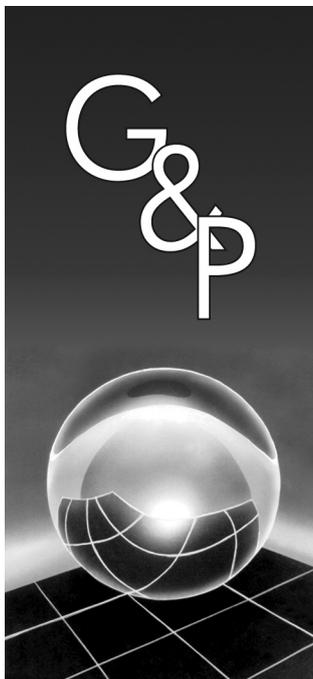


Métodos de identificação de efeitos na dispersão em experimentos fatoriais não replicados

Antonio Fernando de Castro Vieira
Eugenio Kahn Epprecht



Resumo

A identificação de fatores que afetam a média e a dispersão de características de qualidade é essencial para a otimização de produtos e processos produtivos. Para tal identificação, costuma-se utilizar experimentos fatoriais. A análise de efeitos sobre a dispersão, contudo, usualmente demanda repetições do experimento sob as mesmas condições (replicações), que podem ser caras. Assim, vários métodos têm sido propostos na literatura para a identificação de efeitos na dispersão a partir de experimentos fatoriais não replicados. Neste artigo, analisamos alguns desses métodos, ilustrados com um exemplo de um processo produtivo, concluindo pela superioridade de um método iterativo baseado em modelos lineares generalizados. Finalmente, utilizando os modelos para a média e para a dispersão, fornecidos pelo método iterativo, procedemos à otimização do processo descrito no exemplo.

Palavras-chave: Experimentos fatoriais. Modelos lineares generalizados. Otimização de processos.

1 Introdução

Técnicas estatísticas e matemáticas são úteis para o desenvolvimento, melhoria e otimização de processos produtivos. Essas técnicas são igualmente úteis para o projeto e desenvolvimento de novos produtos, bem como para a melhoria de produtos existentes. Nesse contexto, é particularmente importante aplicar essas técnicas na identificação de fatores que influenciam características de qualidade de processos ou produtos.

Na otimização de processos produtivos e de projetos de produtos, os métodos inicialmente utilizados tinham como foco principal a identificação de fatores que afetam a média da característica de qualidade (resposta). Os efeitos transmitidos por esses fatores são denominados efeitos na média (*location effects*). Seu conhecimento possibilita um ajuste nos fatores que propicia um valor desejado para a média da resposta.

Com a modernização dos conceitos de qualidade - quando então variância mínima passou a ser sinônimo de qualidade -, a análise dos fatores que afetam a variância da resposta tornou-se alvo de interesse. Os efeitos transmitidos por esses fatores são denominados efeitos na dispersão (*dispersion effects*). O conhecimento desses efeitos possibilita um ajuste nos fatores que propicia a minimização da variância da resposta. Para uma descrição

abrangente do impacto de tais efeitos na qualidade dos produtos, ver Myers e Montgomery (2002).

Para a identificação dos fatores que afetam a dispersão, é desejável utilizar experimentos replicados, porque, com replicações, pode-se construir um modelo para a dispersão, através da variância das observações replicadas, independentemente do modelo utilizado para a média. Entretanto, experimentos replicados podem ter alto custo ou longo tempo de experimentação. Daí o interesse em experimentos sem replicações. Para uma abordagem de experimentos com replicações, ver McCullagh e Nelder (1989), Grego (1993) e Myers e Montgomery (2002).

Neste artigo descrevemos três métodos de identificação de efeitos na dispersão aplicados a experimentos sem replicações e ilustrados com um exemplo real proveniente da literatura. Nosso objetivo é exemplificar e comparar os três métodos e fornecer uma revisão bibliográfica dos trabalhos relevantes sobre o tema.

Embora tratemos aqui de métodos de identificação de efeitos na dispersão, é importante construir um modelo adequado para a média. Isto porque, não havendo replicações, o modelo para a dispersão é obtido através dos resíduos provenientes do modelo da média. Erros na identificação dos fatores influentes sobre a média afetam tais

resíduos e podem levar à identificação de efeitos espúrios na dispersão ou à não-identificação de efeitos reais.

Na Seção 2, apresentamos uma revisão bibliográfica comentada dos principais artigos sobre a identificação de efeitos na dispersão em experimentos não replicados. Apesar de haver muitos trabalhos sobre o assunto, não encontramos uma descrição detalhada dos métodos reunida em um único artigo, com uma ilustração, como fazemos aqui.

Na Seção 3, apresentamos três métodos para identificar efeitos na dispersão. Os métodos são ilustrados com um exemplo. Na Seção 4, fazemos comentários adicionais comparando a eficácia e a amplitude de aplicabilidade dos métodos descritos na Seção 3. Na Seção 5, apresentamos as conclusões e recomendações, bem como indicações para trabalhos futuros.

2 Revisão da literatura

Nesta seção, apresentamos um apanhado do estado da arte para modelagem da dispersão em experimentos fatoriais não replicados. Aqui fornecemos as referências que descrevem os métodos que consideramos relevantes sobre o assunto.

Quanto a métodos de identificação de modelos para a média, para revisão e comparação, recomendamos Hamada e Balakrishnan (1998).

O primeiro artigo para identificação de efeitos na dispersão em experimentos não replicados é o de Box e Meyer (1986). Eles propuseram um método baseado na variabilidade dos resíduos provenientes do modelo ajustado para a média por mínimos quadrados. Este método é descrito e ilustrado na Seção 3. Bergman e Hynén (1997) também propuseram um método baseado na variabilidade dos resíduos provenientes do modelo ajustado para a média, provendo então uma estatística de teste baseada na distribuição F .

Pan (1999) e McGrath e Lin (2001) analisaram a influência do modelo elaborado para a média na identificação dos efeitos na dispersão, aplicando os métodos de Box e Meyer (1986) e de Bergman e Hynén (1997) a casos reais. Segundo Pan (1999), incluir, no modelo da média, fatores (ou interações) que realmente não afetam a média diminui a eficiência do método usado para identificar os efeitos na dispersão, mas não afeta a sua validade. Isto quer dizer que o método ainda identifica efeitos reais na dispersão, embora esteja mais sujeito a erros na identificação. Por outro lado, não incluir, no modelo da média, fatores (ou interações) que realmente afetam a média pode, cumulativamente, invalidar o método usado para identificar os efeitos na dispersão. Pan afirma que os efeitos na média não identificados podem causar um grande impacto no poder e na probabilidade de erro do método de identificação de efeitos na dispersão, e que a intensidade desse

impacto é imprevisível se o verdadeiro modelo da média é desconhecido. As conclusões de McGrath e Lin (2001) são similares às de Pan (1999). Eles também mostram que, em certos casos, não se pode determinar o que é real: um efeito na dispersão ou dois efeitos na média. Na seção 3, ilustramos, com um exemplo, os problemas detectados por estes autores.

Montgomery (2001) descreve um método gráfico muito simples para identificar efeitos na dispersão. Este método é descrito e ilustrado na Seção 3. Entretanto, como será visto seu caráter subjetivo, eventualmente, leva a dúvidas sobre quais fatores devem afetar a dispersão.

