

# Utilização de variâncias amostrais e resíduos quadráticos na identificação de efeitos de dispersão em projetos experimentais do tipo $2^k$ .

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

## Resumo

*Este trabalho faz algumas considerações sobre a utilização de variâncias amostrais e resíduos quadráticos em métodos não iterativos para identificar efeitos de dispersão em experimentos com poucas replicações. Um estudo de simulação Monte Carlo permitiu concluir que métodos que usam média aritmética na estimativa de efeitos de dispersão apresentam melhor desempenho*

*Palavra Chave: Efeitos de dispersão; Projeto de experimentos; Melhoria da qualidade.*

## 1 Introdução

O crescente aumento da competição industrial faz com que as indústrias busquem, permanentemente, a melhoria da qualidade com a redução de custos. Uma de suas principais ferramentas é a experimentação estatística, que tem sido muito empregada em estudos preliminares de programas de qualidade, por possibilitar o estudo simultâneo dos efeitos individuais de vários fatores sobre as características funcionais de um produto, assim como dos efeitos das interações entre fatores. É considerada muito eficiente, quando conduzida sob um enfoque clássico, com a incorporação das idéias de Taguchi, que relacionam qualidade com variabilidade.

Taguchi mostrou que ajustar um valor médio a um valor alvo é insuficiente para a obtenção de sistemas insensíveis a qualquer fonte de variação, existindo a necessidade de diminuir a variabilidade em torno do valor alvo. Dentro deste enfoque, é necessário identificar os fatores que afetam a média (efeitos de locação) e os fatores que afetam a variabilidade (efeitos de dispersão) da resposta.

Os efeitos de locação e de dispersão são detectados mais facilmente em experimentos com replicações, onde podem ser encontradas, em cada condição experimental, as médias e as variâncias amostrais, que são utilizadas em sua identificação. Entretanto, as técnicas empregadas com esta finalidade supõem normalidade e variância constante, o que não ocorre na distribuição de variâncias amostrais. De acordo com Bartlett e Kendall (1946), a utilização do logaritmo neperiano das variâncias amostrais contorna o problema da normalidade e estabiliza as variâncias, sendo razoavelmente eficiente em projetos com pelo menos 5 replicações (eficiência acima de 80% em relação à variância amostral na estimação da variância de uma distribuição normal).

Este estudo explora o comportamento de métodos para identificação de efeitos de dispersão em experimentos, comparando o desempenho de métodos que mensuram a variabilidade por meio de variâncias amostrais com o desempenho dos que usam resíduos quadráticos, fixando-se em projetos fatoriais do tipo  $2^K$  com poucas replicações.

## 2 Efeitos de dispersão

Diz-se que um fator afeta a variabilidade da resposta, apresentando efeito de dispersão ativo, quando as variâncias nas respostas obtidas nos seus diferentes níveis são diferentes.

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 esta situação alguns métodos foram propostos, entre os quais estão os métodos *R* e *S*, que estimam estes efeitos por meio de quocientes entre médias aritméticas e geométricas, respectivamente, de variâncias amostrais. Entretanto, a utilização destas medidas na estimativa de variâncias, mesmo que sejam estabilizadas por meio da transformação logarítmica, é considerada eficiente apenas em experimentos com mais de 5 replicações (eficiência acima de 80% de acordo com Bartlett e Kendal, 1946), embora apresentem a vantagem de não depender de resultados anteriores. Maiores detalhes sobre os métodos em Nair e Pregibon (1988).

Em função disso, 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. Como o custo é um fator extremamente importante, vários autores dedicaram-se ao estudo de 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 quadráticos, há o inconveniente do resultado da análise depender da qualidade da modelagem do valor esperado. Se esta modelagem, entretanto, é de boa qualidade, o uso de resíduos quadráticos na identificação de fatores com efeitos de dispersão, teoricamente, é mais eficiente que o uso de variâncias amostrais.

