Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 10, n. 1, 2023.

Trabalho apresentado no XLII CNMAC, Universidade Federal de Mato Grosso do Sul - Bonito - MS, 2023

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Modelagem da média e da variância em experimentos com misturas

Edmilson Rodrigues Pinto[1] Leandro Alves Pereira[2]
FAMAT/UFU, Uberlândia, MG

Resumo. Em experimentos industriais, controlar a variabilidade é de suma importância para garantir a qualidade dos produtos. Modelos de regressão clássicos são amplamente utilizados na indústria para experimentos com mistura; porém, quando a suposição de variância constante não é satisfeita, a construção de procedimentos que permitam minimizar a variabilidade torna-se necessária e outros métodos de modelagem estatística devem ser considerados. A abordagem considerada neste artigo utiliza a classe dos modelos lineares generalizados. Esta classe é bem geral e bastante flexível, generalizando algumas das mais importantes distribuições de probabilidade e possibilita modelar a variabilidade através da modelagem conjunta da média e da dispersão (MCMD). O objetivo deste artigo é mostrar como a MCMD pode ser utilizada para modelar a média e a variância em experimentos de mistura para os quais a distribuição da variável resposta não possui dispersão constante. Para obtenção dos modelos da média e da dispersão, será utilizada uma função no R que permite realizar o processo de estimação e seleção de variáveis na MCMD. Uma aplicação a um problema de mistura com variáveis de processo(ruído), proveniente da indústria de alimentos, será utilizada como ilustração. Espera-se que este trabalho possa contribuir para a utilização da MCMD na modelagem de experimentos envolvendo misturas.

Palavras-chave. Experimentos com misturas, modelos lineares generalizados, modelagem conjunta da média e dispersão, seleção de modelos, planejamento robusto.

1 Introdução

Em problemas envolvendo misturas, as respostas a serem observadas e medidas estão funcionalmente relacionadas às proporções dos seus componentes, podendo também existir relação com outros fatores externos à mistura. Modelos de regressão, geralmente com respostas Gaussianas (Normais), são comumente usados para experimentos com misturas.

Um dos grandes desafios da indústria é controlar a variabilidade dos produtos e, desta forma, obter produtos mais robustos, evitando retrabalho, diminuindo desperdícios e custos de produção excessivos. A variabilidade em experimentos com mistura pode acontecer devido à natureza estocástica do processo produtivo e também como consequência da ocorrência de fatores de ruído, os quais são fatores externos que não podem ser controlados pelo experimentador. Portanto, modelar e controlar a variabilidade é de suma importância.

Os modelos de regressão clássicos, geralmente usados em experimentos de mistura, assumem as suposições de normalidade dos erros, variância constante e linearidade dos efeitos sistemáticos. No entanto, podem surgir situações em que tais suposições não possam ser totalmente satisfeitas. De acordo com Nelder e Lee [9], para muitos conjuntos de dados as suposições subjacentes aos modelos lineares clássicos são insatisfatórias e as transformações da variável resposta não podem

[1]edmilson.pinto@ufu.br
[2]leandro.ap@ufu.br

necessariamente satisfazer todas as suposições necessárias aos modelos de regressão clássicos, especialmente normalidade. Por outro lado, os modelos lineares generalizados permitem a análise de dados para os quais as suposições do modelo de regressão clássico não são satisfeitas e fornecem uma estrutura geral para modelagem da variância por meio da modelagem conjunta de média e dispersão (MCMD), veja Nelder e Lee [9] e Pinto e Ponce de Leon [12] para maiores detalhes.

Em experimentos de mistura, a modelagem da variância é considerada apenas em situações onde existem variáveis de ruído. Variáveis de ruído são variáveis de processo que não podem ser controladas, como umidade ou temperatura. Variáveis de processo são fatores em um experimento que não formam nenhuma parte da mistura, mas cujos níveis, quando alterados, podem afetar os resultados finais da mistura, veja Cornell [2] para mais detalhes.

