ANÁLISE CONJUNTA BASEADA EM ESCOLHAS: UM ENFOQUE VIA MODELO LINEAR GENERALIZADO DE POISSON

Eduardo Campana BARBOSA[1]
Carlos Henrique Osório SILVA[1]
Moysés NASCIMENTO[1]
Suzana Maria Della LÚCIA[2]
Valéria Paula Rodrigues MINIM[3]

▪ RESUMO: Neste trabalho apresentou-se uma aplicação prática da metodologia Análise Conjunta Baseada em Escolhas. Foi proposto o modelo Linear Generalizado de Poisson para analisar os dados. O estudo foi conduzido com escolhas referentes à avaliação, por 144 consumidores, de oito amostras de iogurtes *light* sabor morango, variando a informação de três ingredientes (açúcar, gordura e proteína) em um delineamento fatorial completo. São apresentados os resultados e inferências no que se refere à estimação do efeito principal dos atributos, das probabilidades e da razão de escolhas. Verificou-se que o modelo Linear Generalizado de Poisson foi equivalente ao modelo Logit Multinomial e pode ser utilizado para analise em questão. Quanto aos resultados práticos, a maior probabilidade de escolha foi associada ao iogurte *light* sabor morango contendo as informações: 0% de açúcar, 0% de gordura e enriquecido com proteínas bioativas.

▪ PALAVRAS-CHAVE: Preferência do consumidor; iogurte; modelo Logit multinomial.

1 Introdução

Muitos fatores contribuem para que o consumidor faça uma escolha no momento da compra. Dentre estes, merece especial destaque a qualidade e o preço do produto ofertado. Adicionalmente, características intrínsecas como a marca, o design, a embalagem e as informações presentes no rótulo podem interferir nesta decisão, fato explicado devido às expectativas, crenças ou o conhecimento prévio dos consumidores em relação a cada produto (DELIZA *et al*., 2003).

Naturalmente, o sucesso e a sobrevivência de uma empresa estão atrelados à sua capacidade de fidelizar clientes. O bom relacionamento e a comunicação efetiva são princípios importantes para se atingir este objetivo, no entanto, não são suficientes. Logo, estratégias de marketing são desenvolvidas para se tentar compreender o comportamento e

[1] Universidade Federal de Viçosa - UFV, Departamento de Estatística, CEP: 36570-977, Viçosa, MG, Brasil. E-mail: duducampana@hotmail.com; chos@ufv.br; moysesnascm@ufv.br
[2] Universidade Federal do Espírito Santo - UFES, Departamento de Engenharia de Alimentos, CEP: 29500-000, Alegre, ES, Brasil. E-mail: smdlucia@yahoo.com.br
[3] Universidade Federal de Viçosa - UFV - Departamento de Tecnologia de Alimentos, CEP: 36571-000, Viçosa, MG, Brasil. E-mail: vprm@ufv.br

a preferência dos consumidores no momento da compra, visando maximizar os lucros e minimizar os custos e riscos (NATTER e FEURSTEIN, 2002).

É neste contexto que a Análise Conjunta Baseada em Escolhas é aplicada. Tal metodologia permite modelar a escolha do consumidor e estimar probabilidades associadas a diferentes produtos. Moore (2004) observou que a modelagem de escolhas é uma forma realista, simples e prática para avaliar a preferência do consumidor. Na Análise Conjunta Baseada em Escolhas, as variáveis em estudo são atributos específicos dos produtos, definidos por seus respectivos níveis, isto é, valores ou características que os quantificam ou qualificam (AAKER, KUMAR e DAY, 2008).

As aplicações clássicas da Análise Conjunta Baseada em Escolhas englobam o estabelecimento de novas estratégias competitivas (MAHAJAN, GREEN e GOLDBERG, 1982), a avaliação de novos produtos de interesses mercadológicos (LOUVIERE e WOODWORTH, 1983) e a otimização de produtos já existentes (GREEN e KRIEGER, 1993). Trabalhos mais recentes, em diferentes áreas do conhecimento, podem ser verificados em Tempesta *et al*. (2010), Zimmermann *et al*. (2013), Utz *et al*. (2014) e Arenoe *et al*. (2015). No Brasil, na área de tecnologia de alimentos, Deliza *et al*. (2010) utilizaram a Análise Conjunta Baseada em Escolhas para avaliar o efeito do preço, da informação sobre irradiação e da qualidade sensorial na escolha de mamão e Della Lucia *et al.* (2010) utilizaram essa técnica para avaliar a influência de alguns atributos presentes na embalagem de iogurte *light* sabor morango na preferência do consumidor.

Na Análise Conjunta Baseada em Escolhas, o modelo empregado é o Logit Multinomial, que tem seus parâmetros estimados por máxima verossimilhança (MCFADDEN, 1974). Conforme Train (2009), esse método é aplicável se três pressupostos sobre o conjunto de alternativas (produtos) forem verificados, de que estas sejam: i) mutuamente exclusivas, ii) exaustivas e iii) finitas. Portanto, os consumidores precisam escolher apenas uma única alternativa dentre o conjunto finito destas. No entanto, em situações reais, os consumidores podem escolher mais de um ou nenhum dos produtos avaliados, o que viola os pressupostos (i) e (ii), respectivamente.

Neste sentido, outras formas de estimação foram propostas. Uma dessas é por máxima verossimilhança parcial, um método empregado para estimar os parâmetros do modelo de Cox (COX, 1972; 1975), utilizado na área de analise de sobrevivência. Com algumas adaptações (detalhes em SO e KUHFELD, 1995 e KUHFELD, 2010) é possível ajustar o modelo utilizado na Análise Conjunta Baseada em Escolhas pelo referido método. Conforme So e Kuhfeld (1995) e Kuhfeld (2010), essa equivalência ocorre devido à função de verossimilhança do modelo Logit Multinomial ter a mesma forma da função de verossimilhança parcial de Breslow (1974) do modelo de Cox.

