

## **Avaliando aceitação do sabor de suco de caju via modelo de chances proporcionais mistos**

Laura V. T. de Paula<sup>1†</sup>, Ana Maria S. de Araujo<sup>2</sup>, Idemauro Antonio R. de Lara<sup>3</sup>, Tatiana O. Lemos<sup>4</sup>.

<sup>1</sup> Graduanda em Estatística, Universidade Federal do Ceará (UFC)

<sup>2</sup> Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará (UFC)

<sup>3</sup> Departamento de Ciências Exatas, ESALQ/USP

<sup>4</sup> Centro Tecnológico, Centro de Ciências Sociais, Saúde e Tecnologia, Universidade Federal do Maranhão (UFMA)

**Resumo:** *O modelo de chances proporcionais é um método que pode ser utilizado para descrever a relação entre variáveis categóricas ordinais e uma ou mais covariáveis. Sendo este uma extensão dos modelos lineares generalizados, ele permite que se obtenham probabilidades acumuladas para cada uma das categorias de resposta e, por consequência, as probabilidades marginais. Com isso, o presente trabalho tem como finalidade ajustar um modelo de chances proporcionais para verificar aceitação do sabor de suco de caju. Assim, o experimento foi realizado em Fortaleza-CE, onde 100 (cem) provadores não treinados, previamente selecionados, após provarem uma pequena quantidade de suco foram convidados a responder várias questões, dentre elas: “Por favor, utilizando a escala abaixo, descreva o quanto você gostou ou desgostou de cada amostra com relação ao **sabor**”. A resposta a esta questão é composta por várias categorias ordinais para 5 diferentes marcas de suco de caju. Um modelo foi ajustado para esse experimento, e a fim de tratar a possível dependência existente entre as observações foi adicionado ao modelo de chances proporcionais um efeito aleatório, constituindo-se, então em um modelo misto. A análise dos dados bem como o ajuste dos modelos foram efetuados com auxílio do pacote **ordinal** disponível no software **R**, versão 2.15.*

**Palavras-chave:** *Modelos de chances proporcionais, Modelos mistos, Pacote ordinal, Suco de caju.*

**Abstract:** *The proportional odds model is a method that can be used to describe the relationship between ordinal categorical variables and one or more covariates. It is an extension of generalized linear models, it allows to obtain cumulative probabilities for each response category and, consequently, and marginal probabilities. This paper aims to fit a proportional odds model to verify acceptance of flavored cashew juice. The experiment was conducted in Fortaleza-CE, where one 100 (hundred) untrained tasters, previously selected, after proving a small amount of juice were asked to answer several questions, among them: “Please use the scale below, describe how much you liked or disliked of each sample’s **flavor**”. The answer to this question consists of several ordinal categories for five different brands of cashew juice. A model was adjusted for this experiment, and in order to deal with possible dependencies among observations, a random effect was added to the proportional odds model, turning into a mixed model. Data analysis and the fit of the models were made with the aid of **ordinal** package available in **R** software, version 2.15.*

**Keywords:** *Proportional odds models, Mixed models, Ordinal package, Cashew juice .*

---

† Autor correspondente: [lauravtp@hotmail.com](mailto:lauravtp@hotmail.com).

## Introdução

O modelo de regressão de chances proporcionais (“proportional odds”) é uma metodologia que pode ser utilizada para descrever a relação entre variáveis categorizadas ordinais e uma ou mais covariáveis (McCULLAGH, 1980), permitindo que se obtenham probabilidades acumuladas para cada uma das categorias de resposta e, por consequência, as probabilidades marginais. Segundo Stram, Wei e Ware (1988) o ajuste de modelos marginais com dados dessa natureza podem ser utilizados para os casos em que se têm mais de uma ocasião (tempo, dosagem, espaçamento). Uma alternativa para tratar a possível dependência existente entre as observações é a utilização de modelos mistos (COSTA, 2010). Estes modelos são caracterizados pela introdução de um efeito aleatório associado a cada observação, seguindo uma distribuição pré-especificada, permitindo acrescentar uma estrutura de covariância entre as medidas de uma mesma unidade experimental.

