

Identificação de efeitos de dispersão em experimentos fatoriais 2^K e 2^{K-p}

VIVIANE LEITE DIAS DE MATTOS

Prof.^a na Universidade Católica de Pelotas e Doutoranda no Programa de Pós-Graduação em Engenharia de Produção/UFSC.
e-mail: *vldm@atlas.ucpel.tche.br*

PEDRO ALBERTO BARBETTA

Prof. no Departamento de Informática e Estatística/UFSC.
e-mail: *barbetta@inf.ufsc.br*

ROBERT WAYNE SAMOBYL

Prof. no Departamento de Engenharia de Produção/UFSC.
e-mail: *samohyl@deps.ufsc.br*

Resumo

Este artigo apresenta uma estratégia para identificar efeitos de locação e efeitos de dispersão em experimentos fatoriais não-replicados, fracionados ou não. A estratégia é adaptada para projetos experimentais com poucas replicações, sendo ilustrada em um exemplo prático. Além da estratégia proposta, outros métodos típicos de experimentos com replicações foram aplicados nesse exemplo e os resultados encontrados foram bastante parecidos.

Palavras-chave

Efeitos de dispersão, projeto de experimentos, melhoria da qualidade.

Identifying dispersion effects in 2^K e 2^{K-p} factorial experiments

Abstract

This paper comments on a strategy for analyzing location and dispersion effects in unreplicated factorial experiments, concentrating especially on methods for identifying dispersion effects. The strategy is adapted for designs of experiments with few replications. An example is presented. Other methods for replicated experiments are applied in the example and the results found were very similar.

Key words

Dispersion effects, experimental design, quality improvement.

INTRODUÇÃO

Reduzir a variabilidade de um produto ou processo produtivo é fundamental para a obtenção da melhoria da qualidade. Embora Shewhart e Deming já tivessem salientado a necessidade da redução da variabilidade, foi Taguchi quem sugeriu uma metodologia de projeto robusto, por meio da qual produtos e processos tornam-se pouco sensíveis às fontes de variação. De acordo com esta metodologia, realizam-se experimentos planejados, procurando-se identificar os fatores que afetam a média (fatores com efeito de locação) e os fatores que afetam a variabilidade (fatores com efeito de dispersão) do processo. Em muitos casos, também se busca encontrar uma combinação ótima, que aproxima ao máximo a média do processo ao valor-alvo e, ao mesmo tempo, minimiza a variabilidade.

Na indústria, durante as últimas décadas, tem sido dada muita importância à identificação de efeitos de dispersão em experimentos, principalmente em experimentos não-replicados ou com poucas replicações, por serem mais rápidos e econômicos.

De acordo com Montgomery (1997, p.1), um experimento é um teste ou uma série de testes nos quais as variáveis de entrada (fatores controláveis) de um sistema são manipuladas para que sejam identificadas as razões das mudanças nas variáveis de saída (características funcionais do produto ou processo). As variáveis de saída também são chamadas de *respostas*.

Os experimentos normalmente envolvem vários fatores. Uma maneira eficiente de combinar os níveis dos diversos fatores consiste em *cruzá-los* de tal forma que cada nível de um fator seja combinado com todos os níveis dos outros fatores, caracterizando os chamados *projetos fatoriais*. Com a realização de um único projeto fatorial, podem-se avaliar os efeitos sobre a resposta de cada fator e das interações entre fatores.

Quando cada um dos K fatores de um projeto fatorial é ensaiado em apenas dois níveis, o projeto experimental é dito do tipo 2^K . Esse tipo de projeto exige um número relativamente pequeno de ensaios e é adequado quando a quantidade de fatores é elevada. Um projeto 2^K costuma servir de base para outros projetos mais complexos.

Quando se têm muitos fatores, mesmo que eles sejam ensaiados em apenas dois níveis, o número de ensaios necessários ainda pode ser muito grande. Nesse caso, pode-se fazer uso dos projetos fracionários do tipo 2^{K-p} ,

onde K é o número de fatores e p é o número de fracionamentos, resultando em 2^{K-p} combinações de níveis dos fatores efetivamente ensaiados. A idéia dos projetos fracionados é ensaiar apenas parte das possíveis combinações de níveis, sendo planejados de tal forma que garantam a possibilidade de se estimar os efeitos principais (de cada fator isoladamente) e, às vezes, as interações entre dois fatores. Esse tipo de projeto é particularmente importante quando se pretende fazer uma *triagem* de fatores para serem usados em estudos posteriores.

Na análise dos dados de um experimento, bons resultados podem ser obtidos se as técnicas clássicas da experimentação incorporarem as idéias de Taguchi, que relacionam qualidade com variabilidade. Neste contexto, procuram-se identificar os fatores que apresentam efeitos de dispersão por afetar a variabilidade da resposta, além dos tradicionais fatores com efeitos de locação, que afetam o valor esperado da resposta (a Figura 1 ilustra os dois tipos de efeito).

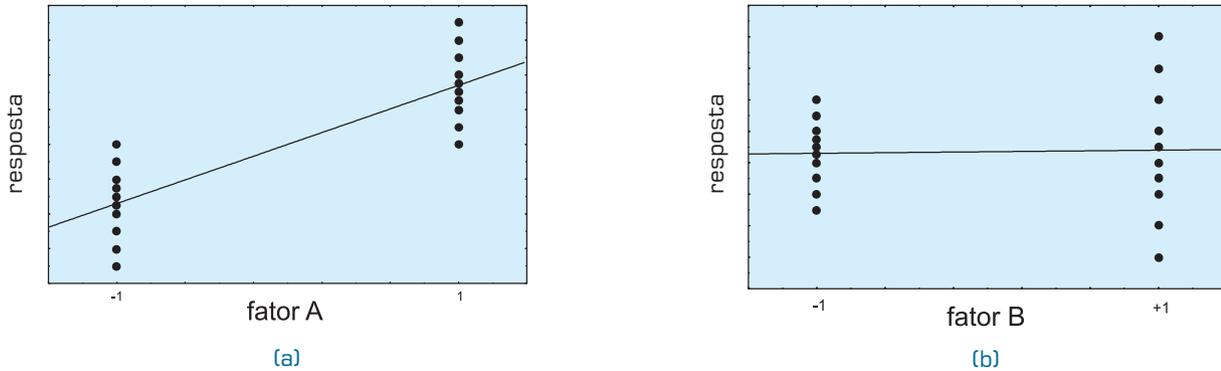
Diz-se que um fator tem *efeito de locação ativo* se o valor esperado da resposta for diferente em seus níveis superior e inferior. Um fator apresenta *efeito de dispersão ativo* se a variância da resposta em seu nível superior for diferente da variância da resposta em seu nível inferior.

