

MODELAGEM CONJUNTA DA MÉDIA E DISPERSÃO DE NELDER E LEE COMO ALTERNATIVA AOS MÉTODOS DE TAGUCHI

Edmilson Rodrigues Pinto *

Faculdade de Matemática
Universidade Federal de Uberlândia (UFU)
Uberlândia – MG
edmilson@famat.ufu.br

Antônio C. M. Ponce de Leon

Instituto de Medicina Social
Universidade do Estado do Rio de Janeiro (UERJ)
Rio de Janeiro – RJ
ponce@ims.uerj.br

* *Corresponding author* / autor para quem as correspondências devem ser encaminhadas

Recebido em 12/2003; aceito em 05/2006 após 2 revisões
Received December 2003; accepted May 2006 after 2 revisions

Resumo

Nos últimos anos, uma coleção de técnicas para a melhoria da qualidade foi desenvolvida, no Japão, por Genichi Taguchi no planejamento, na evolução e na fabricação de produtos industrializados. Esta revolução no campo da qualidade despertou o interesse, no mundo inteiro, de vários engenheiros e estatísticos, que propuseram vários métodos alternativos, mais claros e eficientes do que aqueles propostos por Taguchi. Nelder e Lee observaram que a metodologia de Taguchi conduz a um modelo conjunto para a média e dispersão, usando modelos lineares generalizados. Eles mostraram como esta classe é geral e suficiente para a análise desses modelos. O objetivo deste artigo é fazer uma síntese da modelagem conjunta da média e dispersão, proposta por Nelder e Lee, explicitando, de uma forma concisa, os principais pontos da teoria.

Palavras-chave: modelagem conjunta da média e dispersão; quase verossimilhança estendida; métodos de Taguchi.

Abstract

In the last years, in Japan, Genichi Taguchi developed some techniques to improve the quality of the products during its design, evolution and manufacture. His ideas have raised the attention of engineers and statisticians around the world who have proposed some alternative methods that are simpler and more efficient than Taguchi's methods. Nelder and Lee observed that Taguchi's methodology lead up to a joint model for the mean and dispersion employing generalized linear models. They showed that this class is general and sufficient to analyze Taguchi's Models. The purpose of this work is to describe joint modeling of mean and dispersion that was proposed by Nelder and Lee and the key points of the theory in a concise manner.

Keywords: joint modeling of mean and dispersion; extended quasi-likelihood; Taguchi's methods.

1. Introdução

A intensa competição existente no mercado internacional tem revelado que a qualidade dos produtos é a chave para o sucesso das indústrias. Nos últimos anos Taguchi introduziu uma nova filosofia de qualidade. Esta filosofia tem produzido uma única e poderosa disciplina de melhoramento da qualidade, que difere das práticas tradicionais. A contribuição no campo do controle de qualidade é um dos desenvolvimentos mais significantes das últimas décadas (Phadke, 1989). O objetivo da teoria proposta por Taguchi é o desenvolvimento de produtos que sejam robustos, isto é, que sejam pouco afetados por variações inevitáveis que venham a ocorrer em seus componentes, nas condições de fabricação ou no ambiente em que serão usados. Contrariamente à situação de experimentos clássicos, em que os fatores são fixos durante e no decorrer do experimento, Taguchi introduziu os chamados fatores de ruído. Os fatores de ruído são fixados durante o experimento, mas variam aleatoriamente fora do contexto experimental. A idéia, por trás da metodologia de Taguchi, é minimizar a variabilidade da média de alguma característica de interesse do produto em relação ao seu valor ideal. Essa variabilidade pode ser causada por três tipos de ruídos:

1. *Ruído externo*: causado por condições ambientais, como temperatura e umidade;
2. *Ruído interno*: causado pela deterioração por uso ou por armazenagem;
3. *Ruído de fabricação*: causado por imperfeições no processo de manufatura.

A fim de encontrar um produto que seja insensível aos ruídos, de modo que a média da característica de interesse fique próxima a um valor pré-especificado, os fatores são classificados em dois grupos:

- **Fatores de controle da variabilidade (FCV)**: os quais afetam a variabilidade e, possivelmente, a média da característica de interesse do produto.
- **Fatores de controle do alvo (FCA)**: os quais afetam somente a média da característica de interesse do produto.

A meta dos planejamentos robustos de Taguchi é identificar uma combinação dos níveis de FCV, que minimize a variabilidade, juntamente com uma combinação dos níveis de FCA, que assegure a média da característica de interesse próxima a um valor ideal. FCA são chamados fatores de ajustamento, pois são usados para ajustar a média da característica de interesse do produto próxima a um valor alvo, depois que uma combinação ótima de FCV já tiver sido escolhida.

Ninguém discorda ser de fundamental importância a metodologia proposta por Taguchi, para se ter uma excelente prática industrial (Guedes, 1996). Entretanto, as técnicas utilizadas por ele vêm sofrendo diversas críticas, e procedimentos alternativos têm sido desenvolvidos. Uma discussão editada por Nair (1992) dá uma visão geral sobre esta controvérsia.

Nelder & Lee (1991) introduziram uma modelagem conjunta da média e dispersão, usando Modelos Lineares Generalizados (MLG), com o intuito de dar uma abordagem geral ao tema proposto por Taguchi, baseada em uma sólida teoria estatística. Nesta abordagem são necessários dois MLG interligados, um para a média, outro para a dispersão.

A modelagem conjunta da média e dispersão é um assunto recente e tem atraído a atenção de muitos pesquisadores. Dentre os principais trabalhos nesta área podemos citar: Aitkin (1987), que considerou modelos gaussianos e empregou um modelo de regressão log-linear para a variância; Nelder & Pregibon (1987), que introduziram a quase verossimilhança estendida;

Smyth (1989), que obteve expressões para as equações de verossimilhança, testes assintóticos e algoritmos de estimação para MLG com parâmetro de dispersão variável; Nelder & Lee (1991), que introduziram a modelagem conjunta da média e dispersão para os modelos de Taguchi, usando a quase verossimilhança estendida; Verbyla (1993), que considerou modelos normais heterocedásticos, usando a máxima verossimilhança restrita, e para os quais a variância possuía uma dependência log-linear em relação às covariáveis; Engel & Huele (1996), que propuseram a modelagem simultânea da superfície de resposta para a média e para a variância, usando MLG sob a hipótese de erros normais; Smyth & Verbyla (1996), que forneceram uma abordagem de verossimilhança condicional, para a máxima verossimilhança restrita em MLG com parâmetro de dispersão variável e ligação canônica; Dey, Gelfand & Peng (1997), que consideraram MLG com super-dispersão, usando a família exponencial dupla, e uma abordagem bayesiana; Lee & Nelder (1998), que introduziram o ajustamento para a máxima verossimilhança restrita na análise de dispersão; Smyth & Verbyla (1999), que introduziram métodos ajustados para a verossimilhança na modelagem da dispersão em MLG; Lee & Nelder (2000), que verificaram o relacionamento entre a família exponencial dupla e a quase verossimilhança estendida, e mostraram que elas geram inferências idênticas; Smyth, Huele & Verbyla (2001), que compararam a máxima verossimilhança restrita exata e aproximada para modelos de regressão heterocedásticos, usando os métodos propostos por Lee & Nelder (1998) e Smyth & Verbyla (1999); Cuervo (2001), que introduziu uma abordagem bayesiana à metodologia de modelagem conjunta da média e dispersão; Smyth (2002), que propôs um algoritmo para a máxima verossimilhança restrita em modelos de regressão heterocedásticos; Lee & Nelder (2003), que fizeram uma explanação sobre planejamentos robustos, usando transformação para os dados e usando MLG; e que também mostraram que MLG fornecem uma extensão para a metodologia de superfície de resposta. Neste artigo usaremos a abordagem proposta por Lee & Nelder (1998).

Na seção 2, é apresentada uma breve revisão sobre os conceitos básicos de MLG e sobre Modelagem Conjunta da Média e Dispersão (MCMD). Na seção 3, é mostrado como os problemas propostos por Taguchi podem ser resolvidos usando a MCMD. Na seção 4, a teoria, apresentada nas seções 2 e 3, é ilustrada através de dois exemplos. No primeiro exemplo foram considerados dados contínuos, enquanto que no segundo foram considerados dados de contagem. Finalmente, na seção 5, são feitas considerações finais sobre a MCMD.