O presente trabalho tem por objetivo utilizar o modelo de chances proporcionais com efeitos aleatórios para verificar uma possível relação entre a variável sabor e algumas covariáveis em um experimento, no qual cinco marcas de suco de caju são avaliadas segundo uma escala ordinal de resposta (Likert).

## Material e métodos

### Material

Com o objetivo de comparar a aceitação de cinco marcas comerciais de suco de caju, um estudo foi realizado e, neste, 100 (cem) provadores não treinados, previamente selecionados, após provarem uma pequena quantidade de suco foram convidados a responder várias questões, dentre elas: “Por favor, utilizando a escala abaixo, descreva o quanto você gostou ou desgostou de cada amostra com relação ao **sabor**”. A resposta a esta questão segue a seguinte escala ordinal: (1) Desgostei muitíssimo; (2) Desgostei muito; (3) Desgostei moderadamente; (4) Desgostei ligeiramente; (5) Nem Gosto, nem desgosto; (6) Gostei ligeiramente; (7) Gostei moderadamente; (8) Gostei muito; (9) Gostei muitíssimo. Para efeito deste trabalho as categorias foram reagrupadas em três categorias: (1) Desgosto (junção das categorias 1 a 4); (2) Nem gosto, nem desgosto; (3) Gosto (junção das categorias 6 a 9). O suco provado era proveniente de amostras de suco tropical de caju pronto para beber em embalagens Tetrabrik de 1 litro, pertencentes a um mesmo lote, coletadas em um supermercado de Fortaleza - CE. Foram selecionadas 5 (cinco) embalagens de cada marca.

O experimento foi realizado de acordo com Macfie et al. (1989), de forma que todas as marcas de suco de caju foram provadas o mesmo número de vezes e foram precedidas pelas demais também a mesma quantidade de vezes e ocuparam todas as posições na ordem de prova. Os provadores foram questionados quanto ao sexo, faixa etária, formada pelas categorias menos de 18 anos, de 19 a 25 anos, de 26 a 35 anos, de 36 a 45 anos, de 46 a 55 anos e acima de 56 anos; quanto ao consumo de suco de caju e quanto ao consumo de outros sucos, cujas opções eram (1) Consumo muito frequentemente, (2) Consumo frequentemente, (3) Consumo ocasionalmente, (4) Consumo pouco e (5) Não consumo e também quanto ao gosto por suco de caju e por outros sucos cujas respostas possíveis eram (1) Gosto muitíssimo, (2) Gosto muito e (3) Gosto moderadamente. Essa variáveis podem exercer influência sobre o sabor com relação ao suco de caju e foram aqui consideradas como covariáveis.

### Métodos

Como mencionado anteriormente, a variável resposta em questão é do tipo ordinal, isto é, pode assumir valores no conjunto  $\{1, 2, 3\}$ , tais que,  $1 < 2 < 3$ . A resposta para o  $i$ -ésimo avaliador é denotada pelo vetor  $\mathbf{y}_i = (I_{i1}, I_{i2}, I_{i3})'$ , em que  $I_{ij}$  representam variáveis indicadoras para as categorias de resposta, isto é  $I_{ij} = 1$  quando o provador optou pela  $j$ -ésima categoria ou  $I_{ij} = 0$ , caso contrário. O vetor  $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{7i})'$  diz respeito às 7 covariáveis consideradas no experimento e associadas ao indivíduo  $i$ . Seja  $\pi_j(\mathbf{x})$ ,  $j = 1, 2$  e  $3$ , a probabilidade marginal de ocorrência de  $I_{ij}$  e as probabilidades acumuladas

são representadas por:

$$\gamma_j(\mathbf{x}) = \pi_1(\mathbf{x}) + \dots + \pi_j(\mathbf{x})$$

O modelo de chances proporcionais surgiu na literatura como sendo uma generalização dos modelos logísticos acumulativos e pode ser representado pela seguinte equação:

$$\ln \left[ \frac{\gamma_j(\mathbf{x})}{1 - \gamma_j(\mathbf{x})} \right] = \lambda_j - (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_7 x_7) = \lambda_j - \boldsymbol{\beta} \mathbf{x}, \quad (1)$$

em que  $\lambda_j$  é o intercepto para cada categoria  $j$  e  $\mathbf{x}$  representa o vetor de valores das variáveis explicativas. Nesse modelo os interceptos tem um peso proporcional diferente para cada uma das categorias e estes depende unicamente das categorias e não das variáveis explicativas.