O primeiro método para experimentos não-replicados, empregando resíduos quadráticos, foi proposto por Box e Meyer (1986), sendo a estimativa dos efeitos calculada pela razão entre médias aritméticas destas estatísticas. Um outro método muito citado é o método *H*, proposto a partir de trabalho apresentado por Harvey (1976), que faz estimativas de efeitos de dispersão a partir da razão entre médias geométricas de resíduos quadráticos. Maiores detalhes sobre estes métodos em Brenneman e Nair (2001).

A partir das estimativas dos efeitos de dispersão pelos métodos supracitados, a identificação dos efeitos ativos, é feita, normalmente, pelo gráfico de probabilidade normal. No presente estudo, entretanto, pelas características do estudo, são considerados ativos os fatores ou interações com efeitos padronizados superiores a 2.

## 3 Descrição da simulação

Um estudo do tipo Monte Carlo é realizado para comparar o desempenho dos métodos que usam variâncias amostrais (*R* e *S*) com o desempenho de métodos que usam resíduos quadráticos (*BM* e *H*). É focado um experimento com 3 fatores ( $K=3$ ):  $x_1$ ,  $x_2$  e  $x_3$ , ensaiados a dois níveis ( $-1$  e  $+1$ ), segundo um projeto fatorial completo (projeto tipo  $2^3$  com replicações), contendo efeitos de locação e de dispersão. É considerado que a média dos valores gerados é descrita por uma função aditiva ou linear, de acordo com a expressão:

$$\mu_y = \beta_1 x_1 + \beta_2 x_2, \quad (1)$$

sendo  $\beta_1 = 2$  e  $\beta_2 = 1$ . Também é considerado que dois fatores, identificados por *A* e *B* podem alterar o desempenho dos métodos em estudo.

O fator *A* é utilizado para manipular a quantidade de fatores com efeitos de dispersão. Em um dos níveis ( $-1$ ) considera-se que apenas um fator tem efeitos de dispersão, enquanto que no

outro (+1), dois fatores. A variância é descrita por uma função multiplicativa ou log-linear, de acordo com as expressões:

$$\sigma_y^2 = e^{\theta_1 x_1} \quad \text{e} \quad \sigma_y^2 = e^{\theta_1 x_1 + \theta_2 x_2}, \quad (2) \text{ e } (3)$$

sendo atribuído o valor 0,643 para  $\theta_1$  e  $\theta_2$ , o qual altera o valor da variância, de um nível para outro, em quatro vezes. O fator *B* trata da quantidade de replicações: duas ou quatro. Um terceiro fator *C*, com quatro níveis (métodos *R*, *S*, *H* e *BM*), é utilizado para identificar o método empregado para estimar o efeito de dispersão.

Os níveis destes três fatores do projeto de simulação são combinados segundo um projeto fatorial do tipo  $4 \times 2^2$ . Entretanto, para a geração das amostras são consideradas apenas as quatro condições experimentais que resultam da combinação entre os níveis dos fatores *A* e *B*. A partir das equações (1), (2) e (3), para cada condição experimental do projeto  $2^2$ , são geradas, de acordo com uma distribuição normal, as respostas dos experimentos. A resposta  $y_{nm}$  que corresponde à *m*-ésima replicação ( $m=1,2,\dots,M$ ) da *n*-ésima condição experimental ( $n=1,2,\dots,N$ ) é dada por:

$$y_{nm} = \mu_{y_n} + \sigma_{y_n} \cdot e_n \quad (4)$$

onde  $\mu_{y_n}$  e  $\sigma_{y_n}$  são, respectivamente, o valor esperado e o desvio-padrão da *n*-ésima condição experimental e  $e_n$ , o erro experimental gerado por uma distribuição  $N(0;1)$ .