Ainda neste contexto, Whitehead (1980) e Palmgren (1981) mostraram que alguns modelos log lineares são equivalentes a modelos multinomiais. Mais precisamente, em McCullagh e Nelder (1989), pode-se verificar que o modelo Linear Generalizado de Poisson (com função de ligação logarítmica) é equivalente a um modelo multinomial com probabilidade similar à estimada pelo modelo Logit utilizado na Análise Conjunta Baseada em Escolhas.

Portanto, no presente trabalho propõe-se a utilização do modelo Linear Generalizado de Poisson para estudos de Análise Conjunta Baseada em Escolhas. Embora a equivalência entre esses modelos seja conhecida na literatura, no contexto de Análise Conjunta e modelagem de escolhas esse tópico ainda é recente e pouco explorado, principalmente no Brasil. Logo, tem-se por objetivo apresentar uma alternativa metodológica para analisar dados desta natureza. Adicionalmente, muitos *softwares* estatísticos, com interfaces amigáveis, possuem implementado o ajuste dos modelos Lineares Generalizados, o que favorece para que essa abordagem possa ser utilizada por estudantes e pesquisadores de outras áreas, que não possuem conhecimentos de programação computacional em softwares estatísticos específicos.

2 Material e métodos

2.1 Produtos avaliados. Descrição do conjunto de dados

Para avaliar a escolha dos consumidores por amostras de iogurte *light* sabor morango definiu-se $r = 3$ atributos: informação sobre o conteúdo de açúcar ("açúcar"), informação sobre o conteúdo de gordura ("gordura") e informação sobre o conteúdo de proteína ("proteína"), com os respectivos níveis: açúcar ("0% de açúcar" e "com adoçante"), gordura ("0% de gordura" e "baixo teor de gordura") e proteína ("enriquecido com proteínas do soro do leite" e "enriquecido com proteínas bioativas"). O delineamento fatorial empregado foi o do perfil completo, com $\boldsymbol{J} = 2^3 = 8$ amostras de iogurte, apresentados na Tabela 1.

Tabela 1 - Delineamento experimental do estudo

Amostra	Açúcar	Gordura	Proteína
1	0% de açúcar	0% de gordura	Com proteínas do soro do leite
2	Com adoçante	0% de gordura	Com proteínas do soro do leite
3	0% de açúcar	Baixo teor de gordura	Com proteínas do soro do leite
4	Com adoçante	Baixo teor de gordura	Com proteínas do soro do leite
5	0% de açúcar	0% de gordura	Com proteínas bioativas
6	Com adoçante	0% de gordura	Com proteínas bioativas
7	0% de açúcar	Baixo teor de gordura	Com proteínas bioativas
8	Com adoçante	Baixo teor de gordura	Com proteínas bioativas

O estudo envolveu $n = 144$ consumidores que residiam na cidade de Viçosa (Minas Gerais, Brasil) e o pré-requisito para que o voluntário participasse da pesquisa era que ele tivesse o hábito de consumir iogurte e produtos *light*. Destes, 59,7% foram do sexo feminino e 88,2% encontravam-se com idade entre 15 e 25 anos; 81,2% possuíam curso superior incompleto ou em andamento. Adicionalmente, 40,3% dos participantes afirmaram ter renda entre 1 e 5 salários mínimos; 38,2% entre 6 e 10 salários; 19,4% possuíam renda entre 11 e 20 salários e 2,1% mencionaram uma renda acima de 20 salários mínimos.

A avaliação foi conduzida em um local confortável e sob a orientação dos aplicadores do estudo. As oito amostras de iogurte foram apresentadas aos consumidores por meio de fotografias de suas embalagens. A ordem de apresentação foi balanceada segundo MacFie *et al*. (1989), para permitir a estimação não tendenciosa dos efeitos principais dos atributos e minimizar os efeitos da ordem de apresentação e residual, caracterizados pela influência de uma amostra de iogurte na avaliação da subsequente. Maiores detalhes sobre os procedimentos metodológicos descritos nessa seção podem ser consultados em Della Lúcia *et al*. (2010).

2.2 Metodologia tradicional de análise

A utilidade aleatória é uma variável latente que representa o “benefício” do consumidor ao escolher uma determinada amostra de iogurte. Este termo é composto por um componente fixo e observável e outro aleatório, portanto, não observável. O modelo para a utilidade aleatória é o apresentado em (1).

$$U_{nj} = X_j'\boldsymbol{\beta} + \varepsilon_{nj} \tag{1}$$

em que U_{nj} é o vetor ($\boldsymbol{J}$ x 1) de utilidades aleatórias atribuídas pelo *n*-ésimo consumidor; X_j é o vetor (r x 1) de r atributos referente a j-ésima amostra de iogurte avaliada; $\boldsymbol{\beta}$ é o vetor (r x 1) de parâmetros desconhecidos e ε_{nj} é o vetor ($\boldsymbol{J}$ x 1) de erros aleatórios e não observáveis, referentes ao *n*-ésimo consumidor. Cada ε_{nj} é considerado independente e com distribuição idêntica de valores extremos do tipo I ou Gumbel (TRAIN, 2009).

Supõe-se que o *n*-ésimo consumidor escolhe a *j*-ésima amostra de iogurte, se e somente se, $U_{nj} > U_{nk}, \forall_{j \neq k;\, j,k \in \{1,2,\dots,\boldsymbol{J}\}}$. Sob essa premissa, o modelo em (2) ou Logit Multinomial (MCFADDEN, 1974) é estabelecido.

$$P(Y_n = j|\boldsymbol{X}) = \frac{e^{X_j'\boldsymbol{\beta}}}{\sum_{k=1}^{J} e^{X_k'\boldsymbol{\beta}}} \ \forall_{j = 1,2,\dots,\boldsymbol{J}} \tag{2}$$

em que $P(Y_n = j|\boldsymbol{X})$ é a probabilidade do *n*-ésimo consumidor escolher a *j*-ésima amostra de iogurte, condicional ao conjunto de atributos $\boldsymbol{X}$. Os estimadores de $\boldsymbol{\beta}$ são os que maximizam a função de verossimilhança parcial de Breslow, $L(\boldsymbol{\beta}, \boldsymbol{y}|\boldsymbol{X})$, apresentada em (3). Como a solução do sistema de equações diferenciais não é analítica emprega-se o método numérico de Newton-Raphson (GALLANT, 2009).