O objetivo principal ao ajustar um modelo é estimar o vetor de parâmetros  $\boldsymbol{\beta}$  e as probabilidades associadas a cada categoria de resposta. Essas estimativas são obtidas por meio da teoria da máxima verossimilhança com o uso do processo iterativo de *Newton-Raphson*. Uma vez obtidas as estimativas dos parâmetros do modelo (1), as estimativas das probabilidades acumuladas são dadas por:

$$\hat{\gamma}_j(\mathbf{x}) = \frac{\exp(\hat{\lambda}_j - \hat{\boldsymbol{\beta}}' \mathbf{x})}{1 + \exp(\hat{\lambda}_j - \hat{\boldsymbol{\beta}}' \mathbf{x})}, \quad j = 1, 2 \text{ e } 3.$$

Portanto, a probabilidade marginal associada à  $j$ -ésima categoria é representada por:

$$\hat{\pi}_j(\mathbf{x}) = \hat{\gamma}_j(\mathbf{x}) - \hat{\gamma}_{(j-1)}(\mathbf{x})$$

Uma alternativa para tratar a possível dependência entre as observações é utilizar o modelo de efeitos mistos, ou seja, incluir um efeito aleatório no modelo de chances proporcionais. De acordo com Christensen (2012), o novo modelo será dado por:

$$\ln \left[ \frac{\gamma_j(\mathbf{x})}{1 - \gamma_j(\mathbf{x})} \right] = \lambda_j - \boldsymbol{\beta}' \mathbf{x} - \mathbf{b}_i, \quad (2)$$

em que o termo  $\mathbf{b}_i$  corresponde ao efeito aleatório do avaliador, supondo que  $\mathbf{b}_i \sim \mathcal{N}(0, \sigma_b^2)$ .

Nesse trabalho, a análise dos dados bem como o ajuste dos modelos são efetuados com auxílio do pacote **ordinal** (Christensen, 2011) disponível no *software R*, versão 2.15.

A inclusão do efeito aleatório justifica-se pela necessidade de modelar a possível dependência existente entre as observações de um mesmo avaliador, visto que, todos eles provaram as cinco marcas de suco de caju. A significância deste efeito aleatório foi verificada por meio do intervalo de confiança para o parâmetro  $\sigma_b$  construído a partir da verossimilhança perfilada, em que este procedimento é muito utilizado quando o modelo a ser ajustado envolve parâmetros de perturbação.

## Resultados e discussões

Primeiramente ajustamos um modelo de chances proporcionais misto que continha todas as variáveis independentes. Com base nesse ajuste pode-se observar que as variáveis idade, sexo, consumo de suco de caju, consumo de outros sucos e gosto de suco de caju não foram significativas para o modelo. Assim, o modelo resultante contém apenas duas covariáveis, sendo estas marca e gosto por outros sucos de frutas, que será aqui denotada por “outros”. Assume-se neste modelo um efeito aleatório para avaliador (conforme descrito pela equação 2).

Ao ajustar um novo modelo com as duas variáveis significativas e o efeito aleatório pode-se observar que a estimativa para o desvio padrão do coeficiente aleatório foi 0,8326. Os intervalos de 95% e 99% de confiança para  $\sigma_b$  não contém o valor zero, indicando a existência de efeito aleatório, o que pode ser observado pela Figura 1.