A modelagem estatística da variância em experimentos de mistura com variáveis de ruído foi considerada por Steiner e Hamada [19], que propuseram um modelo combinado de mistura com variáveis de processo e ruído, construíram e resolveram, um problema de otimização para minimizar uma função de perda quadrática, levando em consideração ambos os modelos da média e da variância da resposta. Outra abordagem para modelar a variância em experimentos de mistura com variáveis de ruído é devida a Goldfarb, Borror e Montgomery [3] usando o método delta, que emprega uma aproximação de série de Taylor de primeira ordem do modelo de regressão em um vetor de variáveis de ruído. Para maiores detalhes e mais referências sobre o método delta, veja Pinto [13]. Uma aplicação do método delta para modelar a variabilidade em experimentos de mistura com variáveis de ruído em engenharia de alimentos também é apresentada por [4]. No entanto, em todas as abordagens apresentadas para modelagem da variância, foram assumidas as suposições de normalidade dos erros e parâmetro de dispersão constante, ou seja, o modelo da média foi considerado normal homocedástico e a modelagem da variância se deu apenas em função das variáveis de ruído, presentes no modelo.

A abordagem que consideramos neste artigo é bastante geral porque, ao considerar a classe de modelos lineares generalizados, distribuições diferentes da distribuição normal podem ser consideradas e, além disso, considerando um modelo para a dispersão, o processo de modelagem da variabilidade permite modelar a variância não apenas em situações onde existem variáveis de ruído.

O objetivo deste artigo é mostrar como a modelagem conjunta de média e dispersão pode ser aplicada para obter os modelos da média e da variância em experimentos de mistura, quando a dispersão da distribuição da variável resposta não é constante. Uma função no *software* R (R Core Team [16]), que realiza o processo de estimação de parâmetros e seleção de variáveis na MCMD é apresentada e será utilizada para obter os modelos conjuntos da média e da dispersão.

O artigo está estruturado da seguinte forma. Na Seção 2, apresentamos brevemente os conceitos básicos sobre experimentos com misturas. Na Seção 3, são apresentadas as principais ideias da MCMD. Uma aplicação a um problema da indústria de alimentos é dada na Seção 4 e considerações finais são dadas na Seção 5.

2 Experimentos com misturas

Um experimento de mistura envolve proporções de ingredientes de dois ou mais componentes para fazer diferentes composições de um produto final. Em experimentos envolvendo misturas, o objetivo principal é determinar as proporções dos componentes da mistura que levam a resultados desejáveis da variável resposta com relação a alguma propriedade de interesse. As proporções dos componentes de mistura x_i estão sujeitas às seguintes restrições $0 \leq x_i \leq 1$, $i = 1, 2, \ldots, a$, e $\sum_{i=1}^{a} x_i = 1$, onde a é o número de componentes envolvidos no experimento de mistura. Consequentemente, o espaço de planejamento ou região experimental é um simplex $(a-1)$-dimensional. (veja Boyd e Vandenberghe [1] para um tratamento mais abrangente).

Quando existem outras condições nas proporções das variáveis de mistura, a região experimental

DOI: 10.5540/03.2023.010.01.0066

é parte de um simplex n-dimensional, ou seja, $L_i \leq x_i \leq U_i$ para $i = 1, 2, \ldots, a-1$, com a proporção x_a tomando o valor que compõe o resultado final da mistura. L_i e U_i são, respectivamente, os limites inferior e superior do ingrendiente x_i da mistura.

Considere o modelo linear $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, onde $\boldsymbol{y}^t = (y_1, \ldots y_n)$ é um vetor resposta, $\boldsymbol{X}$ é uma matriz de planejamento de ordem $n \times p$ e posto $p \leq n$, $\boldsymbol{\beta}$ é um vetor de parâmetros desconhecidos, e $\boldsymbol{\epsilon}^t = (\epsilon_1, \ldots \epsilon_n)$ é um vetor de erros aleatórios com média zero e matriz de variância-covariância $\sigma^2\boldsymbol{I}$, onde $\sigma^2 = Var(\epsilon_i)$ para $i = 1, \ldots, n$ e $\boldsymbol{I}$ é a matriz identidade. O valor da resposta em algum ponto $\boldsymbol{x}^t = (x_1, \ldots, x_a)$ na região de interesse $\mathcal{X}$ é dado por $y(\boldsymbol{x}) = \zeta(\boldsymbol{x}, \boldsymbol{\beta}) + \epsilon$, onde $\zeta(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{f}^t(\boldsymbol{x})\boldsymbol{\beta}$ é um modelo polinomial de determinado grau nas variáveis de controle $x_1, \ldots, x_a$. O vetor $\boldsymbol{f}(\boldsymbol{x})$ é avaliado no ponto $\boldsymbol{x}$ e representa a linha da matriz $\boldsymbol{X}$. Os modelos comumente usados em experimentos com mistura são os chamados modelos de Scheffé. Para maiores detalhes, veja Scheffé [17] ou Cornell [2].

