Considerações sobre o uso de resíduos quadráticos para identificar efeitos de dispersão em projetos experimentais dos tipos 2^k e 2^{k-p}.

Viviane Leite Dias de Mattos (UFSC e UCPel) vldm@atlas.ucpel.tche.br
Pedro Alberto Barbetta (UFSC) barbetta@inf.ufsc.br
Robert Wayne Samohyl (UFSC) samohyl@eps.ufsc.br

Resumo

Este trabalho faz algumas considerações sobre a utilização de resíduos quadráticos em métodos não interativos para identificar efeitos de dispersão em experimentos não replicados, estendendo-os à situação de experimentos com poucas replicações. Matematicamente foi possível concluir que, nos métodos que empregam quociente de médias aritméticas (Box e Meyer e Bergman e Hynén), tanto faz trabalhar com todos os resíduos quadráticos ou com as respectivas médias em cada ponto experimental. Entretanto, nos métodos que empregam quociente entre médias geométricas (Harvey e Harvey Modificado), a utilização das médias dos resíduos quadráticos conduzirá a um melhor desempenho.
Palavra Chave: Efeitos de dispersão; Projeto de experimentos; Melhoria da qualidade.

1 Introdução

A importância do estudo da variabilidade nos programas de melhoria da qualidade foi reconhecida a partir da grande ênfase dada por Taguchi à minimização da variabilidade. Embora Shewhart e Deming já tivessem salientado a importância do entendimento e redução desta variabilidade, o fizeram pela remoção de causas assinaláveis ou causas especiais que interferissem no processo. Foi Taguchi quem introduziu a necessidade de desenvolvimento de produtos e processos denominados robustos, insensíveis a qualquer fonte de variação.

A experimentação estatística desempenha importante papel na obtenção deste objetivo. De acordo com esta metodologia, experimentos são planejados para que sejam identificados os fatores que afetam a média ou a locação e os fatores que afetam a variabilidade ou a dispersão do processo.

Este estudo foi realizado com o objetivo de identificar qual é a melhor maneira de considerar os resíduos quadráticos no cálculo de algumas estatísticas empregadas para detectar os fatores que apresentam efeitos de dispersão por afetar a variabilidade de um processo.

2. Efeitos de dispersão

Chama-se efeito de dispersão à existência de variâncias diferentes nos diferentes níveis de um fator.

A identificação de efeitos de dispersão iniciou-se em experimentos com replicações, sendo a variabilidade avaliada por meio da variância amostral, calculada em cada ponto experimental. Para estabilização dos resultados foi aconselhada a utilização de uma transformação matemática, especialmente a logarítmica (Bartlett e Kendal, 1946), que fornece dados que podem ser analisados em um experimento da mesma forma que os dados originais. Embora estes sejam os métodos mais utilizados na engenharia da qualidade por apresentar a vantagem de não depender de resultados anteriores, são considerados eficientes apenas em experimentos com mais de 9 replicações (eficiência acima de 90% de acordo com Bartlett e Kendal, 1946). Maiores detalhes em Nair e Pregibon (1988).

Entretanto, dependendo das características do que está sendo medido ou controlado, o experimento pode apresentar um alto custo por necessitar de um grande número de ensaios. Nos casos de poucas replicações, Barbetta et al. (1999) e Barbetta et al. (2000) já apresentaram alguns procedimentos alternativos. Como o custo é um fator extremamente importante, vários autores, tais como: Box e Meyer (1986) e Bergman e Hynén (1997), entre outros, dedicaram-se a estudar métodos para identificação de efeitos de dispersão em experimentos não replicados, dando maior ênfase aos do tipo 2^{k-p} em função de sua grande aceitação no meio industrial.

Como nesta situação, a identificação dos efeitos de dispersão é feita a partir dos resíduos, há o inconveniente do resultado da análise depender da qualidade da modelagem do valor médio. Se esta modelagem, entretanto, é de boa qualidade, o uso de resíduos na identificação de fatores com efeitos de dispersão, teoricamente, é mais eficiente que o uso de variâncias amostrais. Entre os vários métodos não interativos já propostos para experimentos não replicados, são bastante divulgados os métodos BM, proposto por Box e Meyer (1986), e o método H, proposto a partir de trabalho apresentado por Harvey (1976). São também bastante utilizados os métodos BH, proposto por Bergman e Hynén (1997), que pode ser considerado como uma extensão do método BM, e HM, uma extensão do método H. Em experimentos fatoriais do tipo 2^k, o efeito de determinado fator k pelos métodos BM, H, BH e HM, de acordo com Brenneman e Nair (2001), pode ser calculado pelas expressões (1), (2), (3) e (4), respectivamente.

