



SPOLM 2007

ISSN 2175-6295

Rio de Janeiro- Brasil, 08 e 09 novembro de 2007.

INFLUÊNCIA DA CORRETA IDENTIFICAÇÃO DE EFEITOS DE LOCAÇÃO NO DESEMPENHO DE MÉTODOS IDENTIFICADORES DE EFEITOS DE DISPERSÃO

Viviane Leite Dias de Mattos

Universidade Católica de Pelotas
Félix da Cunha, 412 – Pelotas, RS - Brasil
vldm@atlas.ucpel.tche.br

Pedro Alberto Barbeta

Universidade Federal de Santa Catarina
Bairro Trindade – Florianópolis, SC – Brasil
barbeta@inf.ufsc.br

Suélen dos Santos Hernandes

Universidade Católica de Pelotas
Félix da Cunha, 412 – Pelotas, RS - Brasil
suelen.hernandes@yahoo.com.br

Maurício Mailan Lange

Universidade Católica de Pelotas
Félix da Cunha, 412 – Pelotas, RS - Brasil
mmailange@yahoo.com.br

Renata Peres Dutra

Universidade Católica de Pelotas
Félix da Cunha, 412 – Pelotas, RS - Brasil
renatapdutra@hotmail.com

Resumo

Este estudo foi realizado com o objetivo de avaliar o efeito da correta identificação de efeitos de locação no desempenho de alguns métodos empregados em uma estratégia proposta para identificar efeitos de dispersão em experimentos com poucas replicações. Utilizando simulações de Monte Carlo para gerar respostas de projetos fatoriais do tipo 2^4 , foi possível confirmar o melhor desempenho de métodos que usam resíduos quadráticos em relação aos usuários de variâncias amostrais, mesmo quando os efeitos de locação são determinados a partir dos próprios dados, além do melhor desempenho de métodos que usam média aritmética em relação aos usuários de média geométrica.

Palavras-Chaves: Projeto de experimentos, Efeitos de dispersão, Variabilidade, Qualidade.

Abstract

This study was done to observe the impact of the right identification of location effects in some methods used in a strategy proposed to identify dispersion effects in experiments with few replications. Using Monte Carlo's simulations, it was possible to confirm that the methods that use quadratic residuals are better than the methods that use sample variances, even when the location effects are detected from their own data. Besides, the methods that use arithmetic mean are better than the ones which use geometric mean.

Keywords: Design of Experiments, Dispersion Effects, Variability, Quality.

1. INTRODUÇÃO

Conhecer o comportamento de um sistema produtivo é de fundamental importância num programa de melhoria da qualidade. A metodologia de planejamento estatístico de experimentos é um instrumento bastante poderoso no estudo deste comportamento, possibilitando a identificação de fatores do processo produtivo que afetam a média e a variabilidade de características de qualidade.

Em Mattos (2004) foi proposta uma estratégia para identificar efeitos de dispersão em experimentos com poucas replicações, usados no controle de qualidade *off-line*, sendo avaliado o desempenho de alguns métodos descritos na literatura. Nesse estudo, foram usados métodos propostos para identificar efeitos de dispersão em experimentos replicados, apresentados em Nair e Pregibon (1988), que avaliam a variabilidade a partir de variâncias amostrais, assim como alguns métodos propostos para identificar tais efeitos em experimentos não-replicados, apresentados em Box e Meyer (1986) e Breneman e Nair (2000), que fazem esta avaliação a partir de resíduos quadráticos.

Embora tenha sido feita uma adaptação dos métodos propostos para experimentos não-replicados para a situação de experimentos com replicações, seus desempenhos foram superestimados, pois os fatores com efeitos de locação foram considerados conhecidos *a priori*. De acordo com Pan (1999), em experimentos não-replicados, os procedimentos usuais de identificação de efeitos de locação tanto podem não identificar um efeito existente, como podem identificar algum efeito não existente. Estes equívocos têm conseqüências sobre na identificação de efeitos de dispersão, diminuindo a eficiência dos métodos que fazem sua identificação, embora não afetem sua validade. McGrath e Lin (2001) estudaram o confundimento entre efeitos de locação e efeitos de dispersão em experimentos não-replicados, concluindo que efeitos de dispersão podem produzir correlação entre pares de efeitos de locação, prejudicando a sua correta identificação e, conseqüentemente, a posterior identificação de efeitos de dispersão.