Métodos iterativos para identificação de efeitos na média e na dispersão foram desenvolvidos por McCullagh e Nelder (1989), Engel e Huele (1996) e Lee e Nelder (1998). Em tais métodos, são utilizados os modelos lineares generalizados (MLG), sendo que McCullagh e Nelder (1989) e Lee e Nelder (1998) aplicam MLG para modelar explicitamente tanto a média quanto a dispersão, enquanto Engel e Huele (1999) aplicam MLG somente para modelar a dispersão. Para a modelagem da média, estes últimos assumem um modelo para a média cuja função de ligação é a identidade, o que constitui um caso particular dos modelos MLG propostos por McCullagh e Nelder (1989) e Lee e Nelder (1998). Ademais, Engel e Huele não consideram o uso da técnica de quase-verossimilhança estendida (QVE) (*extended quase-likelihood*), utilizada por McCullagh e Nelder (1989) e Lee e Nelder (1998). Dentre os três métodos, o método de Lee e Nelder (1998) é o único que usa a técnica de máxima verossimilhança restrita (MVR) (*restricted maximum likelihood*), que é um ajuste feito no modelo da média, necessário em experimentos fatoriais altamente fracionados.

3 Métodos de identificação de efeitos na dispersão

Nesta seção, apresentamos três métodos para identificar efeitos na dispersão em experimentos fatoriais não replicados: o método gráfico descrito por Montgomery (2001), o método de Box e Meyer (1986) e o método iterativo de Lee e Nelder (1998), ilustrando o uso e analisando o desempenho de cada um deles pela aplicação de todos a um mesmo conjunto de dados.

3.1 Método gráfico

O gráfico *Resíduos versus Fator* é útil para uma primeira inspeção do efeito de um fator sobre a dispersão. Em alguns casos, pode fornecer indicações conclusivas. A análise é feita observando se há mudanças no espalhamento dos resíduos quando o nível do fator é mudado. Quando a mudança é visível, isto é indício de que a variabilidade é sensível ao fator em questão. Vejamos o método com um exemplo.

Exemplo

Myers e Montgomery (2002) analisaram um experimento de moldagem cuja resposta é a contração do molde. São estudados sete fatores (A-G). Trata-se de um experimento fatorial 2^{7-3} com resolução IV. Os geradores são $I = ABCE$, $II = BCDF$ e $G = ACDG$. Os dados são apresentados na Tabela 1.

Com base no gráfico de probabilidade normal dos efeitos de cada fator e interação entre fatores, mostrado na Figura 1, Myers e Montgomery (2002) consideraram A , B e AB como efeitos importantes. Usando-se o método mínimos quadrados, ajusta-se o seguinte modelo (Equação 1):

$$\text{Modelo M1: } \hat{y} = 27,31 + 6,94A + 17,81B + 5,94AB \quad (1)$$

McGrath e Lin (2001) também analisaram este exemplo e, com base no gráfico de probabilidade normal dos efeitos (Figura 2), propuseram adicionar os efeitos G e CG . Cabe registrar que, devido à estrutura do experimento, a interação AB está confundida com CE e FG , e a interação CG está confundida com AD e EF . Caso não haja conhecimentos práticos sobre a existência ou não dessas interações, há necessidade de um experimento complementar para eliminar as ambiguidades.

Tendo em vista a continuidade do exemplo, vamos supor que a interação ativa seja CG . Entretanto, devemos aplicar o *princípio da hierarquia*. Montgomery (2001), descreve este princípio: se um modelo contém uma interação, deve conter também todos os fatores que a compõem. Ele afirma que isto promove uma consistência interna no modelo e que muitos estatísticos seguem este princípio. Com base nisto, incluímos o fator C .

Usando-se o método de mínimos quadrados ajusta-se o seguinte modelo (Equação 2).

Tabela 1. Dados do experimento de moldagem.

A	B	C	D	E = ABC	F = BCD	G = ACD	Y
-1	-1	-1	-1	-1	-1	-1	6
1	-1	-1	-1	1	-1	1	10
-1	1	-1	-1	1	1	-1	32
1	1	-1	-1	-1	1	1	60
-1	-1	1	-1	1	1	1	4
1	-1	1	-1	-1	1	-1	15
-1	1	1	-1	-1	-1	1	26
1	1	1	-1	1	-1	-1	60
-1	-1	-1	1	-1	1	1	8
1	-1	-1	1	1	1	-1	12
-1	1	-1	1	1	-1	1	34
1	1	-1	1	-1	-1	-1	60
-1	-1	1	1	1	-1	-1	16
1	-1	1	1	-1	-1	1	5
-1	1	1	1	-1	1	-1	37
1	1	1	1	1	1	1	52

$$\text{Modelo M2: } \hat{y} = 27,31 + 6,94A + 17,81B + 5,94AB - 0,44C - 2,44G - 2,69CG \quad (2)$$

Observando os gráficos, é difícil dizer quem tem razão. Com efeito, o caráter subjetivo deste método eventualmente leva a dúvidas sobre quais fatores devem ser incluídos no modelo.

Para avaliar o efeito do fator sobre a dispersão, Montgomery (2001) propõe o uso do gráfico resíduos *versus* fator. Considerando o modelo M1, na Figura 3, apresentamos este gráfico para o fator C . (Os números

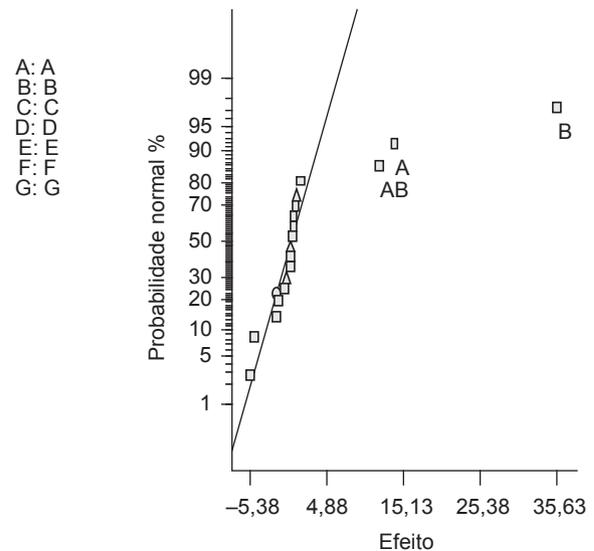


Figura 1. Gráfico de probabilidade normal dos efeitos (reta ajustada por Myers e Montgomery, 2002).

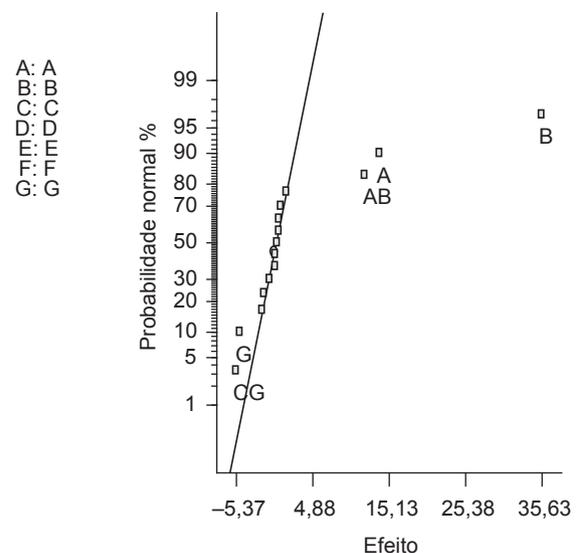


Figura 2. Gráfico de probabilidade normal dos efeitos (reta ajustada por McGrath e Lin, 2001).