$$D_k^{BM} = \frac{1}{2}\left(\log \sum_{S(k+)} r_n^2 - \log \sum_{S(k-)} r_n^2\right) = \frac{1}{2}\log\left(\frac{\sum_{s(k+)} r_n^2}{\sum_{s(k-)} r_n^2}\right) \tag{1}$$

$$D_k^{H} = \frac{1}{N}\left(\sum_{S(k+)} \log r_n^2 - \sum_{S(k-)} \log r_n^2\right) = \frac{1}{n}\log\left(\frac{\prod_{s(k+)} r_n^2}{\prod_{s(k-)} r_n^2}\right) \tag{2}$$

$$D_k^{BH} = \frac{1}{2}\left(\log \sum_{S(k+)} \breve{r}_n^2 - \log \sum_{S(k-)} \breve{r}_n^2\right) = \frac{1}{2}\log\left(\frac{\sum_{s(k+)} \breve{r}_n^2}{\sum_{s(k-)} \breve{r}_n^2}\right) \tag{3}$$

$$D_k^{HM} = \frac{1}{N}\left(\sum_{S(k+)} \log \breve{r}_n^2 - \sum_{S(k-)} \log \breve{r}_n^2\right) = \frac{1}{n}\log\left(\frac{\prod_{s(k+)} \breve{r}_n^2}{\prod_{s(k-)} \breve{r}_n^2}\right) \tag{4}$$

onde:

D_k é o efeito do k-ésimo fator ou k-ésima interação;
N é a quantidade de pontos experimentais considerados;
r_n^2 é o resíduo quadrático do i-ésimo ponto experimental;
$\breve{r}_n^2$ é o resíduo quadrático modificado do i-ésimo ponto experimental;
$s(k+)$ refere-se aos resultados do nível +1 do k-ésimo fator;
$s(k-)$ refere-se aos resultados do nível -1 do k-ésimo fator.

De acordo com Brenneman e Nair (2001), porém, quase todos os métodos não interativos para experimentos não replicados apresentam um certo grau de viés.

De maneira geral, em todos os métodos supracitados, a identificação dos efeitos de dispersão ativos é feita pelo gráfico de probabilidade normal. No presente estudo, entretanto, serão considerados significativos ou ativos os fatores ou interações com efeitos padronizados superiores a 1,96.

3. Resíduo quadrático

Considere que, em um experimento não replicado, os dados, y_n, apresentam distribuição normal com média $\mu_{y,}$ e variância σ_y^2, ou seja, y_n é $N(\mu_y, \sigma_y^2)$, e seguem o modelo: $y_n = \mu_{yn} + \sigma_{yn}.e_n$, onde e_n são $N(0; 1)$, sendo n = 1, 2, ..., N.

A média está relacionada com os efeitos de locação por meio de uma função de ligação que normalmente é linear (aditiva), enquanto que a variância está relacionada com os efeitos de dispersão por meio de uma outra função, normalmente log-linear (multiplicativa).

Neste tipo de experimento, a avaliação dos efeitos de dispersão é feita a partir de resíduos quadráticos, calculados em cada ponto experimental, que são encontrados após a identificação e remoção dos efeitos de locação. A modelagem do valor médio pelo método dos mínimos quadrados ordinários (MQO) permite predizer os resultados esperados para as observações. No n-ésimo ponto experimental o resíduo é encontrado por $r_n = (\hat{y}_n - y_n)$, onde $\hat{y}_n$ é o valor predito e y_n é o valor observado.

Em experimentos replicados, entretanto, existem várias observações em cada ponto experimental, ou seja, $y_{nm} = \mu_{yn} + \sigma_{yn}.e_{nm}$ é a observação da m-ésima replicação do n-ésimo ponto experimental. Portanto, também existirão vários resíduos em cada ponto experimental.

Nair e Pregibon (1988), para inferências sobre efeitos de dispersão em experimentos replicados, usam a soma dos quadrados dos desvios:

$$Z_n = \sum_{m=1}^{M} (y_{nm} - \bar{y}_n)^2 , \qquad (5)$$

onde $\bar{y}_n$ representa a média amostral entre as M replicações do ponto experimental n.

