptado de Greenacre (1984) e Barioni Júnior (1995) apresentam os seguintes passos: seja uma tabela de contingência com $(T \times J)$ em que A é o atributo linha, com T categorias, e B o atributo coluna, com J categorias, com elementos não negativos, de modo que a soma de cada linha ou coluna não seja nula. Seja n_{tj} a frequência observada pela interseção da t -ésima categoria da variável A com a j -ésima categoria da variável B , logo $\mathbf{N} = [n_{tj}]_{T \times J}$ é a matriz de frequências absolutas. A matriz $\mathbf{N}$ pode transformar-se em uma matriz de frequências relativas expressa por: $\mathbf{P} = (1/n)\mathbf{N}$, em que n é a soma das frequências da linha (ou coluna) e $\mathbf{P}$ denomina-se matriz de correspondência.

Em função da matriz de correspondência $\mathbf{P}$, define-se o t -ésimo perfil linha como o vetor: $\mathbf{a}_t = (p_{t1}/p_{t.}, p_{t2}/p_{t.}, \dots, p_{tj}/p_{t.})^\top$, representando cada vetor $\mathbf{a}_t$ ($t = 1, \dots, T$) uma distribuição multinomial, condicionada à t -ésima categoria da variável A . Consequentemente, a matriz do perfil linha é definida por: $\mathbf{R} = \mathbf{D}_r^{-1}\mathbf{P}$ e a matriz do perfil coluna por: $\mathbf{C} = \mathbf{D}_c^{-1}\mathbf{P}^\top$, sendo $\mathbf{D}$ a matriz diagonal de frequências relativas a cada linha ou a cada coluna. Cada perfil (linha ou coluna) é afetado por um peso p_t para $t = 1, \dots, T$ ou p_j para $j = 1, \dots, J$, proporcional aos respectivos totais de linha ou de coluna dos dados originais. Estes pesos são denominados massas, sendo $r_t = p_t. = n_{t.}/n$ para $t = 1, \dots, T$ a massa da t -ésima linha e $c_j = p_{.j} = n_{.j}/n$ para $j = 1, \dots, J$ a massa da j -ésima coluna.

O centroide da linha (ou coluna) de uma tabela de contingência indica, geometricamente, a posição média dos perfis (linha ou coluna), como se fosse o centro de gravidade ou ponto de equilíbrio da matriz de dados. Como cada perfil (linha ou coluna) está associado a uma massa, define-se como centroide a média ponderada dos perfis (linha ou coluna), ou seja, o centroide do perfil linha $\mathbf{Ce}_{(r)} = \sum_t r_t \mathbf{a}_t = \mathbf{R}^\top \mathbf{r} = \mathbf{c}$ e o centroide do perfil coluna $\mathbf{Ce}_{(c)} = \sum_j c_j \mathbf{b}_j = \mathbf{C}^\top \mathbf{c} = \mathbf{r}$.

Cada perfil linha $\mathbf{a}_t$ define um ponto no espaço das linhas R^J . Logo, nesse espaço, acomoda-se um conjunto de pontos representando os T perfis linha. O fato de os elementos de cada perfil linha somarem 1 indica que a nuvem está restrita a uma região, no subespaço de dimensão $J - 1$ do espaço das colunas, conhecida como simplex. A interpretação do espaço do perfil coluna é feita de forma análoga. O formato do conjunto de pontos, ou o aparecimento de pontos isolados, pode indicar a existência ou não de associações entre pontos. Essas associações são medidas por meio das distâncias entre os pontos, utilizando-se a distância Qui-quadrado (χ^2) que é uma distância euclidiana ponderada entre os perfis linha e o centroide $\mathbf{Ce}$, sendo definida por: $D^2\{\mathbf{a}_t, \mathbf{Ce}\} = (\mathbf{R} - \mathbf{Ce}^\top \mathbf{1})^\top \mathbf{D}_{\mathbf{Ce}}^{-1} (\mathbf{R} - \mathbf{Ce}^\top \mathbf{1})$, em que $\mathbf{D}_{\mathbf{Ce}}^{-1}$ é a matriz de ponderação com diagonal representada pelo inverso das coordenadas do centroide. Se a métrica $\mathbf{R} \mathbf{D}_{\mathbf{Ce}}^{-1/2}$ com suas respectivas massas for aplicada, obtém-se a distância euclidiana clássica.

De acordo com a natureza dos dados para a análise, o conjunto de pontos terá maior ou menor dispersão em relação ao centroide. A dispersão pode ser calculada como a média dos desvios dos vários pontos ao centroide. O quantificador normalmente usado da dispersão dos pontos dessa nuvem no espaço R^J é chamado de inércia total do perfil linha.

A análise de correspondência pode ser aplicada como uma técnica exploratória para dados categorizados multivariados (Hoffman e Franke, 1986). Ainda, segundo Fellenberg et al. (2001), a análise de correspondência é um método computacional exploratório para o estudo de associações entre variáveis. O gráfico resultante dessa técnica reduz a dimensão dos dados de um espaço multidimensional para um espaço bidimensional. Tal resultado gráfico mostra a correspondência entre as categorias de variáveis (linha e coluna), analogamente a um gráfico *biplot* (Johnson et al., 2008). Dessa forma, pontos de perfis semelhantes, se localizam mais próximos no plano do que pontos de perfis com características discrepantes.

2.6 Material

Os dados utilizados nesse estudo provêm de uma pesquisa desenvolvida no departamento de Tecnologia de Alimentos da Universidade Federal do Ceará, no ano de 2016 (Rebouças, 2016).

Para elaborar uma bebida com elevado apelo nutricional e funcional foi feita uma combinação das características nutricionais, funcionais e sensoriais das matérias – primas: castanha de caju, suco de uva e substâncias prebióticas. Foram utilizadas duas substâncias prebióticas, inulina (grau de polimerização ≥ 10 , Orafiti GR) e oligofrutose (2 - 8 monômeros, Orafiti P95) e adicionado o açúcar cristal comercial. Para a obtenção do extrato hidrossolúvel utilizaram-se amêndoas da castanha de caju beneficiadas e cruas. Para saborizar as bebidas, foi utilizado concentrado de suco de uva (pH = 2,99; 15,2°Brix), o qual foi definido, por meio de estudos preliminares.

As combinações dos percentuais de suco de uva (16%, 20%, 30%, 40% e 44%) e açúcar (3%, 4%, 6%, 8% e 9%) para as formulações de bebidas prebióticas foram feitas por meio de um fatorial 2^2 em um delineamento composto central rotacional, com cinco réplicas no ponto central, totalizando 13 formulações de bebidas (F_1 a F_{13}), sendo F_9 a F_{13} os de ponto central, como ilustrado na Figura 2.1.

Figura 2.1: Estrutura das 13 formulações de bebidas prebióticas de suco de uva, por meio de um fatorial 2^2 de um delineamento composto central rotacional, conforme estudo desenvolvido na Universidade Federal do Ceará, no ano de 2016

A finalidade do estudo foi desenvolver bebidas prebióticas à base de amêndoa da castanha de caju adicionadas ao suco de uva, a fim de avaliar a aceitação dos atributos sensoriais de impressão global, aroma, corpo, doçura e sabor, utilizando uma escala hedônica de 9 pontos (“9 = gostei extremamente”, “8 = gostei muito”, “7 = gostei moderadamente”, “6 = gostei ligeiramente”, “5 = indiferente”, “4 = desgostei ligeiramente”, “3 = desgostei moderadamente”, “2 = desgostei muito”, “1 = desgostei extremamente”). Nesse trabalho, para efeito de aplicação dos métodos, a escala hedônica original de 9 pontos foi reduzida para 5 pontos em que “5 = gostei extremamente ou muito”, “4 = gostei moderadamente ou ligeiramente”, “3 = indiferente”, “2 = desgostei ligeiramente ou moderadamente”, “1 = desgostei muito ou extremamente”, em razão da ocorrência de dados esparsos.

A avaliação da aceitação sensorial das formulações de bebidas prebióticas de suco de uva foi realizada somente em uma sessão, em que participaram 130 julgadores não treinados, de ambos os sexos e de diferentes faixas etárias. As amostras foram servidas de forma monádica sequencial, seguindo um delineamento de blocos incompletos balanceados, em que cada provador avaliou 4 das 13 formulações das bebidas propostas (Figura 2.2).

Figura 2.2: Esquema experimental em blocos incompletos balanceados com 130 provadores e 13 formulações de bebidas prebióticas de suco de uva propostas, conforme estudo desenvolvido na Universidade Federal do Ceará, no ano de 2016

De acordo com Hinkelmann e Kempthorne (2005), um delineamento em blocos incompletos é dito ser balanceado (ou equilibrado) quando satisfizerem às seguintes condições: (i) o material experimental é dividido em b blocos (provadores) de l unidades cada, aplicando-se diferentes tratamentos (produtos) nas unidades no mesmo bloco; (ii) existem t tratamentos, cada um dos quais ocorre em blocos r vezes e (iii) qualquer dois tratamentos ocorrem em blocos exatamente λ vezes.

Montgomery (2013) relata que quando todas as comparações de tratamento são igualmente importantes, as combinações de tratamento usadas em cada bloco devem ser selecionadas de maneira balanceada, de modo que qualquer par de tratamentos ocorra junto o mesmo número de vezes que qualquer outro par. As seguintes relações são condições necessárias para se ter um delineamento em blocos incompletos balanceados:

$$rt = lb, \quad (2.21)$$

$$\lambda(t-1) = r(l-1), \quad (2.22)$$

$$r > \lambda, \quad (2.23)$$

$$b \geq t. \quad (2.24)$$

Assumindo que existem t tratamentos e b blocos e cada bloco contém l tratamentos, que cada tratamento ocorre r vezes no experimento (ou é replicado r vezes), o número total de parcelas pode ser obtida por $n = lb$, ou $n = rt$, resultando na equação 2.21. Segundo Montgomery (2013), um experimento em blocos incompletos pode ser constituído em $\binom{t}{2}$ pares de tratamento e cada par ocorre em λ blocos e cada bloco tem $\binom{l}{2}$ pares de tratamentos e, então, em b blocos estão $b\binom{l}{2}$ blocos. Desse modo, $\lambda t(t-1) = bl(l-1)$.

2.7 Métodos

Neste estudo, a variável de interesse é categorizada por corresponder a uma escala hedônica: “1 = desgostei extremamente ou muito” a “5 = gostei extremamente ou muito”, para avaliação de uma formulação. Além disso, a variável resposta é ordinal, ou seja, $1 \leq 2 \leq 3 \leq 4 \leq 5$, o que indica a opinião dos provadores em uma escala do menos favorável ao mais favorável.

Dado isso, inicialmente realizou-se uma descrição gráfica, por meio da análise de correspondência para verificar de forma preliminar relações entre as categorias de resposta atribuídas por cada provador, com referência aos atributos sensoriais (impressão global, aroma, corpo, doçura e sabor) e às formulações de bebidas prebióticas (F_1 a F_{13}). No entanto, esse método não leva em consideração o efeito aleatório do provador, em decorrência de uma possível dependência entre as observações de um mesmo provador.

Posteriormente, ajustaram-se modelos independentes dos logitos cumulativos para cada um dos cinco atributos sensoriais, tendo como covariável a formulação de bebida prebiótica de suco de uva. A resposta do i -ésimo indivíduo é denotada pelo vetor $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{ij})^\top$, em que Y_{ij} 's representam as variáveis indicadoras para as categorias de resposta, isto é, $Y_{ij} = 1$ se o i -ésimo indivíduo optou pela j -ésima categoria e $Y_{ij} = 0$, caso contrário e $\theta_j = \mathbb{P}(Y \leq j \mid \mathbf{F})$, $j = 1, 2, \dots, 5$, é a probabilidade acumulada de ocorrência até a j -ésima categoria para uma dada formulação de bebida em relação a um determinado atributo sensorial. Diante disso, o modelo proposto dos logitos cumulativos é representado por:

$$g(\mathbf{F}) = \ln \left[\frac{\theta_j(\mathbf{F})}{1 - \theta_j(\mathbf{F})} \right] = \alpha_j + \beta_j^\top \mathbf{F}, \quad (2.25)$$

sendo, α_j o intercepto da j -ésima categoria de resposta do provador, $\beta_j^\top = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jt})$ o vetor de parâmetros associado às formulações de bebidas, com $j = 1, 2, \dots, 5$ e $t = 1, 2, \dots, 13$, e $\mathbf{F}$ a matriz de delineamento para o efeito fixo ($n \times t$). Nota-se que o modelo dos logitos cumulativos pode ser de chances não proporcionais, em que se assume que o efeito da covariável difere entre os 4 logitos cumulativos (propriedade de chances não proporcionais), sendo necessário 4 vetores de parâmetros de regressão $\beta^\top = (\beta_{1;1}, \beta_{1;2}, \dots, \beta_{1;13}, \dots, \beta_{4;1}, \beta_{4;2}, \dots, \beta_{4;13})$ para descrever tal efeito. Já para o caso de chances proporcionais, assume-se que o vetor de parâmetros de regressão é comum para os 4 logitos, $\beta^\top = (\beta_1, \beta_2, \dots, \beta_{13})$.

Em termos de probabilidades acumuladas, para o modelo (2.25), é definido por:

$$\theta_j(\mathbf{F}) = \frac{\exp(\alpha_j + \beta_j^\top \mathbf{F})}{1 + \exp(\alpha_j + \beta_j^\top \mathbf{F})}, \quad (2.26)$$

send, β_j o vetor de parâmetros que descreve o efeito da covariável formulação de bebida.