2. Modelos Lineares Generalizados

A classe de modelos lineares generalizados é uma extensão da classe de modelos lineares, que permite que os erros venham de várias distribuições de probabilidade ao invés de somente da distribuição normal. Esta classe de distribuições de probabilidade, conhecida pelos estatísticos como família exponencial de distribuições, inclui algumas das principais distribuições de probabilidade como: normal, Poisson, binomial, multinomial, Gama, binomial negativa e a inversa gaussiana. MLG geram uma escala aditiva para o componente sistemático do modelo, proveniente das variáveis explicativas. A escala, na qual os efeitos são assumidos aditivos, é relacionada à média da distribuição da variável aleatória por uma função de ligação. Usando MLG não há necessidade de transformar os dados para obter aditividade. Muitas características dos modelos lineares clássicos são imediatamente estendidas com MLG, entre elas a estrutura do preditor linear, a tabela ANOVA e as idéias de diagnósticos de modelos por meio de análise de resíduos, efeito de alavanca, influência, e

assim por diante. Além disso, um algoritmo razoavelmente simples, uma versão do *Algoritmo Iterativo de Mínimos Quadrados Ponderados*, ajusta todos os MLG.

Sejam y_1, \dots, y_n , n observações independentes da variável resposta Y , e sejam x_1, x_2, \dots, x_k , k covariáveis, cujos efeitos combinam aditivamente para produzir um preditor linear $\eta = \sum_{j=1}^k x_j \beta_j$, onde $\beta_1, \beta_2, \dots, \beta_k$ representam os parâmetros desconhecidos do modelo.

De acordo com McCullagh & Nelder (1989), um MLG é caracterizado por três componentes:

1. O componente aleatório, cuja esperança é denotada por $E(Y) = \mu$;
2. O componente sistemático $\eta = \sum_{j=1}^k x_j \beta_j$;
3. A ligação entre os componentes aleatório e sistemático, dada pela função $\eta = g(\mu)$, onde $g(\cdot)$ é qualquer função monótona diferenciável. A função $g(\cdot)$ é conhecida como função de ligação do MLG e define a escala na qual os efeitos das covariáveis são assumidos serem aditivos.

O logaritmo da função de verossimilhança para um conjunto de y_1, \dots, y_n observações independentes é

$$l(\theta, \phi; \mathbf{y}) = \sum_{i=1}^n \ln \left[\exp \left\{ \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \theta) \right\} \right] = \sum_{i=1}^n \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi). \quad (1)$$

Na equação (1) θ é o parâmetro canônico, $a(\phi)$ tem a forma de ϕ/m , onde ϕ é o parâmetro de dispersão e m é o peso *a priori*. Baseado em dois resultados bem conhecidos,

i.e. $E\left(\frac{\partial l}{\partial \theta}\right) = 0$ e $E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0$, pode-se mostrar que $E(Y) = \mu = \frac{\partial b(\theta)}{\partial \theta} = b'(\theta)$

e $V(\mu) = \frac{\partial \mu}{\partial \theta} = b''(\theta)$, onde $V(\mu)$ é a função de variância. Uma propriedade importante dos MLG é a forma da variância, isto é, $Var(Y) = \phi V(\mu)$, ou seja, a variância é caracterizada por dois componentes: ϕ , a parte independente da média e $V(\mu)$, que descreve como a variância varia com a média.

Para uma observação y_i e considerando ϕ fixado, as equações de máxima verossimilhança para β_j , com $j = 1, \dots, k$, são independentes de ϕ e dadas por:

$$\sum_{i=1}^n w_i (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i} x_{ij} = 0 \quad (2)$$

onde, $w_i = \left[\left(\frac{\partial \eta_i}{\partial \mu_i} \right)^2 V(\mu_i) \right]^{-1}$ é o peso relacionado à i -ésima observação. O conjunto de equações (2) pode ser ajustado pelo *Método Iterativo de Mínimos Quadrados Ponderados*,

usando a variável dependente Z , cujos valores são iguais a $z_i = \eta_i + (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i}$ e cujos respectivos pesos são w_i .

Uma medida de discrepância para os MLG, que generaliza a soma de quadrados dos resíduos do modelo normal, é a função desvio (*deviance*), D , dada pela expressão:

$$D = -2\phi\{l(\hat{\boldsymbol{\mu}}, \phi; \mathbf{y}) - l(\mathbf{y}, \phi; \mathbf{y})\}, \quad (3)$$

onde, $l(\hat{\boldsymbol{\mu}}, \phi; \mathbf{y})$, expresso em função de $\hat{\boldsymbol{\mu}}$, é o máximo do logaritmo da função de verossimilhança, com média $\boldsymbol{\mu}$ e parâmetro de dispersão ϕ fixo; e $l(\mathbf{y}, \phi; \mathbf{y})$ é o máximo da função logarítmica da verossimilhança para o modelo completo, isto é, com n parâmetros, onde n é o número de observações. (Para maiores detalhes, veja McCullagh & Nelder, 1989, pág. 33). D é uma função somente das observações, e para observações independentes a equação (3) pode ser escrita simplesmente como $D = \sum_{i=1}^n d_i$; onde

$$d_i = -2 \int_{y_i}^{\mu_i} \frac{y_i - t}{V(t)} dt \quad (4)$$

Na equação (4) d_i é o componente do desvio para a i -ésima observação e corresponde ao quadrado do resíduo no modelo normal. Portanto, o resíduo do desvio é definido como $r_{D_i} = \text{Sinal}(y_i - \mu_i) \sqrt{d_i}$. Outro tipo de resíduo, frequentemente usado, é o chamado resíduo de Pearson, definido como:

$$r_{P_i} = \frac{y_i - \mu_i}{\sqrt{V(\mu_i)}} \quad (5)$$

2.1 Modelagem conjunta da média e dispersão

Na abordagem de Lee & Nelder (1998), quando média e dispersão são modeladas simultaneamente, dois MLG são necessários, um para a média μ e outro para a dispersão ϕ . Da mesma forma como no modelo da média, o modelo para a dispersão consiste da variável resposta e os três componentes próprios dos MLG. Para a variável resposta pode-se empregar d_i , o componente do desvio (para o *deviance*), dado na equação (4), ou $r_{P_i}^2$, o desvio generalizado de Pearson, com r_{P_i} dado na equação (5); ambos obtidos para cada observação y_i . Sejam u_1, u_2, \dots, u_l as covariáveis que afetam a dispersão, $\zeta = h(\phi)$ a função de ligação, e $\zeta = \sum_{j=1}^l u_j \gamma_j$ o preditor linear, para o qual $\gamma_1, \gamma_2, \dots, \gamma_l$ são os parâmetros desconhecidos. Uma escolha natural para a distribuição do erro é a distribuição Gama (particularmente, quando o componente do desvio é usado como resposta). A função logarítmica é adequada para fazer a ligação entre os componentes sistemático e aleatório no modelo da dispersão, por mapear o intervalo $(0, \infty)$ em $(-\infty, \infty)$. Contudo, outras escolhas também são possíveis.

Considerando y_1, y_2, \dots, y_n n observações independentes da variável resposta Y , x_1, x_2, \dots, x_k as covariáveis que afetam o modelo da média e u_1, u_2, \dots, u_l as covariáveis que afetam o modelo da dispersão. Os modelos para a média e para a dispersão são construídos da seguinte forma.

Modelo para a média (μ_i)

$$E(Y_i) = \mu_i; \quad Var(y_i) = \phi_i V(\mu_i); \quad \eta_i = g(\mu_i) = \sum_{j=1}^k x_{ij} \beta_j.$$

Modelo para a dispersão (ϕ_i)

$$E(d_i) = \phi_i; \quad Var(d_i) = \tau V_D(\phi_i); \quad \zeta_i = h(\phi_i) = \sum_{j=1}^l u_{ij} \gamma_j.$$

No modelo para a dispersão, a média é ϕ , τ representa o parâmetro de dispersão e $V_D(\phi)$ é a função de variância. No caso da distribuição normal para o modelo da média, com $\phi = \sigma^2$ e $V(\mu) = 1$; o modelo da dispersão terá uma distribuição Gama com $\tau = 2$ e $V_D(\phi) = \phi^2$. Veja McCullagh & Nelder (1989).