Ao estender os métodos para identificar efeitos de dispersão em experimentos não replicados para experimentos replicados, foi necessário definir se, no cálculo das diversas estatísticas, deveriam ser considerados os NM resíduos existentes, r_{nm}^2 (n = 1, 2,.., N e m = 1, 2,..., M), isoladamente, ou se poderiam ser substituídos pela média dos resíduos quadráticos em cada ponto experimental, ou seja,

$$\overline{r_n^2} = \frac{\sum_{m=1}^{M} r_{nm}^2}{M} . \qquad (6)$$

Em dois dos métodos supra citados: método BM e método BH, que realizam o cálculo por meio de um quociente entre médias aritméticas, os dois procedimentos conduzem exatamente ao mesmo resultado, desde que o número de replicações seja constante nos diversos pontos experimentais.

Se for examinado o resultado da substituição r_n^2 por $\overline{r_n^2}$ na expressão (1), relativa ao método BM, conforme o demonstrado pelas expressões (7) e (8), verifica-se que seu resultado independe da realização do cálculo a partir de todos os resíduos quadráticos ou de suas médias em cada ponto experimental, podendo-se concluir que os procedimentos são equivalentes, o que vale também para o método BH.

$$D_k^{BM} = \frac{1}{2}\left(\log \sum_{S(k+)} \overline{r_n^2} - \log \sum_{S(k-)} \overline{r_n^2}\right) = \frac{1}{2}\log\left(\frac{\sum_{s(k+)} \overline{r_n^2}}{\sum_{s(k-)} \overline{r_n^2}}\right) \tag{7}$$

$$D_k^{BM} = \frac{1}{2}\log\left(\frac{\sum_{s(k+)}\left(\frac{\sum_{m=1}^{M} r_{nm}^2}{M}\right)}{\sum_{s(k-)}\left(\frac{\sum_{m=1}^{M} r_{nm}^2}{M}\right)}\right) = \frac{1}{2}\log\left(\frac{\sum_{s(k+)}\sum_{m=1}^{M} r_{nm}^2}{\sum_{s(k-)}\sum_{m=1}^{M} r_{nm}^2}\right) = \frac{1}{2}\left(\log \sum_{S(k+)}\sum_{m=1}^{M} r_{nm}^2 - \log \sum_{S(k-)}\sum_{m=1}^{M} r_{nm}^2\right) \tag{8}$$

Entretanto, no método H e no método HM, que realizam o cálculo por meio de um quociente entre médias geométricas, os dois procedimentos não conduzem ao mesmo resultado, não sendo, portanto, equivalentes.

4. Estudo de simulação

O estudo comparativo entre as duas diferentes formas de considerar os resíduos quadráticos no método H e no método HM foi restrito a um projeto fatorial do tipo 2^4 (quatro fatores ensaiados em dois níveis), onde o valor esperado e a variância da resposta satisfazem:

$$\mu_y = \beta_0 + \beta_1 x_1 + \ldots + \beta_4 x_4 + \beta_{12} x_1 x_2 + \ldots. + \beta_{34} x_3 x_4 \tag{9}$$

$$\sigma_y^2 = e^{\theta_0 + \theta_1 x_1 + \ldots. + \theta_{12} x_1 x_2 + \ldots. + \theta_{34} x_3 x_4} \tag{10}$$

Nas simulações do tipo Monte Carlo, as amostras foram geradas a partir de parâmetros especificados. Para que o estudo não se restringisse às condições especificadas, foram consideradas várias alternativas experimentais, tais como: diferente número de efeitos de locação, de efeitos de dispersão e de replicações, sempre variando conforme um projeto experimental adequado (Tabela 1).

No projeto de simulação, os níveis dos fatores considerados foram combinados segundo um experimento fatorial fracionado do tipo 2^{6-1}, o que originou uma matriz de planejamento com 32 pontos experimentais.

Para cada um destes 32 pontos experimentais foram gerados 5.000 experimentos de 4 fatores, também com dois níveis, onde foi avaliado o desempenho dos métodos H e HM com as diferentes formas de considerar os resíduos quadráticos. Para cada método, considerando as 5.000 amostras sob a mesma condição experimental, foram estimadas as probabilidades de identificações corretas e de identificações falsas de efeitos de dispersão a partir das expressões:

$$P(C) = \frac{N_c}{N_{tot}} \quad \text{e} \quad P(F) = \frac{M_F}{M_{tot}} \qquad (11) \text{ e } (12).$$

onde: C representa classificação correta;
F representa classificação incorreta;
N_C é o número de efeitos de dispersão que foram corretamente identificados;
M_F é o número de identificações falsas de efeitos de dispersão;
N_{tot} é o número total de possíveis identificações corretas;
M_{tot} é o número total de possíveis identificações falsas;

A probabilidade de identificação correta de um efeito está relacionada com a ocorrência de erro do tipo I que acontece quando um efeito, embora existente, não consegue ser detectado. Já a probabilidade de identificação falsa de um efeito está relacionada com a ocorrência de erro do tipo II, que acontece quando um efeito não existente é detectado erroneamente.