Este trabalho foi realizado com o objetivo de avaliar a influência da correta identificação de efeitos de locação no desempenho dos métodos que identificam efeitos de dispersão e que fazem uso desta informação, em experimentos com poucas replicações, aferindo a estratégia foco do estudo. Usando a metodologia de simulações de Monte Carlo, os métodos foram comparados em termos da sensibilidade em detectar os efeitos verdadeiros e da especificidade, no sentido de não identificar efeitos falsos.

2. FUNDAMENTAÇÃO TEÓRICA

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 do projeto 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 9, como sugerido por Bartlett e Kendall, 1946) aumenta demasiadamente o custo do projeto experimental.

A estratégia proposta em Mattos (2004) tem por objetivo identificar efeitos de

dispersão em experimentos com poucas replicações, mas também pode ser utilizada para construir modelos de locação e dispersão. Inicialmente, ajusta-se um modelo de locação por meio do método dos mínimos quadrados ordinários (MQO), avaliando-se a qualidade deste e a possibilidade de utilização de transformações matemáticas.

Se o modelo de locação estiver bem ajustado aos dados, estimam-se os efeitos de dispersão pelo método BM (BOX E MEYER, 1986). Se forem identificados mais de um efeito de dispersão, refaz-se a análise pelo método H (HARVEY, 1976).

Estes métodos, propostos inicialmente para experimentos não-replicados, baseiam-se na análise dos resíduos de um modelo ajustado aos dados experimentais (*modelo da média*). Sendo \hat{y}_i o valor predito pelo modelo da média no i -ésimo ponto do projeto experimental e y_i o valor observado nesse ponto ($i = 1, 2, \dots, N$), calculam-se os resíduos por:

$$r_i = y_i - \hat{y}_i \quad (1)$$

Em projetos de experimentos fatoriais do tipo 2^K ou 2^{K-p} , Box e Meyer (1986) propuseram estudar os efeitos de dispersão a partir da razão entre médias aritméticas. Para o cálculo do k -ésimo efeito (considerando a tabela de sinais usual de um projeto 2^K ou 2^{K-p}), a estatística deste método é dada por:

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

onde:

$i(k+)$ indica os ensaios com sinal positivo no k -ésimo efeito;

$i(k-)$ indica os ensaios com sinal negativo no k -ésimo efeito ($k = 1, 2, \dots, K$);

r_i ($i = 1, 2, \dots, N$) são os resíduos do modelo da média, equação (1).

Harvey (1976) fez uma adequação da abordagem de Bartlett e Kendall (1946) para estudar heterocedasticidade em análise de regressão. 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:

$$D_k^H = \frac{1}{n} \left(\sum_{i(k+)} \log r_i^2 - \sum_{i(k-)} \log r_i^2 \right) = \frac{1}{n} \log \left(\frac{\prod_{i(k+)} r_i^2}{\prod_{i(k-)} r_i^2} \right) \quad (3)$$

Nesses métodos, a identificação dos efeitos de dispersão é feita por meio dos resíduos, apresentando o inconveniente de depender da qualidade da modelagem do valor médio. Portanto, se houver dúvidas quanto à adequação do modelo de locação, a estratégia proposta em Mattos (2004) sugere estimar os efeitos de dispersão pelo método R (NAIR e PREGIBON, 1988). Se for identificado mais de um efeito de dispersão, refaz-se a análise pelo método S (NAIR e PREGIBON, 1988).