Sejam X e U as matrizes experimentais dos modelos da média e da dispersão, respectivamente. Note que as covariáveis $\{u_i : i = 1, \dots, l\}$ não são, necessariamente, um subconjunto das covariáveis $\{x_i : i = 1, \dots, k\}$. Desta forma, os fatores experimentais podem ocorrer em X , em U ou em ambas. Usando os fatores experimentais ocorrendo em U (ou em ambas X ou U), procura-se um conjunto de fatores e níveis que minimizem a variância. Usando os fatores ocorrendo somente em X , pode-se ajustar a média próxima a um valor preestabelecido, sem afetar a variância.

Para a estimação conjunta dos parâmetros da média e da dispersão Nelder & Lee (1991) propuseram o uso da *Quase Verossimilhança Estendida (QVE)*. (Introduzida por Nelder & Pregibon, 1987).

2.2 A quase verossimilhança e quase verossimilhança estendida

Wedderburn (1974) propôs a *Quase Verossimilhança (QV)* quando, dado $V(\mu)$, não existe uma distribuição na família exponencial para a variável resposta. A *QV*, definida por:

$$Q(y, \mu) = \int_y^\mu \frac{y-t}{\phi V(t)} dt \quad \text{ou} \quad \frac{\partial Q(y, \mu)}{\partial \mu} = \frac{y-\mu}{\phi V(\mu)}$$

é usada, freqüentemente, para fazer inferência no modelo da média (utilizando ϕ como constante), mas não pode ser usada para fazer inferência no modelo da dispersão.

Supondo que a distribuição de Y seja pertencente à família exponencial, o logaritmo da função de quase verossimilhança é uma função, cuja derivada em relação à média μ , é igual à mesma derivada do logaritmo da função de verossimilhança, isto é, $\frac{\partial Q(y, \mu)}{\partial \mu} = \frac{\partial l(y, \mu)}{\partial \mu}$.

A QV pode ser entendida como uma forma de definir, aproximadamente, distribuições na família exponencial, quando a função de variância não permite uma forma exata. A QV será uma verdadeira verossimilhança, se a distribuição da variável resposta Y pertencer à família exponencial e se $Var(Y) = \phi V(\mu)$. Seu uso como um critério de ajustamento, permite estender a classe de MLG para modelos definidos somente pelas propriedades dos dois primeiros momentos.

Quase verossimilhanças permitem dois tipos de extensões de MLG. Na primeira, MLG com $\phi = 1$, fixo, podem ser estendidos para admitirem ϕ variável; por exemplo, os modelos log-lineares de Poisson, para os quais $Var(Y) = \mu$, podem ser expandidos para admitirem superdispersão com $Var(Y) = \phi\mu$ e $\phi > 1$. Na segunda extensão, $V(\mu)$ pode tomar uma forma que não corresponde àquela, própria de um MLG padrão, por exemplo, $V(\mu) = \mu^\alpha$, com α variável e $\alpha \neq 0, 1, 2, 3$. (Nelder & Lee, 1991).

A quase verossimilhança tem as mesmas equações de estimação que os MLG, gerando estimativas de máxima quase verossimilhança no lugar de estimativas de máxima verossimilhança; e também produzindo um desvio (*deviance*) e um resíduo de Pearson. Os modelos de quase verossimilhanças são muito úteis para modelar quantidades e proporções superdispersas.

McCullagh (1983) mostrou que o estimador de quase verossimilhança, $\hat{\beta}$, para o vetor de parâmetros, β , do modelo, é assintoticamente normal com média β , e que as covariâncias assintóticas podem ser obtidas na forma usual, através da segunda derivada de Q . Adicionalmente, se H_0 e H_a são as hipóteses associadas a dois modelos encaixados de dimensões p e q , respectivamente, com $p < q$, então, sob H_0 , a mudança no quase desvio $D(\hat{\mu}_a, \hat{\mu}_0) = D(y, \hat{\mu}_0) - D(y, \hat{\mu}_a)$ tem uma distribuição assintótica χ_{q-p}^2 ; onde $D(y, \mu) = -2\phi Q(y, \mu)$.

Em MLG a função desvio é usada para medir a discrepância de um ajuste, e também pode ser usada para comparar modelos com diferentes preditores lineares e/ou funções de ligação. A função desvio não pode, contudo, ser usada para comparar modelos com diferentes funções de variância ou diferentes estruturas de dispersão, que aparecem na modelagem conjunta da média e dispersão. Um critério apropriado para este tipo de problema é a *Quase Verossimilhança Estendida* (QVE), Q^+ .

A função Q^+ , para uma simples observação y_i , com média μ_i e variância $Var(Y_i) = \phi_i V(\mu_i)$, é definida por:

$$Q_i^+(y_i, \mu_i) = -\frac{1}{2} \frac{d_i(y_i, \mu_i)}{\phi_i} - \frac{1}{2} \ln[2\pi\phi_i V(y_i)],$$

onde $d_i(y_i, \mu_i)$ é dado em (4). Para y_1, y_2, \dots, y_n , n observações independentes da variável resposta Y , o quase desvio estendido, QD^+ , é dado por:

$$QD^+ = -2 \sum_{i=1}^n Q_i^+ = \sum_{i=1}^n \frac{d_i}{\phi_i} + \sum_{i=1}^n \ln[2\pi\phi_i V(y_i)].$$

Note que Q^+ , assim como Q , não pressupõe uma distribuição completa para Y , mas somente o tipo de relação existente entre a média e a variância das observações. As estimativas dos parâmetros, obtidas pela maximização de Q^+ , são iguais àquelas obtidas maximizando Q (Nelder & Pregibon, 1987).

Quando existe uma distribuição na família exponencial, com uma dada função de variância, a QVE é uma aproximação para o ponto de sela desta distribuição. Para maiores detalhes veja Jorgensen (1987) e (1992).

A quase verossimilhança requer o conhecimento da função de variância multiplicada por uma constante. Usando Q^+ , este requerimento pode ser relaxado. Em certas aplicações ϕ pode não ser constante e o interesse pode residir em modelar ϕ como uma função de covariáveis conhecidas. Neste caso, a quase verossimilhança estendida fornece uma estrutura sobre a qual semelhante análise pode ser feita. Pode-se considerar também a variância pertencente a uma família de funções indexadas por um parâmetro desconhecido λ . Uma família, muito útil, é obtida considerando potências de $\mu : V_\lambda(\mu) = \mu^\lambda$ (Nelder & Pregibon, 1987). Os valores mais comuns de λ são: 0,1,2 e 3; os quais correspondem às funções de variâncias associadas com as distribuições normal, Poisson, Gama e inversa Gaussiana, respectivamente. Para essa família de funções de variância, o desvio é dado por:

$$d_\lambda(y; \mu) = \begin{cases} 2\{y \ln(y/\mu) - (y - \mu)\} & \lambda = 1 \\ 2\{y/\mu - \ln(y/\mu) - 1\} & \lambda = 2 \\ \frac{2\{y^{2-\lambda} - (2-\lambda)y\mu^{1-\lambda} + (1-\lambda)\mu^{2-\lambda}\}}{(1-\lambda)(2-\lambda)} & \text{caso contrário} \end{cases}$$

Outro tipo de função de variância que pode ser usada é a família $V_\lambda(\mu) = \mu^\lambda(1-\mu)^\lambda$ (Nair & Pregibon, 1988), onde para $\lambda = 0$ tem-se a variância constante e para $\lambda = 1$ tem-se uma função de variância do tipo binomial.

Para um dado ϕ_i , a menos de uma constante, a QVE é a QV para um modelo com a função de variância $V(\mu_i)$. Desta forma, maximizando Q^+ , com respeito ao vetor de parâmetros β , dará os mesmos estimadores da QV, com pesos $1/\phi_i$, satisfazendo:

$$\frac{\partial Q^+}{\partial \beta_j} = \sum_{i=1}^n \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0, \text{ para } j = 1, \dots, p, \text{ onde } p \text{ é o número de parâmetros no modelo da}$$

média. A QVE fornece um desvio padronizado, o qual pode ser usado como uma medida de discrepância (Lee & Nelder, 1998).