Em embora Shewhart e Deming já tivessem mostrado a necessidade do entendimento e redução da variabilidade, foi Taguchi quem sugeriu a metodologia de projeto robusto.

Em experimentos com replicações (mais de um ensaio em cada ponto experimental), a identificação dos efeitos de locação pode ser feita através de técnicas estatísticas formais, como o teste t de *Student* ou o teste F da análise de variância. Em experimentos não-replicados é comum usar métodos menos formais, como o gráfico de Pareto e o gráfico de probabilidade normal, pois não existe uma estimativa confiável da variância do erro experimental, que é essencial para a aplicação de um teste estatístico. Nestes processos gráficos, identificam-se efeitos aparentemente significativos.

Embora o método de Taguchi seja bastante controverso, elogiado por uns e criticado por outros, ele foi muito importante para a Engenharia da Qualidade (KACKAR, 1986; BOX, 1988; MONTGOMERY, 1997). Ele conseguiu, com grande objetividade e precisão, detectar e delimitar alguns problemas desta área. Contudo, aparentemente, não teve o mesmo sucesso com algumas técnicas

Figura 1: Representação gráfica de fator com efeito de locação (a) e fator com efeito de dispersão (b) em experimento fictício.



cas propostas para operacionalizar e buscar a solução dos problemas.

Uma das maiores contribuições de Taguchi foi a redefinição de qualidade, relacionando-a com a idéia de perda imposta por um produto à sociedade, descrita matematicamente por meio de uma função quadrática. Dada uma característica funcional y e o seu valor ideal τ , a função perda mede a qualidade em termos da variação de y em relação a τ . Usando uma expansão em série de Taylor, Taguchi *et al.* (1990, p.18) justificam o uso da função perda quadrática, definida da seguinte forma:

$$L(y) = k(y - \tau)^2 \text{ para } y \in [\tau - \delta, \tau + \delta] \quad (1)$$

onde k é uma constante de proporcionalidade, fazendo com que a função L assumira valores em termos monetários, e $[\tau - \delta, \tau + \delta]$ é o intervalo de especificação da característica y .

A equação (1) representa um tipo de característica funcional conhecida como *nominal-é-melhor*, pois existe um valor real τ ideal. Taguchi também definiu a função perda para características funcionais em que se deseja o maior resultado possível (*maior-é-melhor*) ou o menor resultado possível (*menor-é-melhor*) (ver TAGUCHI *et al.*, 1990).

Num programa de melhoria da qualidade, sendo esta medida em termos da função perda quadrática, torna-se natural o objetivo de minimizar seu valor esperado. Fazendo-se a extensão do domínio de $L(y)$ para todos os reais, tem-se a seguinte expressão para a perda esperada:

$$\mu_L = E\{L(y)\} = k[(\mu_y - \tau)^2 + \sigma_y^2] \quad (2)$$

onde μ_y e σ_y^2 são o valor esperado e a variância de y , respectivamente. Fica evidente, pela expressão (2), que a

menor perda esperada é obtida quando se ajusta a média do processo (μ_y) ao valor-alvo (τ) e, simultaneamente, se minimiza a variância do processo (σ_y^2).

Taguchi também chamou a atenção para a necessidade da preocupação com a qualidade iniciar durante a fase de projeto, enfatizando a importância de um trabalho preventivo, além de conscientizar a indústria das vantagens da utilização de projetos de experimentos, bem como do controle, pelo menos indireto, de fatores de ruído para obtenção de produtos e processos robustos.

O objetivo deste artigo é apresentar diversos métodos para identificar efeitos de dispersão, dando ênfase a uma estratégia originalmente proposta por Brenneman (2000) e Brenneman e Nair (2001) para experimentos fatoriais não-replicados. Na segunda seção estão apresentados os principais métodos utilizados para identificar efeitos de dispersão, assim como a estratégia anteriormente mencionada. A terceira seção apresenta um exemplo prático, estendendo a estratégia para a situação de experimentos com poucas replicações. Finalizando, na última seção estão as considerações finais.

MÉTODOS PARA IDENTIFICAÇÃO DE EFEITOS DE DISPERSÃO

Os primeiros métodos para identificar efeitos de dispersão foram desenvolvidos para experimentos com replicações, sendo a variabilidade avaliada por meio da variância amostral, calculada em cada ponto experimental. Embora estes métodos apresentem a vantagem de não depender do ajuste de algum modelo para a média do processo, suas eficiências são baixas em experimentos com poucas replicações, conforme mostra o artigo clássico de Bartlett e Kendall (1946). Por outro lado, o uso de muitas replicações (acima de nove, como sugerido por Bartlett e Kendall, 1946) aumenta demasiadamente o custo do projeto experimental.

Taguchi propôs a utilização de uma estatística de desempenho, denominada razão sinal/ruído, para mensurar a variabilidade provocada por fatores de ruído. Se a situação analisada for do tipo *nominal-é-melhor*, a estatística, de acordo com Vinning e Myers (1990), é calculada pela expressão:

$$\frac{S}{R} = 10 \log \left(\frac{\bar{y}}{s_y^2} \right) \quad (3)$$

onde $\frac{S}{R}$ é a razão sinal/ruído e \bar{y} e s_y^2 são, respectivamente, a média e a variância das observações de cada ponto experimental.