As estimativas desses parâmetros são obtidas pela teoria da máxima verossimilhança, com o uso do processo iterativo de Newton-Raphson, conforme descrito na seção 2.3 (Azzalini, 1996). Para o estudo em questão, a função de verossimilhança é

$$L = \prod_{i=1}^{130} \left\{ \prod_{j=1}^5 [\pi_j(\mathbf{F}_i)]^{Y_{ij}} \right\} = \prod_{i=1}^{130} \left\{ \prod_{j=1}^5 [\theta_j(\mathbf{F}_i) - \theta_{j-1}(\mathbf{F}_i)]^{Y_{ij}} \right\}, \quad (2.27)$$

em que, $Y_{ij} = 1$ se a resposta do i -ésimo indivíduo, $i = 1, 2, \dots, 130$, está na j -ésima categoria de resposta, $j = 1, 2, \dots, 5$ e $Y_{ij} = 0$, caso contrário, com $\sum_{j=1}^5 Y_{ij} = 1$ e F_i é a i -ésima linha de $\mathbf{F}$.

Para a utilização do modelo dos logitos cumulativos de chances proporcionais é necessário verificar a suposição de proporcionalidade das razões de chances. Para isso, usa-se o teste da razão de verossimilhanças (TRV), conforme descrito na seção 2.3. Se o teste for significativo, deve-se optar por usar o modelo dos logitos cumulativos de chances não proporcionais, o que implica um aumento no número de parâmetros.

Uma vez selecionada a estrutura fixa do modelo, deve-se considerar a inclusão de efeitos aleatórios para o provador, em razão de uma possível dependência entre as observações. Assim, para o modelo dos logitos cumulativos com a inclusão de efeito aleatório para provador, tem-se:

$$g(\mathbf{F}, \mathbf{Z}) = \ln \left[\frac{\theta_j(\mathbf{F})}{1 - \theta_j(\mathbf{F})} \right] = \alpha_j + \beta_j^\top \mathbf{F} + \mathbf{u}_i^\top \mathbf{Z}, \quad (2.28)$$

sendo, α_j o intercepto da j -ésima categoria de resposta referente ao um determinado atributo sensorial (impressão global, aroma, corpo, doçura e sabor), β_j o vetor de parâmetros de regressão associado à variável explicativa formulação de bebida prebiótica (F_t), com $j = 1, 2, \dots, 4$ e $t = 1, 2, \dots, 13$, $\mathbf{F}$ e $\mathbf{Z}$ são as matrizes de delineamento para os efeitos fixos e aleatórios, com dimensões $(n \times t)$ e $(n \times q)$, respectivamente, e $\mathbf{u}_i$ é o efeito aleatório associado ao i -ésimo provador, em que $\mathbf{u}_i \sim N(0, \sigma_u^2)$. As estimativas desses parâmetros também são obtidas pela teoria da máxima verossimilhança, com o uso do processo iterativo de Newton-Raphson (Azzalini, 1996).

Para avaliar a significância do efeito aleatório de provador, pode-se efetuar o teste da razão de verossimilhanças, conforme descrito na seção 2.4. A alternativa é construir, a partir da verossimilhança perfilada, o intervalo de confiança para o parâmetro σ_u . Isto é, se o intervalo de 95% de confiança para o desvio padrão do coeficiente aleatório não contiver o valor zero, verifica-se a existência de efeito aleatório para provador.

Depois de selecionado o modelo, calculam-se as estimativas dos parâmetros e se obtêm as probabilidades estimadas. O fluxograma das etapas para o modelo dos logitos cumulativos encontra-se na Figura 2.3.

Figura 2.3: Etapas para o modelo dos logitos cumulativos misto

O procedimento computacional foi desenvolvido usando a função `clmm2(.)` do pacote `ordinal` (Christensen e Christensen, 2015) disponível no *software* R (R Core Team, 2020) para o modelo de efeitos mistos dos logitos cumulativos de chances proporcionais, a referência da categoria da resposta foi “1 = desgostei muito ou extremamente” e para a formulação de bebida foi 4% de açúcar e 20% de suco de uva (F_1). Para análise de correspondência, utilizou-se o pacote `ca` (Nenadic e Greenacre, 2007). O roteiro dos comandos utilizados para fazer as análises encontra-se no Apêndice I.

2.8 Resultados e discussões

Inicialmente, verificou-se associação entre as categorias atribuídas por cada provador aos atributos sensoriais (impressão global, aroma, corpo, doçura e sabor) e as formulações de bebidas prebióticas de suco de uva (F_1 a F_{13}) pelo teste χ^2 . O resultado do teste foi significativo para os atributos sensoriais de impressão global (p-valor = 0,003), aroma (p-valor = 0,003), corpo (p-valor = 0,019), doçura (p-valor = 0,001) e sabor (p-valor = 0,009), ao nível de 5% de significância, ou seja, há evidências de que as

categorias atribuídas por cada provador e as formulações de bebidas são dependentes (estão associadas), com isso pode-se aplicar a análise de correspondência.

(a) Categorias atribuídas por cada provador, com relação ao atributo sensorial impressão global, e as formulações de bebidas prebióticas de suco de uva

(b) Categorias atribuídas por cada provador, com relação ao atributo sensorial aroma, e as formulações de bebidas prebióticas de suco de uva

(c) Categorias atribuídas por cada provador, com relação ao atributo sensorial corpo, e as formulações de bebidas prebióticas de suco de uva

(d) Categorias atribuídas por cada provador, com relação ao atributo sensorial doçura, e as formulações de bebidas prebióticas de suco de uva

(e) Categorias atribuídas por cada provador, com relação ao atributo sensorial sabor, e as formulações de bebidas prebióticas de suco de uva

Figura 2.4: Gráfico entre as categorias atribuídas por cada provador, com relação aos atributos sensoriais e as formulações de bebidas prebióticas de suco de uva, propostas do estudo desenvolvido na Universidade Federal do Ceará, no ano de 2016

Na Figura 2.4, apresentam-se os gráficos obtidos por meio da análise de correspondência entre as categorias atribuídas por cada provador, para cada um dos cinco atributos avaliados e as formulações de bebidas prebióticas de suco de uva. Verificou-se, que a capacidade de representação dos pontos por duas dimensões, para o atributo sensorial de impressão global é de 86,9% (Figura 2.4a), para aroma é de 83,2% (Figura 2.4b), para corpo é de 80,0% (Figura 2.4c), para doçura é de 83,8% (Figura 2.4d) e para sabor é de 84,1% (Figura 2.4e), indicando uma representatividade acima de 80% da associação das formulações de suco de uva (F_1 a F_{13}) com as categorias (“1 = desgostei muito ou extremamente” a “5 = gostei extremamente ou muito”) dos atributos sensoriais (impressão global, aroma, corpo, doçura e sabor). Ainda na Figura 2.4, nota-se que as formulações de bebidas que mais se aproximaram da categoria de resposta 5 (gostei extremamente ou muito), para o atributo sensorial impressão global foram F_3 , F_4 ,

F_6 , F_{11} e F_{13} (Figura 2.4a), com relação ao aroma foi a formulação F_5 (Figura 2.4b), já para corpo foram F_2 , F_3 , F_6 , F_8 e F_{13} (Figura 2.4c), quanto à doçura foram F_6 , F_8 , F_{11} e F_{13} (Figura 2.4d) e, finalmente, para sabor foram F_{10} , F_{11} , F_{12} e F_{13} (Figura 2.4e). Observa-se que a formulação F_{13} encontra-se em quase todos os atributos sensoriais próxima da categoria 5 (gostei extremamente ou muito), com exceção do atributo aroma.

A seguir, foram ajustados os modelos dos logitos cumulativos com e sem chances proporcionais para se aplicar o teste de proporcionalidade de razões de chances. Certificou-se que, para todos os atributos, houve proporcionalidade das razões de chances para os 4 pontos de corte, justificando o uso do modelo de chances proporcionais. Os níveis descritos dos TRV foram: impressão global (p-valor = 0,246), aroma (p-valor = 0,054), corpo (p-valor = 0,067), doçura (p-valor = 0,170) e sabor (p-valor = 0,157).

Adicionalmente, para cada atributo sensorial é aplicado o teste de razão de verossimilhanças para verificar se existe efeito da covariável formulação de bebida, verificando-se, ao nível significância de 5%, efeito significativo (p-valor < 0,01) em todos os casos.

Além disso, a inclusão do efeito aleatório, em provador, se fez necessário. Por meio do critério da verossimilhança perfilada, obteve-se a estimativa para o desvio padrão do coeficiente aleatório e o intervalo de confiança de 95%. Para impressão global $\hat{\sigma}_{uIG} = 2,13$ e o intervalo (1,75 a 2,58). Para aroma $\hat{\sigma}_{uAroma} = 2,40$ e o intervalo (1,99 a 2,89). Para corpo $\hat{\sigma}_{uCorpo} = 2,34$ e o intervalo (1,93 a 2,83). Para doçura $\hat{\sigma}_{uDocura} = 1,72$ e o intervalo (1,39 a 2,12). Para sabor $\hat{\sigma}_{uSabor} = 2,15$ e o intervalo (1,76 a 2,61). Portanto, como os intervalos não contém o valor zero, pode-se considerar a existência de efeito aleatório para provador em cada um dos modelos isolados.

Na Tabela 2.1, são apresentadas as estimativas, os erros padrões, p -valor e o valor da razão de chances associados aos parâmetros do modelo de chances proporcionais de efeito misto (2.28) com relação ao atributo sensorial impressão global. As tabelas referentes aos demais atributos não são apresentadas no corpo deste trabalho, porém são apresentadas as discussões e conclusões.

Tabela 2.1: Estimativas dos parâmetros, erros padrões, estatística do teste e razão de chances para formulações de bebidas prebiótica de suco de uva com a escala hedônica em relação ao atributo sensorial impressão global do estudo desenvolvido na Universidade Federal do Ceará, no ano 2016

Parâmetro	Estimativa	Erro padrão	p -valor	Razão de chances
α_2	-1,67	0,41	< 0,01	0,19
α_3	0,82	0,41	0,04	2,27
α_4	1,65	0,41	< 0,01	5,20
α_5	4,79	0,48	< 0,01	120,05
β_2	1,18	0,49	0,02	3,26
β_3	1,71	0,50	< 0,01	5,54
β_4	3,02	0,52	< 0,01	20,39
β_5	0,70	0,49	0,15	2,02
β_6	2,85	0,52	< 0,01	17,32
β_7	0,19	0,50	0,70	1,21
β_8	2,03	0,51	< 0,01	7,59
β_9	2,09	0,49	< 0,01	8,08
β_{10}	1,86	0,49	< 0,01	6,39
β_{11}	2,04	0,50	< 0,01	7,70
β_{12}	2,31	0,50	< 0,01	10,03
β_{13}	2,76	0,52	< 0,01	15,82

Pela Tabela 2.1, observa-se que o p -valor de β_5 e β_7 são maiores do que o nível de significância de 5%, indicando que há evidências que as estimativas das formulações de bebidas $\beta_1 = \beta_5$ e $\beta_1 = \beta_7$. Portanto, ao paladar dos provadores em relação ao atributo sensorial impressão global não houve diferença entre elas. Verifica-se também, nessa mesma tabela, que entre os valores das estimativas os β_4 , β_6 e β_{13} obtiveram os valores mais altos, indicando uma maior aceitação ao paladar dos provadores em relação ao atributo sensorial impressão global e que a chance do indivíduo ter uma maior aceitabilidade com relação a formulação de bebida F_4 é 20 vezes maior do que a formulação F_1 .

Na Tabela 2.2 são apresentadas as probabilidades estimadas para cada categoria de resposta de impressão global em relação às formulações de bebidas prebióticas de suco de uva, obtidas por meio do modelo dos logitos cumulativos de chances proporcionais de efeito misto (equação 2.28), além dos valores das probabilidades acumuladas das categorias 4 (gostei moderadamente ou ligeiramente) e 5 (gostei extremamente ou muito).

Tabela 2.2: Probabilidades estimadas pelo modelo de chances proporcionais de efeito misto (2.28) de cada categoria de resposta para formulação de bebidas prebiótica de suco de uva com relação ao atributo sensorial impressão global do estudo desenvolvido na UFC, no ano 2016

Formulação de bebida	Impressão Global					$P(Y \geq 4 F)$
	1	2	3	4	5	
F_1	0,16	0,54	0,15	0,15	0,01	0,16
F_2	0,06	0,36	0,20	0,36	0,03	0,39
F_3	0,03	0,26	0,19	0,47	0,04	0,52
F_4	0,01	0,09	0,10	0,65	0,15	0,80
F_5	0,09	0,44	0,19	0,26	0,02	0,28
F_6	0,01	0,11	0,12	0,64	0,13	0,77
F_7	0,14	0,52	0,16	0,18	0,01	0,19
F_8	0,02	0,21	0,18	0,53	0,06	0,59
F_9	0,00	0,20	0,17	0,55	0,06	0,61
F_{10}	0,00	0,00	0,19	0,50	0,05	0,55
F_{11}	0,00	0,20	0,18	0,54	0,06	0,60
F_{12}	0,00	0,17	0,16	0,58	0,08	0,66
F_{13}	0,01	0,11	0,12	0,64	0,12	0,75

O critério utilizado para inferir quais as formulações de bebidas prebióticas foram as mais aceitas pelos 130 julgadores em relação ao atributo sensorial impressão global, foi a probabilidade acumulada das categorias 4 (gostei moderadamente ou ligeiramente) e 5 (gostei extremamente ou muito) estimadas com base no modelo selecionado. Na Tabela 2.2, observa-se que as formulações F_4 , F_6 e F_{13} obtiveram as mais altas probabilidades acumuladas com 0,80, 0,77 e 0,75, respectivamente, as quais já haviam sido sinalizadas como bem aceitas na análise de correspondência (vide Figura 2.4).