Em experimentos com mistura, variáveis de processo aparecem frequentemente. Problemas de experimentos de mistura com variáveis de processo surgem quando no experimento de mistura a característica de interesse é função das proporções dos ingredientes e de outros fatores que não constituem parte alguma da mistura, como temperatura ou tempo, por exemplo. Esses fatores, que dependem das condições do processo, são chamados de variáveis de processo. Desta forma, a variável resposta depende não apenas da proporção dos componentes da mistura, mas também das variáveis de processo. Assim, se em adição aos a componentes de mistura $\boldsymbol{x}^t = (x_1, \ldots, x_a)$ existirem r variáveis de processo $\boldsymbol{z}^t = (z_1, \ldots, z_r)$, podemos considerar modelos aditivos como $\eta(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \zeta(\boldsymbol{x}, \boldsymbol{\beta}) + \vartheta(\boldsymbol{z}, \boldsymbol{\gamma})$ ou modelos completos de produtos cruzados do tipo $\eta(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \zeta(\boldsymbol{x}, \boldsymbol{\beta})\vartheta(\boldsymbol{z}, \boldsymbol{\gamma})$ ou combinações desses modelos como $\eta(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \zeta(\boldsymbol{x}, \boldsymbol{\beta}) + \nu(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, onde $\vartheta(\boldsymbol{z}, \boldsymbol{\gamma})$ representa o modelo para as variáveis de processo e $\nu(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ compreende produto de termos em $\zeta(\boldsymbol{x}, \boldsymbol{\beta})$ e $\vartheta(\boldsymbol{z}, \boldsymbol{\gamma})$. Quando algumas variáveis de processo são incontroláveis ou difíceis de serem controladas, elas devem ser tratadas como variáveis de ruído. Variáveis de ruído são consideradas variáveis aleatórias com uma distribuição de probabilidade supostamente conhecida. Uma apresentação abrangente de vários aspectos da modelagem estatística e planejamento experimental envolvendo experimentos de mistura é fornecida por Cornell [2]. O livro de Smith [18] também é uma excelente referência.

3 Modelagem conjunta da média e da dispersão

De acordo com Nelder e Lee [9], a modelagem conjunta da média e dispersão consiste em encontrar modelos conjuntos para a média e para a dispersão em um MLG. Na abordagem considerada por eles, usando a quase-verossimilhança estendida, dois modelos lineares generalizados interligados são necessários, um para a média e outro para a dispersão.

Sejam $Y_1, \ldots, Y_n$, n variáveis aleatórias independentes com a mesma distribuição de probabilidade, representando as variáveis respostas dependentes, cujos valores observados são dados por $y_1, \ldots, y_n$. Para a i-ésima resposta Y_i é assumido somente serem conhecidos $E(Y_i) = \mu_i$ e $Var(Y_i) = \phi_i V(\mu_i)$, onde ϕ_i é o parâmetro de dispersão e $V(.)$ é a função de variância nos MLGs. Uma vez que a modelagem conjunta da média e dispersão é baseada na quase-verossimilhança estendida, o conhecimento completo da distribuição de probabilidade da variável de resposta não é necessário.

Os modelos da média e da dispersão são construídos da seguinte maneira. Suponha que $\boldsymbol{t}^t = (t_1, \ldots, t_s)$ e $\boldsymbol{u}^t = (u_1, \ldots, u_r)$ sejam os vetores das variáveis independentes que afetam os modelos da média e da dispersão, respectivamente. As variáveis independentes para o modelo da dispersão são comumente, mas não necessariamente, um subconjunto das variáveis independentes para o modelo média. No caso de experimentos com misturas, os vetores $\boldsymbol{t}$ e $\boldsymbol{u}$ podem conter componentes de misturas, variáveis de processo ou ambos.