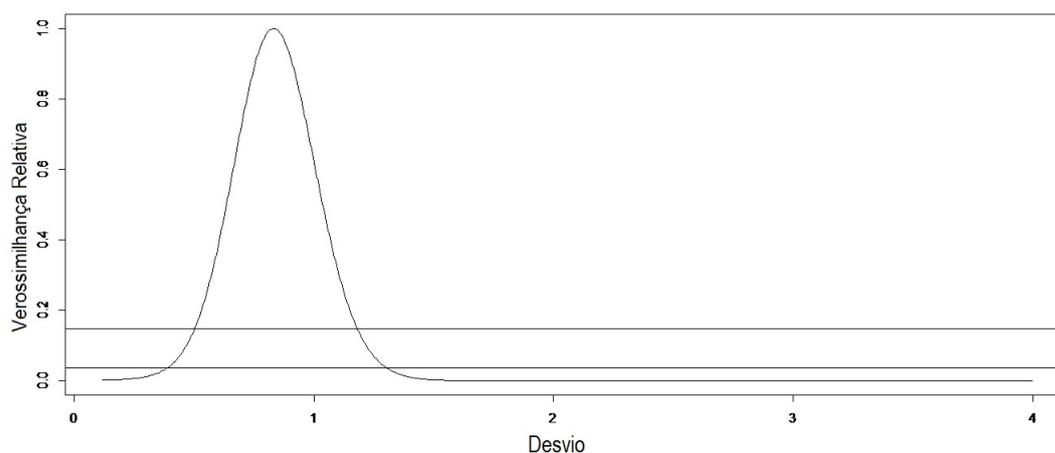


Figura 1: Verossimilhança perfilada de  $\sigma_b$  do modelo (2) para os dados de suco de caju.

Os valores estimados para esse modelo estão apresentados na Tabela 1.

Tabela 1: Estimativas para os parâmetros do modelo de chances proporcionais para sabor do suco de caju.

Parâmetro	Estimativa	Erro-padrão	Valor $z$	Nível descritivo
<b>Interceptos</b>				
$\lambda_2$	0,848	0,260	3,259	
$\lambda_3$	1,253	0,265	4,720	
<b>Covariáveis</b>				
marca(B)	0,340	0,299	1,134	0,256
marca(C)	2,430	0,349	6,959	< 0,001
marca(D)	0,723	0,299	2,419	0,015
marca(E)	0,935	0,301	3,098	0,001
outros (2)	0,615	0,282	2,180	0,029
outros (3)	0,485	0,422	1,151	0,249

A Figura 2 apresenta as probabilidades estimadas de cada categoria de resposta para diferentes combinações de marcas de suco e opções de resposta para “outros”. Por ela observa-se que a marca de suco C obteve a pior avaliação pois apresenta as maiores probabilidades associadas à categoria 3, que corresponde a desgostar do produto, o que não ocorre com as demais marcas.