$$L(\boldsymbol{\beta}, \boldsymbol{y}|\boldsymbol{X}) = \prod_{n=1}^{N}\prod_{j=1}^{J}\left[\frac{e^{X_j'\boldsymbol{\beta}}}{\left(\sum_{k=1}^{J} e^{X_k'\boldsymbol{\beta}}\right)^{d_n}}\right]^{I_{nj}} \tag{3}$$

em que $I_{nj} = 1$ se o *n*-ésimo consumidor escolheu a *j*-ésima amostra de iogurte e $I_{nj} = 0$ caso contrário. O termo n_j indica o número de produtos escolhidos pelo *n*-ésimo consumidor, com $n_j \geq 1$. Posteriormente, as estimativas pontuais de $\boldsymbol{\beta}$ são substituídas no modelo (2) e probabilidades de escolha associadas aos $\boldsymbol{J}$ produtos são estimadas. Para detectar quais atributos influenciam a escolha do consumidor emprega-se o teste de Wald (1943) para os coeficientes. A importância dos níveis de cada atributo é avaliada pela Razão de Escolhas (RE) (GREENE, 2003), conforme em (4).

$$RE(X_p, X_q) = \frac{P(Y_n = p|\boldsymbol{X})}{P(Y_n = q|\boldsymbol{X})} = e^{(X_p' - X_q')\boldsymbol{\beta}} \tag{4}$$

A *RE* é definida como a razão entre a probabilidade de escolha do produto p e a probabilidade de escolha do produto q, onde q é obtido de p fixando $r - 1$ níveis dos r atributos, ou seja, alterando apenas o nível de um único atributo (SILVA e BASTOS, 2010). Logo, por meio dessa medida pode-se inferir o quanto o nível de um atributo é preferível ($RE > 1$) ou não ($RE < 1$) em relação ao outro. Como mencionado anteriormente, a *RE* é constante e depende apenas de características das alternativas p e q. Por isso, a remoção ou a inclusão de alternativas de escolhas não a altera. Mais especificamente, se um produto for excluído ou um novo produto for incluído, a probabilidade de escolha dos demais produtos aumenta ou diminui proporcionalmente. Essa é uma característica particular do modelo Logit Multinomial que é assegurada pela independência dos erros aleatórios (TRAIN, 2009).

2.3 Metodologia proposta de análise

Uma distribuição pertence à família exponencial uniparamétrica se sua função de probabilidade (f.p., caso discreto) ou de densidade de probabilidade (f.d.p., caso contínuo) puder ser escrita conforme em (5). Mais especificamente, dentre os modelos de probabilidade mais conhecidos, pertencem a essa família o da Normal, Binomial, Binomial Negativa, Gama, Poisson, Normal Inversa, Multinomial etc (CORDEIRO *et al*., 1995).

$$f_Y(y|\theta) = h(y)\, exp\{\eta(\theta)\, t(y) - b(\theta)\} \qquad (5)$$

em que $h(y)$, $\eta(\theta)$, $t(y)$ e $b(\theta)$ são funções não únicas e que assumem valores reais. Adicionalmente, a família exponencial na forma canônica é um caso particular de (5) quando $\eta(\theta)$ e $t(y)$ são funções identidade, como apresentado em (6).

$$f_Y(y|\theta) = h(y)\, exp\{\theta\, y - b(\theta)\} \qquad (6)$$

Nelder e Wedderburn (1972) estabeleceram modelos específicos de regressão, fundamentados sob a teoria de família exponencial de distribuições e denominados como Modelos Lineares Generalizados (MLG). A estrutura do MLG é composta pelo trinômio: i) componente aleatório, ii) componente sistemático e iii) função de ligação.

O componente aleatório Y retrata o conjunto de variáveis aleatórias $Y_1, Y_2, \dots, Y_n$, independentes, com distribuição idêntica e pertencente à família exponencial canônica, juntamente com um parâmetro de perturbação positivo ϕ, que pode ou não ser conhecido, e que quantifica a dispersão da referida distribuição (DOBSON, 2001). Portanto, em (7) tem-se a f.p. ou f.d.p. de cada Y_i:

$$f_Y(y_i|\theta,\phi) = exp\{\phi^{-1}[\theta_i y_i - b(\theta_i)] + c(y_i,\phi)\} \qquad (7)$$

Tal que:

$$E(Y_i) = \mu_i = b'(\theta_i)$$

$$VAR(Y_i) = \phi^{-1}\, b''(\theta_i) = \phi^{-1}\, V(\mu_i) = \phi^{-1}\, \frac{d\mu_i}{d\theta_i}\ \forall_{i=1,2,\dots,n}$$

em que ϕ^{-1} e $V(\mu_i)$ representam, respectivamente, a precisão e a função de variância da distribuição.

O componente sistemático $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)'$ ou preditor linear é composto pelo produto da matriz de delineamento $\boldsymbol{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n})'$ do modelo, com dimensão *n* x *r*, onde cada $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots, x_{ir})_{\forall_{i=1,2,\ldots,n}}$, pelo vetor paramétrico desconhecido $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_r)'$, de dimensão $(r+1)$ x 1, cujos efeitos são incorporados como uma soma linear (Firth, 1991), que algebricamente ou em notação matricial é equivalente a:

$$\eta_i = \sum_{j=1}^{r} x_{ij}\, \beta_j, \forall_{i=1,2,\ldots,n}$$

$$\boldsymbol{\eta} = \boldsymbol{X\beta}$$

Por fim, a função de ligação $g(.)$, monótona e diferenciável, tem por finalidade relacionar o componente aleatório ao componente sistemático, isto é, a média da distribuição de cada Y_i ao seu respectivo preditor linear.

$$\eta_i = g(\mu_i)$$

A escolha da função de ligação deve ser realizada de acordo com a distribuição dos dados, avaliando o espaço e o tipo de efeito dos parâmetros, as propriedades estatísticas produzidas e a facilidade de interpretação dos resultados. Dentre as principais funções de ligação destacam-se a Identidade, Logarítmica, Logística, Inversa, Inversa do Quadrado (MCCULLOCH e SEARLE, 2000).