Esses métodos, propostos para experimentos replicados, baseiam-se em variâncias amostrais, as quais são calculadas sobre as replicações de cada ponto do projeto experimental. No método R é calculado o logaritmo da razão entre médias aritméticas das 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 efeito é calculado pelas seguintes expressões (métodos R e S, respectivamente):

$$D_k^R = \frac{1}{2} \left(\log \sum_{i(k+)} s_i^2 - \log \sum_{i(k-)} s_i^2 \right) = \frac{1}{2} \log \left(\frac{\sum_{i(k+)} s_i^2}{\sum_{i(k-)} s_i^2} \right) \quad (4)$$

$$D_k^S = \frac{1}{N} \left(\sum_{i(k+)} \log s_i^2 - \sum_{i(k-)} \log s_i^2 \right) = \frac{1}{N} \log \left(\frac{\prod_{i(k+)} s_i^2}{\prod_{i(k-)} s_i^2} \right) \quad (5)$$

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

..., N).

Em qualquer um dos métodos, a identificação dos efeitos de dispersão é feita pelo gráfico de probabilidade normal dos efeitos. Quando houver efeitos de dispersão aparentemente significativos, pode-se construir um modelo para prever as variâncias e o modelo de locação é feito pelo método dos mínimos quadrados generalizados (MQG), considerando o inverso das variâncias como pesos na expressão dos mínimos quadrados.

3. METODOLOGIA

Este estudo foi realizado seguindo a metodologia de planejamento de experimentos e simulações de Monte Carlo, utilizando como projeto foco um experimento com quatro fatores ($K = 4$): x_1 , x_2 , x_3 e x_4 , ensaiados a dois níveis (-1 e +1), segundo um projeto fatorial completo do tipo 2^4 , com duas ou quatro replicações, contendo efeitos de locação e efeitos de dispersão. As respostas do experimento foram geradas de acordo com uma distribuição normal com média μ_y , descrita por uma função linear, e variância σ_y^2 , descrita por uma função log-linear.

Para o estudo não se restringir a um único contexto, foram consideradas várias alternativas experimentais, tais como: diferente quantidade de efeitos de locação, de efeitos de dispersão e de replicações; variando conforme projeto fatorial fracionado do tipo 2^{6-1} , com seis fatores (codificados por A, B, C, D, E e F e sintetizados no Quadro 1) ensaiados em dois níveis.

Para cada uma das 32 condições experimentais foram geradas respostas para 2.000 experimentos, sendo os efeitos de locação identificados pela já consagrada análise de variância e os efeitos de dispersão, pelos métodos R, S, H e BM, comentados na seção anterior.

Os resultados destas análises foram usadas para definição da sensibilidade, equivalente à proporção de identificações corretas (quociente entre a quantidade de experimentos que identificou corretamente todos os efeitos de dispersão e o total de simulações) e especificidade, equivalente ao complemento da proporção de não ocorrência de identificações falsas (quociente entre a quantidade de experimentos que identificou falsamente pelo menos um efeito de dispersão e o total de simulações).

Fator	Descrição	Nível (-1)	Nível (+1)
A	Fatores com efeitos de locação	β_1 e β_2 (dois efeitos principais)	$\beta_1, \beta_2, \beta_3, \beta_{12}$ e β_{13} (três efeitos principais e duas interações)
B	Intensidade dos efeitos de locação	$\beta_i = 2$ e $\beta_{ij} = 1$ para i e j iguais aos índices especificados no fator A $\beta_i = 0$ e $\beta_{ij} = 0$ para i e j diferentes dos índices especificados no fator A	$\beta_i = 4$ e $\beta_{ij} = 2$ para i e j iguais aos índices especificados no fator A $\beta_i = 0$ e $\beta_{ij} = 0$ para i e j diferentes dos índices especificados no fator A
C	Fatores com efeitos de dispersão	θ_k (um efeito de dispersão), sendo $\theta_2 = 0$	θ_k e $\theta_2 = 0,643$ (dois efeitos de dispersão)
D	Nível de coincidência entre efeito de locação e dispersão	$\theta_k = \theta_i$	$\theta_k = \theta_4$
E	Intensidade dos efeitos de dispersão	$\theta_k = 0,549$ e $\theta_i = 0$ para $i \neq k$ e $i \neq 2$	$\theta_k = 0,896$ e $\theta_i = 0$ para $i \neq k$ e $i \neq 2$
F	Quantidade de replicações	2	4