Seja φ a função de ligação para o modelo da média, isto é, para a i-ésima resposta, $\eta_i = \varphi(\mu_i) = \boldsymbol{f}^t(\boldsymbol{t}_i)\boldsymbol{\beta}$ com $\boldsymbol{f}^t(\boldsymbol{t}_i) = (f_1(\boldsymbol{t}_i), \ldots, f_p(\boldsymbol{t}_i))$ onde $f_j(\boldsymbol{t}_i)$, for $j = 1, \ldots, p$, é uma função conhecida de $\boldsymbol{t}_i$ e $\boldsymbol{\beta}$ é um vetor $p \times 1$ de parâmetros desconhecidos. Para o modelo da dispersão é utilizado como variável resposta o componente do desvio, que, para cada observação y_i, é dado por $d_i = 2\int_{\mu_i}^{y_i} \frac{y_i - l}{V(l)} dl$, veja McCullagh e Nelder [7], p. 360.

De acordo com Lee e Nelder [6], para o modelo da dispersão é assumido uma distribuição Gama com uma função de ligação logarítmica, ou seja, para a i-ésima resposta, $\zeta_i = \log(\phi_i) = \boldsymbol{g}^t(\boldsymbol{u}_i)\boldsymbol{\gamma}$, com $\boldsymbol{g}^t(\boldsymbol{u}_i) = (g_1(\boldsymbol{u}_i), \ldots, g_q(\boldsymbol{u}_i))$, onde $g_j(\boldsymbol{u}_i)$, para $j = 1, \ldots, q$, é uma função conhecida de $\boldsymbol{u}_i$ e $\boldsymbol{\gamma}$ é um vetor $q \times 1$ de parâmetros desconhecidos. Também definimos $\boldsymbol{T} = [\boldsymbol{f}(\boldsymbol{t}_1), \ldots, \boldsymbol{f}(\boldsymbol{t}_n)]^t$ a matriz experimental $n \times p$ para o modelo da média e $\boldsymbol{U} = [\boldsymbol{g}(\boldsymbol{u}_1), \ldots, \boldsymbol{g}(\boldsymbol{u}_n)]^t$ a matriz experimental $n \times q$ para o modelo da dispersão. Veja Pinto e Ponce de Leon [12].

A metodologia para a MCMD tem por base a função de quase verossimilhança estendida proposta por Nelder e Pregibon [10]. O método de estimação para obtenção dos parâmetros foi melhorado por Lee e Nelder [6] com a introdução da quase verossimilhança estendida ajustada. O processo de seleção de variáveis proposto por Pinto e Pereira [15] permitiu uma melhor utilização da MCMD, possibilitando encontrar de forma eficiente os termos no modelo da média e da dispersão.

O processo de seleção de variáveis na MCMD para experimentos de mistura foi proposto por Pinto e Pereira [14] e constitui uma extensão do procedimento proposto por por Pinto e Pereira [15] para o caso de experimentos com misturas. O procedimento de seleção de variáveis proposto por Pinto e Pereira [15] é baseado em testes de hipóteses e na qualidade do ajuste do modelo. Para verificação da qualidade do ajuste do modelo foram considerados os seguintes critérios: $\tilde{R}^2_m$ e $\tilde{R}^2_d$, propostos por Pinto e Pereira [15] para a MCMD; o critério de informação de Akaike estendido ($EAIC$), proposto por Wang e Zhang [21] e o critério de Akaike corrigido (AIC_c), veja Hurvich e Tsai [5]. Desta forma, um critério de verificação da qualidade do ajuste é utilizado, em cada iteração do processo de seleção, como um filtro para a escolha dos termos que serão avaliados por um teste de hipótese. A estratégia de seleção consiste em um esquema de três etapas em que o modelo da dispersão selecionado é utilizado para selecionar o melhor modelo da média e vice-versa. O esquema de seleção utiliza um procedimento recursivo que só é finalizado quando o valor do critério, considerado como medida de seleção para o modelo da média, deixa de melhorar.

O procedimento de estimação e seleção de variáveis pode ser obtido usando a função *stepjglm* do R, criada por Pereira e Pinto [11]. A função é bastante simples de usar, necessitando que o usuário defina, nessa ordem, o modelo inicial para a média; o nível de significância utilizado para testar a inclusão de termos no modelo da média; o nível de significância para testar a inclusão de termos no modelo da dispersão; o conjunto de dados; a família do modelo linear generalizado para o modelo da média (no caso do modelo da dispersão é assumido distribuição Gama com função de ligação logarítmica); a medida de qualidade do ajuste para o modelo da média, podendo ser utilizado $\tilde{R}^2_m$ ou $EAIC$; a medida de qualidade de ajuste para o modelo da dispersão, podendo ser utilizado $\tilde{R}^2_d$ ou AIC_c e os efeitos principais das variáveis de mistura. Também é possível verificar os resultados das iterações passo a passo. O exemplo a seguir mostra a aplicação da MCMD, utilizando a função *stepjglm*, a um experimento de mistura com variáveis de ruído.