2.3.1 Modelo Linear Generalizado de Poisson

Aplica-se o MLG ou modelo de regressão de Poisson (BIRCH, 1963) se a variável dependente possui observações que assumem valores inteiros e não negativos, na forma de contagens. Nessa análise, os dados experimentais serão descritos satisfatoriamente se a variância das observações for proporcional à sua média, devido ao equilíbrio da distribuição de Poisson (HABERMAN, 1970). Naturalmente, em muitas ocasiões essa suposição é violada e duas situações são possíveis: i) dados com subdispersão ou ii) dados com superdispersão. Em (8) verifica-se a função de probabilidade da distribuição de Poisson escrita conforme em (7), dado $\phi = 1$ (conhecido) e $c(y_i, \phi) = -\log(y_i!)$. Adicionalmente, em (9) tem-se a função de ligação logarítmica para o referido modelo.

$$f_Y(y_i|\mu_i, \phi) = exp\{(1)[y_i\, log(\mu_i) - \mu_i] - log\,(y_i!)\} \qquad (8)$$

$$\eta_i = log(\mu_i) \qquad (9)$$

A função de ligação logarítmica assegura que os valores esperados de cada Y_i fiquem restritos ao intervalo $(0; \infty)$ e que efeitos sistemáticos multiplicativos tornem-se aditivos. Além disso, tal função é canônica para o modelo de Poisson, portanto, assegura a existência de um conjunto de estatísticas suficientes e mínimas para $\boldsymbol{\beta}$ e propriedades estatísticas desejáveis para os estimadores de máxima verossimilhança (como facilidade de cálculo e unicidade das estimativas), principalmente em pequenas amostras (CORDEIRO e DEMÉTRIO, 2008).

2.3.1.1 Estimação dos Parâmetros

A estimação do vetor de parâmetros no MLG de Poisson ocorre pelo método da Máxima Verossimilhança (NELDER E WEDDERBURN, 1972). Os estimadores (EMV) ou $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_r)'$ são obtidos pela maximização do logaritmo da função de verossimilhança ou função suporte, $l(\boldsymbol{\beta}) = \ln[L(\boldsymbol{\beta})]$, apresentado em (10).

$$l(\boldsymbol{\beta}) = \phi^{-1}\sum_{i=1}^{n}[y_i\theta_i - b(\theta_i)] + \sum_{i=1}^{n} c(y_i, \phi) \qquad (10)$$

em que $\theta_i = q(\mu_i)$, $\mu_i = \exp(\eta_i)$ e $\eta_i = \sum_{j=1}^{r} x_{ij}\,\beta_j$. Como o sistema formado pelas derivadas parciais, quando igualadas à zero, não possui solução analítica, emprega-se o algoritmo iterativo conhecido como Escore de Fisher, mais simples e similar ao algoritmo de Newton-Raphson se a função de ligação for canônica.

Para avaliar o ajuste do MLG e a importância conjunta das covariáveis, medidas estatísticas denominadas como desvios (ou *deviance*) são utilizadas para indicar o quanto o modelo ajustado se distancia do modelo teórico gerador dos dados (CORDEIRO e DEMÉTRIO, 2008). Mais especificamente, o desvio do modelo nulo refere-se à distância do modelo teórico ao modelo ajustado apenas com o termo intercepto, enquanto o desvio residual, ao modelo ajustado com o intercepto e também com os atributos como covariáveis. Se as covariáveis forem importantes para predizer a variável resposta, após sua inclusão, espera-se que ocorra uma redução do desvio. O teste de Wald (1943) é também aplicado para analisar a significância estatística das estimativas obtidas no MLG de Poisson.

2.4 Aspectos computacionais

Neste trabalho utilizou-se o *software* R (R CORE TEAM, 2016) na condução das análises estatísticas. O procedimento de estimação dos parâmetros do modelo Logit Multinomial, por máxima verossimilhança parcial, ocorre pela função *clogit* do pacote *survival* (AIZAKI e NISHIMURA, 2008). Já na análise proposta, o MLG de Poisson foi ajustado via função *glm* do pacote *stats*. O nível de significância adotado para testes de hipóteses foi $\alpha = 5\%$. Os resultados originais deste estudo podem ser consultados no Apêndice I ou em Della Lúcia *et al*. (2010).

3 Resultados e discussão

3.1 Estimação do vetor de parâmetros

A Tabela 2 apresenta as estimativas de máxima verossimilhança para os parâmetros do MLG de Poisson, com os respectivos erros padrão. O teste de Wald, a significância estatística de cada coeficiente e os desvios do modelo ajustado são também informados.

Tabela 2 - Estimativas, erro padrão, teste de Wald e estatísticas do modelo ajustado

Atributos	$\hat{\beta}$	Erro Padrão	Z	p-valor
Intercepto	-2,2604	0,2072	-10,909	0,0000*
Açúcar	-1,6603	0,2275	-7,299	0,0000*
Proteína	1,5126	0,2166	6,982	0,0000*
Gordura	-0,7885	0,1798	-4,386	0,0000*

Desvio do modelo nulo = 598,88 (G.L. = 1151).
Desvio residual = 441,37 (G.L. = 1148)
* Significativo a 5% de probabilidade

Note que ao incorporar os atributos como covariáveis, o desvio do modelo ajustado reduziu (desvio residual = 441,37) em relação ao desvio do modelo apenas com o termo de intercepto (desvio nulo = 598,88). Portanto, há evidencias de que os atributos influenciem a escolha do consumidor. Essa conclusão foi similar à obtida por Della Lúcia *et al*. (2010), que utilizou o teste da Razão de Verossimilhanças (MOOD, GAYBILL e BOES, 1974) para verificar e inferir sobre a importância conjunta dos atributos na preferência dos consumidores.

Pelo teste de Wald, a hipótese nula $H_0: \beta_i = 0$ foi rejeitada em relação à hipótese alternativa $H_a: \beta_i \neq 0$ para cada um dos três atributos avaliados (p-valor < 0,05), corroborando os resultados anteriores. Portanto, informações referentes ao conteúdo de açúcar, proteína e gordura, presentes na embalagem do iogurte *light* sabor morango, influenciam significativamente a escolha dos consumidores.