Para a exposição de outros métodos descritos na literatura, considere os dados do experimento, y_{ij} ($i = 1, 2, \dots, N$ e $j = 1, 2, \dots, J$), satisfazendo o seguinte modelo linear:

$$y_{ij} = x_{ij}'\beta + \sigma_i e_{ij} \quad (4)$$

$$g(\sigma_i^2) = z_i'\theta \quad (5)$$

onde β e θ são vetores de parâmetros; x_i' e z_i' são as linhas das matrizes de planejamento; e_{ij} é o erro experimental associado à observação y_{ij} ; g é uma função de ligação para a variância; N é o número de pontos experimentais ensaiados e J é o número de replicações. Supostamente e_{ij} tem distribuição normal com média nula e variância unitária. A equação (4), quando vista em termos do valor esperado, é chamada de *modelo da média* ou *modelo de locação*; e a equação (5) de *modelo da variância*. A função g pode ser a função identidade (*modelo linear*), mas é mais comum ser a logarítmica (*modelo log-linear*). A transformação logarítmica na equação (4) induz aditividade e estabiliza a variância (Nair e Pregibon, 1988).

Em experimentos replicados ($J > 1$), além dos procedimentos já apresentados, podem ser usados os métodos denominados S e R por Nair e Pregibon (1988), baseados no modelo log-linear para a variância. No método R é calculado o logaritmo da razão entre médias aritméticas de variâncias amostrais nos diferentes níveis de cada fator, enquanto que no método S , calcula-se o logaritmo da razão entre médias geométricas das variâncias amostrais. Assim, o efeito do k -ésimo fator é calculado pelas seguintes expressões (métodos R e S , respectivamente):

$$D_k^R = \frac{1}{2} \left(\log \sum_{i(k+)} s_i^2 - \log \sum_{i(k-)} s_i^2 \right) = \frac{1}{2} \log \left(\frac{\sum_{i(k+)} s_i^2}{\sum_{i(k-)} s_i^2} \right) \quad (6)$$

$$D_k^S = \frac{1}{N} \left(\sum_{i(k+)} \log s_i^2 - \sum_{i(k-)} \log s_i^2 \right) = \frac{1}{N} \log \left(\frac{\prod_{i(k+)} s_i^2}{\prod_{i(k-)} s_i^2} \right) \quad (7)$$

onde: s_i^2 é a variância amostral no i -ésimo ponto experimental ($i = 1, 2, \dots, N$);

$i(k+)$ indica ensaios realizados no nível superior do k -ésimo fator;

$i(k-)$ indica ensaios realizados no nível inferior do k -ésimo fator ($k = 1, 2, \dots, K$).

Os efeitos das interações são calculados de forma análoga.

O método S é resultante de uma análise por mínimos quadrados no logaritmo das variâncias amostrais, conforme sugere o procedimento clássico de Bartlett e Kendall (NAIR e PREGIBON, 1988).

Para experimentos não-replicados ($J = 1$), os métodos baseiam-se na análise dos resíduos de um modelo ajustado aos dados (*modelo da média*). Numa fase inicial, geralmente o modelo (4) é ajustado aos dados experimentais através do método dos mínimos quadrados ordinários (MQO).

Sendo \hat{y}_i o valor predito pelo modelo no i -ésimo ponto experimental e y_i o valor observado neste ponto ($i = 1, 2, \dots, N$), calculam-se os resíduos por:

$$r_i = y_i - \hat{y}_i \quad (8)$$

A identificação dos efeitos de dispersão é feita através dos resíduos e, portanto, tem o inconveniente de depender da qualidade da modelagem do valor médio. No que segue, são discutidos métodos baseados nos resíduos.

Harvey (1976) fez uma adequação da abordagem de Bartlett e Kendall (1946) para estudar heterocedasticidade em análise de regressão. Sua idéia pode ser usada em experimentos não-replicados, substituindo as variâncias amostrais por resíduos quadráticos. Sob a suposição de um modelo log-linear para a variância, a avaliação de cada efeito é feita pela razão entre médias geométricas. Mais especificamente, a estatística é dada por:

$$D_k^H = \frac{1}{N} \left(\sum_{i(k+)} \log r_i^2 - \sum_{i(k-)} \log r_i^2 \right) = \frac{1}{N} \log \left(\frac{\prod_{i(k+)} r_i^2}{\prod_{i(k-)} r_i^2} \right) \quad (9)$$

onde r_i ($i = 1, 2, \dots, N$) são os resíduos encontrados a partir do modelo da média, conforme equação (8).

Box e Meyer (1986) propuseram estudar os efeitos de dispersão a partir da razão entre médias aritméticas (o que implicitamente supõe o modelo log-linear para a variância). Para o k -ésimo fator, a estatística deste método é dada por:

$$D_k^{BM} = \frac{1}{2} \left(\log \sum_{i(k+)} r_i^2 - \log \sum_{i(k-)} r_i^2 \right) = \frac{1}{2} \log \left(\frac{\sum_{i(k+)} r_i^2}{\sum_{i(k-)} r_i^2} \right) \quad (10)$$

O método de Box e Meyer (1986) é análogo ao método R de Nair e Pregibon (1988) [vide equação (6)], substituindo as variâncias amostrais por resíduos quadráticos. Com a mesma substituição, o método de Harvey (1976) é análogo ao método S de Nair e Pregibon (1988), [vide equação (7)].

Outro método foi descrito por Wang (1989) para projetos fatoriais em geral. Para projetos dos tipos 2^k e 2^{k-p} , a estatística de Wang tem a forma:

$$D_k^W = \frac{1}{2N} \left(\sum_{i(k+)} \frac{r_i^2}{\hat{\sigma}^2} - \sum_{i(k-)} \frac{r_i^2}{\hat{\sigma}^2} \right) \quad (11)$$

onde $\hat{\sigma}^2$ é uma estimativa da variância, obtida pelo método usual de um modelo linear (supondo variância constante). Sob a hipótese de um modelo log-linear, com apenas um efeito ativo e dados provenientes de uma distribuição normal, a estatística D_k^W tem distribuição Qui-quadrado com um grau de liberdade. Porém, a suposição de que existe apenas um efeito de dispersão significativo, em geral, não é razoável em termos práticos.