4 Exemplo

Nesta seção, a MCMD, utilzando a função *stepjglm*, será aplicada a um problema da indústria de alimentos. O problema consistiu de um experimento com três ingredientes de mistura e duas variáveis de ruído; e teve como objetivo investigar e avaliar a qualidade da farinha final, composta por diferentes misturas de farinha de trigo, para produção de pães. Os ingredientes de mistura consistiram de três tipos de farinha de trigo: duas norueguesas, *Tjalve* (x_1) e *Folke* (x_2) e uma americana, *Hard Red Spring* (x_3), que foram consideradas como variáveis de mistura, e dois tipos de variáveis de processo: tempo de mistura (z_1) e tempo de descanso da massa (z_2), consideradas como variáveis de ruído. A variável resposta foi considerada como o volume do pão após assado, com valor-alvo de 530 ml. As restrições nos ingredientes de mistura foram $0,25 \leq x_1 \leq 1,0$; $0 \leq x_2 \leq 0,75$ e $0 \leq x_3 \leq 0,75$. Para as variáveis de ruído, foram consideradas três situações para o tempo de mistura: 5, 15 e 25 minutos e três situações para o tempo de descanso da massa: 35, 47,5 e 60 minutos.

Um desenho fatorial completo 3^2 foi usado para as variáveis de ruído e as 10 corridas experimentais, correspondente a um planejamento em rede simplex (*simplex lattice design*), foram replicadas em cada uma das nove combinações dos tempos de mistura (tempo em que os ingredientes são misturados) e de descanso (tempo de fermentação da massa), de modo que o planejamento completo envolveu $n = 90$ corridas experimentais. As variáveis de ruído foram codificadas como $z_1 = (\text{tempo de mistura} - 15)/10$ e $z_2 = (\text{tempo de descanso} - 47.5)/12.5$, fornecendo as seguintes codificações: -1, 0, 1, de acordo com seus valores mínimo, médio e máximo. Para maiores detalhes sobre o experimento, veja [8].

Para a aplicação do processo de estimação e seleção de variáveis, consideramos $V(\mu) = 1$ e função de ligação identidade para o modelo da média. Para o modelo de dispersão consideramos a distribuição Gama com função de ligação logarítmica. Os termos considerados em ambos os modelos da média e da dispersão consistiram dos termos do modelo cúbico nos componentes de mistura cruzado com o modelo quadrático nas variáveis de ruído. (Veja Pinto e Pereira [14]).

A medida de seleção da qualidade do ajuste, considerada para o modelo da média, foi o $\tilde{R}^2_m$ com medida de penalização $\lambda_n = \sqrt{n}$ e para o modelo da dispersão a medida de seleção foi o AIC_c. Para maiores detalhes, veja Pinto e Pereira [14]. Desta forma, a função *stepjglm* teve os seguintes argumentos:
$objeto = stepjglm(modelo, 0.1, 0.1, Dados, gaussian, sqrt(90), \text{“}AIC\text{”}, \text{“}-1 + x1 + x2 + x3\text{”})$,
onde $modelo = as.formula(y \sim x1 : x2 + x1 : x3 + x2 : x3 + x1 : x2 : (x1 - x2) + x1 : x3 : (x1 - x3) + +x1 : z1 + x2 : z1 + x3 : z1 + x1 : x2 : z1 + x1 : x3 : z1 + x1 : x2 : (x1 - x2) : z1 + x1 : x3 : (x1 - x3) : z1 + x1 : z2 + x2 : z2 + x3 : z2 + x1 : x2 : z2 + x1 : x3 : z2 + x1 : x2 : (x1 - x2) : z2 + x1 : x3 : (x1 - x3) : z2)$.