.

Fator	Descrição	Nível (–1)	Nível (+1)
A	Fatores com efeitos de locação	β_1 e β_2 (dois efeitos principais)	β_1, β_2, β_3, β_{12} e β_{13} (três efeitos principais e duas interações)
B	Intensidade dos efeitos de locação	$\beta_i = 1$ e $\beta_{ij} = 0{,}5$ 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	$\beta_i = 2$ 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
C	Fatores com efeitos de dispersão	θ_k (um efeito de dispersão)	θ_k e θ_2 (dois efeitos de dispersão)
D	Coincidência ou não de efeito de locação e dispersão	$\theta_k = \theta_1$	$\theta_k = \theta_4$
E	Intensidade dos efeitos de dispersão	$\theta_k = 0{,}549$ e $\theta_i = 0$ para $i \neq k$	$\theta_k = 0{,}643$, $\theta_2 = 0{,}549$ e $\theta_i = 0$ para $i \neq k$ e $i \neq 2$
F	Quantidade de replicações	2	4

Tabela 1 – Descrição de fatores e níveis utilizados no estudo de simulação.

Os resultados encontrados, apresentados por meio de medidas descritivas na Tabela 2, mostram que o desempenho médio dos métodos é melhor quando o cálculo é realizado a partir da média dos resíduos quadráticos, pois apresenta maior probabilidade de identificação de efeitos corretamente e menor probabilidade de identificação de efeitos falsamente. Em relação aos métodos H e HM, não foi detectada a existência de diferença significativa.

resíduos	Identificações corretas		Identificações falsas	
	Método H	Método HM	Método H	Método HM
r_{nm}^2	26,73	26,81	2,18	2,27
$\overline{r_n^2}$	41,61	40,73	1,64	1,67

Tabela 2 – Percentagem média de identificações corretas e falsas de efeitos de dispersão, por método e por tipo de resíduo.

5. Considerações Finais

Para experimentos com poucas replicações, não existe ainda uma teoria consagrada para identificação de efeitos de dispersão, tendo em vista que, em experimentos replicados, os métodos existentes são comprovadamente eficientes apenas se forem utilizadas muitas replicações.

Nos métodos em que o efeito de dispersão é avaliado por estatísticas que resultam de um quociente entre médias aritméticas, método BM e método BH, a utilização dos *NM* resíduos quadráticos ou das *N* médias de resíduos quadráticos conduz ao mesmo resultado, o mesmo não acontecendo com os métodos que utilizam quociente entre médias geométricas, método H e método HM. Nestes casos, o presente trabalho mostrou através de simulações de Monte Carlo, que o uso das médias de resíduos quadráticos, calculadas para cada ponto experimental, oferece melhor desempenho, embora este procedimento também apresente maior variabilidade de resultados.

6. Agradecimentos

Este trabalho foi apoiado pela Escola de Educação da Universidade Católica de Pelotas. – http:/www.ucpel.tche.br e 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.

Referências

Barbetta, P.A.; Ribeiro, J.L.D.; Samohyl, R.W. “Uma nova fórmula de modelar variâncias em experimentos com poucas replicações”, Enegep98, (1999):pp.56-63, Rio de Janeiro:Brasil.

Barbetta, P.A.; Ribeiro, J.L.D.; Samohyl, R.W. “Variance Regression Models in Experiments with Few Replications”. Quality and Reliability Engineering International, 16 (2000): pp.397-404.

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 (1946): pp.128-138.

Bergman, B.; Hynén, A., “Dispersion Effects From Unreplicated Designs in 2^{k-p} Series”, Technometrics, 39 (1997): pp.191-198.

Box, G, E. P.; Meyer, R.D., “Dispersion Effects From Fractional Design”, Technometrics, 28 (1986): pp. 19-27.

Brenneman, W.A.; Nair, V.N., “Methods for Identifying Dispersion Effects in Unreplicated Factorial Experiments: A Critical Analysis and proposed Strategies”, Technometrics, 43 (2001): pp. 388-405.

Harvey, A. C., “Estimating Regression Models with Multiplicative Heterocedasticity”, Econometrica, Vol. 44, N°3 (1976): pp. 461-465.

Nair, V. N.; Pregibon, D., “Analysing Dispersion Effects From Replicated Factorial Experiments”. Technometrics, 30, N°3 (1988): pp. 247-257.