QUADRO 1 - Descrição dos fatores do projeto de simulação

4. RESULTADOS

A sensibilidade e a especificidade foram analisadas separadamente. Inicialmente foi utilizada uma análise de variância para detectar os efeitos de locação. Após, foram usados contrastes com o objetivo de avaliar o desempenho dos métodos que usam diferentes medidas

para avaliar variabilidade (variância e resíduos quadrático) e dos que fazem uso de diferentes tipos de média para estimar o efeito de dispersão (média aritmética e média geométrica).

Com relação à sensibilidade, foram encontradas evidências de que os métodos apresentam diferentes comportamentos ($F = 0,178$; $gl_1 = 3$; $gl_2 = 85$; $p = 0,00000$), conforme evidenciado na Figura 1, que sugere a existência de superioridade dos métodos H e BM. Além disso, confirma-se a existência de interação entre o método empregado na identificação de efeitos de dispersão com a quantidade de tais efeitos ($F = 0,064$; $gl_1 = 3$; $gl_2 = 85$; $p = 0,00000$), assim como com a quantidade de replicações ($F = 0,100$; $gl_1 = 3$; $gl_2 = 85$; $p = 0,00000$). De acordo com a Figura 2, quando existe apenas um efeito de dispersão, a maior sensibilidade média esperada é para o método BM, seguido dos métodos H e R, enquanto que quando existem dois efeitos de dispersão, a maior sensibilidade média esperada é para o método H, seguido do método BM. Já quando existem menos replicações (duas), a maior sensibilidade média esperada é para o método BM, seguido do método H. Quando existem mais replicações, esta maior sensibilidade é esperada para o método H, seguido do método BM (Figura 3).

Na análise realizada por contrastes, foi possível constatar a diferença no desempenho de métodos que fazem uso de diferentes formas de mensuração da variabilidade ($F = 124,58$; $gl_1 = 1$; $gl_2 = 85$; $p = 0,000000$). Os métodos usuários de resíduos quadráticos (0,58) apresentaram maior sensibilidade em relação aos que fazem uso da variância (0,51). Com relação à utilização de diferentes tipos de média na estimativa do efeito de dispersão, também foram encontradas evidências de diferença significativa ($F = 15,71$; $gl_1 = 1$; $gl_2 = 85$; $p = 0,000153$). Os métodos que utilizam a média aritmética (0,55) apresentaram maior sensibilidade do que os que fazem uso da média geométrica (0,53).

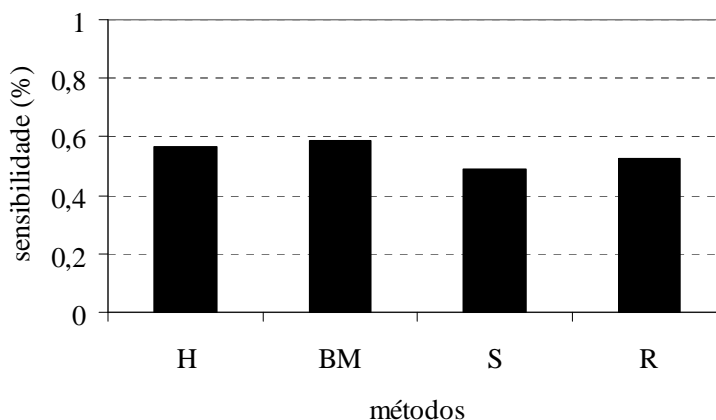


FIGURA 1 – Sensibilidade esperada de alguns métodos que identificam efeitos de dispersão