Agora, para um dado μ_i , a QVE toma a forma de uma verossimilhança para a distribuição Gama, com variável resposta d_i . Desta forma: $E(d_i) = \phi_i$ e $Var(d_i) = 2\phi_i^2$. A justificativa para fixar o modelo para a dispersão como Gama é que o desvio tem uma distribuição próxima da distribuição Gama até mesmo quando Y não tem uma distribuição normal (Nair, 1992). Note que a distribuição para dispersão é Gama exata se Y tem uma distribuição normal. No modelo da dispersão é comum tomar a função de ligação logarítmica. Analogamente ao

modelo da média, maximizando Q^+ , com respeito ao vetor de parâmetros γ , as equações de estimação desses parâmetros serão dadas por $\frac{\partial Q^+}{\partial \gamma_j} = \sum_{i=1}^n \frac{d_i - \phi_i}{\phi_i^2} \frac{\partial \phi_i}{\partial \gamma_j} = 0$, para $j = 1, \dots, q$,

onde q é o número de parâmetros no modelo da dispersão. Lee & Nelder (1998) sugeriram o ajustamento da máxima verossimilhança restrita em experimentos altamente fracionados, onde o número de parâmetros no modelo da média é uma grande fração do tamanho da amostra. Para esses experimentos os estimadores dos parâmetros do modelo da dispersão podem ser altamente viesados.

A técnica de máxima verossimilhança restrita, desenvolvida para fazer o ajustamento de graus de liberdade para modelos lineares mistos, foi estendida por Cox & Reid (1987) para uma ampla classe de modelos através de uma QVE-ajustada. Lee & Nelder (1998) afirmam que a QVE-ajustada, proposta por Cox & Reid (1987), envolve uma pesada computação com baixa convergência. Assim, eles sugerem estimar γ minimizando o quase desvio estendido ajustado, QD^{+A} , dado por:

$$QD^{+A} = \sum_{i=1}^n \frac{d_i^*}{\phi_i} + \sum_{i=1}^n \ln \{2\pi\phi_i V(y_i)\},$$

onde, $d_i^* = \frac{d_i}{(1-h_i)}$, sendo h_i o i -ésimo elemento da diagonal da matriz

$H = W^{1/2} X (X^T W X)^{-1} X W^{1/2}$, a matriz de projeção do estimador de mínimos quadrados ponderados iterativos de β , onde W é uma matriz diagonal com entradas

$w_i = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{\phi_i V(\mu_i)}$. Lee & Nelder (1998) afirmam que a minimização de QD^{+A} é rápida

e pode ser feita ajustando um MLG para o modelo da dispersão, assumido ter distribuição Gama, com d_i^* como resposta. Nesse caso, $E(d_i^*) = \phi_i$ e $Var(d_i^*) = 2\phi_i^2$.

Outros tipos de ajustamentos também são possíveis. McCullagh & Nelder (1989) sugerem ajustamentos para diferentes situações dos modelos. Para uma exposição detalhada da teoria de MLG consulte os livros de Cordeiro (1986), McCullagh & Nelder (1989), Dobson (1990) ou Myers, Montgomery & Vining (2002).

3. Aplicação da Modelagem Conjunta da Média e Dispersão aos Modelos Propostos por Taguchi

Em MLG, ϕ e μ são medidas de desempenho para o ruído e para a média, respectivamente. A dependência funcional entre a média e a variância é eliminada através de uma escolha apropriada da função de variância $V(\mu)$. As medidas de desempenho do alvo, $MDA(\mu)$, e de desempenho do ruído $MDR(\phi)$, são modeladas através de especificações apropriadas para as funções de ligação da média e da dispersão. Neste estágio, o importante é encontrar modelos aditivos simples para a média e para a dispersão.

Box (1988) considera dois critérios para análise de dados, em experimentos para melhoramento da qualidade, como sendo de grande importância, separação e parcimônia. Separação é a eliminação de algumas complicações desnecessárias no modelo, devido à dependência funcional entre a média e a variância; e parcimônia é a provisão de um modelo aditivo mais simples possível. O critério de separação de Box pode ser obtido por MLG através da descoberta de uma função apropriada para a variação dos dados. De maneira similar, a modelagem da variância é generalizada pela modelagem da dispersão. Variância e dispersão são iguais somente para erros normais. O segundo critério de Box, parcimônia, é interpretado como a descoberta de uma função de ligação apropriada para produzir aditividade dos efeitos das covariáveis, juntamente com um conjunto parcimonioso de covariáveis, que têm poder explicativo sobre a variável resposta.

Uma covariável de dispersão pode ou não ser a mesma que uma covariável para a média. Com isto, na formulação do MLG, as duas metas: separação e parcimônia, são interpretadas da seguinte forma:

- Separação: será encontrada escolhendo a função de variância correta para a média, de modo que ϕ fique livre das influências de μ ;
- Parcimônia: será encontrada escolhendo corretamente a função de ligação e o preditor linear para os modelos da média e da dispersão.

Usando a QVE, como critério de otimização, necessita-se conhecer somente as expressões da $E(Y) = \mu$ e da $Var(Y) = \phi V(\mu)$. A Tabela 1 dá um resumo da modelagem conjunta da média e da dispersão.

Tabela 1 – Sumário da modelagem conjunta da média e dispersão.

Componente	Modelo para média	Modelo para dispersão
Resposta	Y	$d^* = d / (1 - h)$
Média	μ	ϕ
Variância	$\phi V(\mu)$	$2\phi^2$
Função de Ligação	$\eta = g(\mu)$ (g qualquer)	$\zeta = \ln(\phi)$ (usualmente \ln)
Preditor Linear	$\eta = X\beta$	$\zeta = U\gamma$
Componente do desvio	$d = -2 \int_y^\mu \frac{y-t}{V(t)} dt$	$2 \left\{ -\ln\left(\frac{d^*}{\phi}\right) + \frac{(d^* - \phi)}{\phi} \right\}$
Peso <i>a priori</i>	$1/\phi$	$1 - h$

Nelder & Lee (1991) sugerem os seguintes passos para a modelagem conjunta da média e dispersão.

- **Passo 01:** Identificação de $V(\mu)$
 Neste estágio o critério de separação deve ser o mais importante. Separação é interpretada como a descoberta de uma função de variância apropriada para a média.

- **Passo 02:** Modelagem conjunta da média e dispersão

Modelos saturados são ajustados para a média e a dispersão; e procura-se por um máximo da quase verossimilhança estendida. Neste estágio o critério de parcimônia é mais importante, e pode ser encontrado, escolhendo, para cada modelo, uma função de ligação apropriada, juntamente com um conjunto parcimonioso de covariáveis no preditor linear.

- **Passo 03:** Verificação do modelo

Verifique o ajuste dos modelos (veja McCullagh & Nelder, 1989). Se a verificação do ajuste é satisfatória, isto é, se os modelos estão bem ajustados, vá ao passo 04. Caso contrário, volte para o passo 01.

- **Passo 04:** Predição dos modelos

Primeiro, minimize a variabilidade, encontrando um conjunto ótimo de níveis dos fatores de controle da variabilidade. Depois ajuste a média próxima ao valor alvo, escolhendo um conjunto apropriado de níveis dos fatores de controle da média.