O intercepto do modelo de Poisson foi significativo, no entanto, esse termo não possui interpretação prática na Análise Conjunta Baseada em Escolhas, pois os principais resultados envolvem apenas as estimativas dos efeitos principais dos atributos. Adicionalmente, termos de interação entre os atributos podem ser incluídos no modelo, porém, conforme Siqueira (2000), o percentual de explicação da variabilidade total associada a estes é pequeno, cerca de 5 a 10%, e por isso muitas vezes são negligenciados.

Destaca-se que nos dados analisados, a média e a variância amostral da variável escolha foram muito próximas, respectivamente, 0,1250 e 0,1094. Verificou-se a equivalência estatística dessas medidas via teste *t* de Student (t = 1,5931 e valor-p = 0,1114), o que corroborou a escolha do MLG de Poisson. Para avaliar o ajuste do MLG de Poisson em situações adversas da Análise Conjunta Baseada em Escolhas, foram realizadas algumas simulações, cujos códigos em linguagem R e resultados podem ser consultados nos apêndices II e III, respectivamente. Algumas informações sobre as avaliações:

i) Para dados com subdispersão, as estimativas do efeito principal dos atributos e de erro padrão dos estimadores, pelo ajuste do MLG de Poisson, foram semelhantes às obtidas pelo ajuste por máxima verossimilhança parcial do modelo Logit Multinomial. Isto é, mesmo violando o equilíbrio da distribuição de Poisson os resultados não divergiram, o que é assegurado pela equivalência dos modelos (MCCULLAGH e NELDER, 1989). No entanto, segundo Winkelmann (1995), na ausência de equidispersão, inferências sobre os parâmetros do modelo, produzidas por testes de hipóteses e intervalos de confiança podem ser comprometidas, já que as estimavas de erro padrão não serão consistentes. Portanto, os autores sugerem o estudo de metodologias corretivas como o modelo Quase-Poisson (WEDDERBURN, 1974) ou o modelo contagem Gama (WINKELMANN, 1995).

ii) Dados com superdispersão não foram analisados, pois nos estudos de Análise Conjunta Baseada em Escolhas a variável dependente é representada por um conjunto de valores zero e um, isto é, $\boldsymbol{Y} \sim Bin(N = 1, p)$, com $0 \leq p \leq 1$. Portanto, apenas se $p \rightarrow 0$ tem-se que $V(\boldsymbol{Y}) \geq E(\boldsymbol{Y})$. Neste cenário nenhum produto seria escolhido por nenhum consumidor, pois apenas valores nulos (0) seriam gerados. Logo, essa é uma situação improvável e o leitor não precisa se preocupar com o problema de superdispersão e, consequentemente, com a utilização do MLG Binomial Negativo (HINDE & DEMÉTRIO, 1998b) ou de métodos mais complexos, como o de variância robusta (LIANG e ZEGUE, 1986), modelo Quase-Poisson (WEDDERBURN, 1974) ou modelo Linear Generalizado Duplo (SMYTH, 1989).

iii) Adicionalmente, nos dados simulados, alguns consumidores hipotéticos não escolheram nenhum dos produtos e outros consumidores escolheram mais de um dos produtos. Os resultados demonstraram que as estimativas de efeito principal dos atributos e de erro padrão dos estimadores, pelo ajuste do MLG de Poisson, coincidiram com as obtidas pelo ajuste do modelo Logit Multinomial por máxima verossimilhança parcial. Portanto, o modelo proposto pode ser utilizado para as diferentes variações da metodologia Análise Conjunto Baseada em Escolhas.

Neste sentido, o MLG de Poisson é uma alternativa metodológica que pode auxiliar estudantes e pesquisadores no desenvolvimento de novos estudos de Análise Conjunta Baseada em Escolhas, ampliando o emprego dessa técnica estatística, principalmente no Brasil. Adicionalmente, como a classe dos MLG já é amplamente difundida em diversas áreas do conhecimento, o ajuste desses modelos encontra-se implementado em *softwares* estatísticos com *interfaces* amigáveis, como o Minitab versão 17 (*Stat – Regression – Poisson Regression – Fit Poisson model*), IBM SPSS (*Analyze – Generalized Linear Models – Type of Model – Poisson loglinear*) e StatSoft Statistica (*Statistics – Generalized Linear/Nonlinear Models – Distribution Poisson*).

3.2 Probabilidades de escolha

Obtidas as estimativas de máxima verossimilhança para os parâmetros do MLG de Poisson, as probabilidades de escolha foram estimadas substituindo estes valores na modelo (2). A Tabela 3 apresenta as probabilidades para as oito amostras de iogurte *light* sabor morango.

Concluiu-se que a amostra de iogurte com maior probabilidade de escolha, de acordo com os consumidores avaliados, foi a amostra 5, com probabilidade 0,4734 e características: 0% de açúcar, 0% de gordura e com proteínas bioativas. As amostras de iogurte 7 (0% de açúcar, baixo teor de gordura e enriquecido com proteínas bioativas) e 1 (0% de açúcar, 0% de gordura e Enriquecido com proteínas de soro do leite) foram, respectivamente, a segunda (0,2152) e a terceira (0,1043) com as maiores probabilidades de escolha. A amostra de iogurte 4, com adoçante, baixo teor de gordura e proteínas do soro de leite, foi a que apresentou a menor probabilidade de escolha, igual a 0,0090.