“2” que aparecem no gráfico indicam as ocorrências de dois resíduos com um mesmo valor). Observamos que o espalhamento dos resíduos fica visivelmente maior com a mudança do nível. Isto indica que o fator *C* afeta a dispersão: a variabilidade aumenta quando mudamos *C* de -1 para +1.

Na Figura 4, apresentamos o gráfico equivalente para o Modelo M2. Neste gráfico, observamos que o espalhamento dos resíduos não muda significativamente com a mudança do nível. Isto é um indício de que a variabilidade não muda quando o fator *C* muda de -1 para +1. Portanto, temos evidências de que a não-inclusão de *C*, *G* e *CG* no modelo para a média resulta em identificar um efeito espúrio do fator *C* sobre a dispersão. Entretanto, analisando os gráficos do Modelo M2 para os outros fatores, encontramos indicação de que os fatores *F* e *G* afetam a dispersão. Os gráficos referentes a estes fatores estão na Figura 5. Observamos que, em ambos os gráficos, o espalhamento dos resíduos fica menor com a mudança do nível. Isto é um indício de que a variabilidade diminui quando os fatores *F* e *G* mudam do nível -1 para +1.

Com o Modelo M1, identificou-se *C* como fator que afeta a dispersão. Com o Modelo M2, identificou-se *F* como fator que afeta a dispersão. Portanto, com a análise do Exemplo, constatamos que diferentes modelos para a média podem levar à identificação de diferentes efeitos na dispersão. Evidentemente, este método gráfico não é um teste formal. Mas ele pode, em muitos casos, fornecer alguma evidência dos fatores que afetam a dispersão.

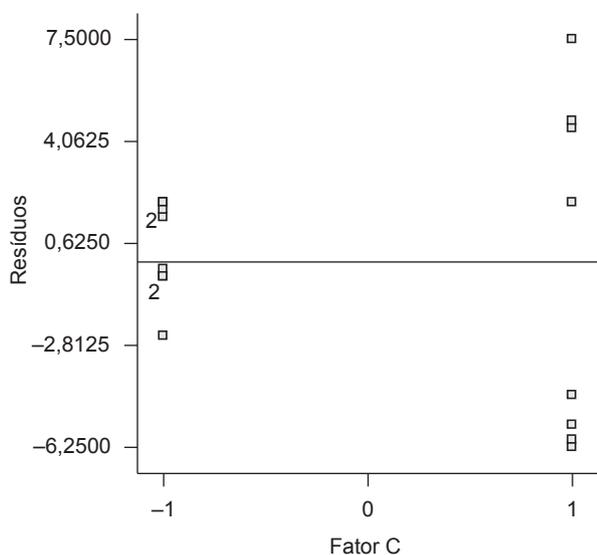


Figura 3. Gráfico dos Resíduos versus Fator C - Modelo M1. *Obs.: os números “2” indicam que o ponto corresponde a dois resíduos de igual valor.*

3.2 Método de Box e Meyer

Box e Meyer (1986) propuseram um método baseado na variabilidade dos resíduos provenientes do modelo ajustado para a média por mínimos quadrados. Este método só é aplicável a experimentos fatoriais com dois níveis, completos ou fracionados. Portanto, as variáveis x_{ij} são geralmente codificadas em dois níveis: -1 e +1.

Seja o modelo linear (Equação 3)

$$y = X\beta + \varepsilon \tag{3}$$

sendo (Equação 4)

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \tag{4}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{e} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

e, por sua vez, *k* corresponde ao número de variáveis regressoras e *n*, ao número de respostas. A variável que aparece em cada coluna da matriz *X* pode representar tanto um fator principal quanto uma interação entre fatores. Definem-se então, para cada variável, os seguintes conjuntos M_j e P_j (Equação 5):

$$M_j = \{i: x_{ij} = -1\} \quad P_j = \{i: x_{ij} = +1\} \tag{5}$$

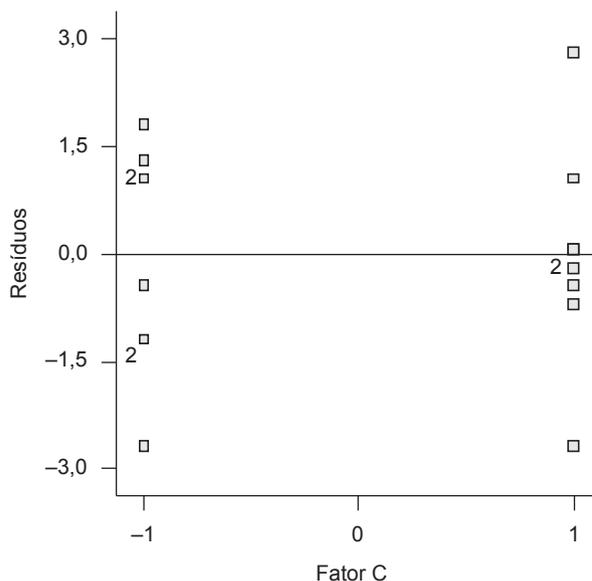


Figura 4. Gráfico dos Resíduos versus Fator C - Modelo M2. *Obs.: os números “2” indicam que o ponto corresponde a dois resíduos de igual valor.*