3.1 Método iterativo para a modelagem conjunta da média e dispersão

- **Modelo para a média**

Sejam y_1, \dots, y_n , n observações independentes da variável resposta $Y^T = (Y_1, \dots, Y_n)$, x_1, x_2, \dots, x_p , as p covariáveis que afetam a média e $\beta_1, \beta_2, \dots, \beta_p$ os parâmetros desconhecidos do modelo. Considere que $\mu^T = (\mu_1, \dots, \mu_n)$, $\phi^T = (\phi_1, \dots, \phi_n)$ e suponha as expressões da $E(Y_i) = \mu_i$ e da $Var(Y_i) = \phi_i V(\mu_i)$ conhecidas. Assuma $k = 1$, $\beta_0^T = (0, \dots, 0)$, $\mu_0^T = (y_1, \dots, y_n)$ e $\phi_0^T = (1, \dots, 1)$. Agora, o método iterativo dos mínimos

quadrados ponderados pode ser usado para obter $\beta_{(j)} = (X^T W_{(j-1)} X)^{-1} X^T W_{(j-1)} z_{(j-1)}$, um vetor $(p \times 1)$, sendo X a matriz de planejamento para o modelo da média, de dimensão $(n \times p)$; $W_{(j-1)} = \text{Diag}(w_{(j-1)1}, \dots, w_{(j-1)n})$, a matriz $(n \times n)$ dos pesos para o MLG, onde $\text{Diag}(w_{(j-1)1}, \dots, w_{(j-1)n})$ representa uma matriz diagonal com as entradas

$$w_{(j-1)i} = \left(\frac{\partial \mu_{(j-1)i}}{\partial \eta_{(j-1)i}} \right)^2 \frac{1}{V(\mu_{(j-1)i})} \text{ na diagonal, e } z_{(j-1)}^T = (z_{(j-1)1}, \dots, z_{(j-1)n}) \text{ um vetor } (n \times 1),$$

$$\text{com } z_{(j-1)i} = \eta_{(j-1)i} + \frac{\partial \eta_{(j-1)i}}{\partial \mu_{(j-1)i}} (y_i - \mu_{(j-1)i}), \text{ para } i = 1, \dots, n \text{ e } j = 1, 2, \dots.$$

Em cada iteração j ($j = 1, 2, \dots$) um novo $\beta_{(j)}$ é obtido e o processo continua até que um critério de convergência seja satisfeito. Um possível critério de convergência pode ser:

$$\|\beta_{(j)} - \beta_{(j-1)}\|^2 < \delta, \text{ onde } \|\cdot\| \text{ representa a norma de um vetor e } \delta \in \mathbb{R}.$$

Após a convergência ter sido encontrada, faça $W_k = W_{(j-1)}$, guarde o último $\beta_{(j)}$ como β_k , e utilize-o para o cálculo do vetor $(n \times 1)$ da média μ_k , isto é, $\mu_k = g_m^{-1}(X \beta_k)$; g_m é uma

função conhecida que possui inversa. Com o valor estimado de μ_k , calcula-se o vetor $(n \times 1)$, d_k^* , com $d_{ki}^* = \frac{d_{ki}}{1-h_{ki}}$, onde $d_{ki} = 2 \int_{\mu_{ki}}^{y_i} \frac{y_i - t}{V(t)} dt$ e h_{ki} é o i -ésimo elemento da diagonal de $H_k = W_k^{1/2} X (X^T W_k X)^{-1} X W_k^{1/2}$. De posse do vetor d_k^* ajuste o modelo da dispersão, considerando pesos iguais a $1-h_{ki}$ para cada valor d_{ki}^* .

• **Modelo para a dispersão**

Dados $d_k^* = (d_{k1}^*, \dots, d_{kn}^*)$ e $(1-h_{k1}, \dots, 1-h_{kn})$. Sejam $\gamma_{(0)}^T = (0, \dots, 0)$, $\phi_0^T = (d_{k1}^*, \dots, d_{kn}^*)$, u_1, u_2, \dots, u_q , as q covariáveis que afetam a dispersão e $\gamma_1, \gamma_2, \dots, \gamma_q$ os parâmetros desconhecidos do modelo. Considerando uma distribuição Gama para o modelo da dispersão e usando o método iterativo de mínimos quadrados, obtém-se $\gamma_{(j)} = (U^T V_{(j-1)} U)^{-1} U^T V_{(j-1)} s_{(j-1)}$, um vetor $(q \times 1)$, sendo U a matriz de planejamento para o modelo da dispersão, de dimensão $(n \times q)$; $V_{(j-1)} = \text{Diag}(v_{(j-1)1}, \dots, v_{(j-1)n})$, a matriz $(n \times n)$ dos pesos para o MLG, representando uma matriz diagonal com as entradas $v_{(j-1)i} = \left(\frac{\partial \phi_{(j-1)i}}{\partial \zeta_{(j-1)i}} \right)^2 \frac{1}{2\phi_{(j-1)i}^2} (1-h_{(j-1)i})$ na diagonal, e $s_{(j-1)}^T = (s_{(j-1)1}, \dots, s_{(j-1)n})$, um vetor $(n \times 1)$, com $s_{(j-1)i} = \zeta_{(j-1)i} + \frac{\partial \zeta_{(j-1)i}}{\partial \phi_{(j-1)i}} (d_{(j-1)i} - \phi_{(j-1)i})$, para $i = 1, \dots, n$ e $j = 1, 2, \dots$.

Da mesma forma como feito no modelo da média, em cada iteração j ($j = 1, 2, \dots$) um novo $\gamma_{(j)}$ é obtido e o processo continua até que um critério de convergência seja satisfeito. Após a convergência, o último $\gamma_{(j)}$ é guardado como γ_k e é utilizado para o cálculo da média ϕ_k , isto é, $\phi_k = g_d^{-1}(U \gamma_k)$, onde g_d é uma função conhecida que possui inversa. No caso do modelo da dispersão, g_d é, geralmente, tomada como a função logarítmica.

Agora, com o valor de $\phi_k = (\phi_{k1}, \dots, \phi_{kn})$ estimado, volta-se para o modelo da média e usa-se, novamente, o método iterativo dos mínimos quadrados ponderados, agora com os novos pesos $\frac{1}{\phi_{ki}}$; desta forma, no modelo da média, para cada $j-1$, as entradas da matriz

diagonal $W_{(j-1)}$ serão $w_{(j-1)i} = \left(\frac{\partial \mu_{(j-1)i}}{\partial \eta_{(j-1)i}} \right)^2 \frac{1}{\phi_{ki} V(\mu_{(j-1)i})}$. Assim, alterna-se entre os modelos

da média e da dispersão até que um critério de convergência seja satisfeito, por exemplo,

$$\frac{|QD_k^{+A} - QD_{k-1}^{+A}|}{|QD_k^{+A}|} < \varepsilon, \tag{6}$$

com $\varepsilon \in \mathbb{R}$, onde QD_k^{+A} é o valor do quase desvio estendido ajustado obtido no k -ésimo ciclo e $||$ representa o operador valor absoluto. Considere $QD_0^{+A} = 0$.

Observe que no k -ésimo ciclo deve-se averiguar se o modelo para a média ou para a dispersão está bem ajustado, bem como se algum parâmetro no modelo da média, ou no modelo da dispersão, deve ou não ser excluído do modelo conjunto. Para a seleção dos parâmetros no modelo da média pode-se usar a estatística $t = \hat{\beta}_{kj} / \sqrt{c_{jj}}$, onde c_{jj} é o elemento pertencente à diagonal da matriz $(X^T W_k X)^{-1}$; considerando que um valor de $|t| \geq 3$ indica que o parâmetro deve ser significativo e que um valor de $|t| \leq 1$ indica que o parâmetro não deve ser significativo (Vieira, 2004). Para valores de $1 < |t| < 3$ deve-se verificar o quase desvio estendido, ajustado com e sem este parâmetro, e se houver uma diferença desses quase desvios $> 4,0$ ($\cong \chi_1^2$), o parâmetro será considerado significativo. No modelo da dispersão pode-se usar o mesmo procedimento, mas agora com $t = \hat{\gamma}_{kj} / \sqrt{b_{jj}}$ e com b_{jj} o elemento pertencente à diagonal da matriz $(U^T V_k U)^{-1}$. O teste da estatística t serve para fazer uma filtragem no número de parâmetros, fornecendo uma indicação dos parâmetros que devem ou não ser verificados na análise do quase desvio estendido.

No último ciclo, isto é, quando a convergência tiver sido alcançada, o vetor de parâmetros β , que interfere na média, e o vetor de parâmetros γ , que interfere na dispersão, terão sido encontrados.

O próximo passo agora é verificar quais são os melhores níveis das covariáveis para os parâmetros no vetor γ , de modo que a variância da resposta seja mínima; depois verificar quais são os melhores níveis das covariáveis para os parâmetros no vetor β , de modo que o valor da média fique o mais próximo possível do valor de interesse pré-especificado (valor alvo). O algoritmo para a modelagem conjunta da média e dispersão é mostrado no apêndice A.

4. Exemplos Numéricos

A seguir são dados dois exemplos numéricos para ilustrar a teoria descrita nas seções anteriores. O primeiro exemplo trata de dados contínuos, enquanto que o segundo, considera dados de contagem.