Bergman e Hynén (1997) propuseram uma modificação ao método de Box e Meyer (1986). Para reduzir o viés no cálculo dos efeitos de dispersão, eles propõem utilizar os resíduos sob um modelo expandido de média. Para calcular o efeito de dispersão do fator k , é definido um novo conjunto de fatores. Este conjunto deve ser composto pelos elementos identificados inicialmente pelo método dos mínimos quadrados ordinários (MQO), pelo fator k e pelas interações do fator k com os fatores identificados como tendo efeito de locação. Por exemplo, suponha que tenha sido identificado o seguinte conjunto de efeitos significativos no modelo da média: $L = \{I, A, B\}$. Para calcular o efeito de dispersão do fator C , os resíduos são calculados em relação a um modelo com o seguinte conjunto de efeitos: $L_E = \{I, A, B, C, AC, BC\}$. A estatística deste método, relativa ao k -ésimo fator, é dada por:

$$D_k^{BH^{(1)}} = \frac{\sum_{i(k+)} \bar{r}_i^2}{\sum_{i(k-)} \bar{r}_i^2} \quad (12)$$

onde \bar{r}_i ($i = 1, 2, \dots, N$) são os resíduos do modelo expandido da média. Sob a hipótese de que os dados provêm de distribuição normal e que todos os demais

efeitos são nulos, a estatística D_k^{BH} segue uma distribuição de Fisher-Snedecor com (m, m) graus de liberdade, onde m é o número de elementos de L_E . Alternativamente, pode-se utilizar a transformação logarítmica, conforme expressão (13), e identificar os efeitos ativos considerando a aproximação à distribuição normal.

$$D_k^{BH^{(2)}} = \frac{1}{2} \left(\log \sum_{i(k+)} \bar{r}_i^2 - \log \sum_{i(k-)} \bar{r}_i^2 \right) = \frac{1}{2} \log \left(\frac{\sum_{i(k+)} \bar{r}_i^2}{\sum_{i(k-)} \bar{r}_i^2} \right) \quad (13)$$

O cálculo de resíduos a partir de um modelo expandido da média também pode ser feito com o método de Harvey (1976), aqui denominado de método *HM*, conforme sugerem Brenneman (2000) e Brenneman e Nair (2001) [vide equação (14)].

$$D_k^{HM} = \frac{1}{N} \left(\sum_{i(k+)} \log \bar{r}_i^2 - \sum_{i(k-)} \log \bar{r}_i^2 \right) = \frac{1}{N} \log \left(\frac{\prod_{i(k+)} \bar{r}_i^2}{\prod_{i(k-)} \bar{r}_i^2} \right) \quad (14)$$

Após o cálculo dos efeitos de dispersão por métodos previamente discutidos, podem-se colocar os valores em um gráfico de probabilidade normal, conforme sugerido por Daniel (1959) *apud* Montgomery (1997, p.318 e p.335), para identificar efeitos aparentemente significativos (ou ativos). Usando o método de Box e Meyer (1986), Montgomery observou que os fatores e interações que não têm efeito de dispersão apresentam distribuição aproximadamente normal com média zero e variância constante, razão pela qual ficam ao longo de uma linha reta sobre o gráfico de probabilidade normal. Os fatores ou interações que apresentam efeitos de dispersão não têm média zero e, portanto, distanciam-se da linha reta formada pelos outros fatores e interações.

A partir de uma reflexão teórica e avaliação de resultados por meio de simulações, Brenneman (2000) e Brenneman e Nair (2001) realizaram uma análise crítica dos métodos descritos anteriormente e desenvolveram uma estratégia para identificar efeitos de dispersão em projetos do tipo 2^{k-p} não-replicados.

Sobre os métodos eles concluíram:

- o método proposto por Box e Meyer (1986) funciona bem apenas quando existe um único fator com efeito de dispersão; caso contrário, apresenta *viés estrutural* e *viés de estimação*. O viés de estimação tende a se reduzir com o aumento do tamanho do experimento (mais pontos experimentais), mas o mesmo não acontece com o viés estrutural;
- o método proposto por Bergman e Hynén (1997) pode apresentar o mesmo tipo de viés estrutural. Além disso, a estatística criada por estes autores só terá uma distribui-

- ção F se a hipótese nula considerar todos os efeitos de dispersão nulos e não apenas o efeito do fator analisado;
- o método proposto por Wang (1989) é viesado para modelos de dispersão log-linear, além de não seguir uma distribuição Qui-quadrado quando existir mais de um efeito de dispersão ativo;
 - o método proposto por Harvey (1976) apresenta viés que depende da estrutura do efeito de dispersão ativo e da adequação do modelo, diminuindo com o aumento do tamanho do experimento. Entretanto, se o conjunto dos fatores com efeito de locação (L) for fechado e apresentar menos do que $N/2$ elementos, então esta estatística não apresentará viés para fatores que não pertençam ao conjunto fechado ($D - L$), onde D representa o conjunto de fatores com efeito de dispersão. Se, por exemplo, ($D \subset L$), os estimadores de todos os efeitos de dispersão são não-viesados. Brenneman (2000) mostrou que a estatística utilizada por este método pode subestimar o efeito de dispersão de um fator k se este fator e suas respectivas iterações com os fatores que apresentam efeito de locação (L) não fazem parte do modelo de locação utilizado;
 - o método de Harvey modificado corrige em parte o viés, ao usar o modelo expandido de locação, empregando o fator analisado (k) e suas iterações com os elementos de L nas diversas modelagens da média, além dos fatores ativos para efeitos de locação.

Nos métodos analisados, os autores constataram a existência de um certo grau de viés, que pode ser decorrente da inadequação do modelo de locação, bem como da presença de múltiplos efeitos de dispersão. Estes métodos podem identificar não apenas falsos efeitos de dispersão, como também podem subestimar os existentes.