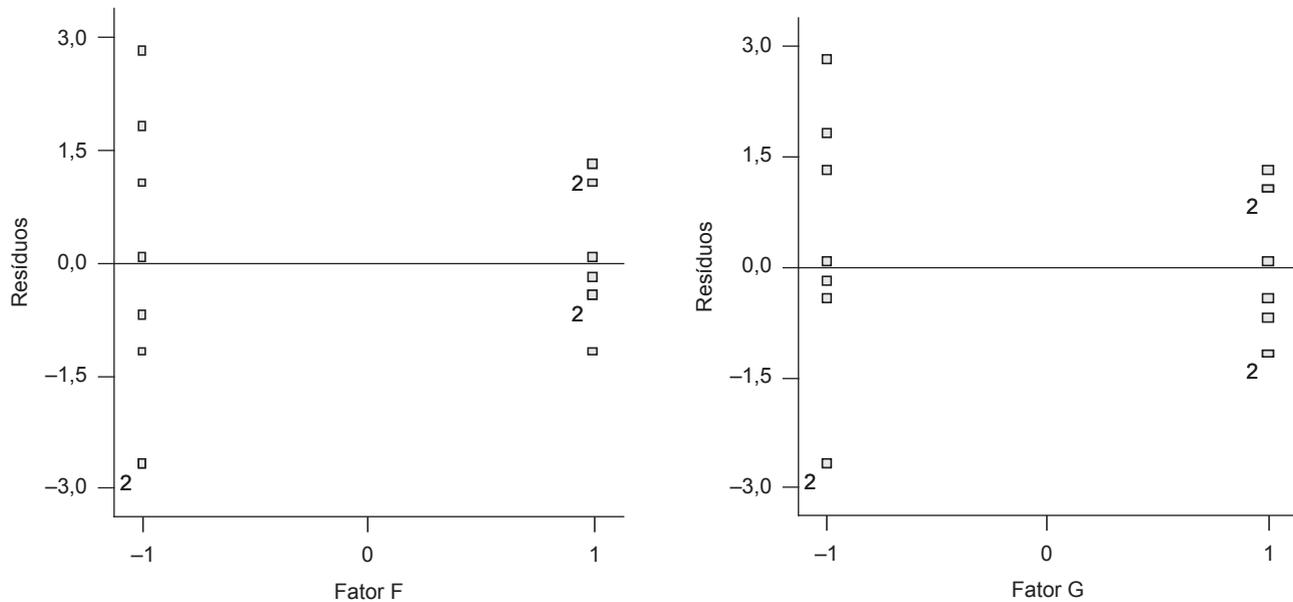


Figura 5. Gráficos dos Resíduos versus Fatores F e G - Modelo M2. Obs.: os números “2” indicam que o ponto corresponde a dois resíduos de igual valor.

Para estimativa dos parâmetros, é aplicado o método dos mínimos quadrados.

Sejam $e_i = (y_i - \hat{y}_i)$ os resíduos, $i = 1, 2, \dots, n$, do modelo ajustado. Para cada coluna (variável j), define-se:

$s_{j-}^2 = \text{var}_{i \in P_j}(e_{ij})$ a variância amostral dos resíduos e_{ij} tais que $i \in M_j$

$s_{j+}^2 = \text{var}_{i \in P_j}(e_{ij})$ a variância amostral dos resíduos e_{ij} tais que $i \in P_j$

Se os valores dos desvios-padrão s_{j+} e s_{j-} não são muito diferentes, tem-se a indicação de que, quando o fator j muda de um nível para outro, a variância dos resíduos não é afetada por esse fator, ou seja, a variância da resposta não se altera com o nível do fator j . Por outro lado, se um valor é muito maior do que o outro, tem-se a indicação de que a variância da resposta é afetada por esse fator. Vejamos o método aplicado ao Exemplo.

Exemplo (continuação)

Considerando o Modelo M1, na Tabela 2 apresentamos o cálculo dos desvios-padrão s_{j+} e s_{j-} associados aos fatores B e C.

Para o fator B, temos $s_{B+} = 4,01$ e $s_{B-} = 4,41$ – valores que diferem em dez por cento –, enquanto que, para o fator C, temos $s_{C+} = 5,70$ e $s_{C-} = 1,63$ – uma diferença muito maior. Entretanto, não sabemos se esta diferença é significativa. Para uma melhor verificação da magnitude relativa dos efeitos sobre a dispersão, Box e Meyer (1986) definiram a estatística (Equação 6),

$$F_j = \ln \frac{s_{j+}^2}{s_{j-}^2} = \ln(s_{j+}^2) - \ln(s_{j-}^2) \quad j = 1, 2, \dots, r \quad (6)$$

em que r é o número de variáveis (fatores e interações).

Montgomery (2001) assegura que, se os erros forem normalmente distribuídos e se a variância dos resíduos para as respostas em que o nível do fator j é (-1) for igual à variância dos resíduos para as respostas em que o nível do fator j é (+1), então F_j segue aproximadamente a distribuição normal. Assim sendo, um gráfico de probabilidade normal dos valores de F_j pode indicar os fatores j que afetam a dispersão.

Considerando o Modelo M1, na tabela abaixo apresentamos os valores da estatística F_j (Tabela 2).

Na Figura 6, apresentamos o gráfico de probabilidade normal de F_j , em que, efetivamente, o ponto correspondente ao fator C está afastado da reta em torno da qual se alinham os demais pontos. Isto indica que o fator C afeta a dispersão.

Considerando o Modelo M2, na tabela abaixo apresentamos os valores da estatística F_j (Tabela 3).

Verificamos que o valor de F_j para o fator C diminuiu. Na Figura 7, apresentamos o gráfico de probabilidade

A	B	AB	C	AC	BC	E
-0,38	-0,18	0,11	2,5	-0,41	-0,24	-0,03

D	AD	BD	ABD	CD	G	F	DE
0,51	0,42	-0,18	0,52	0,51	0,23	-0,3	0,72

Tabela 2. Valores da estatística F_j Modelo M1.

<i>F</i>	<i>G</i>	<i>B</i>	<i>AD</i>	<i>BD</i>	<i>ABD</i>	<i>A</i>	
-1,5	-1,31	-0,38	-0,38	-0,36	-0,3	-0,22	
<i>D</i>	<i>C</i>	<i>AC</i>	<i>CD</i>	<i>BC</i>	<i>DE</i>	<i>E</i>	<i>AB</i>
-0,22	-0,06	-0,02	0,13	0,25	0,33	0,39	1,8

Tabela 3. Valores da estatística F_j Modelo M2.

normal de F_j para o Modelo M2. Observando o gráfico, não temos mais indicação de que a dispersão seja afetada pelo fator *C* (o que confirma a indicação do gráfico da Figura 4), mas sim pelos fatores *F* e *G* (o que confirma a indicação dos gráficos da Figura 5) e pela interação *AB*. Agora temos um dilema. Temos um modelo simples para a média (M1) com um fator de dispersão (*C*) ou modelo mais complexo para a média (M2), com dois fatores (*F* e *G*) e uma interação (*AB*) que afetam a dispersão.

Na Figura 8, apresentamos o gráfico dos resíduos *versus* nível da interação *AB*. Esse gráfico também indica que a interação *AB* afeta a dispersão.

Os experimentos fracionados (como no exemplo) são usados nos estágios iniciais da investigação. Se optarmos pelo modelo mais simples (M1), estaremos eliminando fatores potencialmente importantes, que poderiam ser identificados em experimentos posteriores. Então, uma estratégia recomendável, neste caso, é seguir com o modelo mais complexo (M2), pelo menos até ter confirmação ou não em estágios posteriores.

Tabela 2. Cálculo dos Desvios-padrão s_{j+} e s_{j-} associados aos fatores *B* e *C*.