4.1 Mistura para bolo

Este exemplo, retirado de Atkinson & Donev (1992), concentra-se no desenvolvimento de uma nova mistura para bolo a ser apresentada no mercado. O produto necessita ser robusto às condições inadequadas de cozimento, representadas pelos fatores ambientais, como temperatura do forno, x_4 , e tempo em que o bolo permanece assando, x_5 . Os três fatores sob controle do fabricante são: a quantidade de farinha, de açúcar e de ovo, denotadas, respectivamente, por x_1 , x_2 e x_3 . O planejamento experimental, dado na Tabela 2, consiste de um fatorial 2^3 , com ponto central nos fatores de planejamento, cruzado com um fatorial 2^2 ,

mais o ponto central nos fatores ambientais. Os níveis dos fatores, iguais a zero, correspondem à composição pretendida da mistura e às condições ideais de cozimento sugeridas pelo fabricante. A resposta é um índice de predileção, isto é, uma nota variando de 0,0 a 10,0 e, obviamente, quanto maior melhor.

Inspecionando os resultados da Tabela 2 pode-se notar que os ensaios 7 e 9 produzem misturas que são menos suscetíveis às variações nos fatores ambientais x_4 e x_5 , mas o ensaio 7 tem a média mais alta e assim deve ser a melhor mistura para o mercado. Esta análise informal dos resultados deste experimento é suficiente para extrair informação relevante. Contudo, em experimentos mais complicados, uma análise mais sofisticada seria necessária.

Tabela 2 – Exemplo – mistura para bolo.

Ensaio	Fatores de planejamento			Fatores ambientais							
	Quantidade de			x_4	x_5	Temperatura do forno	0	-1	1	-1	1
	Farinha	Açúcar	Ovo								
	x_1	x_2	x_3								
1	0	0	0			6,7	3,4	5,4	4,1	3,8	
2	-1	-1	-1			3,1	1,1	5,7	6,4	1,3	
3	1	-1	-1			3,2	3,8	4,9	4,3	2,1	
4	-1	1	-1			5,3	3,7	5,1	6,7	2,9	
5	1	1	-1			4,1	4,5	6,4	5,8	5,2	
6	-1	-1	1			5,9	4,2	6,8	6,5	3,5	
7	1	-1	1			6,9	5,0	6,0	5,9	5,7	
8	-1	1	1			3,0	3,1	6,3	6,4	3,0	
9	1	1	1			4,5	3,9	5,5	5,0	5,4	

Para este exemplo, seguiu-se o procedimento descrito em Engel & Huele (1996), isto é, considerou-se: y_1, y_2, \dots, y_n , n observações independentes da variável resposta Y ; $\mathbf{s}^T = (s_1, \dots, s_k)$, k fatores de planejamento, e $\mathbf{r}^T = (r_1, \dots, r_l)$, l fatores de ruído. Neste exemplo $k = 3$ e $l = 2$. Seja $\mathbf{x}^T = (x_1, \dots, x_p)$ o vetor contendo as p covariáveis que afetam o modelo da média e que $\mathbf{u}^T = (u_1, \dots, u_q)$ é o vetor com as q covariáveis que afetam o modelo da dispersão. Os fatores ocorrendo em \mathbf{x}^T e \mathbf{u}^T podem ser efeitos lineares, quadráticos ou interações dos fatores de \mathbf{s}^T e \mathbf{r}^T .

O objetivo em um problema de planejamento robusto é obter um modelo para $E(Y)$ e para $Var(Y)$, a média e a variância do processo, respectivamente. Durante o experimento, os valores das respostas são observados condicionalmente aos níveis dos fatores de ruído, desta forma, define-se $\mu_i = E(Y_i | \mathbf{r}_i^T)$ e $\sigma_i^2 = Var(Y_i | \mathbf{r}_i^T)$, onde \mathbf{r}_i^T é a i -ésima linha da matriz \mathbf{R} ($n \times l$) dos fatores de ruído. O modelo proposto para o problema em questão foi: $Y_i = \mu_i + \varepsilon_i$, sendo $E(\varepsilon_i) = 0$ e $Var(\varepsilon_i) = \sigma_i^2$; onde $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$, com \mathbf{x}_i^T a i -ésima linha da matriz de planejamento \mathbf{X} ($n \times p$), do modelo da média, e $\boldsymbol{\beta}$ o vetor ($p \times 1$); $\sigma_i^2 = \exp(\mathbf{u}_i^T \boldsymbol{\gamma})$, com \mathbf{u}_i^T a i -ésima linha da matriz de planejamento \mathbf{U} ($n \times q$), do modelo da dispersão; e $\boldsymbol{\gamma}$ o vetor ($q \times 1$).

Para aplicar a modelagem conjunta da média e dispersão aos dados da Tabela 2, suponha $V(\mu_i) = 1$, função de ligação identidade para o modelo da média e distribuição Gama com função de ligação logarítmica para o modelo da dispersão. Note que não está sendo suposto que o modelo da média é conhecido, pois para usar a modelagem conjunta da média e dispersão necessita-se somente do conhecimento das funções de variância e de ligação. Observe também que outras funções de variância e de ligação, para o modelo da média, poderiam ter sido usadas. Para as principais distribuições conhecidas, pertencentes à família exponencial, existem funções de ligação e de variância apropriadas, entretanto podem existir situações em que não se está certo sobre qual distribuição usar, ou seja, não se sabe qual a distribuição correta. Nessas situações podem-se usar as funções de ligação e de variância conhecidas (veja McCullagh & Nelder, 1989); sendo as melhores aquelas que fornecerem o melhor ajuste aos dados, ou seja, que gerarem o menor quase desvio estendido.

Voltando ao exemplo, como $\mu_i = E(Y_i | \mathbf{r}_i^T)$, tem-se o valor esperado da resposta condicional aos fatores de ruído. Da mesma forma, $\sigma_i^2 = Var(Y_i | \mathbf{r}_i^T)$ é uma resposta para a variância, também condicional aos fatores de ruído. Assim, pode-se encontrar $E(Y_i)$ e $Var(Y_i)$ da seguinte forma: $E(Y_i) = E(E(Y_i | \mathbf{r}_i^T))$ e $Var(Y_i) = Var(E(Y_i | \mathbf{r}_i^T)) + E(Var(Y_i | \mathbf{r}_i^T))$.

Para o problema da mistura de bolo, o que interessa são os efeitos principais e as interações de primeira ordem, exceto a interação entre os dois fatores de ruído, $x_4 x_5$. As matrizes de planejamento \mathbf{X} , para o modelo da média, e \mathbf{U} , para o modelo da dispersão, foram consideradas iguais e estão mostradas no apêndice B.

O programa da modelagem conjunta da média e dispersão convergiu, usando o critério de convergência dado na equação (6), após 6 iterações e o resultado para os modelos da média e da dispersão, após os testes para verificação da significância dos parâmetros, é mostrado na Tabela 3.

Tabela 3 – Estimativa dos parâmetros para as variáveis significativas no modelo da média e da dispersão para o exemplo da mistura de bolo.

Modelo	Coefficiente	Estimativa	Erro padrão
Média	Intercepto	4,7	0,16
	x_2	0,11	0,14
	x_3	0,46	0,14
	$x_2 x_3$	-0,63	0,14
Dispersão	x_1	-0,74	0,17

No modelo da média, o parâmetro β_2 , correspondente à variável x_2 , não foi significativo, entretanto, decidiu-se deixá-lo no modelo, pois uma interação desta covariável com a covariável x_3 foi significativa. Assim,

$$\hat{E}(Y) = 4,7 + 0,11x_2 + 0,46x_3 - 0,63x_2x_3 \quad \text{e} \quad \hat{Var}(Y) = \exp\{-0,74x_1\}, \quad (7)$$

pois nenhuma interação com os fatores de ruído foi significativa, em nenhum dos modelos. Caso algum fator de ruído fosse considerado significativo, ter-se-ia que assumir uma média, possivelmente zero, e uma variância para este fator, pois os fatores de ruído são considerados variáveis aleatórias e não podem ser controlados. (para maiores detalhes, veja Engel & Hule, 1996). Na equação (7), para que a variância seja mínima, deve-se ter $x_1 = 1$ (nível alto de farinha) e para que a média fique a mais alta possível, deve-se ter $x_2 = -1$ (nível baixo de açúcar) e $x_3 = 1$ (nível alto de ovo). Logo, a melhor combinação possível é: 1, -1 e 1, correspondendo, exatamente, à linha 7 da Tabela 2; como era esperado.