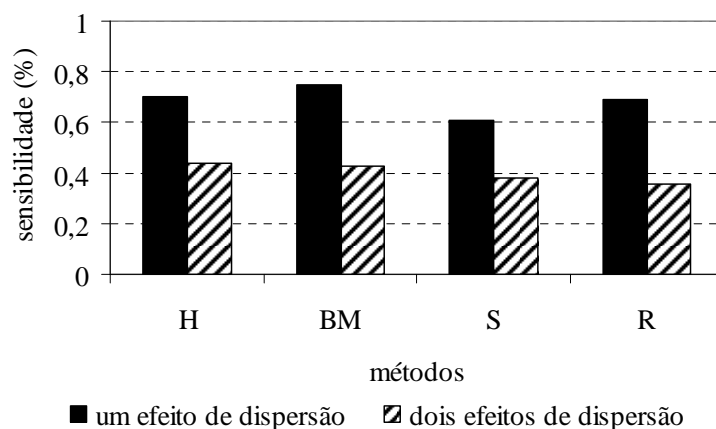


FIGURA 2 – Sensibilidade esperada, por quantidade de efeitos de dispersão, de alguns métodos que identificam efeitos de dispersão

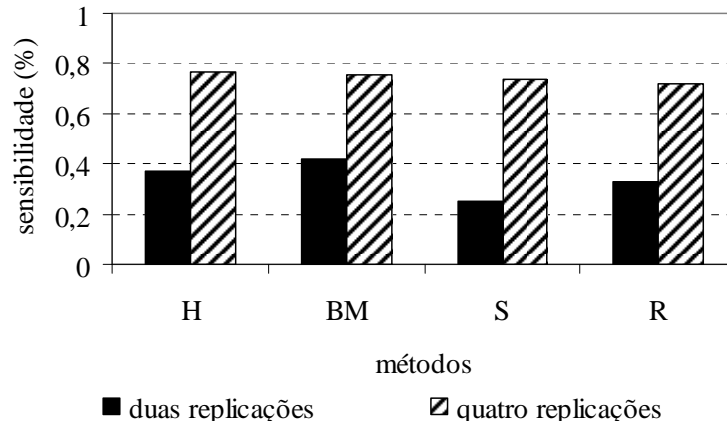


FIGURA 3 – Sensibilidade esperada, por quantidade de replicações, de alguns métodos que identificam efeitos de dispersão

Com relação à especificidade, também foram encontradas evidências de que os métodos apresentam diferentes comportamentos ($F = 0,0298$; $gl_1 = 3$; $gl_2 = 85$; $p = 0,000003$), conforme evidenciado na Figura 4, sugerindo existência de superioridade dos métodos BM e R. Também confirma-se a existência de interação entre o método empregado com a quantidade de efeitos de dispersão ($F = 0,068$; $gl_1 = 3$; $gl_2 = 85$; $p = 0,00000$), assim como com a quantidade de replicações ($F = 0,038$; $gl_1 = 3$; $gl_2 = 85$; $p = 0,00000$), conforme o evidenciado nas Figuras 5 e 6. Quando existem menos efeitos de dispersão, os métodos BM e R apresentam maior especificidade média, enquanto que quando existem mais efeitos, o método H é ligeiramente superior aos demais. Quando são usadas duas replicações, as maiores especificidades médias esperadas são para os métodos BM e R, enquanto que quando forem usadas mais replicações, os métodos apresentem especificidades médias esperadas semelhantes.

Na análise realizada por contrastes, não foi possível constatar a diferença na especificidade de métodos que fazem uso de diferentes formas de mensuração da variabilidade ($F = 0,57$; $gl_1 = 1$; $gl_2 = 85$; $p = 0,451201$). Com relação à utilização de diferentes tipos de média na estimação dos efeitos de dispersão, foram encontradas evidências de diferença significativa ($F = 31,02270$; $gl_1 = 1$; $gl_2 = 85$; $p = 0,000000$): os métodos que utilizam a média aritmética (0,55) apresentaram maior especificidade do que os que fazem uso da média geométrica (0,52).