Tabela 3 - Probabilidades de Escolha para as 8 amostras de iogurte

Amostra	Açúcar	Gordura	Proteína	Probabilidades Estimadas
1	0% de açúcar	0% de gordura	Proteínas do soro do leite	0,1043
2	Com adoçante	0% de gordura	Proteínas do soro do leite	0,0198
3	0% de açúcar	Baixo teor de gordura	Proteínas do soro do leite	0,0474
4	Com adoçante	Baixo teor de gordura	Proteínas do soro do leite	0,0090
5	0% de açúcar	0% de gordura	Proteínas bioativas	0,4734
6	Com adoçante	0% de gordura	Proteínas bioativas	0,0900
7	0% de açúcar	Baixo teor de gordura	Proteínas bioativas	0,2152
8	Com adoçante	Baixo teor de gordura	Proteínas bioativas	0,0409
				1,0000

3.3 Razão de escolha

São apresentados os cálculos para a Razão de Escolha. No intuito de padronizar as interpretações, tais medidas foram estabelecidas visando superar a unidade.

$$RE(X_1, X_2) = \frac{P(Y_n = 1|\boldsymbol{X})}{P(Y_n = 2|\boldsymbol{X})} = \frac{0{,}1043}{0{,}0198} = 5{,}26$$

Mantendo fixos os níveis dos atributos Gordura (0% de Gordura) e Proteína (soro do leite), e modificando apenas os níveis do atributo Açúcar (0% de Açúcar para Com adoçante), conclui-se que a amostra de iogurte 1, com 0% de Açúcar, é 5,26 vezes mais provável de ser escolhida do que a amostra 2, com adoçante. Tal conclusão foi similar à relatada por Reis *et al*. (2009), de que o nível "Com adoçante" causou impacto negativo em relação à "Sem adição de açúcar", em um estudo de Análise Conjunta com avaliação das embalagens de iogurte baseada em notas.

$$RE(X_1, X_3) = \frac{P(Y_n = 1|\boldsymbol{X})}{P(Y_n = 3|\boldsymbol{X})} = \frac{0{,}1043}{0{,}0474} = 2{,}20$$

Mantendo fixos os níveis dos atributos Açúcar (0% de Açúcar) e Proteína (soro do leite), e modificando apenas os níveis do atributo Gordura (0% de gordura para baixo teor de gordura), conclui-se que a amostra de iogurte 1, com 0% de gordura, é 2,20 vezes mais provável de ser escolhida do que a amostra 3, com baixo teor de gordura. Portanto, os consumidores entrevistados desejam consumir um produto isento de gordura.

$$RE(X_5, X_1) = \frac{1}{RE(X_1, X_5)} = \frac{P(Y_n = 5|\boldsymbol{X})}{P(Y_n = 1|\boldsymbol{X})} = \frac{0{,}4734}{0{,}1043} = 4{,}53$$

Mantendo fixos os níveis dos atributos Açúcar (0% de Açúcar) e Gordura (0% de Gordura), e modificando apenas os níveis do atributo Proteína (de bioativa para soro do leite), conclui-se que a amostra de iogurte 5, com proteínas bioativa, é 4,53 vezes mais provável de ser escolhida do que a amostra 1, com proteínas do soro do leite. Conforme relatado por Della Lúcia *et al*. (2010), o consumidor associou ao termo "proteínas do soro do leite" um aspecto negativo, possivelmente devido às características sensoriais percebidas como pouco agradáveis. Já o termo "bioativas" foi associado a um alimento mais saudável.

4 Conclusões

O MLG de Poisson foi equivalente ao modelo Logit Multinomial e pode ser empregado nos estudos de Análise Conjunta Baseada em Escolhas. A análise proposta foi válida para as diferentes variações da metodologia em estudo, principalmente quando os consumidores não escolhem nenhum e/ou escolhem mais de um dos produtos avaliados. Em caso de subdispersão os resultados não se alteraram, no entanto, por questões teóricas, metodologias corretivas foram sugeridas e podem ser avaliadas em trabalhos futuros.

Quanto aos aspectos práticos do estudo, concluiu-se que os três atributos manipulados na análise (açúcar, gordura e proteína) foram importantes na escolha dos consumidores. A amostra de iogurte light sabor morango contendo as informações "0% de açúcar", "0% de gordura" e "enriquecido com proteínas bioativas" foi a que obteve maior probabilidade de escolha para o grupo de consumidores participantes deste estudo.

BARBOSA, E. C.; SILVA, C. H. O.; NASCIMENTO, M.; DELLA LUCIA, S. M.; MINIM, V. P. R. Choice-based conjoint analysis: An approach by generalized linear Poisson model. *Rev. Bras. Biom*. Lavras, v.35, n.1, p.194-212, 2017.

- *ABSTRACT: This paper presented a practical application of the choice-based conjoint analysis methodology. It was proposed the generalized linear Poisson model to analyze the data. The study was conducted with choices regarding the evaluation done by 144 consumers, eight samples of light yoghurt strawberry flavor, varying the information of three ingredients (sugar, fat and protein) in a full factorial design. Inferences and results are presented in terms of estimating the main effect of the attributes, probability and ratio of choice. It was found that the generalized linear Poisson model was equivalent to the multinomial logit model and can be used for the analysis in question. Practical results: the most likely choice was associated with yoghurt containing strawberry flavor light information: 0% of sugar, 0% fat and enriched with bioactive proteins.*

- *KEYWORDS: Consumer preference; yogurt; multinomial logit model.*

Referências

AAKER, D. A.; KUMAR, V.; DAY, G. S. *Marketing research*. John Wiley & Sons, 2008.

AIZAKI, H.; NISHIMURA, K. Design and analysis of choice experiments using R: a brief introduction. *Agricultural Information Research*, v.17, n.2, p.86-94, 2008.

ARENOE, B.; VAN DER REST, J. P. I.; KATTUMAN, P. Game theoretic pricing models in hotel revenue management: An equilibrium choice-based conjoint analysis approach. *Tourism Management*, v.51, p.96-102, 2015.

BRESLOW, N. Covariance analysis of censored survival data. *Biometrics*, p.89-99, 1974.

BIRCH, M. W. Maximum likelihood in three-way contingency tables. *Journal of the Royal Statististical Society B*, v.25, p.220-233, 1963.

CORDEIRO, G. M.; CRIBARI-NETO, F.; AUBIN, E. Q.; FERRARI, S. L. P. Bartlett corrections for one-parameter exponential family models. *Journal of Statistical Computation and Simulation*, v.53, p.211-231, 1995.

CORDEIRO, G. M.; DEMÉTRIO, C. G. *Modelos lineares generalizados e extensões*. São Paulo, 2008.

COX, D. R. Regression Models and Life-tables (with discussion). *Journal of Royal Statistical Society*, B, v.34, p.187-220, 1972.