<i>y</i>	\hat{y}	Resíduos	Nível	<i>Fator B</i>		<i>Fator C</i>		
				$i \in P_j$	$i \in M_j$	$i \in P_j$	$i \in M_j$	
6	8,50	-2,50	-1		-2,50	-1	-2,50	
10	10,50	-0,50	-1		-0,50	-1	-0,50	
32	32,25	-0,25	1	-0,25		-1	-0,25	
60	58,00	2,00	1	2,00		-1	2,00	
4	8,50	-4,50	-1		-4,50	1	-4,50	
15	10,50	4,50	-1		4,50	1	4,50	
26	32,25	-6,25	1	-6,25		1	-6,25	
60	58,00	2,00	1	2,00		1	2,00	
8	8,50	-0,50	-1		-0,50	-1	-0,50	
12	10,50	1,50	-1		1,50	-1	1,50	
34	32,25	1,75	1	1,75		-1	1,75	
60	58,00	2,00	1	2,00		-1	2,00	
16	8,50	7,50	-1		7,50	1	7,50	
5	10,50	-5,50	-1		-5,50	1	-5,50	
37	32,25	4,75	1	4,75		1	4,75	
52	58,00	-6,00	1	-6,00		1	-6,00	
			Variância	16,11	19,43	Variância	32,44	2,66
			Desvio-padrão	4,01	4,41	Desvio-padrão	5,70	1,63

3.3 Método iterativo de Lee e Nelder

A seguir, descrevemos o método de Lee e Nelder (1998). O método envolve um modelo para a média e um modelo para a dispersão, que se alimentam mutuamente (cada um usa o outro) de maneira iterativa. A descrição a seguir se divide, para maior clareza, em subseções que tratam de cada um dos modelos e do procedimento iterativo. Uma subseção própria descreve também o método de máxima verossimilhança restrita, que é usado no caso em que o número de dados é pequeno em relação ao número de parâmetros do modelo para a média.

3.3.1 Modelo para a média

O modelo proposto para a média é (Equação 7),

$$\eta_i = g(\mu_i) = x'_i \beta \quad (7)$$

A média e a variância de y_i são $E(y_i) = \mu_i$ e $\text{var}(y_i) = \phi_i V(\mu_i)$ em que ϕ_i é o parâmetro de dispersão (que não depende da média) e $V(\mu_i)$ é a função de variância, que depende da média. Observe que o parâmetro de dispersão, que é constante para os membros da família exponencial, é agora considerado diferente para cada combinação dos níveis dos fatores. Sendo o parâmetro de dispersão diferente para cada resposta (y_i), Lee e Nelder (1998) propõem então a função de quase-verossimilhança estendida (*extended quasi-likelihood*; vide McCULLAGH; NELDER, 1989) (Equação 8):

$$-2Q^+ = \sum_{i=1}^n \left\{ \frac{d_i}{\phi_i} + \ln [2\pi\phi_i V(y_i)] \right\} \quad (8)$$

em que d_i é o quase-deviance, dado por $d_i = 2 \int_{\mu_i}^{y_i} \frac{y_i - \mu_i}{V(t)} dt$ e

$V(t)$ é a função de variância para um valor t da média de y_i .

Para estimar o vetor β dos coeficientes do modelo, procede-se à maximização da função de quase-verossimilhança estendida (QVE), obtendo-se as equações-escore (Equação 9):

$$\sum_{i=1}^n \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_i = 0 \quad (9)$$

3.3.2 Máxima verossimilhança restrita

Em experimentos altamente fracionados (como no exemplo), nos quais o número de dados é pequeno em relação ao número de parâmetros do modelo para a média, os estimadores dos parâmetros do modelo de dispersão podem ter vieses acentuados, devido à escassez de graus de liberdade para modelagem da dispersão (LEE; NELDER, 1998). Neste caso, procedimentos de ajuste devem ser adotados para o modelo da média. Os autores citados propõem uso da técnica de máxima verossimilhança restrita (MVR) com a maximização da seguinte função (Equação 10),

$$-2Q_B = \sum_{i=1}^n \left\{ \frac{d_i^*}{\phi_i} + \ln [2\pi\phi_i V(y_i)] \right\} \quad (10)$$

em que $d_i^* = d_i / (1 - h_i)$ e, por sua vez, h_i , $i = 1, 2, \dots, n$, são os elementos da diagonal da “matriz chapéu” H , que é a matriz de projeção dos valores ajustados sobre os valores observados para a média. Então, $\hat{y} = Hy$. A matriz H é dada por (MCCULLAGH; NELDER, 1989) Equação 11,

$$H = W^2 X (X'WX)^{-1} X'W^2 \quad (11)$$

em que a matriz W é diagonal, com os elementos dados por (Equação 12),

$$w_i = \frac{1}{\text{var}(\mu_i)} \left(\frac{d\mu_i}{d\eta_i} \right)^2 \quad (12)$$

Em síntese, o ajuste indicado consiste apenas em substituir d_i por d_i^* , na função de quase-verossimilhança estendida, ou seja, substituir a expressão (3) por (5).

3.3.3 Modelo para a dispersão

O modelo proposto para a dispersão é (Equação 13),

$$\xi_i = h(\phi_i) = z_i' \gamma \quad (13)$$

Neste modelo, a variável de resposta é o resíduo quase-deviance d_i . Para representar d_i , Lee e Nelder (1998) recomendam a distribuição gama, com parâmetro de dispersão igual a 2. Portanto, supõe-se que $E(d_i) = \phi_i$ e $\text{var}(d_i) = 2\phi_i^2$. Os vetores x , do modelo da média, e z podem ter elementos comuns. Ou seja, fatores que afetam a média podem afetar ou não a dispersão e vice-versa. Lee e Nelder (1998) afirmam que a escolha usual da função de ligação é $h(\phi_i) = \ln(\phi_i)$, ou seja, eles consideram que os efeitos dos fatores sobre a dispersão são multiplicativos. Na Tabela 3, é apresentado um resumo dos elementos dos modelos para a média e para a dispersão.

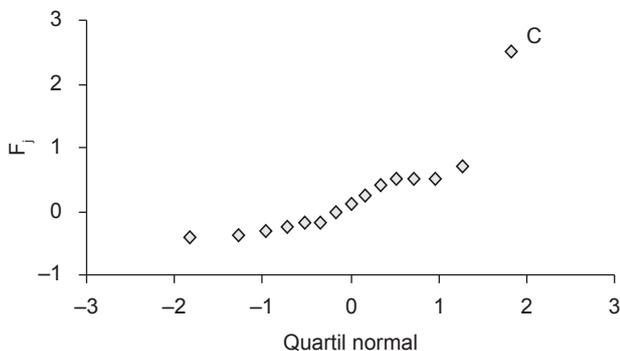


Figura 6. Gráfico de Probabilidade Normal de F_j – Modelo M1.

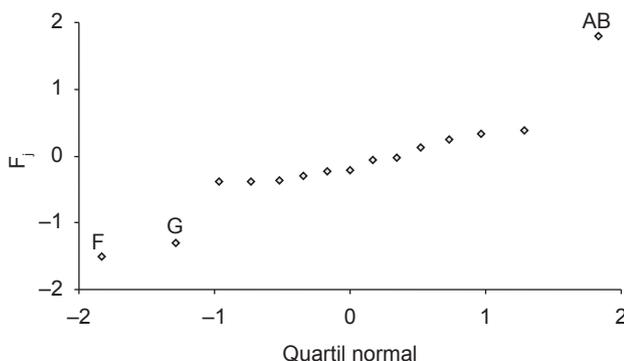


Figura 7. Gráfico de Probabilidade Normal de F_j – Modelo M2.