Em relação aos demais atributos sensoriais, seguiram-se os mesmos procedimentos metodológicos adotados para o atributo sensorial impressão global. Os principais resultados estão sintetizados, na Tabela 2.3, em que são apresentadas as probabilidades acumuladas estimadas das categorias 4 e 5 obtidas por meio do modelo dos logitos cumulativos de chances proporcionais de efeito misto.

Pela Tabela 2.3, constata-se que as formulações de bebidas prebióticas de suco de uva melhores avaliadas, para o atributo sensorial aroma foram F_6 , F_{12} e F_{13} com probabilidades de 0,49, 0,47 e 0,48, respectivamente. Para o atributo sensorial corpo foram as formulações de suco de uva F_6 , F_9 e F_{13} , com probabilidades 0,81, 0,77 e 0,78, respectivamente. Para o atributo sensorial doçura foram as formulações de suco de uva F_6 , F_{11} e F_{13} , com probabilidades de 0,84, 0,79 e 0,80, respectivamente. Para o atributo sensorial sabor foram as formulações de suco de uva F_4 , F_6 e F_{13} , com probabilidades de 0,85, 0,75 e 0,74, respectivamente.

Tabela 2.3: Probabilidades acumuladas estimadas para as categorias 4 e 5 das formulações de bebidas prebióticas de suco de uva dos atributos sensoriais do estudo desenvolvido na Universidade Federal do Ceará, no ano 2016

Formulação de bebida	Atributo Sensorial			
	Aroma	Corpo	Doçura	Sabor
	$P(Y \geq 4 F)$	$P(Y \geq 4 F)$	$P(Y \geq 4 F)$	$P(Y \geq 4 F)$
F_1	0,10	0,39	0,48	0,15
F_2	0,19	0,53	0,74	0,38
F_3	0,36	0,55	0,53	0,57
F_4	0,38	0,64	0,62	0,85
F_5	0,01	0,55	0,52	0,47
F_6	0,49	0,81	0,84	0,75
F_7	0,09	0,33	0,17	0,14
F_8	0,35	0,75	0,72	0,73
F_9	0,38	0,77	0,74	0,59
F_{10}	0,33	0,63	0,62	0,51
F_{11}	0,43	0,67	0,79	0,64
F_{12}	0,47	0,68	0,78	0,62
F_{13}	0,48	0,78	0,80	0,74

Fazendo uma associação da Tabela 2.3 (levando em consideração apenas as três formulações de suco de uva com valores mais altos das probabilidades acumuladas nas categorias de resposta 4 e 5) com os gráficos obtidos por meio da análise de correspondência, verificou-se que as formulações de suco de uva melhor avaliadas pelos 130 provadores, para o atributo sensorial corpo foram F_6 e F_{13} , doçura F_6 , F_{11} e F_{13} e sabor F_{13} .

As proporções observadas e as previstas das categorias atribuídas por cada provador para as formulações de bebidas prebióticas de suco de uva com relação ao atributo sensorial impressão global são apresentadas na Figura 2.5.

Figura 2.5: Proporções observadas e proporções previstas fornecida pelo o modelo dos logitos cumulativos de chances proporcionais mistos para formulação de suco com relação ao atributo sensorial impressão global do estudo desenvolvido pela Universidade Federal do Ceará, no ano 2016

Na Figura 2.5, em geral, as probabilidades observadas estão próximas das previstas com raras exceções, mostrando que o modelo dos logitos cumulativos de chances proporcionais de efeito misto (2.28) tem uma boa capacidade preditiva.

2.9 Conclusão

Neste trabalho, descreve-se uma alternativa de análise de dados sensoriais, com ênfase no modelo dos logitos cumulativos de chances proporcionais, com a inclusão de efeito aleatório para provador. Além disso, apresentou-se uma técnica exploratória preliminar baseada na análise de correspondência.

O modelo proposto, neste capítulo, forneceu um bom ajuste aos dados e respondeu aos objetivos do estudo de propor a formulação de bebida prebiótica com maior aceitabilidade para os consumidores com relação à cada atributo sensorial.

As vantagens do uso desses modelos são: trabalhar com a distribuição original dos dados, permitir ensaios desbalanceados e com dependência entre observações sistemáticas, aplicabilidade em demais áreas correlatas. Porém ainda há a necessidade de desenvolvimento de métodos e técnicas de diagnósticos.

Referências

- Abd El-Salam, M. H., Hippen, A. R., El-Shafie, K., Assem, F. M., Abbas, H., Abd El-Aziz, M., Sharaf, O., e El-Aassar, M. (2011). Preparation and properties of probiotic concentrated yoghurt (labneh) fortified with conjugated linoleic acid. *International journal of food science & technology*, 46(10):2103–2110.
- Agresti, A. (2010). *Analysis of ordinal categorical data*, volume 656. John Wiley & Sons.
- Azzalini, A. (1996). *Statistical inference based on the likelihood*, volume 68. CRC Press.
- Barioni Júnior, W. (1995). *Análise de correspondência na identificação dos fatores de risco associados à diarreia e à performance de leitões na fase de lactação*. PhD thesis, Universidade de São Paulo.
- Christensen, R. H. B. e Brockhoff, P. B. (2013). Analysis of sensory ratings data with cumulative link models. *Journal de la Societe Francaise de Statistique & Revue de Statistique Appliquee*, 154(3):58–79.
- Christensen, R. H. B. e Christensen, M. R. H. B. (2015). Package “ordinal”. *Stand*, 19:2016.
- Coldebella, A. (2000). *Análise de correspondência na identificação de variáveis associadas com intervalo entre partos em bovinos leiteiros*. PhD thesis, Universidade de São Paulo.
- Cordeiro, G. M. e Demétrio, C. G. (2008). Modelos lineares generalizados e extensões. *Piracicaba: USP*.
- Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S., et al. (2002). *Analysis of longitudinal data*. Oxford University Press.
- Fellenberg, K., Hauser, N. C., Brors, B., Neutzner, A., Hoheisel, J. D., e Vingron, M. (2001). Correspondence analysis applied to microarray data. *Proceedings of the National Academy of Sciences*, 98(19):10781–10786.
- Giolo, S. R. (2017). *Introdução à análise de dados categóricos com aplicações*. Editora Blucher.
- Greenacre, M. J. (1984). Theory and applications of correspondence analysis.
- Hartzel, J., Agresti, A., e Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, 1(2):81–102.

- Hicks, C. R. (1964). Fundamental concepts in the design of experiments.
- Hinkelmann, K. e Kempthorne, O. (2005). *Design and analysis of experiments: Advanced experimental design*, volume Volume 2 of *Wiley series in probability and mathematical statistics. Applied probability and statistics*. Wiley, rev. ed. edition.
- Hoffman, D. L. e Franke, G. R. (1986). Correspondence analysis: graphical representation of categorical data in marketing research. *Journal of marketing Research*, 23(3):213–227.
- Johnson, R. A., Wichern, D. W., et al. (2008). *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ.
- Laird, N. M. e Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.
- Lemos, T. D. O., Rodrigues, M. D. C. P., De Lara, I. A. R., De Araújo, A. M. S., De Lemos, T. L. G., Pereira, A. L. F., e De Paula, L. V. T. (2015). Modeling the acceptability of cashew apple nectar brands using the proportional odds model. *Journal of Sensory Studies*, 30(2):136–144.
- Li, B., Hayes, J. E., e Ziegler, G. R. (2014). Interpreting consumer preferences: Physicohedonic and psychohedonic models yield different information in a coffee-flavored dairy beverage. *Food quality and preference*, 36:27–32.
- Los, P. R., Marson, G. V., Dutcosky, S. D., Nogueira, A., Marinho, M. T., e Simões, D. R. S. (2020). Optimization of beef patties produced with vegetable oils: a mixture design approach and sensory evaluation. *Food Science and Technology*, (AHEAD).
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127.
- McCulloch, C., Searle, S., e Neuhaus, J. (2011). Generalized, linear, and mixed models. second.
- Montgomery, D. C. (2013). *Design and analysis of experiments*. John wiley & sons.
- Nelder, J. A. e Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384.
- Nenadic, O. e Greenacre, M. (2007). Correspondence analysis in r, with two-and three-dimensional graphics: the ca package. *Journal of statistical software*, 20(3).
- Paulino, C. D. M. e Singer, J. d. M. (2006). *Análise de dados categorizados*. Edgard Blücher São Paulo.
- Pereira, A. L. F., Almeida, F. D. L., de Jesus, A. L. T., da Costa, J. M. C., e Rodrigues, S. (2013). Storage stability and acceptance of probiotic beverage from cashew apple juice. *Food and bioprocess technology*, 6(11):3155–3165.
- Peterson, B. e Harrell Jr, F. E. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 39(2):205–217.
- Pinheiro, J. C. e Bates, D. M. (2000). Mixed-effects models in s and s-plus.,(springer-verlag: New york). *Mixed-effects models in S and S-Plus*. Springer Verlag, New York.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 44, pages 50–57. Cambridge University Press.
- Rebouças, M., Rodrigues, M., e Freitas, S. (2018). Utilization of mathematical models to evaluate the acceptance and physicochemical parameters for the development of a beverage made from cashew nut. *International Food Research Journal*, 25(2).
- Rebouças, M. C. (2016). Bebida prebiótica à base de amêndoa da castanha de caju: estudos com consumidores em diferentes abordagens para avaliação de fatores sensoriais e externos ao produto.
- Sharif, M. K., Masoos, S., Hafiz, R., e Muhammmad, N. (2017). Sensory evaluation and consumer acceptability. *Faisalabad: University of Agriculture*.
- Stiger, T. R., Barnhart, H. X., e Williamson, J. M. (1999). Testing proportionality in the proportional odds model fitted with gee. *Statistics in Medicine*, 18(11):1419–1433.
- Tutz, G. (2012). *Regression for categorical data*. Cambridge University Press.
- Verbeke, G., Fieuws, S., Molenberghs, G., e Davidian, M. (2014). The analysis of multivariate longitudinal data: A review. *Statistical methods in medical research*, 23(1):42–59.
- Williams, O. D. e Grizzle, J. E. (1972). Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association*, 67(337):55–63.

3 MODELO CATEGORIZADO ORDINAL MULTIVARIADO PARA ANÁLISE SIMULTÂNEA DE ATRIBUTOS SENSORIAIS

Resumo

Alimentos prebióticos costumam fazer parte da recomendação nutricional pelos reconhecidos benefícios à saúde e ao funcionamento gastrointestinal. Nesse trabalho, a motivação envolveu um estudo de 13 bebidas prebióticas à base de amêndoa da castanha de caju, adicionada ao suco de uva, segundo um delineamento em blocos incompletos, com 130 potenciais consumidores. Os atributos avaliados foram impressão global, aroma, corpo, doçura e sabor, por meio de uma escala hedônica de 9 pontos, caracterizando uma variável resposta categorizada ordinal. Estudos dessa natureza avaliam os atributos de forma separada, por meio de modelos individuais, para obter conclusões para cada atributo. A proposta neste estudo é apresentar um método de análise conjunta dos atributos, por meio de um modelo “multi” multivariado, visto que cada modelo em separado já corresponde a uma análise multivariada pela natureza da variável, com inclusão de efeitos aleatórios para considerar a estrutura do delineamento experimental. Os métodos multivariados para dados categorizados e a teoria da máxima verossimilhança formam a base dessa proposta. O modelo único para a análise conjunta dos atributos sensoriais selecionado foi o dos logitos cumulativos de chances não proporcionais de efeito misto. De acordo com esse modelo, as três formulações de bebidas prebióticas com maiores probabilidades de aceitação são composta por: 8% de açúcar e 40% de suco de uva (F_4), 6% de açúcar e 44% de suco de uva (F_6), e 9% de açúcar e 30% de suco de uva (F_{13}). Os resultados obtidos corroboram com as análises individuais em separado para cada atributo, porém, como vantagens do método proposto tem-se a unicidade de análise e economia de tempo computacional.

Palavras-chave: Bebida prebiótica; Escala hedônica; Blocos incompletos balanceados; Máxima verossimilhança.

3.1 Introdução

Atualmente, há um interesse crescente dos consumidores por alimentos que, além da função básica de nutrir, promovam efeitos benéficos à saúde (Prates e Mateus, 2002; Franco, 2006). Assim, consumidores esclarecidos procuram, no mercado alimentício, produtos que correspondam aos seus novos hábitos de vida e anseios de bem-estar e longevidade. Esse novo perfil de consumo estimulou as indústrias a realizarem pesquisas e desenvolverem produtos que atendam a esses critérios. Para avaliar a aceitação dos consumidores por determinados produtos, geralmente é feita uma pesquisa denominada análise sensorial, em que os atributos (ou características) são classificados segundo uma escala de resposta previamente estabelecida com base em categorias ordinais, conhecida como escala hedônica. No teste de escala hedônica, a aceitabilidade do produto é medida, por meio de uma hierarquia baseada em categorias ordinais que medem os diferentes graus de gostar (Dutcosky, 2019). Nguyen e Wismer (2019) utilizaram essa técnica para fazer uma comparação dos perfis dos atributos sensoriais e do gosto entre produtos alimentares regulares com redução de sódio.