Atkinson & Donev (1992), utilizando modelos separados para a média e para a dispersão, obtiveram os mesmos resultados, com relação à escolha dos fatores, e valores praticamente idênticos em relação às estimativas dos parâmetros. Vale ressaltar também que usando a análise de Taguchi (não mostrada), a qual procura o máximo de uma relação Sinal Ruído (SR), o ensaio 7, na Tabela 2, é o melhor, pois possui o maior valor de SR. Para esse experimento simples, a escolha dos níveis dos fatores seria a mesma, tanto usando a razão sinal ruído de Taguchi, quanto usando MLG. Entretanto, não se teria um modelo para a média da resposta nem para a dispersão e, portanto, não se saberia quais fatores afetam a média e quais afetam a variância; informação que pode ser muito importante num processo de produção. Neste artigo, não se quer entrar em detalhes nem fazer críticas sobre métodos de Taguchi, o que se pretende é apresentar uma teoria alternativa, com uma sólida base estatística.

4.2 Análise de contagem usando MLG com parâmetro de dispersão variável

A título de ilustração da teoria para dados discretos serão gerados dados de contagem de uma distribuição discreta com parâmetro de dispersão variável. A modelagem conjunta da média e dispersão será aplicada a esses dados a fim de verificar se essa abordagem é capaz de identificar as covariáveis que afetam os modelos da média e da variância e se os valores encontrados para os parâmetros, correspondentes às covariáveis selecionadas, são próximos aos usados inicialmente para gerar os dados.

Em analogia ao exemplo 4.1, considere as covariáveis x_1 , x_2 e x_3 como fatores de controle, de modo que x_2 e x_3 afetam a média e que somente x_1 afeta a dispersão. Não serão considerados fatores de ruído afetando a média ou a dispersão. Desta forma, será considerado o preditor linear para o modelo da média como $\eta(\mu) = 2,2 + 0,46x_3 - 0,63x_2x_3$, e para o modelo da dispersão $\zeta(\phi) = -0,74x_1$. Observe que o valor 4,7, na equação (7), foi trocado por 2,2. Isto foi feito apenas para que os valores gerados para a resposta não fossem valores grandes, ou seja, maiores que 100. Note também que a covariável x_2 não foi colocada no modelo da média, embora ela seja importante (veja seção 4.1). Optou-se por não colocá-la, pois, a título de ilustração, o que se quer é apenas testar se o algoritmo funciona para o modelo usado para gerar os dados, independente de x_2 estar ou não no modelo. Suponha também que a função de variância seja $V(\mu_i) = \mu_i$, desta forma, os dados de contagem serão gerados de modo que:

$$E(Y_i) = \mu_i \quad \text{e} \quad Var(Y_i) = \phi_i \mu_i. \quad (8)$$

Será considerado também que ambas as funções de ligação, para a média e para a variância, sejam a função logarítmica. Assim, $\phi_i = \exp(-0,74x_1)$ e $\mu_i = \exp(2,2 + 0,46x_3 - 0,63x_2x_3)$.

A geração dos dados foi adaptada de McCullagh & Nelder (1989), página 198. Considere a variável aleatória $Y|N = Z_1 + \dots + Z_N$, onde Z_i são variáveis aleatórias independentes e identicamente distribuídas; e N é uma variável aleatória com distribuição de Poisson, independente de Z_i , para $i = 1, \dots, N$. Logo,

$$E(Y) = E(N)E(Z) \text{ e } Var(Y) = E(N)E(Z^2). \tag{9}$$

Seja $E(Z_i) = V_i$, então, para que se tenha $E(Y_i) = \mu_i$ na equação (9), deve-se fazer $E(N_i) = \mu_i / V_i$. Desta forma, deve-se gerar $N_i \sim \text{Poisson}(\mu_i / V_i)$. Da equação (8) sabe-se

que $Var(Y_i) = \frac{\mu_i}{V_i} [Var(Z_i) + E^2(Z_i)] = \phi_i \mu_i$; neste caso, tomando $Z_i \sim \text{Poisson}(V_i)$,

$Var(Y_i) = \frac{\mu_i}{V_i} (V_i + V_i^2) = \mu_i \phi_i$ e $V_i = |\phi_i - 1|$, pois $V_i > 0$. Portanto, se gerar

$N_i \sim \text{Poisson}(\mu_i / V_i)$ e com o valor N_i gerado, somar N_i distribuições de Poisson com parâmetro V_i , garante-se, por construção, que $E(Y_i) = \mu_i$ e $Var(Y_i) = \mu_i \phi_i$. Note que, por estar usando a quase verossimilhança estendida, não é preciso conhecer a distribuição de Y . A Tabela 4 mostra os valores de V_i e μ_i / V_i usados para gerar os dados.

Tabela 4 – Valores de V_i e μ_i / V_i usados para gerar dados de contagem com parâmetro de dispersão variável.

Ensaio	x_1	x_2	x_3	μ_i	ϕ_i	V_i	μ_i / V_i
1	0	0	0	9,015	1,000	1,000	9,015
2	-1	-1	-1	3,046	2,140	1,140	2,670
3	+1	-1	-1	3,046	0,467	0,533	5,710
4	-1	+1	-1	10,720	2,140	1,140	9,400
5	+1	+1	-1	10,720	0,467	0,533	20,110
6	-1	-1	+1	26,680	2,140	1,140	23,400
7	+1	-1	+1	26,680	0,467	0,533	50,060
8	-1	+1	+1	7,580	2,140	1,140	6,650
9	+1	+1	+1	7,580	0,467	0,533	14,220

As observações foram geradas uma única vez e para cada linha da Tabela 4, geraram-se 5 resultados, num total de 45 observações. Os dados gerados foram usados como resposta para a modelagem conjunta da média e dispersão, considerando o modelo para a média com ligação logarítmica e função de variância $V(\mu) = \mu$. O modelo para a dispersão foi

considerado Gama com função de ligação também logarítmica. A matriz de planejamento, considerada tanto para o modelo da média, quanto para o modelo da dispersão, foi a mesma do exemplo 4.1 e é mostrada no apêndice B.

O programa convergiu, usando o critério de convergência, dado na equação (6), após 4 iterações, e o resultado para os modelos da média e da dispersão, depois de realizados os testes para verificação da significância dos parâmetros, é mostrado na Tabela 5.

Tabela 5 – Estimativa dos parâmetros para as variáveis significativas no modelo da média e da dispersão para os dados de contagem.

Modelo	Coefficiente	Estimativa	Erro padrão
Média	Intercepto	2,212	0,082
	x_3	0,381	0,065
	x_2x_3	-0,580	0,102
Dispersão	x_1	-0,753	0,191

Assim, $\hat{\mu} = \exp(2,212 + 0,381x_3 - 0,580x_2x_3)$ e $\hat{\phi} = \exp\{-0,753x_1\}$. Observe que a modelagem conjunta da média e da dispersão conseguiu identificar as covariáveis que afetam a média e a dispersão. Além disso, as estimativas dos parâmetros encontradas foram próximas dos valores usados para gerar os dados. Vale salientar que, nesse exemplo, o objetivo não foi fazer um estudo de simulação, mas apenas aplicar a MCMD a um conjunto de dados, gerados a partir das hipóteses de interesse.

5. Considerações Finais

Nelder & Lee (1991) afirmam que a modelagem conjunta da média e dispersão é geral e suficiente para ajustar os modelos de Taguchi. Usando MLG não é preciso usar transformação para os dados. Modelos com resposta contínua, ou na forma de contagem e proporção, podem ser ajustados usando o mesmo algoritmo. Além disso, o critério de separação pode ser satisfeito pela especificação correta da função de variância no MLG; e parcimônia pode ser encontrada escolhendo funções apropriadas de ligação e covariáveis para os parâmetros dos modelos da média e da dispersão, respectivamente.