O resultado da função *stepjglm* foi o seguinte: $E(Y|Z_1, Z_2) = 488,961x_1 + 432,21x_2 + 574,124x_3 + 174,216x_1x_3Z_1 + (56,621x_1 + 35,904x_2 + 79,146x_3)Z_2$ e $\log(\phi) = 6,9984x_1 + 5,94x_2 + 7,325x_3 - 7,9662x_2x_3$. Desta forma, considerando Z_1 e Z_2 variáveis aleatórias independentes com $E(Z_i) = \mu_i$ e $Var(Z_i) = \sigma_i^2$, para $i = 1, 2$, o modelo da média, incondicional às variáveis de ruído, é obtido como $E(Y) = E[E(Y|Z_1, Z_2)]$, e dado por

$$E(Y) = 488,961x_1 + 432,21x_2 + 574,124x_3 + 174,216x_1x_3\mu_1 + (56,621x_1 + 35,904x_2 + 79,146x_3)\mu_2.$$

O modelo da variância, incondicional às variáveis de ruído, é obtido como $Var(Y) = E[Var(Y|Z_1, Z_2)] + Var[E(Y|Z_1, Z_2)]$. Assim, como $Var(Y|Z_1, Z_2) = \phi V(\mu)$ e como supomos que $V(\mu) = 1$ e conhecemos o modelo para $\log(\phi)$, temos que $Var(Y|Z_1, Z_2) = \exp\{6,9984x_1 + 5,94x_2 + 7,325x_3 - 7,9662x_2x_3\}$. Desta forma, como $Var[E(Y|Z_1, Z_2)] = (174,216x_1x_3)^2\sigma_1^2 + (56,621x_1 + 35,904x_2 + 79,146x_3)^2\sigma_2^2$ e $E[Var(Y|Z_1, Z_2)] = \exp\{6,9984x_1 + 5,94x_2 + 7,325x_3 - 7,9662x_2x_3\}$ obtemos que

$$Var(Y) = \exp\{6,9984x_1 + 5,94x_2 + 7,325x_3 - 7,9662x_2x_3\} + (174,216x_1x_3)^2\sigma_1^2 + (56,621x_1 + 35,904x_2 + 79,146x_3)^2\sigma_2^2.$$

Seguindo as ideias de Taguchi para a melhoria da qualidade, veja Taguchi [20], após encontrarmos as equações para $E(Y)$ e $Var(Y)$, podemos considerar o problema de minimizar a variância sujeito às seguintes condições: i) média para a variável resposta (valor alvo) igual a 530 ml (volume do pão após assado); ii) condição de mistura e iii) restrições nas variáveis de mistura. Além disso, como o problema de otimização também envolve a média e a variância das variáveis de ruído, alguns cenários ou prioris envolvendo esses parâmetros deverão ser considerados, para maiores detalhes veja [4].

5 Considerações finais

A aplicação da MCMD em problemas com misturas é um resultado novo, embora a teoria da MCMD já exista há algum tempo, não existe aplicação alguma em problemas envolvendo misturas. O procedimento proposto, utilizando a função *stepjglm*, é bastante inovador, pois além de fornecer como resultado final os modelos ótimos para a média e para dispersão, permite verificar também se o modelo considerado para a variável resposta tem dispersão constante ou não.

A utilização da quase-verossimilhança estendida na MCMD, permite que os modelos, bem como os desenhos experimentais ótimos, sejam obtidos sem o conhecimento completo da verossimilhança, necessitando somente do conhecimento da relação entre a média e a variância da resposta. Além disso, a MCMD permite que a dispersão seja modelada de forma muito eficaz em problemas de planejamento robusto, mesmo sem replicação, e o modelo da variância pode ser obtido de uma forma simples a partir do modelo de dispersão. Para um entendimento detalhado da modelagem da variabilidade em experimentos com misturas, incluindo planejamento ótimo de experimentos, veja Pinto [13].

Os resultados apresentados neste artigo através da aplicação da MCMD em experimentos com misturas são bastante animadores, indicando que a teoria promete ser de grande utilidade na modelagem e na condução de experimentos industriais envolvendo misturas. Contudo, ainda existe pouco entrosamento entre os pesquisadores acadêmicos e os profissionais da indústria. Acreditamos que a função *stepjglm*, desenvolvida por Pereira e Pinto [11], poderá contribuir para que a teoria desenvolvida por Pinto e Pereira [15] e Pinto e Pereira [14], não fique restrita somente ao meio acadêmico e se torne parte importante do planejamento industrial.

Agradecimentos

À FAPEMIG, pelo apoio financeiro - Projetos APQ-03365-18 e RED-00133-21.

Referências

[1] S Boyd e L Vandenberghe. **Convex Optimization**. Cambridge: Cambridge University Press, 2004. ISBN: 9780521833783.