Mediante a aplicação de testes sensoriais de aceitação, os julgadores podem dar sua opinião sobre diferentes características do produto relacionadas a: aparência, aroma, sabor e textura utilizando os órgãos dos sentidos (olfato, tato, paladar, visão e audição). Dessa forma, por se tratar dos órgãos do sentido humano, tem-se uma análise sensorial. A prática da análise sensorial é considerada milenar nas indústrias de cerveja, vinho e destilados da Europa. Segundo Chaves (1998), no Brasil, a prática se iniciou em 1954 com degustadores para a classificação do café brasileiro no laboratório de Tecnologia do Instituto Agrônomo de Campinas - SP. Seu uso estende-se desde o monitoramento, melhoramento

ou lançamento de novos produtos no mercado até a análise do efeito da embalagem (Rebouças, 2016; Borém et al., 2019). No contexto do desenvolvimento de novos produtos, é de fundamental importância avaliar a resposta dos consumidores ainda nos estágios iniciais, e não somente no produto final. Para tanto, faz-se necessário realizar uma otimização da formulação do produto, para que se avalie as diferentes concentração dos ingredientes utilizados que podem levar a aceitação do mesmo. Desta forma, é possível obter a formulação que permite alcançar a maior aceitabilidade possível (Rebouças et al., 2016, 2018).

De acordo com Dutcosky (2019), a avaliação sensorial fornece suporte técnico para a pesquisa, industrialização, *marketing* e controle de qualidade. Desse modo, percebe-se a importância da aplicação da análise sensorial e o quanto ela influencia no desenvolvimento de um determinado produto. A qualidade sensorial do alimento e a manutenção da mesma favorecem a fidelidade do consumidor a um produto específico em um mercado cada vez mais exigente (Teixeira, 2009). Assim, um produto que é desenvolvido baseado no conhecimento prévio dos atributos sensoriais, bem como, que tenha um controle preciso dessa qualidade sensorial, geralmente torna-se mais atraente, acarretando uma vantagem para empresa, ao produzir um produto que atenda aos novos anseios do consumidor.

Em algumas pesquisas sensoriais, a quantidade de amostras a serem avaliadas pode ser numerosa, inviabilizando a formação de blocos completos homogêneos. Como também, pode haver o interesse de utilizar um número pequeno de produtos por indivíduo por razões de ordem econômica, ou logística, ou da própria natureza do experimento. Ademais, essa diversidade do produto pode comprometer a indicação do melhor e há possibilidade de promover uma fadiga aos participantes. Nesse contexto, sugere-se um experimento em blocos incompletos, em que os provadores não degustam todas as opções do produto em teste. De acordo com Montgomery (2013), situações como essa, geralmente, ocorrem em razão das instalações experimentais ou ao tamanho físico do bloco.

Montgomery (2013) relatou que, quando todas as comparações de tratamento são igualmente importantes, as combinações de tratamento usadas em cada bloco devem ser selecionadas de maneira balanceada, de modo que qualquer par de tratamentos ocorra junto o mesmo número de vezes que qualquer outro par. As seguintes relações são condições necessárias para se ter um delineamento em blocos incompletos balanceados:

$$rt = hb, \quad (3.1)$$

$$\lambda(t - 1) = r(h - 1), \quad (3.2)$$

$$r > \lambda, \quad (3.3)$$

$$b \geq t, \quad (3.4)$$

Assumindo que existem t tratamentos e b blocos e cada bloco contém h tratamentos e que cada tratamento ocorre r vezes no experimento (ou é replicado r vezes). O delineamento em blocos incompletos balanceados foram usados por Idolo e Marshall (2019) para avaliar o efeito da temperatura de envelhecimento no vinho *hibiscus sabdariffa* (roselle).

Outro ponto que pode ser destacado são os estudos que utilizam variáveis categóricas (escalas hedônicas) para mensurar atributos sensoriais analisados de forma isolada por atributo. Como no estudo de Aaby et al. (2019) que determinaram as relações entre os atributos sensoriais de oito genótipos de frutos de framboesa vermelha (*Rubus idaeus* L.) e a qualidade por regressão e correlação parcial por mínimos quadrados. Fatoretto et al. (2018) propuseram o modelo de chances proporcionais de efeitos mistos para o atributo sensorial sabor na escolha entre dois tipos de tomates. Rebouças et al. (2018) aplicaram a metodologia de superfície de resposta para a seleção de formulação de bebidas prebióticas e Li et al. (2014) sugerem o uso do modelo de regressão linear múltipla para a tomada de decisão do melhor sabor de café.

A proposta de contribuição dessa pesquisa é realizar uma análise conjunta dos 5 atributos sensoriais envolvidos, por meio de modelo “multi” multivariado, visto que cada modelo em separado já corresponde a uma análise multivariada pela natureza da variável, com inclusão de efeitos aleatórios para considerar a estrutura do delineamento experimental. Este trabalho tem como motivação um estudo sensorial realizado pelo departamento de Tecnologia de Alimentos da Universidade Federal do Ceará (UFC) no ano de 2016 e, dessa forma, também apresentar contribuições para os pesquisadores potenciais da sensometria, por meio de um conjunto de técnicas que auxiliam na tomada de decisão (seleção do produto).

3.2 Modelo cumulativos

Um modelo usado para variáveis categorizadas ordinais são os modelos cumulativos baseados em categorias de limite (*category boundaries*), definido por Edwards e Thurstone (1952). Supõe-se que a variável observada $\mathbf{Y}$ é uma versão categorizada de uma variável latente $\mathbf{U}$. Por exemplo, quando a variável resposta é agrupada ou quando a variável avaliada é ordenada por uma escala. Em ambos os casos, a variável latente é usada como base para a construção deste tipo de modelo.

Para um determinado vetor $\mathbf{x}^\top = (x_1, x_2, \dots, x_p)$ de variáveis explicativas, a categoria de limite postula que a variável observável $Y \in \{1, \dots, c\}$ e a variável latente $\mathbf{U}$ estão conectadas por

$$Y = j \iff \tau_{j-1} < \mathbf{U} < \tau_j, \text{ com } j = 1, \dots, c \quad (3.5)$$

em que $-\infty \leq \tau_0 < \tau_1 < \dots < \tau_c \leq \infty$. Isso significa que $\mathbf{Y}$ é uma versão categorizada de $\mathbf{U}$ determinada pelos limites de $\tau_1, \dots, \tau_{c-1}$. Além disso, assume-se que a variável latente $\mathbf{U}$ é determinada pelas variáveis explicativas de forma linearizada por

$$\mathbf{U} = -\mathbf{x}^\top \boldsymbol{\beta} + \epsilon, \quad (3.6)$$

em que $\boldsymbol{\beta}^\top = (\beta_1, \dots, \beta_p)$ é o vetor de parâmetros associados e ϵ é uma variável aleatória com função de distribuição H . O sinal de “-” pode ser incorporado ao vetor de parâmetros, $\mathbf{U} = \mathbf{x}^\top (-\boldsymbol{\beta}) + \epsilon$, apenas para obter uma forma mais simples do modelo.

A partir dessas suposições, segue-se imediatamente que a variável observada $\mathbf{Y}$ é determinada pelo modelo

$$\mathbb{P}(\mathbf{Y} \leq j | \mathbf{x}) = \mathbb{P}(\mathbf{U} \leq \alpha_j) = F(\alpha_j + \mathbf{x}^\top \boldsymbol{\beta}). \quad (3.7)$$

O modelo (3.7) é chamado de modelo cumulativo com função de distribuição H e o lado esquerdo da equação é a soma das probabilidades $\mathbb{P}(\mathbf{Y} = 1 | \mathbf{x}) + \dots + \mathbb{P}(\mathbf{Y} = j | \mathbf{x})$.

3.3 Modelo dos logitos cumulativos

A escolha da função de distribuição caracteriza um modelo cumulativo específico. Uma escolha comum é a função de distribuição logística $H(x) = 1/(1 + \exp(-x))$. Consequentemente, o modelo passa a chamar modelo dos logitos cumulativos definido por:

$$\mathbb{P}(\mathbf{Y} \leq j|\mathbf{x}) = \frac{\exp(\alpha_j + \mathbf{x}^\top \boldsymbol{\beta})}{1 + \exp(\alpha_j + \mathbf{x}^\top \boldsymbol{\beta})}, \quad (3.8)$$

com $j = 1, \dots, q = c - 1$. Outras duas formas de reescrever a (3.8) são

$$\ln \left\{ \frac{\mathbb{P}(\mathbf{Y} \leq j|\mathbf{x})}{\mathbb{P}(\mathbf{Y} > j|\mathbf{x})} \right\} = \alpha_j + \mathbf{x}^\top \boldsymbol{\beta} \quad (3.9)$$

ou

$$\frac{\mathbb{P}(\mathbf{Y} \leq j|\mathbf{x})}{\mathbb{P}(\mathbf{Y} > j|\mathbf{x})} = \exp(\alpha_j + \mathbf{x}^\top \boldsymbol{\beta}). \quad (3.10)$$

Um caso particular do modelo dos logitos cumulativos foi chamado de modelo de chances proporcionais por McCullagh (1980). Esse nome se deve a uma propriedade de proporcionalidade do modelo (3.8), visto que

$$\frac{\mathbb{P}(\mathbf{Y} \leq j|\mathbf{x})}{\mathbb{P}(\mathbf{Y} > j|\mathbf{x})} = \frac{\mathbb{P}(\mathbf{Y} \leq j|\mathbf{x})}{1 - \mathbb{P}(\mathbf{Y} \leq j|\mathbf{x})}.$$

O lado esquerdo da equação (3.10) representa a probabilidade de $\mathbf{Y} \leq j$ ocorra em vez de $\mathbf{Y} > j$ e a utilização da função logística no modelo (3.9), lineariza nos parâmetros.

Várias funções de ligações foram propostas para o modelo com resposta binária como em Prentice (1976), Pregibon (1980), Aranda-Ordaz (1983), Morgan (1985), Stukel (1988), Friedl (1993), e Czado (1992). Em princípio, essas funções de ligações também podem ser usadas para variáveis de resposta categorizadas ordinais. Genter e Farewell (1985) sugerem uma classe de funções de ligações dentro da abordagem dos modelos cumulativos.

3.4 Versão ampliada do modelo dos logitos cumulativos

O modelo dos logitos cumulativos de chances proporcionais é baseado na suposição de que a variável explicativa causa uma mudança na escala latente, mas não alteram os limites de $\tau_1, \dots, \tau_q$. Em um modelo mais geral (de chances não proporcionais), pode haver também uma dependência das variáveis explicativas $\mathbf{w}^\top = (w_1, \dots, w_m)$ na forma linear dada por:

$$\tau_j = \alpha_j + \mathbf{w}^\top \boldsymbol{\delta}_j$$

em que $\boldsymbol{\delta}_j^\top = (\gamma_{j1}, \dots, \gamma_{jm})$ é um vetor de parâmetros específicos das categorias (Terza, 1985). Logo, o modelo ampliado é dado por:

$$\mathbb{P}(\mathbf{Y} \leq j | \mathbf{x}, \mathbf{w}) = H(\alpha_j + \mathbf{w}^\top \boldsymbol{\delta}_j + \mathbf{x}^\top \boldsymbol{\beta}), \quad (3.11)$$

em que $\mathbf{w}$ são definidas como variáveis de limite e $\mathbf{x}$ as variáveis de deslocamento. As variáveis de limite são sempre ponderadas por um vetor de parâmetros específicos por categoria $\boldsymbol{\delta}_j$ e as variáveis de deslocamento são sempre ponderadas por um vetor de parâmetro global $\boldsymbol{\beta}$. Naturalmente, as variáveis $\mathbf{x}$ podem ser ponderadas por parâmetros específicos da categoria, definido por

$$\mathbb{P}(\mathbf{Y} \leq j | \mathbf{x}) = H(\alpha_j + \mathbf{x}^\top \boldsymbol{\delta}_j), \quad (3.12)$$

que generaliza o modelo (3.7) assumindo parâmetros específicos da categoria em $\mathbf{x}$, sendo este um caso especial em que $\boldsymbol{\delta}_1 = \dots = \boldsymbol{\delta}_q = \boldsymbol{\beta}$.

3.5 Função de ligação e matrizes de delineamento para o modelo dos logitos cumulativos

A estrutura dos modelos cumulativos, torna-se especifica quando se define a função de ligação e o preditor linear, que depende da matriz do delineamento. A função de ligação $g = (g_1, \dots, g_q)$ para os modelos (3.7) ou (3.11) é dada por

$$g_j(\pi_1, \dots, \pi_q) = H^{-1}(\pi_1 + \dots + \pi_j),$$

com $j = 1, \dots, q$. Para a matriz de delineamento, deve-se distinguir entre o modelo (3.7) e o modelo mais geral (3.11). Para ambos os modelos, o preditor linear é $\boldsymbol{\eta} = \mathbf{X}\mathbf{B}$, sendo a matriz de delineamento do modelo (3.7):

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 & \mathbf{x}^\top \\ 0 & 1 & \dots & 0 & \mathbf{x}^\top \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \mathbf{x}^\top \end{bmatrix}$$

e o vetor de parâmetro $\mathbf{B} = (\alpha_1, \dots, \alpha_q, \boldsymbol{\beta})^\top$. Para o modelo ampliado (3.11) com a variável de limite $\mathbf{w}$ e variáveis de deslocamento $\mathbf{x}$, a matriz de delineamento é definida por

$$\mathbf{X} = \begin{bmatrix} 1 & \mathbf{w}^\top & 0 & 0 & \cdots & 0 & \mathbf{x}^\top \\ 0 & 0 & 1 & \mathbf{w}^\top & \cdots & 0 & \mathbf{x}^\top \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \mathbf{x}^\top \\ 0 & 0 & \cdots & 0 & 1 & \mathbf{w}^\top & \mathbf{x}^\top \end{bmatrix}$$

e o vetor de parâmetro $\mathbf{B} = (\alpha_1, \boldsymbol{\delta}_1, \alpha_2, \boldsymbol{\delta}_2, \dots, \alpha_q, \boldsymbol{\delta}_q, \boldsymbol{\beta})^\top$. Com isso, nem a variável de limite nem a variável de deslocamento contêm uma constante. A constante α_j é especificado para cada categoria da variável resposta.