A partir desta análise crítica, Brenneman (2000) e Brenneman e Nair (2001) sugerem um algoritmo para identificar efeitos de locação e efeitos de dispersão em projetos experimentais do tipo 2^{K-p} não-replicados, conforme os seguintes passos:

- 1) ajustar um modelo de locação pelo método dos mínimos quadrados ordinários (MQO), identificando o conjunto L de efeitos de locação;
- 2) se L contém um conjunto fechado L' com $N/2$ ou mais elementos, então a identificação dos efeitos de dispersão é difícil, sendo necessário aumentar o número de ensaios (alterar o projeto experimental);
- 3) se não, usar o método de Harvey modificado para estimar os efeitos de dispersão e identificar os ativos (por exem-

plo, por meio do gráfico de probabilidade normal). Seja D^0 o conjunto de efeitos de dispersão ativos identificados nesta etapa;

- 4) incluir o efeito k no conjunto dos efeitos, mesmo quando foi identificado como não-ativo, mas corresponde à interação de dois outros efeitos, k_1 e k_2 , identificados como ativos, isto é, $k = k_1 * k_2$, onde $*$ indica o operador de interação;
- 5) construir uma equação para a variância com os efeitos de dispersão D^0 , usando o modelo log-linear [$\log(\sigma_i^2) = z_i\theta$], onde as estimativas dos elementos de θ podem ser obtidas por MQO;
- 6) refazer o modelo de locação usando mínimos quadrados generalizados (MQG), sendo a cada observação atribuído um peso fornecido pelo inverso da variância predita pelo modelo definido no passo (5);
- 7) calcular novos resíduos e novos efeitos de dispersão. Efeitos de dispersão não identificados como ativos poderão ser eliminados e as estimativas refeitas;
- 8) repetir os passos (5), (6) e (7) até que o processo se estabilize.

Alguns autores sugerem apenas uma ou duas iterações devido ao problema da convergência na presença de valores discrepantes. De acordo com Hooper (1993), valores muito pequenos como estimativas das variâncias geram pesos muito grandes, o que faz com que as estimativas das iterações seguintes sejam piores.

Uma discussão mais teórica sobre a construção de modelos para a variância está descrita em Carroll e Ruppert (1988).

Na análise dos experimentos, bons resultados podem ser obtidos se as técnicas clássicas da experimentação incorporarem as idéias de Taguchi, que relacionam qualidade com variabilidade.

Em experimentos com poucas replicações ($2 \leq J \leq 4$), o método clássico de Bartlett e Kendall é pouco eficiente e este trabalho mostra a possibilidade de usar o procedimento proposto por Brenneman e Nair (2001). Neste caso, têm-se JN observações, produzindo JN resíduos. Sugere-se que sejam calculadas as médias aritméticas, em cada ponto experimental, com os quadrados dos resíduos, resultando em N resíduos quadráticos médios. A justificativa de usar resíduos quadráticos médios é feita por Barbetta *et al.* (1999). Foi constatado, por meio de simulações de Monte Carlo, que agregar fatores não-

significativos e trabalhar com os resíduos quadráticos médios fornece estimadores mais eficientes de efeitos de dispersão do que trabalhar com resíduos quadráticos individuais.

EXEMPLO

Para ilustrar a estratégia proposta, foram utilizados dados de um experimento desenvolvido por Silva (2002). Trata-se de uma central de comutação telefônica para serviços especializados, tais como: auxílio à lista, hora certa, teledespertador automático, etc. O experimento foi realizado com a ajuda de um simulador e tinha como objetivo avaliar o efeito dos fatores: número de troncos (*A*), número de filtros digitais (*B*), número de canais de máquina de mensagem (*C*) e número de atendentes (*D*) no tempo de resposta de uma consulta.

A identificação de efeitos de dispersão iniciou-se em experimentos com replicações, sendo a variabilidade avaliada por meio da variância amostral.

O número de troncos refere-se à quantidade de linhas ligadas à rede telefônica e determina a quantidade de usuários que podem ser atendidos simultaneamente por qualquer um dos serviços oferecidos. O filtro digital é um dispositivo responsável pelo reconhecimento das cifras que o usuário discar a fim de navegar no *menu* de um serviço solicitado e só pode ser utilizado por um único usuário. Já os canais de máquina de mensagem são responsáveis pela geração das mensagens requisitadas pelos usuários, só podendo ser utilizado um canal por mensagem. O número de atendentes refere-se à quantidade de posições de atendimento personalizado.

O tempo de resposta (*Y*), medido em segundos, é o intervalo de tempo desde que o usuário gerou a chamada até o momento que obteve acesso à informação solicitada.

Tabela 1 mostra os níveis de cada fator, que foram cruzados de acordo com um projeto fatorial 2^4 , com quatro replicações. Os níveis inferiores de cada fator correspondem às condições atuais de uma certa central no horário de maior movimento, enquanto que os níveis superiores correspondem ao que pode ser feito com determinado investimento. O experimento simulado foi realizado para subsidiar o estudo da capacidade do sistema.

Tabela 2 apresenta os resultados do experimento na ordem natural da seqüência dos níveis dos fatores. (A fonte original não cita se a ordem dos ensaios foi aleatorizada ou não, embora, em experimentos simulados por computador, normalmente a aleatorização não seja fundamental para a validade do experimento.)

Para identificar possíveis efeitos de locação e dispersão, foi utilizada a estratégia descrita na Seção 2. Como esta estratégia foi originalmente proposta para experi-

mentos não-replicados (BRENNEMAN e NAIR, 2001), tornou-se necessária uma adaptação: em cada ponto experimental foram calculados quatro resíduos quadráticos, referentes às quatro re-

plicações, que foram, posteriormente, substituídos pela média aritmética. A avaliação dos efeitos de dispersão foi realizada pelo método Harvey (*H*) e pelo método Harvey modificado (*HM*), sendo empregado o gráfico de probabilidade normal para identificação dos efeitos de dispersão ativos.

Na primeira iteração, buscou-se identificar os efeitos de locação ativos. Da análise da variância na Tabela 3, estão identificados os seguintes fatores com efeitos de locação ativos: *A*, *B* e *D*, além das interações *AB*, *AD* e *BD*. Assim, o conjunto de fatores com efeitos de locação é dado por: $L = \{I, A, B, D, AB, AD \text{ e } BD\}$, onde *I* representa a média geral.

Aplicando o método dos mínimos quadrados ordinários (*MQO*) aos dados, obteve-se o modelo inicial para a média:

Tabela 1: Descrição dos níveis dos fatores envolvidos no experimento.

FATORES	NÍVEL INFERIOR (-1)	NÍVEL SUPERIOR (+1)
Troncos (<i>A</i>)	1320 unidades	1800 unidades
Filtros (<i>B</i>)	176 unidades	240 unidades
Canais (<i>C</i>)	900 unidades	1260 unidades
Atendentes (<i>D</i>)	520 unidades	760 unidades

Fonte: Silva (2002).

$$\hat{\mu}_y = 52,50 + 5,68A - 0,07B - 6,59D - 0,06AB - 1,97AD - 0,09BD \quad (15)$$

Calculando-se os dezesseis resíduos quadráticos médios e aplicando (9), foi avaliada a existência de efeitos de dispersão pelo método *H* por meio do gráfico de probabilidade normal [Figura 2 (a)]. Foram sendo identificados os efeitos de dispersão ativos *A* e *AD*. Assim, pelo método *H*, tem-se o conjunto inicial de efeitos de dispersão: $D^0 = \{I, A, AD\}$, onde *I* representa a média geral.