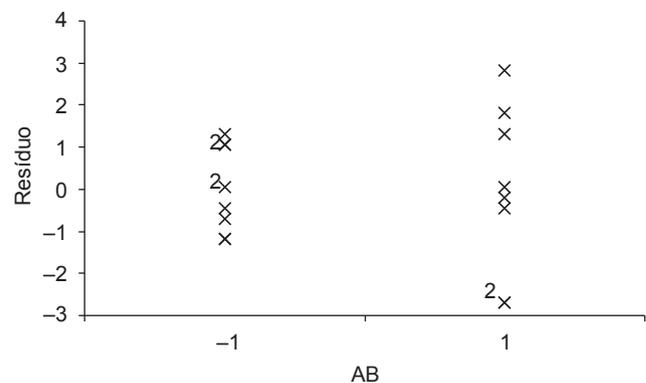


Figura 8. Gráfico dos Resíduos versus o Nível da Interação AB . Obs.: os números “2” indicam que o ponto corresponde a dois resíduos de igual valor.

3.3.4 Procedimento iterativo

Para a modelagem da média e da dispersão, temos os seguintes passos:

- a) Ajuste do modelo da média. Escolher a função de ligação, a função de variância e selecionar os coeficientes significativos. Os graus de liberdade restantes serão usados para o ajuste do modelo da dispersão. Para a primeira iteração, o parâmetro de dispersão ϕ_i é considerado constante. Nas iterações seguintes, usa-se $1/\phi_i$, proveniente do modelo da dispersão, para estimar os coeficientes com as equações de quase-verossimilhança estendida. O desvio $d_i^* = d_i/(1 - h_i)$ resultante do ajuste do modelo será a resposta do modelo da dispersão.
- b) Ajuste do modelo da dispersão. Com d_i^* como resposta, a escolha comum é a distribuição gama com função de ligação logarítmica e parâmetro de dispersão igual a 2. Podem-se usar os métodos já descritos para identificar os fatores que podem afetar a dispersão. O valor ajustado ϕ_i resultante será usado no modelo da média.
- c) O processo iterativo termina quando os parâmetros do modelo da dispersão são iguais em duas iterações consecutivas, a menos de uma tolerância.

Exemplo (continuação)

Neste exemplo, foi usado um experimento fatorial 2^{7-3} com sete fatores A-G sem replicações. Se considerarmos o Modelo M1 como o adequado, os fatores A e B afetam a média e temos indicação de que o fator

C afeta a dispersão. Tratamos então os outros fatores como sendo inertes, de modo que podemos considerar um experimento 2^3 com duas replicações. Na Tabela 4, apresentamos este experimento.

Vamos considerar os modelos para a média e para a dispersão (Equação 14),

$$y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB + \epsilon \tag{14}$$

em que $E(\epsilon) = 0$ e $\text{var}(\epsilon) = \phi = \exp(\gamma_0 + \gamma_1 A + \gamma_2 B + \gamma_3 C)$, sendo ϕ com distribuição gama.

Como agora temos replicações, podemos estimar a variância de cada resposta e usá-la para fazer o ajuste de um modelo independentemente do modelo da média. Na Tabela 5, apresentamos a estimativa da variância de cada resposta.

Sem nenhuma análise formal, observamos que a variância de cada resposta entre os valores de y é muito menor no nível -1 do fator C. Isto confirma as indicações dos métodos descritos nas seções 3 e 4, de que C tem um efeito considerável na dispersão. Na Tabela 5, temos que a estimativa da variância da resposta 4 é zero. Para este valor da resposta, o valor da função de log-verossimilhança da distribuição gama é infinito, o que torna inviável a estimativa dos coeficientes do modelo. Entretanto podemos atribuir um valor muito pequeno para esta resposta, digamos 0,01, para contornar este problema e poder ajustar o modelo.

Na Tabela 6, apresentamos as estimativas dos coeficientes para o modelo da dispersão, fornecidos pelo software ARC.

Tabela 3. Modelos para a média e para a dispersão.

Componente	Modelo	
	Média	Dispersão
Resposta	y	$d^* = d/(1-h)$
Valor esperado da resposta	μ	ϕ
Variância da resposta	$\phi V(\mu)$	$2\phi^2$
Função de ligação	$\eta = g(\mu)$	$\zeta = h(\mu) = \ln \phi$
Preditor linear	$\eta = x'\beta$	$\zeta = z'\gamma$
Deviance	d^*	$2\{-\ln(d^*/\phi) + (d^* - \phi)/\phi\}$
Peso	$1/\phi$	$1-h$

Tabela 4. Experimento 2^3 com duas replicações.

A	B	C	y_{i1}	y_{i2}
-1	-1	-1	6	8
1	-1	-1	10	12
-1	1	-1	32	34
1	1	-1	60	60
-1	-1	1	4	16
1	-1	1	15	5
-1	1	1	26	37
1	1	1	60	52

Tabela 5. Estimativa da variância da resposta.

A	B	C	S ²
-1	-1	-1	2.0
1	-1	-1	2.0
-1	1	-1	2.0
1	1	-1	0.0
-1	-1	1	72.0
1	-1	1	50.0
-1	1	1	60.5
1	1	1	32.0

Tabela 6. Coeficientes e erro-padrão do modelo para a dispersão.

	Estimativa	Erro-padrão	Est./Erro-padrão
Constante	2,09321	0,468632	4,46664
A	-0,41828	0,468632	-0,89255
B	-0,37885	0,468632	-0,80843
C	1,88688	0,468632	4,02636

Observando a tabela, vemos que apenas C é significativo. Ajustando o modelo apenas com C , obtemos o resultado, fornecido pelo software *ARC* (Equação 15):

$$\hat{\phi} = \exp(2,19 + 1,79C) \quad (15)$$

Com este modelo, podemos estimar os valores de ϕ para cada resposta e utilizar $1/\hat{\phi}$ como peso para ajustar o modelo da média. O modelo ajustado, fornecido pelo software *ARC*, foi (Equação 16),

$$\hat{y} = 27,73 + 7,71A + 18,70B + 5,76AB \quad (16)$$

Entretanto, podemos aplicar o método de Lee e Nelder (1998) para modelar a média e a variância, obtendo modelos similares aos das Equações (8) e (9). Para dar início ao procedimento iterativo, ajusta-se o modelo da média sem considerar as replicações. Para os cálculos do modelo da média, usamos a planilha *Excel* (apenas as operações com matrizes) e, para o modelo da dispersão, usamos o software *Arc*. Adotamos a técnica de máxima verossimilhança restrita, uma vez que temos 6 coeficientes para estimar e apenas 16 observações. Houve convergência após 4 iterações. Na Tabela 7, apresentamos as estimativas dos coeficientes para o modelo da média.

Na Tabela 8, apresentamos as estimativas dos coeficientes para o modelo da dispersão.

Os modelos para a média e para a dispersão são, portanto (Equações 17 e 18):

$$\hat{y} = 27,71 + 7,68A + 18,67B + 5,77AB \quad (17)$$

$$\hat{\phi} = \exp(1,95 + 1,57C) \quad (18)$$

Observe que as equações (9) e (10) são bem parecidas, bem como as equações (8) e (11). Portanto, os modelos da média e da dispersão obtidos pelos dois procedimentos devem produzir as mesmas estimativas da resposta. No modelo da dispersão, os resultados obtidos pelos dois procedimentos estão de acordo com a indicação de que o fator C afeta a dispersão. Pode haver casos em que haja desacordo entre os dois procedimentos na indicação do fator de dispersão. Nesses casos, uma explicação possível é que as medidas de variabilidade dos dois procedimentos medem variabilidades de diferentes tipos ou de fontes diferentes (MCCULLAGH; NELDER, 1989). Entretanto, para estimar o efeito do fator C sobre a média e sobre a dispersão, o modelo da dispersão ajustado com a variância das observações replicadas é mais confiável porque o valor esperado da variável de resposta deste modelo é a própria variância da resposta, enquanto que a estimativa de ϕ a partir dos resíduos do modelo da média é uma estimativa com viés, cujo efeito se manifesta principalmente sobre o termo constante (LEE; NELDER, 1988).