3.6 Material

Os dados usados, neste estudo, provêm de uma pesquisa desenvolvida no departamento de Tecnologia de Alimentos da Universidade Federal do Ceará (UFC), no ano de 2016. Neste trabalho, realizou-se uma combinação das características nutricionais, funcionais e sensoriais das matérias-primas: castanha de caju, suco de uva, açúcar e substâncias prebióticas para o desenvolvimento de uma bebida. Para as formulações das bebidas utilizou-se o extrato hidrossolúvel de amêndoas da castanha de caju (ACC) beneficiadas e cruas, duas substâncias prebióticas, inulina (grau de polimerização ≥ 10 , Orafti GR) e oligofrutose (2 - 8 monômeros, Orafti P95) e açúcar cristal comercial. Para dar sabor, utilizou-se suco de uva concentrado (pH = 2,99; 15,2°Brix), o qual foi definido por meio de estudos preliminares. Mediante o delineamento composto central rotacional fatorial 2^2 , com cinco repetições no ponto central, utilizando combinações dos percentuais do suco (16%, 20%, 30%, 40% e 44%) e açúcar (3%, 4%, 6%, 8% e 9%), formularam-se 13 bebidas prebióticas (F_1 a F_{13}), sendo F_9 a F_{13} os do ponto central (Rebouças, 2016).

A aceitação das formulações desenvolvidas por Rebouças (2016) foram avaliadas por meio da escala hedônica de 9 pontos (“9 = gostei extremamente”, “8 = gostei muito”, “7 = gostei moderadamente”, “6 = gostei ligeiramente”, “5 = não gostei, nem desgostei”, “4 = desgostei ligeiramente”, “3 = desgostei moderadamente”, “2 = desgostei muito”, “1 = desgostei extremamente”) em relação a aceitação aos atributos sensoriais de impressão global, aroma, corpo, doçura e sabor. Entretanto, em razão da ocorrência de dados esparsos e para evitar a super parametrização, neste trabalho, a escala hedônica original de 9 pontos foi reduzida para 5 pontos, em que “5 = gostei extremamente ou muito”, “4 = gostei moderadamente ou ligeiramente”, “3 = não gostei, nem desgostei”, “2 = desgostei ligeiramente ou moderadamente”, “1 = desgostei muito ou extremamente”.

A avaliação da aceitação sensorial das formulações de bebidas prebióticas foi realizada em somente uma sessão, com 130 participantes não treinados. As amostras das bebidas foram servidas de forma monádica sequencial, seguindo um delineamento de blocos incompletos balanceados, em que cada provador avaliou 4 das 13 formulações de bebidas propostas, estabelecendo 10 grupos de repetições.

A avaliação da aceitação sensorial das formulações de bebidas prebióticas (tratamentos), realizada por Rebouças (2016), utilizou $b = 130$ participantes (blocos) não treinados, em que cada provador avaliou $l = 4$ das $t = 13$ formulações de bebidas propostas sendo cada um deles repetidas $r = 4 \times 10$ vezes. Pode-se verificar que as condições necessárias (equações 3.1, 3.2, 3.3 e 3.4) para que seja um delineamento em blocos incompletos balanceados são satisfeitas:

$$rt = hb \Rightarrow 40 \times 13 = 4 \times 130 \Rightarrow 520 = 520,$$

$$\lambda(t-1) = r(h-1) \Rightarrow 10 \times 12 = 40 \times 3 \Rightarrow 120 = 120,$$

$$r > \lambda \Rightarrow 40 > 10,$$

$$b \geq t \Rightarrow 130 \geq 13.$$

Portanto, o experimento é considerado em blocos incompletos balanceados (ou equilibrado).

3.7 Método

3.7.1 Análise exploratória por meio da análise de correspondência múltipla

A análise de correspondência múltipla, neste trabalho, foi usada como uma técnica exploratória de dados para evidenciar associações entre a categoria de resposta atribuída por cada provador (“1 = desgostei muito ou extremamente” a “5 = gostei extremamente ou muito”) e as formulações de bebidas prebióticas (F_1 a F_{13}), com referência aos atributos sensoriais conjuntamente (impressão global, aroma, corpo, doçura e sabor). Para a utilização desse método é necessário verificar se existe associação entre as formulações de bebidas prebióticas e as categorias das respostas dos participantes. Um teste que verifica essa associação é o teste χ^2 , no qual a hipótese nula é que as categorias de respostas atribuídas por cada provador e as formulações de bebidas prebióticas não estão associadas. Essa técnica permite a redução de um espaço multidimensional para um bidimensional, permitindo melhor visualização em um diagrama de dispersão. Nesse gráfico, os pontos de perfis semelhantes se localizam mais próximos no plano do que pontos de perfis com características diferentes. Tal resultado gráfico mostra a correspondência entre as categorias de variáveis (categorias de resposta atribuída por cada provador e formulações de bebidas prebióticas), analogamente a um gráfico *biplot* (Johnson et al., 2008).

3.7.2 Proposta do modelo dos logitos cumulativos “multi” multivariado

Nesta seção, apresenta-se a proposta de um modelo “multi” multivariado para a seleção de formulações de bebidas prebióticas à base de amêndoa da castanha de caju adicionada ao suco de uva (F_1 a F_{13}), baseado em um modelo linear generalizado para respostas categorizadas e ordinais, em que a variável resposta é uma escala hedônica de 5 pontos (“1 = desgostei muito ou desgostei extremamente” a “5 = gostei extremamente ou gostei muito”) composta por um conjunto de atributos sensoriais (impressão global, aroma, corpo, doçura e sabor). Por esse modelo ter variável resposta multicategórica é considerado multivariado, por ter mais de uma covariável (formulações de bebidas e atributos sensoriais) ele é considerado múltiplo e por ser um único modelo para todos os atributos sensoriais é definido como “multi”.

Seja $\mathbf{Y}$, o vetor que representa a variável resposta, na escala hedônica com 5 categorias, ao qual se associam às covariáveis de formulação de bebida ($\mathbf{F}$) e atributo sensorial ($\mathbf{A}$), ou seja, a resposta do i -ésimo provador é denotada pelo vetor $\mathbf{Y}_i = (Y_{i11}, Y_{i12}, \dots, Y_{ijl})^\top$, em que Y_{ijl} ’s representam as variáveis indicadoras para as categorias de resposta, isto é, $Y_{ijl} = 1$ se o i -ésimo provador optou pela j -ésima categoria de resposta e do l -ésimo atributo sensorial e $Y_{ijl} = 0$, caso contrário.

O modelo proposto é baseado nas probabilidades acumuladas das categorias de resposta, no qual se assume proporcionalidade ou não das razões de chances. Seja, $\theta_j = P(Y \leq j \mid \mathbf{F}, \mathbf{A})$, $j = 1, 2, \dots, 5$, a probabilidade acumulada de ocorrência até a j -ésima categoria de resposta para uma dada formulação de bebidas ($\mathbf{F}$) de um determinado atributo sensorial ($\mathbf{A}$). Diante disso, o modelo proposto dos logitos cumulativos “multi” multivariado é representado por:

$$\ln \left[\frac{\theta_j(\mathbf{F}, \mathbf{A})}{1 - \theta_j(\mathbf{F}, \mathbf{A})} \right] = \alpha_j + \beta_j^\top \mathbf{F} + \delta_j^\top \mathbf{A}, \quad (3.13)$$

sendo, α_j o intercepto da j -ésima categoria de resposta, $\beta_j^\top = (\beta_{j1}, \beta_{j2}, \dots, \beta_{j13})$ o vetor de parâmetros associado às formulações de bebidas, $\delta_j^\top = (\delta_{j1}, \delta_{j2}, \dots, \delta_{j5})$ o vetor de parâmetros associado ao atributo sensorial (em que, “1 = aroma”, “2 = corpo”, “3 = sabor”, “4 = impressão global” e “5 = doçura”), com $j = 1, 2, \dots, 5$. Nota-se que o modelo dos logitos cumulativos “multi” multivariado pode ser de chances não proporcionais, em que se assume que o efeito da covariável difere entre os q logitos cumulativos (propriedade de chances não proporcionais), sendo necessário q vetores de parâmetros de regressão $\beta^\top = (\beta_{11}, \beta_{12}, \dots, \beta_{113}, \dots, \beta_{41}, \beta_{42}, \dots, \beta_{413})$ e $\delta^\top = (\delta_{11}, \delta_{12}, \dots, \delta_{15}, \dots, \delta_{41}, \delta_{42}, \dots, \delta_{45})$, para descrever tais efeitos. Já para o caso de chances proporcionais assume-se que o vetor de parâmetro de regressão é comum para os q logitos, $\beta^\top = (\beta_1, \beta_2, \dots, \beta_{13})$ e $\delta^\top = (\delta_1, \delta_2, \dots, \delta_5)$. As estimativas desses parâmetros são obtidas pela teoria da máxima verossimilhança, com o uso do processo iterativo de Newton-Raphson. Em termos de probabilidades acumuladas, o modelo (3.13) é definido por:

$$\theta_j(\mathbf{F}, \mathbf{A}) = \frac{\exp(\alpha_j + \beta_j^\top \mathbf{F} + \delta_j^\top \mathbf{A})}{1 + \exp(\alpha_j + \beta_j^\top \mathbf{F} + \delta_j^\top \mathbf{A})}, \quad (3.14)$$

com, β_j e δ_j sendo os vetores de parâmetros que descrevem os efeitos das formulações de bebidas e dos atributos sensoriais, respectivamente.

Para a utilização do modelo dos logitos cumulativos “multi” multivariado, para análise conjunta dos atributos sensoriais, é necessário verificar a suposição de proporcionalidade das razões de chances. Para isso, usa-se o teste da razão de verossimilhanças (TRV) na qual a hipótese nula é que $\beta_j = \beta$ e $\delta_j = \delta$, ou seja, os dados podem ser ajustados pelo modelo dos logitos cumulativos “multi” multivariado de chances proporcionais e a hipótese alternativa é que $\beta_j \neq \beta$ e $\delta_j \neq \delta$ com $j = 1, \dots, 4$, isto é, as covariáveis de formulações de bebidas prebióticas (F_1 a F_{13}) e atributo sensorial (A_1 a A_5 , em que “1 = aroma”, “2 = corpo”, “3 = sabor”, “4 = impressão global” e “5 = doçura”) podem ser ajustadas utilizando-se modelo dos logitos cumulativos “multi” multivariado de chances não proporcionais. A estatística do teste é dada por:

$$\Lambda = -2 \log \left(\frac{L_{H_0}}{L_{H_A}} \right) \quad (3.15)$$

em que, L_{H_0} é a função de verossimilhança sob a hipótese nula, assumindo chances proporcionais, e L_{H_A} representa a função de verossimilhança sob a hipótese alternativa, isto é, assumindo não proporcionalidade. Para a tomada de decisão, se $\Lambda < \chi_{(d, 1-\alpha)}^2$, em que d é a diferença do número de parâmetros dos modelos sob hipótese nula e hipótese alternativa e $(1 - \alpha)$ é a confiabilidade do teste, então hipótese nula não é rejeitada. Caso contrário se o teste for significativo, deve-se optar por usar o modelo dos logitos cumulativos “multi” multivariado de chances não proporcionais, o que implica em um aumento no número de parâmetros no modelo.

Além do mais, se faz necessário verificar o efeito das covariáveis, formulação de bebida e atributo sensorial, para isso realiza-se TRV. A hipótese nula é que não há efeito de formulação de bebida prebiótica ($\beta_j = \mathbf{0}$) e/ou do atributo sensorial ($\delta_j = \mathbf{0}$), e a hipótese alternativa é que existe efeito das covariáveis, isto é, $\beta_j \neq \mathbf{0}$ e/ou $\delta_j \neq \mathbf{0}$. Se a hipótese nula for rejeitada, há indício de que existe pelo menos um β_j e/ou δ_j (com $j = 1, 2, \dots, 4$) estatisticamente diferente de zero.

Como os participantes não são treinados e o experimento é em blocos incompletos, é fundamental considerar a inclusão de efeito aleatório para provador. Assim, o modelo dos logitos cumulativos para análise unificada dos atributos sensoriais seria denominado como de efeito misto descrito por:

$$\ln \left[\frac{\theta_j(\mathbf{F}, \mathbf{A}; \mathbf{Z})}{1 - \theta_j(\mathbf{F}, \mathbf{A}; \mathbf{Z})} \right] = \alpha_j + \boldsymbol{\beta}_j^\top \mathbf{F} + \boldsymbol{\delta}_j^\top \mathbf{A} + \mathbf{u}^\top \mathbf{Z}, \quad (3.16)$$

sendo, α_j o intercepto da j -ésima categoria de resposta do provador, $\boldsymbol{\beta}_j^\top = (\beta_{j1}, \beta_{j2}, \dots, \beta_{j13})$ o vetor de parâmetros associado a formulação de bebidas, com $j = 1, 2, \dots, 5$, $\boldsymbol{\delta}_j^\top = (\delta_{j1}, \delta_{j2}, \dots, \delta_{j5})$ o vetor de parâmetros associado ao atributo sensorial (em que, “1 = aroma”, “2 = corpo”, “3 = sabor”, “4 = impressão global” e “5 = doce”) e $\mathbf{u}$ é o efeito aleatório associado ao provador, em que $\mathbf{u} \sim N(0, \sigma_u^2)$. Para certificar que provador tem efeito aleatório, constrói-se a partir da verossimilhança perfilada o intervalo de confiança para o parâmetro σ_u . Se o intervalo de 95% de confiança para a estimativa do desvio padrão do coeficiente aleatório, $\hat{\sigma}_u$, não contém o valor zero, confirma-se a existência de efeito aleatório para provador.