Para o método de Harvey modificado, *HM*, é necessário expandir o modelo da média. Essa expansão deve ser diferente para cada efeito de dispersão a ser estimado. Entretanto, como foram consideradas apenas as interações de segunda ordem, não houve muita variação entre os termos utilizados nos modelos, que se reduziram a dois tipos, identificados na Tabela 4. Para cada modelo são calculados os dezesseis resíduos quadráticos médios. Aplicando (14), estima-se o efeito de dispersão. A Figura 2(b) apresenta o gráfico de probabilidade normal desses efeitos, o qual sugere o mesmo conjunto $D^0 = \{I, A, AD\}$ obtido pelo método *H*.

Tabela 2: Resultado do projeto experimental e estatísticas do tempo de resposta.

FATORES					TEMPO DE RESPOSTA (EM SEGUNDOS)					
PE	A	B	C	D	REPLICAÇÃO				MÉDIA	DESVIO PADRÃO
					I	II	III	IV		
1	-1	-1	-1	-1	51,414	51,576	51,330	51,443	51,441	0,102
2	-1	-1	-1	1	42,221	42,310	42,224	42,164	42,230	0,060
3	-1	-1	1	-1	51,414	51,576	51,330	51,443	51,441	0,102
4	-1	-1	1	1	42,221	42,310	42,224	42,164	42,230	0,060
5	-1	1	-1	-1	51,414	51,576	51,330	51,443	51,441	0,102
6	-1	1	-1	1	42,240	42,185	42,213	42,044	42,171	0,087
7	-1	1	1	-1	51,414	51,576	51,330	51,443	51,441	0,102
8	-1	1	1	1	42,240	42,185	42,213	42,044	42,171	0,087
9	1	-1	-1	-1	66,575	66,869	66,594	66,809	66,712	0,149
10	1	-1	-1	1	49,622	50,229	49,951	49,930	49,933	0,248
11	1	-1	1	-1	66,575	66,869	66,594	66,809	66,712	0,149
12	1	-1	1	1	49,622	50,229	49,951	49,930	49,933	0,248
13	1	1	-1	-1	66,809	66,869	66,881	66,535	66,774	0,162
14	1	1	-1	1	49,051	49,419	49,408	49,399	49,319	0,179
15	1	1	1	-1	66,809	66,869	66,881	66,535	66,774	0,162
16	1	1	1	1	49,051	49,419	49,408	49,399	49,319	0,179

OBS: PE = ponto experimental; A = tronco; B = filtro; C = canal; D = atendente.

Fonte: Silva (2002).

Tabela 3: Resultados de testes F para identificação dos fatores de locação (apenas os efeitos significativos ao nível de significância de 5%).

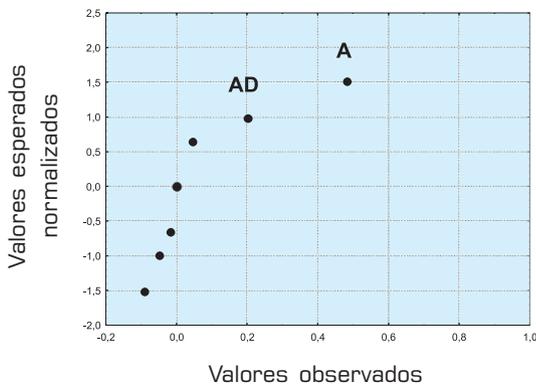
FATORES	SOMA DOS QUADRADOS	GRAUS DE LIBERDADE	QUADRADOS MÉDIOS	F	p-valor
A	2066,839	1	2066,839	93996,2	0,00000
B	0,369	1	0,369	16,8	0,00016
D	2779,135	1	2779,135	126390,1	0,00000
AB	0,243	1	0,243	11,0	0,00172
AD	248,141	1	248,141	11285,0	0,00000
BD	0,537	1	0,537	24,4	0,00001
Falta de ajuste	0,381	9	0,042	1,9	0,07037
Erro puro	1,055	48	0,022		
Total	5096,700	63			

Tabela 4: Composição do conjunto L_E , no modelo expandido da média, para estimativa dos efeitos de dispersão.

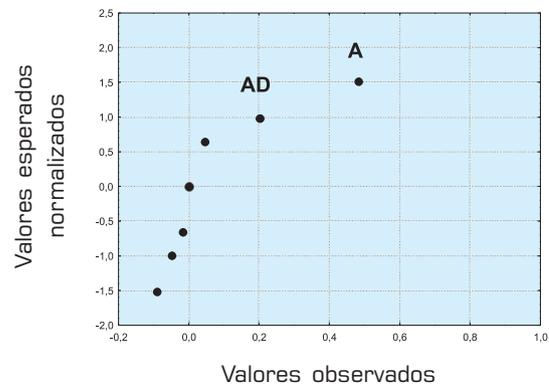
EFEITO OU INTERAÇÃO	COMPOSIÇÃO DO CONJUNTO L_E
A	I, A, B, D, AB, AD e BD
B	I, A, B, D, AB, AD e BD
C	I, A, B, C, D, AB, AD, AC, BD, BC e CD
D	I, A, B, D, AB, AD e BD
AB	I, A, B, D, AB, AD e BD
AC	I, A, B, C, D, AB, AD, AC, BD, BC e CD
AD	I, A, B, D, AB, AD e BD
BC	I, A, B, C, D, AB, AD, AC, BD, BC e CD
BD	I, A, B, D, AB, AD e BD
CD	I, A, B, C, D, AB, AD, AC, BD, BC e CD

OBS: Foram desconsideradas interações de grau superior a dois.

Figura 2: Gráficos de probabilidade normal para efeitos de dispersão medidos pelos métodos (a) H e (b) HM .