Foram simulados 5.000 resultados experimentais (amostras) em cada combinação dos níveis dos fatores *A* e *B*, onde foram calculadas a proporção de identificação correta de *todos* os efeitos de dispersão (*PIC*) e a proporção de identificação falsa de *algum* efeito de dispersão (*PIF*), por meio das expressões:

$$PIC = \frac{QIC}{QS} \quad \text{e} \quad PIF = \frac{QIF}{QS}, \quad (5) \text{ e } (6)$$

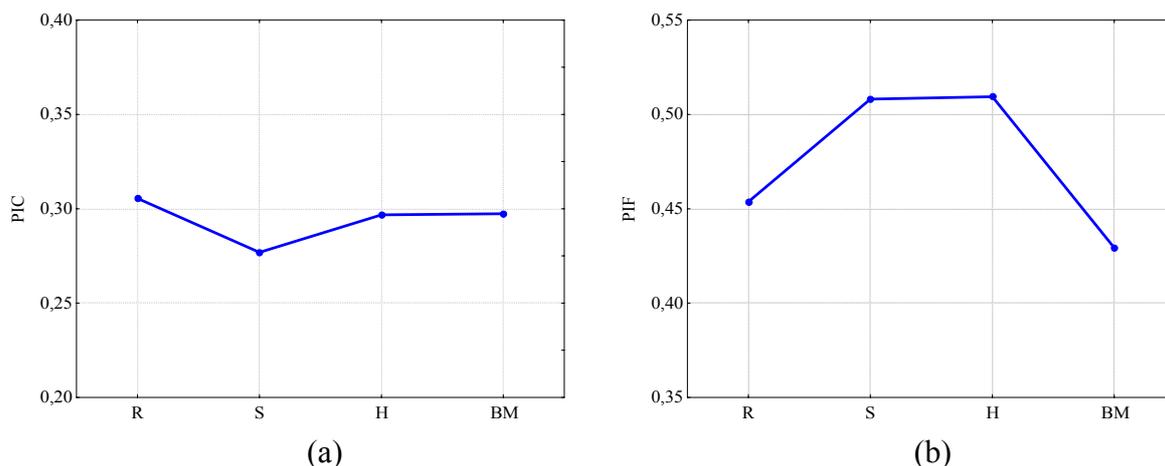
onde: *QIC* é a quantidade de resultados experimentais em que se identificou corretamente todos os efeitos de dispersão; *QIF* é a quantidade de resultados experimentais em que se identificou falsamente de algum efeito de dispersão e *QS* é a quantidade de resultados experimentais.

#### 4 Resultados

Inicialmente, as variáveis resposta *PIC* e *PIF*, encontradas em cada uma das 5.000 amostras geradas sob a mesma condição experimental, são organizadas e representadas pela média na Tabela 1 e nos gráficos (a) e (b) da Figura 1, sendo constatado que o método *R* apresenta maior *PIC*, enquanto o método *BM*, menor *PIF*.

Método	<i>PIC</i>	<i>PIF</i>
<i>R</i>	0,3055	0,4537
<i>S</i>	0,2769	0,5081
<i>H</i>	0,2968	0,5094
<i>BM</i>	0,2974	0,4293

**Tabela 1** – *PIC*'s e *PIF*'s médias, por método, nas condições experimentais simuladas.



**Figura 1** – *PIC's* e *PIF's* médias, por método, nas condições experimentais simuladas.

Após, são realizadas análises de variância sobre estas variáveis, com a variância estabilizada para proporções de amostras de tamanho 5.000, pela utilização das transformações  $\arcsen\sqrt{PIC}$  e  $\arcsen\sqrt{PIF}$ , o que possibilita realizar os testes com a distribuição Qui-quadrado. Sendo a verdadeira variância do erro conhecida ( $\frac{1}{4n}$ , onde  $n$  representa a quantidade de experimentos - amostras), a significância dos resultados pode ser avaliada por meio da estatística:

$$Q^2 = \frac{SS_{\text{Fonte de variação}}}{\sigma^2} \quad (7)$$

que apresenta distribuição Qui-quadrado com  $df_{\text{Fonte de variação}}$  graus de liberdade. Nesta expressão,  $SS$  representa soma de quadrados e  $\sigma^2$ , a variância estabilizada. As fontes de variação podem ser os fatores  $A$ ,  $B$  e  $C$ , além de suas respectivas interações.