Em termos de probabilidades acumuladas, o modelo (3.16) é definido por:

$$\theta_{jl}(\mathbf{F}, \mathbf{A}; \mathbf{Z}) = \frac{\exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{F} + \boldsymbol{\delta}_j^\top \mathbf{A} + \mathbf{u}^\top \mathbf{Z})}{1 + \exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{F} + \boldsymbol{\delta}_j^\top \mathbf{A} + \mathbf{u}^\top \mathbf{Z})}, \text{ com } j = 1, 2, \dots, 4 \text{ e } l = 1, 2, \dots, 4. \quad (3.17)$$

O procedimento computacional foi desenvolvido usando os pacotes `ca` (Nenadic e Greenacre, 2007) para a análise de correspondência múltipla e o pacote `ordinal` (Christensen e Christensen, 2015) para o modelo dos logitos cumulativos (“multi” multivariado de efeito misto para análise conjunta dos atributos sensoriais, em que a referência da categoria de resposta foi “1 = desgostei muito ou extremamente”, para formulação de bebida foi F_1 e para atributo sensorial foi “1 = aroma”), disponível no *software* R (R Core Team, 2020). O roteiro dos comandos utilizados para fazer as análises encontra-se no Apêndice II.

3.8 Resultados e discussões

Inicialmente, fez-se uma análise exploratória, utilizando uma análise de correspondência múltipla, nos dados com o objetivo de ter um resultado preliminar em relação as formulações de bebidas prebióticas (F_1 a F_{13}), as categorias de resposta dos provadores (“1 = desgostei muito ou desgostei extremamente” a “5 = gostei extremamente ou gostei muito”) e os atributos sensoriais (impressão global, aroma, corpo, doçura e sabor). Verificou-se associação entre as categorias atribuídas por cada provador e as formulações de bebidas pelo teste χ^2 . O resultado do teste foi significativo, ao nível de 5% de significância, (p -valor $< 0,001$), ou seja, existe uma associação entre o parecer dos provadores e as formulações de bebidas com relação os atributos sensoriais conjuntamente.

Posteriormente, por meio da análise de correspondência múltipla representou-se graficamente a distribuição das formulações de bebidas (F_1 a F_{13}), dos atributos sensoriais (IG = impressão global, A = aroma, C = corpo, D = doçura e S = sabor) e as notas associadas (1 a 5), conforme Figura 3.1.

Figura 3.1: Representação gráfica bidimensional por meio da análise de correspondência múltipla entre formulações de bebidas prebióticas, atributos sensoriais e as categorias de respostas dos provadores, do estudo desenvolvido na Universidade Federal do Ceará, no ano de 2016

Observa-se, pela Figura 3.1 que as formulações F_3 (4% de açúcar e 40% de suco de uva), F_6 (6% de açúcar e 44% de suco de uva) e F_{13} (6% de açúcar e 30% de suco de uva) de bebidas prebióticas foram as que mais se aproximaram da categoria de resposta 5 (gostei extremamente ou muito) com relação a todos os atributos sensoriais. As formulações F_3 e F_6 são as que tem uma das maiores concentrações de suco de uva (com 40% e 44%, respectivamente), apontando assim, de forma prévia, que os degustadores simpatizaram por formulações com maiores concentrações de suco de uva. Agora, a concentração da bebida prebiótica F_{13} é de ponto central, ou seja, a média das porcentagem dos níveis de suco uva (30%) e açúcar (6%).

Ajustando-se o modelo dos logitos cumulativos “multi” multivariado, conforme descrito na seção (3.7.2), constatou-se a não proporcionalidade entre as razões de chances. Conforme o teste da razão de verossimilhanças, ao nível de 5% de significância, rejeita-se a hipótese nula (p -valor $< 0,001$), ou seja, há indícios de que não exista proporcionalidade entre as chances das categorias de respostas dos julgadores. Assim, recomenda-se o uso do modelo dos logitos cumulativos “multi” multivariado de chances não proporcionais, o que implica em um aumento no número de parâmetros no modelo (3.13).

Em seguida, constatou-se, ao nível de 5% de significância, que existe efeito de formulação de bebida (p -valor $< 0,01$) e atributo sensorial (p -valor $< 0,01$). Assim, esses dois fatores foram considerados

no modelo.

Para comprovar que bloco (provador) apresenta um componente, utilizou-se o teste da verossimilhança perfilada. Obteve-se a estimativa para o desvio padrão: $\hat{\sigma}_u = 1,89$ e o intervalo de confiança (com 95% de confiabilidade) foi $IC_{95\%} = [1,70; 2,24]$, que não contém o valor zero, confirmando a existência de efeito aleatório para provador. Consequentemente, adota-se o modelo dos logitos cumulativos “multi” multivariado de chances não proporcionais com efeito aleatório em provador (bloco), cujas estimativas dos parâmetros, os erros padrões, os p-valores e as razões de chances (R.C) são apresentados na Tabela 3.1.

Tabela 3.1: Estimativas dos parâmetros, erros padrões, p-valores e razões de chances do modelo dos logitos cumulativos “multi” multivariado de chances não proporcionais mistos para análise conjunta dos atributos sensoriais referente ao estudo desenvolvido na Universidade Federal do Ceará, no ano 2016

Parâmetro	Estimat.	E.P.	p-valor	R.C.	Parâmetro	Estimat.	E.P.	p-valor	R.C.
α_2	-1,92	0,31	< 0,01	0,15	α_4	2,02	0,27	< 0,01	7,51
α_3	0,04	0,27	0,87	1,04	α_5	4,29	0,38	< 0,01	72,73
β_{22}	-1,36	0,34	< 0,01	0,26	β_{210}	-2,28	0,45	< 0,01	0,10
β_{32}	-0,84	0,25	< 0,01	0,43	β_{310}	-1,16	0,27	< 0,01	0,31
β_{42}	-1,08	0,26	< 0,01	0,34	β_{410}	-1,26	0,26	< 0,01	0,28
β_{52}	-0,97	0,37	0,01	0,38	β_{510}	-0,87	0,39	0,03	0,42
β_{23}	-0,61	0,35	0,08	0,54	β_{211}	-1,93	0,42	< 0,01	0,14
β_{33}	-0,75	0,27	0,01	0,47	β_{311}	-1,65	0,29	< 0,01	0,19
β_{43}	-1,05	0,27	< 0,01	0,35	β_{411}	-1,75	0,27	< 0,01	0,17
β_{53}	-0,94	0,37	0,01	0,39	β_{511}	-1,17	0,38	< 0,01	0,31
β_{24}	-1,22	0,32	< 0,01	0,30	β_{212}	-3,28	0,71	< 0,01	0,04
β_{34}	-1,84	0,27	< 0,01	0,16	β_{312}	-1,80	0,31	< 0,01	0,17
β_{44}	-1,94	0,27	< 0,01	0,14	β_{412}	-1,88	0,29	< 0,01	0,15
β_{54}	-1,66	0,39	< 0,01	0,19	β_{512}	-0,91	0,42	0,03	0,40
β_{25}	-0,50	0,31	0,11	0,61	β_{213}	-2,31	0,42	< 0,01	0,10
β_{35}	-0,92	0,27	< 0,01	0,40	β_{313}	-2,00	0,29	< 0,01	0,14
β_{45}	-0,99	0,27	< 0,01	0,37	β_{413}	-2,09	0,27	< 0,01	0,12
β_{55}	-0,87	0,40	0,03	0,42	β_{513}	-1,91	0,37	< 0,01	0,15
β_{26}	-1,78	0,34	< 0,01	0,17	δ_{21}	-0,05	0,23	0,84	0,93
β_{36}	-1,96	0,27	< 0,01	0,14	δ_{31}	-0,26	0,16	0,11	0,86
β_{46}	-2,25	0,27	< 0,01	0,11	δ_{41}	0,81	0,16	< 0,01	0,30
β_{56}	-2,44	0,38	< 0,01	0,09	δ_{51}	0,35	0,22	0,12	0,34
β_{27}	-0,21	0,30	0,47	0,81	δ_{22}	-0,12	0,23	0,60	2,37
β_{37}	0,03	0,26	0,92	1,03	δ_{32}	-0,41	0,16	0,01	1,63
β_{47}	0,24	0,28	0,38	1,28	δ_{42}	-0,38	0,15	0,01	0,43
β_{57}	-0,61	0,44	0,17	0,54	δ_{52}	-0,72	0,20	< 0,01	0,58
β_{28}	-1,75	0,35	< 0,01	0,17	δ_{23}	0,15	0,23	0,51	1,05
β_{38}	-1,58	0,27	< 0,01	0,21	δ_{33}	-0,42	0,16	0,01	1,29
β_{48}	-1,84	0,27	< 0,01	0,16	δ_{43}	-0,52	0,16	< 0,01	0,45
β_{58}	-1,23	0,40	< 0,01	0,29	δ_{53}	-0,84	0,20	< 0,01	0,71
β_{29}	-1,60	0,36	< 0,01	0,20	δ_{24}	0,81	0,21	< 0,01	1,22
β_{39}	-1,45	0,27	< 0,01	0,23	δ_{34}	0,23	0,16	0,14	0,85
β_{49}	-1,65	0,27	< 0,01	0,19	δ_{44}	-0,03	0,15	0,85	0,26
β_{59}	-1,27	0,38	< 0,01	0,28	δ_{54}	-0,19	0,21	0,35	0,31

O critério utilizado para inferir quais as formulações de bebidas prebióticas foram as mais aceitas pelos 130 julgadores em relação aos atributos sensoriais conjuntamente foi a probabilidade acumulada entre as categorias 4 (gostei moderadamente ou ligeiramente) e 5 (gostei extremamente ou muito) preditas de acordo com o modelo ajustado, que são apresentadas na Tabela 3.2.

Tabela 3.2: Probabilidades das categorias de respostas com relação aos atributos sensoriais e probabilidades acumuladas previstas das categorias 4 e 5, do estudo desenvolvido na Universidade Federal do Ceará, no ano 2016

Atributo Sensorial	Categoria de Resposta	Formulação de Bebida												
		F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}
Aroma	1	0,13	0,04	0,08	0,04	0,09	0,03	0,11	0,03	0,03	0,02	0,02	0,01	0,02
	2	0,44	0,33	0,31	0,13	0,26	0,13	0,47	0,19	0,21	0,28	0,18	0,18	0,14
	3	0,20	0,16	0,15	0,16	0,21	0,10	0,23	0,13	0,15	0,19	0,17	0,14	0,13
	4	0,21	0,42	0,41	0,58	0,40	0,56	0,16	0,59	0,54	0,47	0,57	0,62	0,59
	5	0,02	0,05	0,05	0,09	0,04	0,18	0,03	0,06	0,07	0,04	0,06	0,05	0,12
	$P(Y \geq 4)$	0,23	0,47	0,46	0,67	0,44	0,74	0,19	0,65	0,61	0,51	0,63	0,67	0,71
Corpo	1	0,13	0,04	0,07	0,04	0,08	0,02	0,11	0,02	0,03	0,01	0,02	0,01	0,01
	2	0,38	0,27	0,26	0,10	0,21	0,10	0,41	0,15	0,17	0,23	0,15	0,14	0,11
	3	0,38	0,41	0,40	0,38	0,45	0,32	0,39	0,37	0,39	0,44	0,40	0,39	0,36
	4	0,10	0,25	0,24	0,41	0,23	0,42	0,07	0,41	0,36	0,29	0,39	0,43	0,43
	5	0,01	0,03	0,03	0,07	0,03	0,14	0,02	0,05	0,05	0,03	0,04	0,03	0,09
	$P(Y \geq 4)$	0,11	0,28	0,27	0,48	0,26	0,56	0,09	0,46	0,41	0,32	0,43	0,46	0,52
Doçura	1	0,15	0,04	0,09	0,05	0,10	0,03	0,13	0,03	0,03	0,02	0,03	0,01	0,02
	2	0,32	0,23	0,21	0,07	0,16	0,08	0,35	0,12	0,14	0,20	0,12	0,12	0,09
	3	0,20	0,13	0,11	0,10	0,16	0,06	0,23	0,09	0,10	0,14	0,10	0,10	0,09
	4	0,29	0,49	0,49	0,59	0,48	0,49	0,21	0,63	0,59	0,54	0,62	0,67	0,57
	5	0,04	0,11	0,10	0,19	0,10	0,34	0,08	0,13	0,14	0,10	0,13	0,10	0,23
	$P(Y \geq 4)$	0,33	0,60	0,59	0,78	0,58	0,83	0,29	0,76	0,73	0,64	0,75	0,77	0,80
Sabor	1	0,12	0,03	0,07	0,04	0,08	0,02	0,10	0,02	0,03	0,01	0,02	0,01	0,01
	2	0,35	0,25	0,23	0,09	0,19	0,09	0,38	0,13	0,15	0,21	0,13	0,12	0,10
	3	0,22	0,15	0,15	0,12	0,19	0,08	0,28	0,12	0,13	0,17	0,14	0,13	0,11
	4	0,27	0,47	0,46	0,58	0,45	0,49	0,19	0,61	0,57	0,52	0,60	0,65	0,57
	5	0,04	0,10	0,09	0,17	0,09	0,31	0,07	0,12	0,12	0,09	0,11	0,09	0,21
	$P(Y \geq 4)$	0,31	0,57	0,55	0,75	0,54	0,80	0,26	0,73	0,69	0,61	0,71	0,74	0,78
IG	1	0,26	0,08	0,16	0,09	0,17	0,06	0,22	0,06	0,07	0,03	0,05	0,01	0,03
	2	0,37	0,34	0,29	0,12	0,23	0,14	0,42	0,20	0,22	0,31	0,20	0,21	0,15
	3	0,14	0,10	0,08	0,11	0,15	0,06	0,17	0,08	0,09	0,14	0,11	0,11	0,10
	4	0,21	0,42	0,41	0,57	0,40	0,53	0,15	0,59	0,54	0,47	0,57	0,61	0,58
	5	0,02	0,06	0,06	0,11	0,05	0,21	0,04	0,07	0,08	0,05	0,07	0,06	0,14
	$P(Y \geq 4)$	0,23	0,48	0,47	0,68	0,45	0,74	0,19	0,66	0,62	0,52	0,64	0,67	0,72