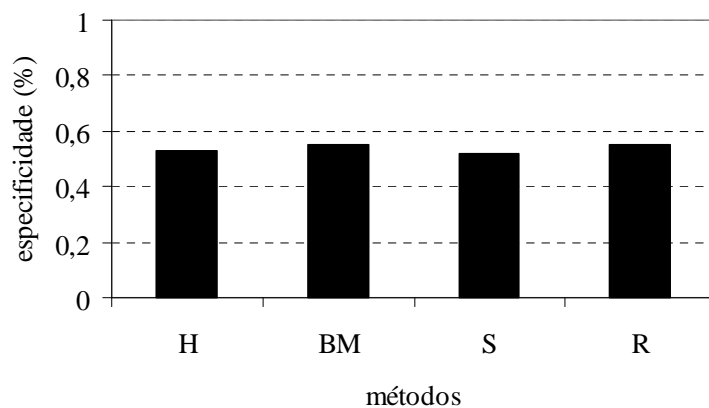


FIGURA 4 – Especificidade esperada de alguns métodos que identificam efeitos de dispersão

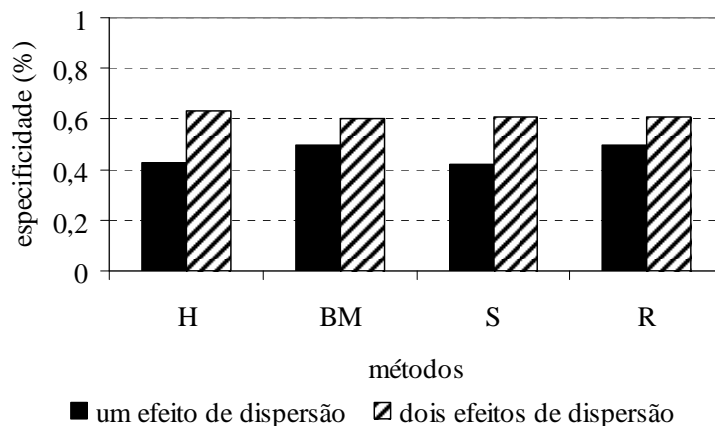


FIGURA 5 – Especificidade esperada, por quantidade de efeitos de dispersão, de alguns métodos que identificam efeitos de dispersão

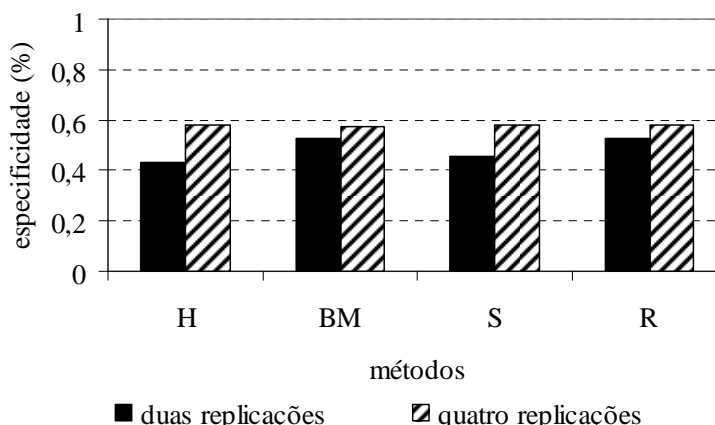


FIGURA 6 – Especificidade esperada, por quantidade de replicações, de alguns métodos que identificam efeitos de dispersão

5. CONSIDERAÇÕES FINAIS

A estratégia definida em Mattos (2004) utilizou simulações de Monte Carlo para avaliar o desempenho de alguns métodos identificadores de efeitos de dispersão. Em experimentos com poucas replicações (menos que cinco), os resultados encontrados sugeriram que os métodos de identificar efeitos de dispersão em experimentos não-replicados, adaptados para a situação de poucas replicações, apresentavam melhor desempenho (maior sensibilidade média esperada e maior especificidade média esperada) em relação aos métodos típicos de experimentos com replicações. Dentre os métodos baseados em resíduos quadráticos, os métodos que usam razões de médias aritméticas apresentaram melhor desempenho.