COX. D. R. Partial Likelihood. *Biometrika*, v.62, 1975.

DELLA LUCIA, S. M.; MINIM, V. P. R.; SILVA, C. H. O.; MINIM, L. A.; SILVA, R. C. S. N. Análise conjunta de fatores baseada em escolhas no estudo da embalagem de iogurte light sabor morango. *Brazilian Journal of Food technology*, p.11-18, 2010.

DELIZA, R.; ROSENHAL, A.; SILVA, A. L. S. Consumer atitude towards information on non conventional technology. *Trends in Food Science & Technology*, v.14, n.1-2, p.43-49, 2003.

DELIZA, R.; ROSENTHAL, A.; HEDDERLEY, D.; JAEGER, S. R. Consumer perception of irradiated fruit: a case study using choice-based conjoint analysis. *Journal of Sensory Studies*, v.25, n.2, p.184-200, 2010.

DOBSON, A. J. *An introduction to generalized linear models*. 2.ed. London: Chapman & Hall/CRC, 2001.

GALLANT, A. R. *Nonlinear statistical models*. John Wiley & Sons, 2009.

GREEN, P. E.; KRIEGER, A. M. Conjoint analysis with product-positioning applications. *Handbooks in operations research and management science*, v.5, p.467-515, 1993.

GREENE, W. H. *Econometric analysis*. Pearson Education India, 2003.

HABERMAN, S. *The general log-linear model*. PhD dissertation. University of Chicago Press, Chicago, Illinois, 1970.

HOLFORD, T. R. The analysis of rates and survivorship using log-linear models. *Biometrics*, v.36, p.299-305, 1980.

HINDE, J.; DEMÉTRIO, C. G. B. Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, v.27, p.151-170, 1998b.

KUHFELD, W. F. *Marketing research methods in SAS. Experimental design, choice, conjoint, and graphical techniques*. Cary, NC, SAS-Institute TS-722, 2010.

LAIRD, N.; OLIVIER, D. Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, v.76, n.23, p.1-240, 1981.

LEE, J. Odds ratio or relative risk for cross-sectional data?. *International Journal of Epidemiology*, v.23, n.1, p.201-203, 1994.

LIANG, K. Y.; ZEGER, S. L. Longitudinal data analysis using generalized linear models. *Biometrika*, v.73, n.1, p.13-22, 1986.

LOUVIERE, J. J.; WOODWORTH, G. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of marketing research*, p.350-367, 1983.

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

MCCULLOCH, C. E.; SEARLE, S. R (2000). *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons, 2000.

MAHAJAN, V.; GREEN, P. E.; GOLDBERG, S. M. A conjoint model for measuring self-and cross-price/demand relationships. *Journal of Marketing Research*, p.334-342, 1982.

MCFADDEN, D. Conditional Logit Analysis of Qualitative Choice Behavior, *Frontiers in Econometrics*, P. Zarembka Eds., New York: Academic Press, p.105-142, 1974.

MCCULLAGH, P.; NELDER, J. A. *Generalized linear models* (Vol. 37). CRC press, 1989.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction of the theory of statistics*. New York: McGraw-Hill, 1974.

MOORE, W. L. A cross-validity comparison of rating-based and choice-based conjoint analysis models. *International Journal of Research in Marketing*, v.21, n.3, p.299-312, 2004.

NATTER, M.; FEURSTEIN, M. Real world performance of choice-based conjoint models. *European Journal of Operational Research*, v.137, n.2, p.448-458, 2002.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society*, A, v.135, p.370-384, 1972.

PALMGREN, J (1981). The Fisher information matrix for log linear models arguing conditionally on observed explanatory variable. *Biometrika*, v.68, n.2, p.563-566, 1981.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

REIS, R. C.; MINIM, V. P. R.; DIAS, B. R. P.; CHAVES, J. B. P.; MINIM, L. A. Impact of the use of different sweeteners in the acceptability of strawberry light yogurt. *Alimentos e Nutrição*, v.20, n.1, p.53-60, 2009.

SILVA, C.; BASTOS, F. *Introdução à conjoint analysis*. IX Encontro Mineiro de Estatística (MGEST), Viçosa, 2010.

SIQUEIRA, J. O. *Mensuração da Estrutura de Preferência do Consumidor: Uma Aplicação de Conjoint Analysis em Marketing*. 230 p. Dissertação (Mestrado em Administração, concentração em Métodos Quantitativos) Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo - SP, 2000.

SMYTH, G. K. Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society*. Series B, v.51, n.47–60, 1989.

SO, Y.; WARREN F, K. Multinomial logit models. *SUGI 20 Conference Proceedings*, 1995.

TEMPESTA, T.; GIANCRISTOFARO, R. A.; CORAIN, L.; SALMASO, L.; TOMASI, D.; BOATTO, V. The importance of landscape in wine quality perception: An integrated approach using choice-based conjoint analysis and combination-based permutation tests. *Food Quality and Preference*, v.21, n.7, p.827-836, 2010.

TRAIN, K. E. *Discrete choice methods with simulation*. Cambridge: Cambridge University Press, 2009.

UTZ, K. S.; HOOG, J.; WENTRUP, A.; BERG, S.; LÄMMER, A.; JAINSCH, B.; ..., & SCHENK, T. Patient preferences for disease-modifying drugs in multiple sclerosis therapy: a choice-based conjoint analysis. *Therapeutic advances in neurological disorders*, v.7, n.6, p.263-275, 2014.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc*., v.54, p.426-482, 1943.

WEDDERBURN, R. W. Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. *Biometrika*, v.61, n.3, p.439-447, 1974.

WINKELMANN, R. Duration Dependence and Dispersion in Count-Data Models. *Journal of Business & Economic Statistics*, v.13, n.4, p.467–474, 1995.

WHITEHEAD, J. Fitting Cox's regression model to survival data using GLIM. *Applied Statistics*, p.268-275, 1980.

ZIMMERMANN, T. M.; CLOUTH, J.; ELOSGE, M.; HEURICH, M.; SCHNEIDER, E.; WILHELM, S.; WOLFRATH, A. Patient preferences for outcomes of depression treatment in Germany: A choice-based conjoint analysis study. *Journal of affective disorders*, v.148, n.2, p.210-219, 2013.