Na Tabela 3.2, observa-se que as formulações de bebidas prebióticas de ponto central (F_9 a F_{13} , com 6% de açúcar e 30% de suco de uva) obtiveram os valores de probabilidades acumuladas nas categorias 4 (gostei moderadamente ou gostei ligeiramente) e 5 (gostei extremamente ou gostei muito) entre 0,32 e 0,80 em relação a todos os atributos sensoriais. Vale ressaltar que a formulação F_{13} se destaca mais entre as outras, de ponto central, em todos os atributos com os valores de probabilidade acumuladas: 0,71 para aroma, 0,52 para corpo, 0,80 para doçura, 0,78 para sabor e 0,72 para impressão global, ou seja, os participantes julgaram a formulação como sendo de “4 - gosto moderado ou ligeiramente” ou “5 - extremamente ou muito”. Usando esse mesmo critério, para as demais formulações que não são de ponto central, as bebidas F_4 (8% de açúcar e 40% de suco de uva) e F_6 (6% de açúcar e 44% de suco de uva) obtiveram as mais altas probabilidades acumuladas com 0,67 e 0,74, respectivamente, para aroma, com 0,48 e 0,56, respectivamente, para corpo, com 0,78 e 0,83, respectivamente, para doçura, com 0,75 e 0,80, respectivamente, para sabor, com 0,68 e 0,74, respectivamente, para impressão global. Observa-se que a formulação mais bem aceita com relação a todos os atributos sensoriais é composta por 6% de açúcar e 44% de suco de uva (F_6) e a menos aceita foi F_7 com 3% de açúcar e 30% de suco de uva.

As probabilidades observadas e as previstas pelo modelo estão próximas, com raras exceções,

mostrando que modelo dos logitos cumulativos “multi” multivariado de chances não proporcionais de efeito misto tem uma boa capacidade preditiva (Figura 3.2).

Figura 3.2: Probabilidades observadas e previstas para as formulações de bebidas prebióticas com relação aos atributos sensoriais conjuntamente do estudo desenvolvido pela Universidade Federal do Ceará, no ano 2016

Como resultado, as formulações com maior aceitabilidade por meio do modelo proposto dos logitos cumulativos “multi” multivariado de chances não proporcionais de efeito misto, foram F_4 (8% de açúcar e 40% de suco de uva), F_6 (6% de açúcar e 44% de suco de uva) e F_{13} (6% de açúcar e 30% de suco de uva), podendo dar um destaque maior para a formulação F_6 , a bebida com maior concentração de suco de uva (44%) e um nível intermediário de açúcar (6%).

Além disso, identificou-se que as três formulações de bebidas prebióticas mais aceitas encontram-se no primeiro quadrante do delineamento composto central rotacional fatorial 2^2 , ou seja, as formulações selecionadas são a que tiveram as maiores concentrações de suco de uva (com 30%, 40% e 44%) e de açúcar (com 6% e 8%), indicando assim um favoritismo por sucos mais doces e com maior concentração no suco da fruta.

3.9 Conclusão

Em alguns estudos sensoriais, as variáveis respostas são categorizadas, e normalmente, são analisadas como contínuas. Além disso, os atributos sensoriais são avaliados estatisticamente de forma separada e concluem o produto com maior aceitabilidade por atributo.

No presente trabalho, apresenta-se uma análise simultânea dos atributos sensoriais, levando em consideração a multiplicidade e a natureza da variável resposta além de uma possível dependência entre os indivíduos (por conta da estrutura do delineamento), por meio do modelo proposto dos logitos cumulativos “multi” multivariado de chances não proporcionais de efeito misto. Dessa forma, a avaliação

do produto ocorre de maneira unificada, facilitando um parecer sensorial. Este método apresenta técnica multivariada compatível aos métodos utilizados na área de alimentos.

Referências

- Aaby, K., Skaret, J., Røen, D., e Sønsteby, A. (2019). Sensory and instrumental analysis of eight genotypes of red raspberry (*rubus idaeus* l.) fruits. *Journal of Berry Research*, (Preprint):1–16.
- Aranda-Ordaz, F. J. (1983). An extension of the proportional-hazards model for grouped data. *Biometrics*, pages 109–117.
- Borém, F. M., Ribeiro, F. C., Figueiredo, L. P., Giomo, G. S., Siqueira, V. C., e Dias, C. A. (2019). Sensory analysis and fatty acid profile of specialty coffees stored in different packages. *Journal of food science and technology*, 56(9):4101–4109.
- Chaves, J. (1998). Análise sensorial: glossário. *UFV, Viçosa-MG*, 31.
- Christensen, R. H. B. e Christensen, M. R. H. B. (2015). Package “ordinal”. *Stand*, 19:2016.
- Czado, C. (1992). On link selection in generalized linear models. In *Advances in GLIM and Statistical Modelling*, pages 60–65. Springer.
- Dutcosky, S. D. (2019). *Análise sensorial de alimentos*. Champagnat - Pucpress, 5^a edition.
- Edwards, A. L. e Thurstone, L. L. (1952). An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 17(2):169–180.
- Fatoretto, M. B., de Lara, I. A. R., Loro, A. C., e Spoto, M. H. F. (2018). Sensory evaluation of dehydrated tomatoes using the proportional odds mixed model. *Journal of food processing and preservation*, 42(11):e13822.
- Franco, R. (2006). *Análise comparativa de legislações referentes aos alimentos funcionais*. PhD thesis, Dissertação (Mestrado em Nutrição Humana Aplicada)–Faculdade de Ciências.
- Fiedl, H. (1993). Verallgemeinerte logistische modelle in der analyse von zervix-karzinomen.
- Genter, F. C. e Farewell, V. T. (1985). Goodness-of-link testing in ordinal regression models. *Canadian Journal of Statistics*, 13(1):37–44.
- Idolo, I. e Marshall, L. (2019). The effect of ageing temperature on the sensory qualities of hibiscus sabdariffa (roselle) wine. *African Journal of Food, Agriculture, Nutrition and Development*, 19(3):14726–14738.
- Johnson, R. A., Wichern, D. W., et al. (2008). *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ.
- Li, B., Hayes, J. E., e Ziegler, G. R. (2014). Interpreting consumer preferences: Physicohedonic and psychohedonic models yield different information in a coffee-flavored dairy beverage. *Food quality and preference*, 36:27–32.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127.
- Montgomery, D. C. (2013). *Design and analysis of experiments*. John wiley & sons.

- Morgan, B. J. (1985). The cubic logistic model for quantal assay data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34(2):105–113.
- Nenadic, O. e Greenacre, M. (2007). Correspondence analysis in r, with two-and three-dimensional graphics: the ca package. *Journal of statistical software*, 20(3).
- Nguyen, H. e Wismer, W. V. (2019). A comparison of sensory attribute profiles and liking between regular and sodium-reduced food products. *Food Research International*, 123:631–641.
- Prates, J. A. e Mateus, C. (2002). Componentes com atividade fisiológica dos alimentos de origem animal. *Revista portuguesa de ciências veterinárias*, 97(541):3–12.
- Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(1):15–24.
- Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics*, pages 761–768.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rebouças, M., Rodrigues, M., e Freitas, S. (2018). Utilization of mathematical models to evaluate the acceptance and physicochemical parameters for the development of a beverage made from cashew nut. *International Food Research Journal*, 25(2).
- Rebouças, M. C. (2016). Bebida prebiótica à base de amêndoa da castanha de caju: estudos com consumidores em diferentes abordagens para avaliação de fatores sensoriais e externos ao produto.
- Rebouças, M. C., Rodrigues, M. D. C., De Freitas, S. M., e Ferreira, B. (2016). The physicochemical optimization and acceptability of a cashew nut-based beverage varying in mango juice and sugar: A pilot study. *Beverages*, 2(3):23.
- Stukel, T. A. (1988). Generalized logistic models. *Journal of the American Statistical Association*, 83(402):426–431.
- Teixeira, L. V. (2009). Análise sensorial na indústria de alimentos. *Revista do Instituto de Laticínios Cândido Tostes*, 64(366):12–21.
- Terza, J. V. (1985). Ordinal probit: a generalization. *Communications in Statistics-Theory and Methods*, 14(1):1–11.

4 CONSIDERAÇÕES FINAIS

Muitas vezes, quando as variáveis de interesse de um estudo são categorias ordinais, como no caso da classificação dada a um atributo sensorial por provadores, na análise estatística é relevante respeitar a verdadeira natureza dessa variável. Dada essa importância, uma possível opção para variáveis respostas politômicas ordinais são os modelos lineares generalizados em que os modelos dos logitos cumulativos de efeito misto são uma alternativa. A partir dessa classe de modelos é possível verificar as relações entre covariáveis ou fatores que influenciam essa variável resposta de interesse, bem como fazer previsões das probabilidades para cada uma das categorias de resposta dos indivíduos. Além disso, utilizando o modelo dos logitos cumulativos de efeito misto na análise sensorial é possível inferir qual produto é mais apreciado e quais covariáveis podem estar influenciando no gosto de diferentes tipos de consumidores. Este estudo é feito, na maioria das vezes, antes desse produto começar a ser comercializado, evitando, assim, prejuízos.

No presente trabalho, contempla-se uma análise individual para cada atributo sensorial por meio do modelo dos logitos de chances proporcional de efeito misto e uma análise simultânea dos atributos sensoriais que permite extrair uma conclusão singular da aceitabilidade dos participantes com relação à composição de bebida integrado por extrato de amêndoa da castanha de caju, substâncias prebióticas e suco de uva, além de reduzir o tempo computacional, por meio do modelo proposto “multi” multivariado dos logitos de chances não proporcional de efeito misto. Vale destacar que o resultado do modelo proposto para a análise conjunta dos atributos foi em consonância com os modelos individuais (separado para cada atributo), além da exploratória obtida pela análise de correspondência (simples e múltipla), as formulações indicadas foram as mesmas: F_4 (com 8% de açúcar e 40% de suco de uva), F_6 (6% de açúcar e 44% de suco de uva) e F_{13} (6% de açúcar e 30% de suco de uva). Uma possível limitação para essa classe de modelos é quando se aumenta a escala da variável resposta, pois o modelo se torna mais complexo, por apresentar mais parâmetros, dificultando a interpretação do mesmo.

Outras áreas das Ciências também realizam estudos em que a variável resposta é categorizada e ordinal, e os modelos descritos, neste trabalho, também podem ser utilizados. Contudo, este campo de pesquisa (dados multicategóricos) e em particular para variável ordinal ainda é emergente na literatura, como, por exemplo, a análise de resíduos.

Como previsão para trabalhos futuros, pode-se estudar e buscar a construção de resíduos mais apropriados para a análise de dados politômicos, ampliar a verificação da qualidade do ajuste do modelo, explorar outras funções de ligações e incluir dados esparsos na análise.