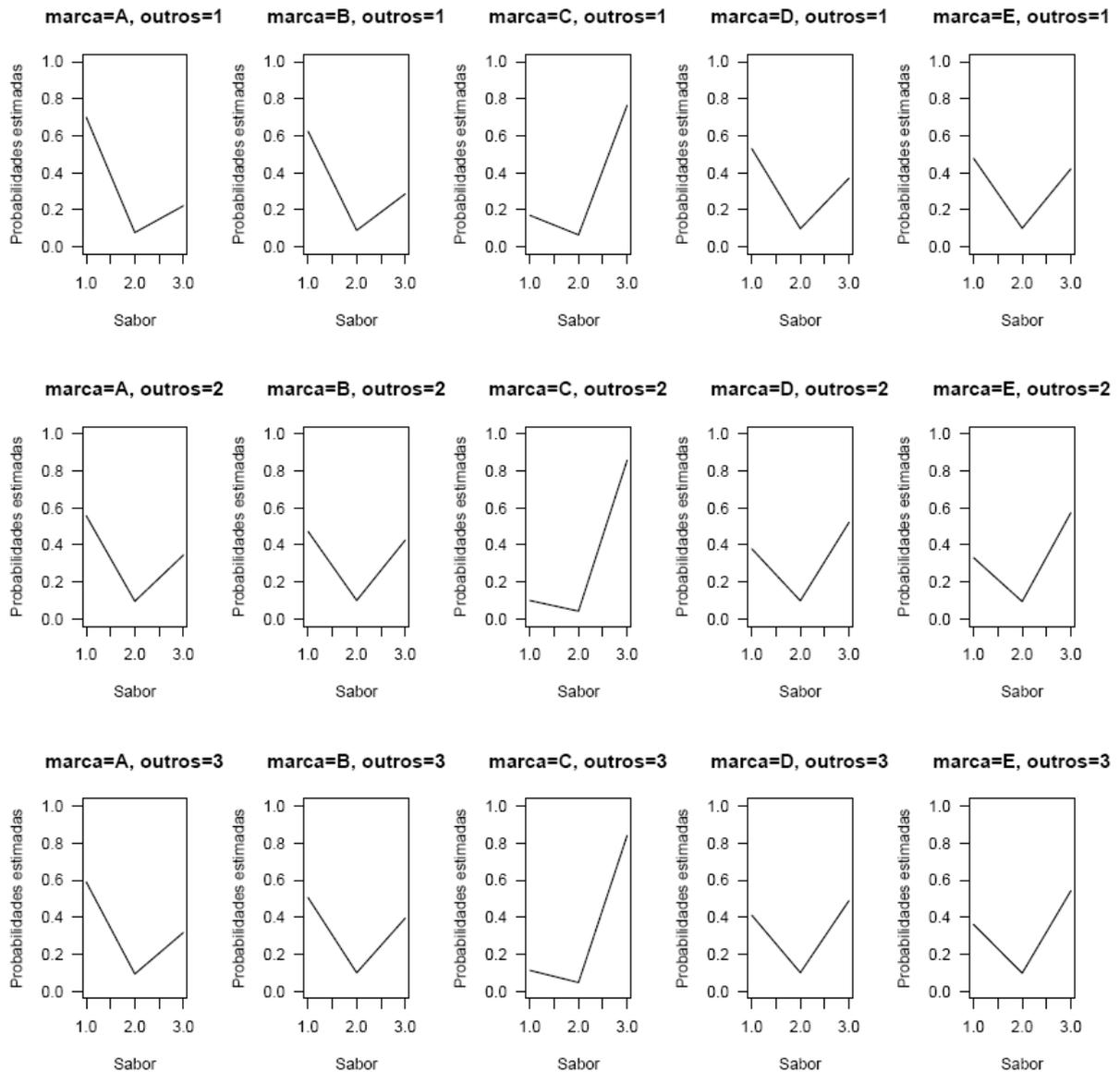


Figura 2: Estimativas das probabilidades referentes ao modelo ajustado.

## Conclusões

A metodologia utilizada propiciou a seleção de variáveis e de um modelo com menos parâmetros e, além disso, a inclusão do efeito aleatório possibilitou a modelagem da dependência entre as observações repetidas de um mesmo provador. Com relação aos resultados obtidos, as probabilidades estimadas indicaram uma baixa aceitação à marca C. Quanto às demais marcas, houve uma divisão na opinião dos provadores.

## Referências

CHRISTENSEN, R.H. Analysis of ordinal data with cumulative link models estimation with the R-package ordinal. Disponível em:

[http://cran.r-project.org/web/packages/ordinal/vignettes/clm\\_intro.pdf](http://cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf). Acesso em: 10 abril 2013.

CHRISTENSEN, R.H. Analysis of sensory ratings data with cumulative link models. *Journal de la Société Française de Statistique*. (Submetido)

COSTA, T. R. *Modelos lineares mistos: uma aplicação na produção de leite de vacas da raça sindi*. 2010. 79 p. Dissertação (Mestrado em Biometria e Estatística Aplicada) - Universidade Federal Rural de Pernambuco, UFRPE, Recife, 2010.

MACFIE, H. J.; BRATCHELL, N.; GREENHOFF, K.; VALLIS, L. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies* v. 4, n. 2, p. 129 – 148, 1989.

McCULLAGH, P. Regression Methods for Ordinal Data. *Journal of The Royal Statistical Society, Series B*, London, v. 42, p. 109 – 142, 1980.

STRAM, D.O.; WEI, L.J.; WARE, J.H. Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates. *Journal of American Statistical Association*, Boston, v. 83, p. 631 – 637, 1988.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2012. Disponível em: <http://www.R-project.org/>. Acesso em: 10 abril 2013.