Recebido em 30.03.2016

Aprovado após revisão em 14.07.2016

Apêndice I - – Resultados originais do trabalho. Fonte: Della Lúcia et al. (2010).

Tabela 4. Resumo da análise de estimação dos coeficientes do modelo por máxima verossimilhança.

Variável	Estimativa do Coeficiente ($\hat{\beta}$)	Valor Hazard Ratio (HR)
Açúcar	-1,66007*	0,190
Gordura	-0,78846*	0,455
Proteína	1,51251*	4,538

*significativo ($p < 0,0001$) pelo teste de qui-quadrado.

Tabela 5. Probabilidades estimadas pela ANCFE para os oito tratamentos.

Tratamento	Probabilidades Estimadas pela ANCFE
1	0,10431
2	0,01983
3	0,04741
4	0,00901
5	0,47336
6	0,09000
7	0,21516
8	0,04091

Apêndice II – Código em R do trabalho de simulação.

```
### Definindo atributos, níveis e os produtos para o exemplo hipotético.

install.packages("conjoint")
library("conjoint")

# Definindo os atributos e níveis

experiment = expand.grid(
A = c("a1","a2"),
B = c("b1","b2"),
C = c("c1","c2"),
D = c("d1","d2"))

# Definindo os produtos em um Esquema Fatorial Completo

design = caFactorialDesign(data = experiment, type = "full")
print(design)

# Codificação dos níveis

code = caEncodedDesign(design)
code

# O número de produtos

(p = dim(code)[1])

# Definindo o número de consumidores hipotéticos e a variável indicadora de consumidores

(n = 48)

(id = rep(1:n, each = p))

# Definindo uma variável indicadora para cada produto hipotético

(prod = rep(1:p, n))

# Replicando a matriz de codificação dos níveis dos atributos para cada um dos consumidores.
```

```
(X = as.matrix(do.call("rbind", replicate(n, code, simplify = F))))

# Definindo a variável resposta ou as escolhas feitas por cada consumidor

set.seed(123456)

n_0 = 10 # número de consumidores que não vão escolher nenhum dos 16 produtos

(choice_0 = rep(0, n_0*p)) # Estabelecendo a variável escolha para esses n = 10 consumidores
(choice_1 = rbinom((n - n_0)*p, size = 1, prob= 0.7))  # Estabelecendo a variável escolha para os
demais consumidores, elevados valores de "p" para gerar subdispersão

(choice = c(choice_1, choice_0)) # Definindo o conjunto final de escolhas

# Média e variância da variável escolha, teste t para concluir se E(Y)>V(Y) ⟹ V(Y)<E(Y) ou
subdispersão

mean(choice)
var(choice)
t.test(choice, mu=var(choice), alternative = c("greater"))

# Contabilizando o número de vezes que cada produto hipotético foi escolhido

ind=rep(seq(1:p),n)
esc=NULL
for (i in 1:p)
{
esc[i]=sum(1*(ind==i&choice==1))
}
esc
(sum_esc=sum(esc)) # número total de escolhas feitas pelos n = 48 consumidores

# Definindo o conjunto de dados final

dados=as.data.frame(cbind(id, prod, X, choice))
colnames(dados)<-(c("id","Produto","A","B","C","D","choice"))
dados
head(dados,16)
attach(dados)

### Ajuste do Modelo Logit Multinomial por máxima verossimilhança parcial (função clogit,
pacote survival)
```

```
install.packages("survival")
library("survival")

m1=clogit(choice ~ A + B + C + D, data=dados, method="breslow")
summary(m1)

### Ajuste do Modelo Linear Generalizado de Poisson por máxima verossimilhança

library("stats")

m2=glm(choice ~ A + B + C + D, data=dados, family=poisson(link="log"))
summary(m2)
```

Apêndice III - Ajuste dos modelos Logit Multinomial e de Poisson por máxima verossimilhança parcial.

```
R Console
> ### Ajuste do Modelo Logit Multinomial por máxima verossimilhança parcial (fu$
> m1=clogit(choice ~ A + B + C + D, data=dados, method="breslow")
> summary(m1)
Call:
coxph(formula = Surv(rep(1, 768L), choice) ~ A + B + C + D, data = dados,
    method = "breslow")

  n= 768, number of events= 424

      coef exp(coef) se(coef)      z Pr(>|z|)
A  0.07551   1.07843  0.09720  0.777    0.437
B  0.01887   1.01905  0.09713  0.194    0.846
C  0.05662   1.05825  0.09717  0.583    0.560
D -0.06606   0.93607  0.09718 -0.680    0.497

  exp(coef) exp(-coef) lower .95 upper .95
A    1.0784     0.9273    0.8914     1.305
B    1.0190     0.9813    0.8424     1.233
C    1.0583     0.9450    0.8747     1.280
D    0.9361     1.0683    0.7737     1.132

Concordance= 0.539  (se = 0.031 )
Rsquare= 0.002   (max possible= 0.999 )
Likelihood ratio test= 1.44  on 4 df,   p=0.8366
Wald test            = 1.44  on 4 df,   p=0.8367
Score (logrank) test = 1.44  on 4 df,   p=0.8366
```

```
R Console
> ### Ajuste do Modelo Linear Generalizado de Poisson por máxima verossimilhanç$
> m2=glm(choice ~ A + B + C + D, data=dados, family=poisson(link="log"))
> summary(m2)

Call:
glm(formula = choice ~ A + B + C + D, family = poisson(link = "log"),
    data = dados)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1084  -1.0376   0.4668   0.5537   0.6305

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.72316    0.29656  -2.439   0.0147 *
A            0.07551    0.09720   0.777   0.4373
B            0.01887    0.09713   0.194   0.8460
C            0.05662    0.09717   0.583   0.5601
D           -0.06606    0.09718  -0.680   0.4966
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 503.76  on 767  degrees of freedom
Residual deviance: 502.32  on 763  degrees of freedom
AIC: 1360.3

Number of Fisher Scoring iterations: 5
```