(a)



(b)

As diferenças entre os resíduos quadráticos (método H) e os resíduos quadráticos modificados (método HM) foram infinitesimais, assim, adotou-se o método H pela sua relativa simplicidade.

Os efeitos da interação AD e do fator A foram identificados como ativos. Conforme passo (4) do algoritmo, o efeito do fator D também deve ser incluído ao conjunto, resultando em $D^0 = \{I, A, D \text{ e } AD\}$. A partir dele, a modelagem foi feita nos logaritmos neperianos dos resíduos quadráticos médios (aplicando MQO), resultando no modelo:

$$\hat{\sigma}_y^2 = \exp\{-3,96 + 0,48A + 0,05D + 0,20AD\} \quad (16)$$

A partir da equação (16) é possível estimar a variância em cada uma das condições experimentais. Os modelos podem ser refinados usando o método dos mínimos quadrados generalizados (MQG) para estimar um novo modelo para a média. Cada observação é pondera-

da pelo inverso da variância estimada pela expressão (16) e os novos resíduos permitem reavaliar os efeitos de dispersão.

Os resultados desta segunda análise não modificaram a composição do conjunto L (efeitos de locação). Com o refinamento do modelo, parece ser confirmada a não existência de efeito de dispersão no fator D e acentuada a importância do efeito de dispersão da interação AD e, principalmente, do efeito do fator A , conforme Figura 3(a). Portanto, um novo conjunto dos fatores com efeitos de dispersão foi gerado de acordo com esta estratégia: $D^1 = \{I, A, AD\}$.

Os resultados de uma terceira iteração não modificaram a composição do conjunto L , além de confirmar a composição do conjunto D , acentuando ainda mais os efeitos de dispersão de A e AD .

Como o experimento tem replicações, foi realizada uma nova análise, empregando os métodos R e S , de Nair e Pregibon (1988). O método S corresponde à abordagem

Figura 3: Gráficos de probabilidade normal para efeitos de dispersão estimados pelo método *H* na segunda (a) e terceira (b) iteração.

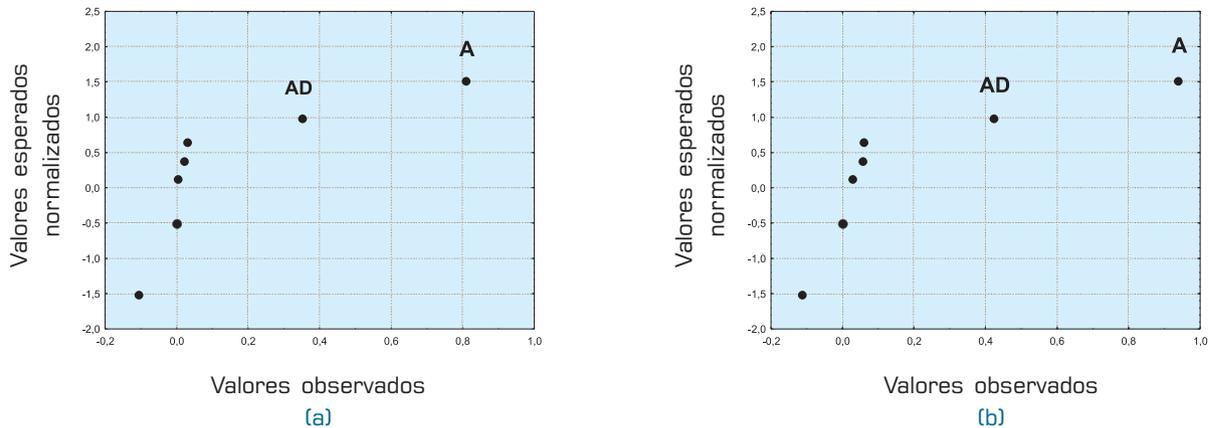
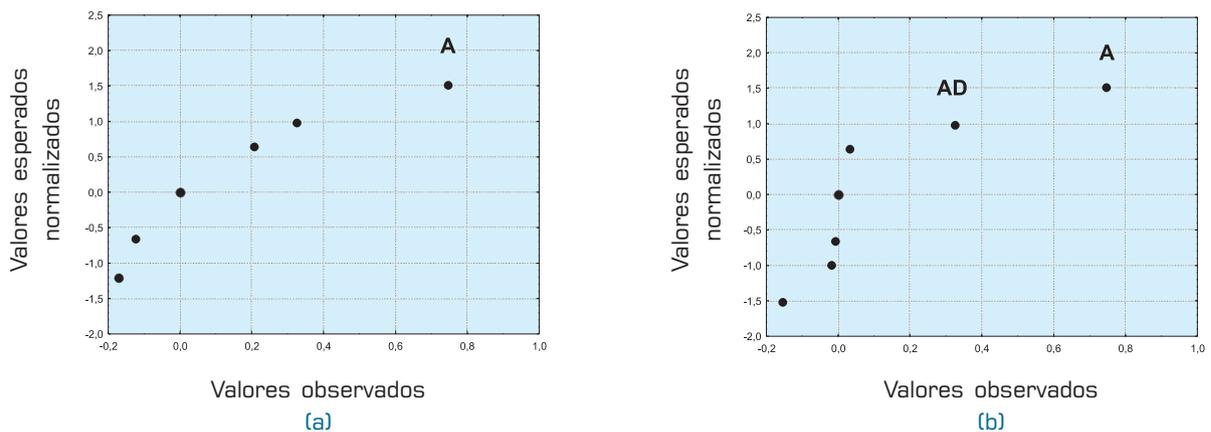


Figura 4: Gráficos de probabilidade normal para efeitos de dispersão estimados pelos métodos (a) *R* e (b) *S*.



clássica de Bartlett e Kendall (1946), em que a dispersão é avaliada em cada ponto experimental através da variância amostral. Esse método poderia ser considerado como o método padrão se houvessem 10 ou mais replicações. Os resultados, mostrados na Figura 4, evidenciam que, de acordo com o método *R*, parece ser ativo apenas o fator *A* [Figura 4 (a)], enquanto que pelo método *S*, os fatores *A* e a interação *AD* [Figura 4 (b)].

Considerando o exposto, há fortes evidências de que o fator *A* (número de troncos) altera a variância do tempo de resposta. Como a interação *AD* apareceu ativa em alguns dos métodos utilizados, então há uma suspeita que a influência do fator *A* no o tempo de resposta possa ser diferente, dependendo do nível do fator *D* (número de atendentes).