Continuando a análise da média e da dispersão no Exemplo, vamos agora considerar o Modelo M2 (Equação 19).

$$y = \beta_0 + \beta_1A + \beta_2B + \beta_3AB + \beta_4C + \beta_5G + \beta_6CG + \varepsilon \quad (19)$$

Como foi detectado que F , G e AB afetam a dispersão, o modelo para a dispersão é (Equação 20),

$$\text{var}(\varepsilon) = \phi = \exp(\gamma_0 + \gamma_1F + \gamma_2G + \gamma_3AB) \quad (20)$$

Houve convergência após 6 iterações. Na Tabela 9, apresentamos as estimativas dos coeficientes para o modelo da média.

Na Tabela 10, apresentamos as estimativas dos coeficientes para o modelo da dispersão. Observamos que o fator G não deve ser considerado significativo para a dispersão porque o quociente entre a estimativa e o erro-padrão de seu coeficiente tem módulo menor do que 1. O modelo para a dispersão passa a ser $\phi = \exp(\gamma_0 + \gamma_1F + \gamma_2AB)$

Repetindo o processo, após 5 iterações, temos os seguintes modelos para a média e para a dispersão (Equações 21 e 22):

$$\hat{y} = 27,05 + 7,53A + 17,77B + 6,02AB - 0,56C - 2,57G - 2,57CG \quad (21)$$

Tabela 7. Coeficientes e erro-padrão do modelo da média.

	Estimativa	Erro-padrão	Est./Erro-padrão
Constante	27,7139	0,4188	66,17
A	7,6829	0,4188	18,34
B	18,6726	0,4188	44,58
AB	5,7655	0,4188	13,77

Tabela 8. Coeficientes e erro-padrão do modelo da dispersão.

	Estimativa	Erro-padrão	Est./Erro-padrão
Constante	1,95373	0,3396	5,75
C	1,57280	0,3396	4,63

Tabela 9. Coeficientes e erro-padrão do modelo da média.

	Estimativa	Erro-padrão	Est./Erro-padrão
Constante	27,4466	0,5366	51,15
A	7,1289	0,5337	13,36
B	17,8070	0,5337	33,37
C	-0,5567	0,4394	-1,27
G	-2,5765	0,5152	-5,00
AB	6,0462	0,6258	9,66
CG	-2,5761	0,4394	-5,86

Tabela 10. Coeficientes e erro-padrão do modelo da dispersão.

Coefficiente	Estimativa	Erro-Padrão	Est./Erro-padrão
Constante	1,1536	0,3505	0,438
F	0,7708	0,3505	2,199
G	0,1876	0,3505	-0,535
AB	0,8832	0,3505	2,377

$$\phi = \exp(0,17 - 0,80F + 0,83AB) \quad (22)$$

3.4 Otimização do processo de moldagem

Para otimizar o processo, é desejável minimizar a contração média \hat{y} do molde, bem como a sua dispersão. Vamos considerar inicialmente o Modelo M1 e, em seguida, o Modelo M2.

3.4.1 Considerando o Modelo M1

Na Seção 5.3, obtivemos os seguintes modelos para a média e para a dispersão (Equação 23):

$$\begin{aligned} \hat{y} &= 27,73 + 7,71A + 18,70B + 5,76AB \\ \hat{\phi} &= \exp(2,19 + 1,79C) \end{aligned} \quad (23)$$

Podemos minimizar a média com $A = -1$ e $B = -1$, e minimizar a variância com $C = -1$. A estimativa da contração média será então $\hat{y} = 27,73 - 7,71 - 18,70 + 5,76 = 7,08$.

O efeito de diminuir o fator C de $+1$ para -1 é reduzir a variância em $\exp(1,79) = 5,99$ unidades.

3.4.2 Considerando o Modelo M2

Na Seção 5.3 obtivemos os seguintes modelos para a média e para a dispersão (Equação 24):

$$\begin{aligned} \hat{y} &= 27,05 + 7,53A + 17,77B + 6,02AB - \\ &- 0,56C - 2,57G - 2,57CG \\ \phi &= \exp(0,17 - 0,80F + 0,83AB) \end{aligned} \quad (24)$$

O efeito de aumentar o fator F de -1 para $+1$ é reduzir a variância em $\exp(-0,80) = 0,45$.

Se A e B tiverem sinais contrários, a redução na variância será $\exp(-0,83) = 0,44$, em relação a quando A e B tiverem sinais iguais. Como já vimos, para otimizar o processo, é desejável minimizar a contração média \hat{y} do molde e minimizar $\text{var}(y)$. Com o modelo acima, podemos minimizar a média com $A = -1$, $B = -1$, $G = +1$ e $C = +1$, e minimizaríamos a variância com $F = +1$ e a interação $AB = -1$ (o que só pode ser conseguido com A e B de sinais diferentes). Temos então um conflito para atingir os objetivos.

Uma maneira de analisar esta situação é minimizar a dispersão com $F = +1$ e a interação $AB = -1$. Como o coeficiente de B no modelo da média é maior do que o de A , com $A = +1$ e $B = -1$ conseguimos uma menor contração média mantendo $G = +1$ e $C = +1$. O resultado é uma contração média estimada em 4,53. Se essa contração média não for satisfatória, podemos estabelecer um valor alvo para a média e, então, resolver o seguinte problema:

Minimizar $0,83AB$

Sujeito a:

- $\hat{y} = 27,44 + 7,15A + 17,77B + 6,02AB - 0,56 - 2,57 - 2,57 = \text{um valor alvo}$

- $-1 \leq A \leq 1$

- $-1 \leq B \leq 1$

4 Comentários e recomendações

Nesta seção, faremos comentários e recomendações sobre os métodos descritos na Seção 3, comparando a sua eficácia e amplitude de aplicabilidade.

4.1 Método gráfico

Evidentemente, o método gráfico descrito na Seção 3.1 não é um teste formal, podendo em muitos casos não ser conclusivo. Entretanto, em muitos casos, ele pode fornecer alguma evidência dos fatores que afetam a dispersão, que pode ser confirmada pela aplicação do método iterativo descrito na Seção 3.3.

4.2 Método Box e Meyer

Como foi visto, o método de Box e Meyer (1986) pode levar à identificação incorreta dos efeitos sobre a dispersão se o modelo para a média não for bem ajustado.

Brenneman e Nair (2001) analisaram, além do método de Box e Meyer (1986), vários outros métodos não iterativos. Vamos denominar todos esses métodos de "métodos diretos". Eles são aplicáveis apenas em experimentos fatoriais com dois níveis, completos (2^k) ou fracionados (2^{k-p}), e só admitem, para a modelagem da média, os modelos lineares, com estimação pelo método de mínimos quadrados.