[2] J A Cornell. **Experiments with mixtures: designs, models, and the analysis of mixture data**. 3ª ed. New York: Wiley, 2002. ISBN: 9780471393672.

[3] H Goldfarb, C Borror e D Montgomery. "Mixture-process variable experiments with noise variables". Em: **Journal of Quality Technology** 35 (2003), pp. 393–405. DOI: 10.1080/00224065.2003.11980237.

[4] D Granato, V Calado e E R Pinto. "Mathematical and Statistical Applications in Food Engineering". Em: **Optimization of food processes using mixture experiments - some applications**. Ed. por S Sevda e A Singh. 1ª ed. Boca Raton: CRC Press, 2020. Cap. 3, pp. 21–35. ISBN: 139781138347670.

[5] C M Hurvich e C L Tsai. “Regression and time series model selection in small samples”. Em: **Biometrika** 76 (1989), pp. 297–307. DOI: 10.1093/biomet/76.2.297.

[6] Y Lee e J A Nelder. “Generalized linear models for analysis of quality improvement experiments”. Em: **The Canadian Journal of Statistics** 26 (1998), pp. 95–105. DOI: 10.2307/3315676.

[7] P McCullagh e J A Nelder. **Generalized Linear Models**. 2ª ed. London: Chapman & Hall, 1989. ISBN: 9780412317606.

[8] T Naes, E M Faergestad e J A Cornell. “A comparison of methods for analyzing data from a three component mixture experiment in the presence of variation created by two process variables”. Em: **Chemometrics and Inteligence Laboratory Systems** 41 (1998), pp. 221–235. DOI: 10.1016/S0169-7439(98)00056-2.

[9] J A Nelder e Y Lee. “Generalized linear models for the analysis of Taguchi-type experiments”. Em: **Applied Stochastic Models and Data Analysis** 7 (1991), pp. 107–120. DOI: 10.1002/asm.3150070110.

[10] J A Nelder e D Pregibon. “An extended quasi-likelihood function”. Em: **Biometrika** 74 (1987), pp. 221–232. DOI: 10.2307/2336136.

[11] L. A. Pereira e E. P. Pinto. **stepjglm: variable selection for joint modeling of mean and dispersion**. R package version 0.0.1, https://cran.r-project.org/package=stepjglm, acessado em 30/03/2023. 2021.

[12] E R Pinto e A Ponce de Leon. “Modelagem conjunta da média e dispersão de Nelder e Lee como alternativa aos métodos de Taguchi”. Em: **Pesquisa Operacional** 26 (2006), pp. 203–224. DOI: 10.1590/S0101-74382006000200002.

[13] E. R. Pinto. “Modelagem da variabilidade em experimentos com misturas”. Tese para Professor Titular na área de Estatística. Tese de doutorado. Universidade Federal de Uberlândia, 2021. DOI: 10.14393/ufu.te.2021.5536.

[14] E. R. Pinto e L. A. Pereira. “Estimation and variable selection in joint mean and dispersion models applied to mixture experiments”. Em: **Chemometrics and Intelligent Laboratory Systems** 227 (2022), p. 104590. DOI: DOI:10.1016/j.chemolab.2022.104590.

[15] E. R. Pinto e L. A. Pereira. “On variable selection in joint modeling of mean and dispersion”. Em: **Brazilian Journal of Probability and Statistics** 35.4 (2021), pp. 875–894. DOI: DOI:10.1214/21-BJPS512.

[16] R Core Team. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing. Vienna, Austria, 2019. URL: http://www.R-project.org/.

[17] H Scheffé. “Experiments with Mixture”. Em: **Journal of the Royal Statistical Society B** 20 (1958), pp. 344–366. DOI: jstor.org/stable/2983895.

[18] W F Smith. **Experimental design for formulation**. New York: Society for Industrial e Applied Mathematics - SIAM, 2005. ISBN: 0898715806.

[19] S H Steiner e M Hamada. “Making mixtures robust to noise and mixing measurement errors”. Em: **Journal of Quality Technology** 29 (1997), pp. 441–450. DOI: 10.1080/00224065.1997.11979795.

[20] G Taguchi. **Introduction to Quality Engineering**. New York: Unipub-Kraus International Publications, White Plains, 1986. ISBN: 978283310846.

[21] D Wang e Z Zhang. “Variable selection in joint generalized linear models”. Em: **Chinese Journal of Applied Probability and Statistics** 25 (2009), pp. 245–256.