APÊNDICES

Apêndice I

```
#####
##                                R SCRIPT                                ##
#####

## Packages
library("ordinal", "ca", "factoextra")

## Reading the database
data.Grape <- choose.files()

## Attaching the database
attach(data.Grape)

## Putting variables as factors
# a) Independent variables
data.Grape$Provador      <- as.factor(data.Grape$Provador)
data.Grape$Formulação    <- as.factor(data.Grape$Formulação)

# b) Dependent variables
data.Grape$ImpressaoGlobal <- as.factor(data.Grape$ImpressaoGlobal)
data.Grape$Aroma          <- as.factor(data.Grape$Aroma)
data.Grape$Cor            <- as.factor(data.Grape$Cor)
data.Grape$Corpo          <- as.factor(data.Grape$Corpo)
data.Grape$Doçura         <- as.factor(data.Grape$Doçura)
data.Grape$Sabor          <- as.factor(data.Grape$Sabor)

#####
##                                CORRESPONDENCE ANALYSIS                                ##
#####

# a) Contingency table (beverage formulation x sensory attribute)
tab.cont.OI <- table(data.Grape$Formulação, data.Grape$ImpressaoGlobal)
tab.cont.CO <- table(data.Grape$Formulação, data.Grape$Cor)
tab.cont.AR <- table(data.Grape$Formulação, data.Grape$Aroma)
tab.cont.BO <- table(data.Grape$Formulação, data.Grape$Corpo)
tab.cont.SW <- table(data.Grape$Formulação, data.Grape$Doçura)
tab.cont.FL <- table(data.Grape$Formulação, data.Grape$Sabor)

# b) Chi-square association test
# H0: the variables are not associated (independent)
# HA: the variables are associated (dependent)

chisq.test(tab.cont.OI)$p.value > 0.05 # (FALSE: Rejeita H_0)
```

```

chisq.test(tab.cont.AR)$p.value > 0.05 # (FALSE: Rejeita H_0)
chisq.test(tab.cont.CO)$p.value > 0.05 # (FALSE: Rejeita H_0)
chisq.test(tab.cont.BO)$p.value > 0.05 # (FALSE: Rejeita H_0)
chisq.test(tab.cont.SW)$p.value > 0.05 # (FALSE: Rejeita H_0)
chisq.test(tab.cont.FL)$p.value > 0.05 # (FALSE: Rejeita H_0)

# c) Simple Correspondence Analysis
ca.OI <- ca(tab.cont.OI)
ca.AR <- ca(tab.cont.AR)
ca.CO <- ca(tab.cont.CO)
ca.BO <- ca(tab.cont.BO)
ca.SW <- ca(tab.cont.SW)
ca.FL <- ca(tab.cont.FL)

# d) Simple Correspondence Analysis plot
fviz_ca_biplot(ca.OI, title = " ", ylim = c(-.5,.5), xlim = c(-.75,.75),
               arrow = c(FALSE, TRUE))
fviz_ca_biplot(ca.AR, title = " ", ylim = c(-.5,.5), xlim = c(-.75,.75),
               arrow = c(FALSE, TRUE))
fviz_ca_biplot(ca.CO, title = " ", ylim = c(-.5,.5), xlim = c(-.75,.75),
               arrow = c(FALSE, TRUE))
fviz_ca_biplot(ca.BO, title = " ", ylim = c(-.5,.5), xlim = c(-.75,.75),
               arrow = c(FALSE, TRUE))
fviz_ca_biplot(ca.SW, title = " ", ylim = c(-.5,.5), xlim = c(-.85,.85),
               arrow = c(FALSE, TRUE))
fviz_ca_biplot(ca.FL, title = " ", ylim = c(-.5,.5), xlim = c(-.75,.75),
               arrow = c(FALSE, TRUE))

#####
##                CUMULATIVE LOGIT MIXED MODELS                ##
#####

# Assigning variables
Block      <- data.Grape$Provador
Treatment  <- data.Grape$Formulação
Response.Overall.Impression <- data.Grape$ImpressaoGlobal

#####
## A) Overall Impression ##
#####

# A.1) Testing if grape juice formulation is significant for the model
fit.clmm.po.OI.with <- clmm2(ordered(Response.Overall.Impression) ~ 1,
                             data = data.Grape,
                             random = Block,
                             Hess = TRUE) # CLMMPO
fit.clmm.po.OI.without <- clmm2(ordered(Response.Overall.Impression) ~ Treatment,

```

```

data = data.Grape,
random = Block,
Hess = TRUE) # CLMMPO

anova(fit.clmm.po.OI.without,
      fit.clmm.po.OI.with)[2,7] > 0.05 #(TRUE: not reject H_0)

# A.2) Testing the assumption of proportional odds model
fit.clmm.po.OI <- clmm2(ordered(Response.Overall.Impression) ~ Treatment,
                        data = data.Grape,
                        random = Block,
                        Hess = TRUE) # CLMMPO
fit.clmm.npo.OI <- clmm2(ordered(Response.Overall.Impression) ~ 1,
                        nominal = ~ Treatment,
                        data = data.Grape,
                        random = Block,
                        Hess = TRUE) # CLMMNPO

anova (fit.clmm.po.OI, fit.clmm.npo.OI)[2,7] > 0.05 # (TRUE: not reject H_0)

# A.3) Selected model
fit.clmm.po.OI <- clmm2(ordered(Response.Overall.Impression) ~ Treatment,
                        data = data.Grape,
                        random = Block,
                        Hess = TRUE, nAGQ = 5)

summary(fit.clmm.po.OI)

# A.4) Standard deviation of the random coefficient
confint(profile(fit.clmm.po.OI, range = c(0.12, 4), nSteps = 20, trace = 0))

# A.5) Predicted odds and Observed odds
pred.OI <- round(predict(fit.clmm.po.OI, newdata = data.Grape),3)
table.pred.OI <- data.frame(cbind(data.Grape[,c(3,10)], pred.OI))
table.obs.OI <- addmargins(prop.table(table(data.Grape$Formulação,
                                             data.Grape$ImpressaoGlobal), margin = 1))

# A.6) Diagnostic analysis
# (Graph of probabilities observed verses probabilities predicted)
# Table of observed and predicted odds
date.pred.obs <- choose.files()

# Reshaping table of observed and predicted odds
date.OI.pred.obs = data.frame(Formulation = date.pred.obs$Formulação,
Level = as.factor(date.pred.obs$ImpressaoGlobal),
Value = c(date.pred.obs$pred.OI, date.pred.obs$obs.OI),
Probability = rep(c("Predicted", "Observed"),

```

```

each = length(date.pred.obs$Formulação)))

# Graph: predicted and observed probabilities
ggplot2::ggplot(date.OI.pred.obs,
  aes(x = Level, y = Value, shape = Probability, colour = Probability)) +
  geom_point(size = 2) +
  facet_wrap(~factor(Formulation, levels = c("F1", "F2", "F3", "F4",
                                             "F5", "F6", "F7", "F8",
                                             "F9", "F10", "F11", "F12",
                                             "F13")), ncol = 4) +
  xlab("\nRating Scale\n") + ylab("Probability\n") +
  scale_colour_manual(name="Probability",
    breaks=c("Predicted","Observed"),
    values=c("blue", "red")) +
  theme(legend.position="top")

# A.7) Odds ratio
OR <- exp(fit.clmm.po.OI$coefficients)
ICbeta <- confint.default(fit.clmm.po.OI,level=0.95)
ICOR <- exp(ICbeta)
round((cbind(OR, ICOR)),3)

```

Apêndice II

```

#####
##                               R SCRIPT                               ##
#####
## Packages
library("ordinal";"ca";"factoextra")

## Reading the database
Grape.data <- choose.files()

## Attaching the database
attach(Grape.data)

## Multivariate grape database
Mult.Grape.data <- data.frame(Juice      = as.factor(Grape.data$Juice),
                             Panelist   = as.factor(Grape.data$Panelist),
                             Formulation = as.factor(Grape.data$Formulation),
                             Attribute   = as.factor(rep(c("Overall Impression",
                                                             "Aroma", "Body", "Sweetness", "Flavor"),
                                                             each = length(Grape.data$Formulation))),
                             Response    = as.factor(c(Grape.data$OverallImpression,
                                                         Grape.data$Aroma, Grape.data$Body,
                                                         Grape.data$Sweetness, Grape.data$Flavor)))

```

```
#####
##          MULTIPLE CORRESPONDENCE ANALYSIS          ##
#####
# A.1) Multiple contingency table (beverage formulation x sensory attribute)
tab.cont.mult <- table(Mult.Grape.data$Formulation, Mult.Grape.data$Response,
                      Mult.Grape.data$Attribute)
addmargins(tab.cont.mult)

# A.2) Multiple probability table with marginal totals
# (beverage formulation x sensory attribute)
tab.cont.prop <- round(addmargins(prop.table(tab.cont.mult, margin = 1)),3)
tab.cont.prop

# A.3) Chi-square association test
# H0: the variables are not associated (independent)
# HA: the variables are associated (dependent)

chisq.test(table(Mult.Grape.data$Formulation,
                 Mult.Grape.data$Response))$p.value > 0.05 # (FALSE: Reject H_0)

# A.4) Multiple Correspondence Analysis
Juice.data.matrix <- data.frame(
  Mult.Grape.data$Formulation[Mult.Grape.data$Attribute=="Overall Impression"],
  Mult.Grape.data$Response[Mult.Grape.data$Attribute=="Overall Impression"],
  Mult.Grape.data$Response[Mult.Grape.data$Attribute=="Aroma"],
  Mult.Grape.data$Response[Mult.Grape.data$Attribute=="Body"],
  Mult.Grape.data$Response[Mult.Grape.data$Attribute=="Sweetness"],
  Mult.Grape.data$Response[Mult.Grape.data$Attribute=="Flavor"])
colnames(Juice.data.matrix) <- c("Form", "IG", "Aroma", "Corpo", "Doçura", "Sabor")

cats <- apply(Juice.data.matrix, 2, function(x) nlevels(as.factor(x)))
mca.juice <- mjca(Juice.data.matrix, lambda = "indicator")
mca.data.frame <- data.frame(mca.juice$colcoord, Variable = rep(names(cats), cats))
rownames(mca.data.frame) <- mca.juice$levelnames

# A.5) Multiple Correspondence Analysis plot
ggplot(data = mca.data.frame, aes(x = X1, y = X2, label = rownames(mca.data.frame))) +
  xlab("Dimension 1") + ylab("Dimension 2") +
  coord_cartesian(xlim = c(-2.3,2.3), ylim = c(-2.3,2.3)) +
  geom_hline(yintercept = 0, colour = "gray70") +
  geom_vline(xintercept = 0, colour = "gray70") +
  geom_density2d(colour = "gray90") +
  geom_text(aes(colour = Variable)) +
  theme_bw() + theme(legend.position="none") +
  scale_colour_manual(values=c("black", "black", "black", "blue", "black", "black"))
```



```
#####
## CUMULATIVES LOGIT MIXED MULTIVARIATE WITH PROPORTIONAL ODDS ##
#####

# B.1) Testing the assumption of proportional odds model
fit.clmm.po <- clmm2(ordered(Response) ~ Formulation,
                    data = Mult.Grape.data,
                    random = Panelist, Hess = TRUE) # POM
fit.clmm.npo <- clmm2(ordered(Response) ~ 1,
                    nominal = ~ Formulation, data = Mult.Grape.data,
                    random = Panelist, Hess = TRUE) # NPOM

anova (fit.clmm.po, fit.clmm.npo)[2,7] > 0.05 # (FALSE: reject H_0)

# B.2) Testing if grape juice formulation is significant for the model
fit.clmm.npo.with.Formulation <- clmm2(ordered(Response) ~ 1,
                                       data = Mult.Grape.data, random = Panelist,
                                       Hess = TRUE) # NPOM with Formulation
fit.clmm.npo.without.Formulation <- clmm2(ordered(Response) ~ 1,
                                       nominal = ~ Formulation,
                                       data = Mult.Grape.data, random = Panelist,
                                       Hess = TRUE) #NPOM without Formulation
anova(fit.clmm.npo.without.Formulation, fit.clmm.npo.with.Formulation)

# B.3) Testing if grape juice Attribute is significant for the model
fit.clmm.npo.with.Formulation.Attribute <- clmm2(ordered(Response) ~ 1,
                                                nominal = ~ Formulation,
                                                data = Mult.Grape.data,
                                                random = Panelist, Hess = TRUE)
                                                # NPOM with Formulation and Attribute
fit.clmm.npo.without.Formulation.Attribute <- clmm2(ordered(Response) ~ 1,
                                                nominal = ~ Formulation + Attribute,
                                                data = Mult.Grape.data,
                                                random = Panelist, Hess = TRUE)
                                                # NPOM without Formulation and Attribute

anova(fit.clmm.npo.without.Formulation,
      fit.clmm.npo.with.Formulation,
      fit.clmm.npo.without.Formulation.Attribute)

# B.4) Selected model
fit.clmm.npo <- clmm2(ordered(Response) ~ 1,
                    nominal = ~ Formulation + Attribute,
                    data = Mult.Grape.data,
                    random = Panelist, Hess = TRUE, nAGQ = 5)

summary(fit.clmm.npo)
```

```

coef(summary(fit.clmm.npo))

# B.5) Standard deviation of the random coefficient
confint(profile(fit.clmm.npo, range = c(0.12, 4), nSteps = 20 , trace = 0))

# B.6) Predicted odds and Observed odds
pred      <- round(predict(fit.clmm.npo, newdata = Mult.Grape.data),3)
table.pred <- data.frame(cbind(Mult.Grape.data, pred))
table.obs  <- round(addmargins(prop.table(table(Mult.Grape.data$Formulation,
                                                Mult.Grape.data$Response,
                                                Mult.Grape.data$Attribute),
                                                margin = 1)),2)

# B.7) Diagnostic analysis
# (Graph of probabilities observed verses probabilities predicted)
# Table of observed and predicted odds
date.pred.obs <- choose.files()

# Reshaping table of observed and predicted odds
date.pred.obs <- data.frame(Formulation = date.pred.obs$Formulation,
                           Level       = as.factor(date.pred.obs$Response),
                           Value       = c(date.pred.obs$pred, date.pred.obs$obs),
                           Probability = rep(c("Predicted", "Observed"),
                           each        = length(date.pred.obs$Formulation)))

# Graph: predicted and observed probabilities
ggplot(date.pred.obs,
       aes(x = Level, y = Value, shape = Probability, colour = Probability)) +
  geom_point(size = 2) +
  facet_wrap(~factor(Formulation, levels = c("F1", "F2", "F3", "F4",
                                             "F5", "F6", "F7", "F8",
                                             "F9", "F10", "F11", "F12",
                                             "F13")), ncol = 4) +
  xlab("\nRating Scale\n") + ylab("Probability\n") +
  scale_colour_manual(name="Probability",
                     breaks=c("Predicted","Observed"),
                     values=c("blue", "red")) +
  theme(legend.position="top")

# B.8) Odds ratio
OR <- exp(fit.clmm.npo$coefficients)
ICbeta <- confint.default(fit.clmm.npo,level=0.95)
ICOR <- exp(ICbeta)
round((cbind(OR, ICOR)),2)

```