Brenneman e Nair (2001), concluíram, com resultados de estudos de simulação, que *todos* os métodos diretos apresentam algum viés, que depende do modelo ajustado para a média e do número de efeitos ativos na dispersão. Essas simulações evidenciaram que, *mesmo que o modelo para a média seja o correto*, esses métodos podem identificar efeitos espúrios na dispersão. Entretanto, esses métodos podem fornecer alguma evidência dos fatores que afetam a dispersão, que pode ser confirmada pela aplicação do método iterativo descrito na Seção 3.3.

4.3 Método iterativo de Lee e Nelder

Brenneman e Nair (2001) analisaram também os métodos iterativos, concluindo pela superioridade destes em relação aos métodos diretos. Eles asseguram que o método de Lee e Nelder (1998) com máxima verossimilhança restrita (MVR) é superior aos demais. Ademais, enquanto os métodos diretos, como foi visto, são aplicáveis apenas em experimentos fatoriais com dois níveis e só consideram, para a modelagem da média, modelos lineares com estimação pelo método dos mínimos quadrados, o método iterativo de Lee e Nelder (1998) é aplicável também a experimentos com vários níveis, e utiliza modelos lineares generalizados (MLG), o que

permite considerar um número muito maior de possibilidades de modelagem. Portanto, recomendamos o método iterativo descrito na Seção 3.3.

5 Conclusões

Neste artigo foram apresentados e analisados os principais métodos de identificação de fatores que afetam a dispersão da resposta, aplicados a experimentos fatoriais sem replicações. A análise permitiu concluir pela superioridade do método iterativo de Lee e Nelder (1998) sobre os métodos diretos e sobre o método gráfico. Isto porque aquele método é aplicável a experimentos com vários níveis para cada fator e utiliza MLG para a modelagem da média, permitindo, portanto, considerar um número muito maior de possibilidades de modelagem. Ele é, portanto, o método geral a recomendar. Além disso, os demais métodos mostraram que podem levar à identificação incorreta dos efeitos sobre a dispersão. Por outro lado, eles podem fornecer alguma evidência de fatores que afetem a dispersão; de modo que são úteis se empregados em conjunto com o método de Lee e Nelder, mas não devem ser utilizados isoladamente.

Dado o fato de que os métodos de identificação de fatores aqui analisados, como grande parte dos métodos de análise estatística, não podem ser automatizados, e requerem a interação, o julgamento e a intervenção humana, torna-se difícil, senão impossível, realizar metaexperimentos (experimentos de uso de métodos de análise de experimentos) com um grande número de repetições, para poder fazer afirmações sobre a significância das conclusões obtidas. O procedimento usual na literatura especializada para a análise de desempenho de um ou mais de tais métodos é o que aqui foi adotado: aplicar o(s) método(s) a um ou dois problemas-exemplo que sirva(m) para ilustrar ou diagnosticar seu poder e suas limitações (o exemplo aqui utilizado já foi usado por vários autores em seus artigos). Seria então interessante, e fica aqui indicado como possibilidade para trabalhos futuros, aplicar os métodos a outros conjuntos de dados, reais ou gerados aleatoriamente segundo modelos com distintas características, para tornar a análise mais abrangente. Ainda que a presente análise tenha evidenciado deficiências de alguns métodos (e, neste aspecto, seja conclusiva), a experimentação com um conjunto mais vasto de casos pode indicar em que situações tais métodos são mais (ou menos) propensos à falha.

Methods for the identification of dispersion effects in unreplicated factorial experiments

Abstract

The identification of factors that affect the location and dispersion of the probability distribution of quality characteristics is essential in the optimization of production processes and product design. For this purpose, factorial experiments are typically used. The analysis of dispersion effects, however, usually requires replications of the experiment, which can be expensive. Therefore, several methods have been proposed for the identification of the dispersion effects with unreplicated factorial experiments. In this article, we analyze some of these methods concluding by the superiority of an interactive method based on generalized linear models and which provides a model for the mean together with a model for the dispersion of the quality characteristic. Finally, the use of the resulting model in an optimization process is presented through an example.

Keywords: Factorial experiments; Generalized linear models; Process optimization

Referências Bibliográficas

- BERGMAN, B.; HYNÉN, A. Dispersion Effects from Unreplicated Designs in the 2^{k-p} Series. **Technometrics**, v. 39, n. 2, p. 191-198, 1997.
- BOX, G. E. P.; MEYER, R. D. Dispersion Effects from Factorial Designs. **Technometrics**, v. 28, n. 1, p. 19-27, 1986.
- BRENNEMAN, W.; NAIR, V. N. Methods for Identifying Dispersion Effects in Unreplicated Factorial Experiments: A Critical Analysis and Proposed Strategies. **Technometrics**, v. 43, n. 4, p. 388-405, 2001.
- ENGEL, J. E.; HUELE, A. F. A Generalized Linear Modelling Approach to Robust Design. **Technometrics**, v. 38, n. 4, p. 365-373, 1996.
- GREGO, J. M. Generalized Linear Models and Process Variation. **Journal of Quality Technology**, v. 25, n. 4, p. 288-295, 1993.

- HAMADA, M. E.; BALAKRISHNAN, N. Analyzing Unreplicated Factorial Experiments: A Review with Some New Proposals (with discussion). **Statistica Sinica**, v. 8, n. 1, p. 1-38, 1998.
- LEE, Y.; NELDER, J. A. Generalized Linear Models for the Analysis of Quality Improvement Experiments. **The Canadian Journal of Statistics**, v. 26, n. 1, p. 95-105, 1998.
- McGRATH, R. N.; LIN, D. K. Confounding of Location and Dispersion Effects in Unreplicated Fractional Factorials. **Journal of Quality Technology**, v. 33, n. 2, p. 129-139, 2001.
- MCCULLAGH, P. E.; NELDER, J. A. **Generalized Linear Models**. London: Chapman-Hall, 1989.
- MONTGOMERY, D. C. **Design and Analysis of Experiments**. 5 ed. New York, NY: John Wiley & Sons, 2001.
- MYERS, R. H.; MONTGOMERY, D. C. **Response Surface Methodology**. 2 ed. New York, NY: Wiley & Sons, 2002.
- MYERS, R. H.; MONTGOMERY, D. C.; VINING, G. G. **Generalized Linear Models with Applications in Engineering and the Sciences**. New York, NY: John Wiley & Sons, 2002.
- PAN, G. The Impact of Unidentified Location Effects on Dispersion Effects From Unreplicated Factorial Designs. **Technometrics**, v. 41, n. 4, p. 313-326, 1999.

Sobre os autores

Antonio Fernando de Castro Vieira

Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio),
Rua Marquês de São Vicente, 225, Gávea, CEP 22453-900, Rio de Janeiro, RJ, Brasil
e-mail: afcv@puc-rio.br

Eugenio Kahn Epprecht

Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio),
Rua Marquês de São Vicente, 225, Gávea, CEP 22453-900, Rio de Janeiro, RJ, Brasil
e-mail: eke@puc-rio.br

Agradecimentos: O segundo autor agradece o apoio do CNPq.

Recebido em 13/2/2006
Aceito em 29/12/2008