Entretanto, nessas simulações, os efeitos de locação foram considerados conhecidos *a priori*, fazendo com que o desempenho dos métodos usuários de resíduos quadráticos, identificados como apresentando melhor desempenho, fossem superestimados. No presente estudo foram realizadas simulações com a identificação dos efeitos de locação feitas a partir dos próprios dados, sendo possível confirmar as evidências encontradas de melhor desempenho de métodos que usam resíduos quadráticos em relação àqueles que usam variâncias amostrais e de métodos que usam média aritmética em relação aos que usam média geométrica, especialmente quando existe apenas um efeito de dispersão.

Portanto, os resultados encontrados nessas simulações, em que os efeitos de locação foram determinados *a posteriori*, concordaram com os resultados encontrados em Mattos (2004), em que tais efeitos foram determinados *a priori*, confirmando a estratégia posposta. Tal fato sugere a eficiência do método empregado para identificação de efeitos de locação (Análise de Variância), tendo em vista a importância da correta identificação de efeitos de

locação no desempenho de métodos identificadores de efeitos de dispersão que fazem uso de resíduos quadráticos, conforme mostraram Bergman e Hynén (1987), Pan (1999) e McGrath e Lin (2001), entre outros.

De acordo com Montgomery (1997), a qualidade de um processo produtivo pode ser considerada como inversamente proporcional à sua variabilidade. Esforços no sentido de avaliar o desempenho de métodos que podem ser empregados com o objetivo de minimizá-la, devem ser bem aceitos em toda empresa que tenha a qualidade como uma de suas metas.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] BARBETTA, P.A.; RIBEIRO, J.L.D. & SAMOHYL, R.W. Variance Regression Models in Experiments with Few Replications. *Quality Reliability Engineering*, v.16, p.397-404, 2000.
- [2] BARTLETT, M.S. & KENDALL, D.G. The Statistical Analysis Variance-Heterogeneity and the Logarithmic Transformation. *Journal of the Royal Statistics Society, Ser.B*, 8, p.128-138, 1946.
- [3] BERGMAN, B. & HYNÉN, A. Dispersion Effects from Unreplicated Designs in the 2^k p Series. *Technometrics*, v.39, n.2, 191-198, 1997.
- [4] BOX, G.E.P. & MEYER, R.D. Dispersion Effects From Fractional Design. *Technometrics*, v.28, n.1, 19-27, 1986.
- [5] HARVEY, A.C. Estimating Regression Models with Multiplicative Heterocedasticity. *Econometrica*, v.44, n.3, p.461-465, 1976.
- [6] MATTOS, V.L.D; BARBETTA, P.A. & SAMOHYL, R.W. Identificação de Efeitos de Dispersão em Experimentos Fatoriais 2^K e 2^{K-p} . *Produção*. Vol. 14, n.4, p.36-46, 2004.
- [7] MATTOS, V. L. D. Identificação de Efeitos de Dispersão em Experimentos com Poucas Replicações. Tese (Engenharia de Produção). Florianópolis, 2004. 166p.
- [8] MCGRATH, R.N. & LIN, K.J. Confounding of Location and Dispersion in Unreplicated Fractional Factorials, *Journal of Quality Technology*, v.33, n.1, 129-139, 2001.
- [9] MONTGOMERY, D.C. Design and analysis of experiments. 4 ed. USA: John Wiley&Sons, 1997. 704p.
- [10] NAIR, V. N. & PREGIBON, D. Analysing Dispersion Effects from Replicated Factorial Experiments. *Technometrics*, v.30, n.3, p.247-257, 1988.
- [11] PAN, G. The Impact of Unidentified Location Effects on Dispersion-Effects Identification From Unreplicated Factorial Designs. *Technometrics*, v.41, n.4, 313-326, 1999.