Um outro aspecto muito importante da modelagem conjunta da média e dispersão é que esta abordagem permite encontrar, além dos fatores que afetam a média, aqueles que afetam a dispersão. Dessa forma, pode-se escolher valores das covariáveis de modo que a resposta para o modelo da dispersão seja mínima.

Agradecimentos

Os autores agradecem as críticas e sugestões de um revisor anônimo, que possibilitaram melhorar sobremaneira a qualidade e a objetividade deste artigo.

Referências Bibliográficas

- (1) Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using GLIM. *Applied Statistics*, **36**, 332-339.
- (2) Atkinson, A.C. & Donev, A.N. (1992). *Optimum experimental designs*. Oxford Science Publications, Clarendon Press, UK.
- (3) Box, G.E.P. (1988). Signal to noise, performance criteria and transformations. *Technometrics*, **30**, 1-17.
- (4) Cuervo, E.C. (2001). Modelagem da variabilidade em modelos lineares generalizados. Tese de D. Sc., IM – UFRJ, Rio de Janeiro, RJ, Brasil.
- (5) Cordeiro, G.M. (1986). *Modelos lineares generalizados*. VII SINAPE–Campinas – SP.
- (6) Cox, D.R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of The Royal Statistical Society – Serie B*, **49**(1), 1-39.
- (7) Dey, K.D.; Gelfand, A.E. & Peng, F. (1997). Overdispersed generalized linear models. *Journal of Statistical Planning and Inference*, **64**, 93-107.
- (8) Dobson, A.J. (1990). *An introduction to generalized linear models*. Chapman & Hall, New York.
- (9) Engel, J. & Huele, A.F. (1996). A generalized linear modeling approach to robust design. *Technometrics*, **38**(4), 365-373.
- (10) Guedes, T.A. (1996). Procedimentos de otimização no planejamento e controle da qualidade de produtos e processos. Tese de D. Sc., Depto. de Engenharia de Produção – UFSC, Florianópolis, SC, Brasil.
- (11) Jorgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of The Royal Statistical Society – Serie B*, **49**(2), 127-162.
- (12) Jorgensen, B. (1992). The theory of exponential dispersion models and analysis of deviance. Monografia de Matemática, nº 51 – IMPA.
- (13) Lee, Y. & Nelder, J.A. (1998). Generalized linear models for analysis of quality improvement experiments. *The Canadian Journal of Statistics*, **26**(1), 95-105.
- (14) Lee, Y. & Nelder, J.A. (2000). The relationship between double-exponential families and extended quasi-likelihood families, with application to modelling Geissler's human Sex ratio data. *Applied Statistics*, **49**(3), 413-419.
- (15) Lee, Y. & Nelder, J.A. (2003). Robust design via generalized linear models. *Journal of Quality Technology*, **35**(1), 2-12.
- (16) McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics*, **11**(1), 59-67.
- (17) McCullagh, P. & Nelder, J.A. (1989). *Generalized linear models*. 2ª edição, Chapman & Hall, Londres.

- (18) Myers, R.H.; Montgomery, D.C. & Vining, G.G. (2002). *Generalized linear models with applications in engineering and the sciences*. Wiley, New York.
- (19) Nair, V.N. (1992). Taguchi's parameter design: a panel discussion. *Technometrics*, **34**(2), 127-161.
- (20) Nelder, J.A. & Lee, Y. (1991). Generalized linear models for the analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis*, **7**, 107-120.
- (21) Nelder, J.A. & Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, **74**, 221-232.
- (22) Phadke, M.S. (1989). *Quality engineering using robust design*. Prentice-Hall, Englewood Cliffs, New Jersey.
- (23) Pinto, E.R. (2005). Planejamento de experimentos ótimos para modelos lineares generalizados com parâmetro de dispersão variando. Tese de D. Sc. COPPE – UFRJ, Rio de Janeiro, RJ, Brasil.
- (24) Smyth, G.K. (1989). Generalized linear models with varying dispersion. *Journal of The Royal Statistical Society – Serie B*, **51**(1), 47-60.
- (25) Smyth, G.K. (2002). An efficient algorithm for REML in heteroscedastic regression. *Journal of Graphical and Computational Statistics*, **11**, 15-27.
- (26) Smyth, G.K.; Huele, A.F. & Verbyla, A.P. (2001). Exact and approximate REML for heteroscedastic regression. *Statistical Modelling*, **1**, 161-175.
- (27) Smyth, G.K. & Verbyla, A.P. (1996). A conditional likelihood approach to REML in generalized linear models. *Journal of the Royal Statistical Society – Series B*, **58**, 565-572.
- (28) Smyth, G.K. & Verbyla, A.P. (1999). Adjusted likelihood methods for modeling dispersion in generalized linear models. *Environmetrics*, **10**, 696-709.
- (29) Vieira, A.F.C. (2004). Análise da média e dispersão em experimentos fatoriais não replicados para otimização de processos industriais. Tese de D. Sc. Depto. de Engenharia de Produção – PUC – Rio, Rio de Janeiro, Brasil.
- (30) Verbyla, A.P. (1993). Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *Journal of the Royal Statistical Society – Series B*, **55**, 493-508.
- (31) Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, **61**, 439-447.

Apêndice A – Algoritmo para a Modelagem conjunta da média e dispersão

O método iterativo da modelagem conjunta da média e dispersão pode ser resumido no seguinte algoritmo. Considere as suposições e definições apresentadas na seção 3.1.

Inicie

Coloque $k = 1$

Faça $\beta_0^T = (0, \dots, 0)$, $\mu_0^T = (y_1, \dots, y_n)$ e $\phi_0^T = (1, \dots, 1)$

Enquanto a convergência não for satisfeita (modelo conjunto)

Coloque $j = 1$

Enquanto a convergência não for satisfeita (modelo da média)

Calcule $\beta_{(j)} = (\mathbf{X}^T \mathbf{W}_{(j-1)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_{(j-1)} \mathbf{z}_{(j-1)}$

Se $\|\beta_{(j)} - \beta_{(j-1)}\|^2 < \delta$, pare (convergência satisfeita)

Faça $\beta_k = \beta_{(j)}$; $\mathbf{W}_k = \mathbf{W}_{(j-1)}$; $\mu_k = \mathbf{g}_m^{-1}(\mathbf{X}\beta_k)$

Senão

Faça $j = j + 1$

Fim Se

Fim Enquanto

Calcule d_k^*

Faça $\gamma_0^T = (0, \dots, 0)$ e $\phi_0^T = (d_{k1}^*, \dots, d_{kn}^*)$

Coloque $j = 1$

Enquanto a convergência não for satisfeita (modelo da dispersão)

Calcule $\gamma_{(j)} = (\mathbf{U}^T \mathbf{V}_{(j-1)} \mathbf{U})^{-1} \mathbf{U}^T \mathbf{V}_{(j-1)} \mathbf{s}_{(j-1)}$

Se $\|\gamma_{(j)} - \gamma_{(j-1)}\|^2 < \delta$, pare (convergência satisfeita)

Faça $\gamma_k = \gamma_{(j)}$; $\mathbf{V}_k = \mathbf{V}_{(j-1)}$; $\phi_k = \mathbf{g}_d^{-1}(\mathbf{U}\gamma_k)$

Senão

Faça $j = j + 1$

Fim Se

Fim Enquanto

Faça $QD_0^{+A} = 0$

Se $\frac{|QD_k^{+A} - QD_{k-1}^{+A}|}{|QD_k^{+A}|} < \varepsilon$, pare (convergência atingida).

Senão

Faça $k = k + 1$

Fim Se

Fim Enquanto.

Faça $\beta = \beta_k$, $\gamma = \gamma_k$

Fim

A listagem do programa para a modelagem conjunta da média e dispersão, em linguagem de programação FORTRAN 90, é dada em Pinto (2005).

Apêndice B

Matriz de planejamento para o modelo da média e da dispersão do exemplo 4.1. A primeira linha da matriz representa as variáveis consideradas inicialmente no modelo.

$$X = U = \begin{bmatrix} 1 & x_1 & x_2 & x_3 & x_4 & x_5 & x_1x_2 & x_1x_3 & x_1x_4 & x_1x_5 & x_2x_3 & x_2x_4 & x_2x_5 & x_3x_4 & x_3x_5 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$