CONSIDERAÇÕES FINAIS

A metodologia de análise de efeitos de dispersão proposta por Brenneman (2000) e Brenneman e Nair (2001), para experimentos não-replicados, inicia com a identificação de efeitos de locação pelos métodos clássicos. Usa-se, inicialmente, o método de mínimos quadrados ordinários (*MQO*) para construir um modelo para a média. Sugere o método *HM* (Harvey modificado) para avaliar a existência de efeitos de dispersão. Como alguns efeitos falsos podem ser detectados, existe necessidade de refinar o modelo através do método dos mínimos quadrados generalizados (*MQG*).

O presente artigo mostra a possibilidade de usar esta estratégia em experimentos com poucas replicações.

Nesse caso, tem-se vários resíduos quadráticos em cada ponto experimental e uma sugestão é utilizar a média desses resíduos quadráticos, em cada ponto experimental, para identificar o efeito e modelar a variância.

O procedimento foi ilustrado em um experimento com quatro replicações. Foram usados inicialmente os métodos *H* (Harvey) e *HM* (Harvey modificado) para identificação de efeitos de dispersão. Como os resultados encontrados foram praticamente os mesmos, optou-se pelo método *H* pela sua simplicidade. De acordo com essa estratégia, inicialmente foram detectados o fator *A* e a interação *AD*, concordando com o resultado do método *S*, também denominado por alguns autores de método de Bartlett e Kendall. O refinamento do modelo da média por *MQG*, confirmou o resultado.

Identificar efeitos de dispersão é uma tarefa extremamente difícil, principalmente em experimentos não-re-

plicados ou com poucas replicações. Embora não exista ainda uma teoria consagrada a respeito, a estratégia utilizada pareceu eficiente para a situação de experimentos com poucas replicações, pois os resultados do exemplo, com a aplicação de vários métodos, levaram a resultados parecidos. Embora o exemplo apresentado seja um experimento completo (2^K), os procedimentos são análogos para projetos fracionados (2^{K-p}), com particular atenção ao problema de *confundimentos*, tanto entre os efeitos de locação como entre os efeitos de dispersão.

Salienta-se, ainda, a importância de estudos nesta área, tendo em vista que projetos de experimentos do tipo 2^K ou 2^{K-p} , não-replicados ou com poucas replicações, têm grande aceitação na área industrial por apresentarem custo relativamente baixo. Além disso, são muito utilizados como ponto de partida para estudos mais detalhados.

Artigo recebido em 21/07/2003

Aprovado para publicação em 02/08/2004

■ Referências Bibliográficas

- BARBETTA, P. A. ; RIBEIRO, J. L. D.; BORNIA, A. C. Construção de Modelos para a Variância em Estudos Experimentais. *Produto & Produção*, v. 3, n. 2, p. 56-65, 1999.
- BARTLETT, M. S.; KENDALL, D. G. The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation. *Journal of the Royal Statistics Society*, Ser. B, 8, p.128-138, 1946.
- BERGMAN, B.; HYNÉN, A. Dispersion Effects from Unreplicated Designs in 2^{k-p} Series. *Technometrics*, 39, p.191-198, 1997.
- BOX, G. E. P. Signal-to-noise Ratios, Performance Criteria, and Transformation (com discussão). *Technometrics*, v. 30, n. 1, p. 1 – 40, 1988.
- BOX, G. E. P.; MEYER, R. D. Dispersion Effects from Fractional Design. *Technometrics*, 28, p.19-27, 1986.
- BRENNEMAN, W. A. *Inference for Location and Dispersion Effects in Unreplicated Factorial Experiments*. Tese de doutorado. University of Michigan. 108p. USA, 2000.
- BRENNEMAN, W. A.; NAIR, V. N. Methods for Identifying Dispersion Effects in Unreplicated Factorial Experiments: a Critical Analysis and Proposed Strategies. *Technometrics*, 43, p. 388-405, 2001.
- CARROLL, R. J.; RUPPERT, D. - *Transformation and Weighting in Regression*. USA: Chapman and Hall, 1988. 249 p.
- HARVEY, A. C. Estimating Regression Models with Multiplicative Heterocedasticity. *Econometrica*, v. 44, n. 3, p. 461-465, 1976.
- HOOPER, P. N. Iterative weighted least square estimation in heterocedastic linear models. *Journal of the American Statistical Association*, v. 93, n. 144, p. 36-45, 1998.
- KACKAR, R. N. - Taguchi's Quality Philosophy: Analysis and Commentary. *Quality Progress*, December, p. 21-29, 1986.
- MONTGOMERY, D. C. *Design and Analysis of Experiments*. 4 ed. USA: John Wiley&Sons, 1997. 704 p.
- NAIR, V. N.; PREGIBON, D. Analysing Dispersion Effects from Replicated Factorial Experiments. *Technometrics*, 30, p. 247-257, 1988.
- SILVA, M. V. D. Avaliação de Desempenho de uma Plataforma de Comunicação Telefônica para Serviços Especializados de Atendimento ao Cliente. Dissertação de mestrado do Programa de Pós-Graduação em Ciências da Computação da UFSC. 101 p. Florianópolis, 2002.
- TAGUCHI, G.; ELSAYED, E. A.; HSIANG, T. *Engenharia da Qualidade em Sistemas de Produção*. São Paulo: McGraw-Hill, 1990.
- VINNING, G. G.; MYERS, R. H. Combining Taguchi and Response Surface Philosophies: A Dual Response Approach. *Journal of Quality Technology*, v. 22, n. 1, p.38-45, 1990.
- WANG, P. C. Tests for Dispersion Effects from Orthogonal Array. *Computational Statistics and Data Analysis*, 8, p. 109-117, 1989.

■ Agradecimentos

Este trabalho foi apoiado pelo Núcleo de Normalização e Qualimetria (NNQ) do Programa de Pós-Graduação em Engenharia de Produção (PPGEP) da Universidade Federal de Santa Catarina (UFSC) – <http://www.qualimetria.ufsc.br> e pela Escola de Educação da Universidade Católica de Pelotas – <http://www.ucpel.tche.br>.