Esta análise realizada nos valores das *PIC's* transformadas (Tabela 2), entretanto, não consegue detectar diferença significativa entre os métodos, indicando, porém, a interação do fator  $C$  (método) com o fator  $A$  (quantidade de efeitos de dispersão).

Fator ou interação	Soma dos quadrados (SS)	Graus de liberdade (df)	$Q^2$	Valor-p
$G$	0,001899	3	3,7989	$P > 0,05$
$CG$	0,004317	3	8,6335	$P < 0,05$
$FG$	0,000012	3	0,0237	$P > 0,05$

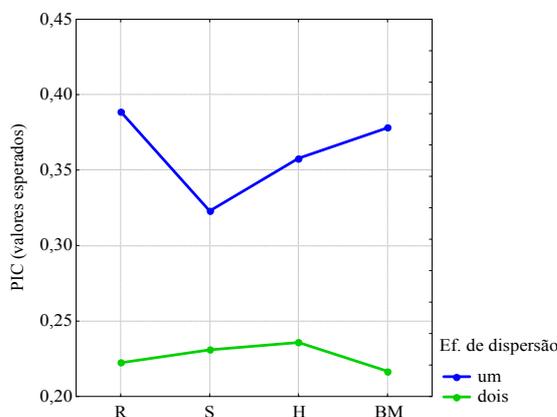
Nota:  $\chi^2_{(1;0,05)} = 7,81$ .

**Tabela 2** – Análise do efeito do fator  $G$  (tratamento do resíduo quadrático) e suas interações com os fatores  $C$  e  $F$  na variável transformada  $\arcsen\sqrt{PIC}$ .

A Figura 2 mostra que, se existe apenas um fator com efeito de dispersão, o método  $R$  apresenta melhor desempenho, mas que seu desempenho é muito afetado pela quantidade de

efeitos, diminuindo consideravelmente com o seu aumento. Nesta situação, o método *H* parece apresentar melhor desempenho.

Esta mesma análise aplicada aos valores das *PIF's* transformadas (Tabela 3), entretanto, detecta diferença significativa entre os métodos, mas não detecta interação deste fator com os demais considerados no experimento. O método *BM* apresenta menor *PIF*.



**Figura 2** – Efeito da interação do fator *G* (método) com o fator *C* (efeitos de dispersão) sobre a variável resposta *PIC*, nas condições experimentais simuladas.

Fator ou interação	Soma dos quadrados (SS)	Graus de liberdade (df)	$Q^2$	Valor-p
<i>G</i>	0,021701	3	43,4024	$P < 0,001^*$
<i>CG</i>	0,001364	3	2,7288	$P > 0,05$
<i>FG</i>	0,000190	3	0,3790	$P > 0,05$

Nota:  $\chi^2_{(3;0,001)} = 16,27$ ;  $\chi^2_{(1;0,05)} = 7,81$ .

**Tabela 3** – Análise do efeito do fator *G* (tratamento do resíduo quadrático) e suas interações com os fatores *C* e *F* na variável transformada  $\arcsen\sqrt{PIF}$ .

Para complementar a análise, são usados contrastes ortogonais para comparar as *PIC's* e *PIF's* entre: (a) métodos que usam média aritmética (*R* e *BM*) e métodos que usam média geométrica (*S* e *H*); (b) métodos que usam variâncias amostrais (*R* e *S*) e métodos que usam resíduos quadráticos (*H* e *BM*).

Os resultados da análise nas *PIC's*, mostrados na Tabela 4, indicam que não existem evidências de diferença significativa entre métodos que utilizam diferentes tipos de média, nem entre métodos que usam diferentes medidas para estimar a variabilidade.

Para as *PIF's*, entretanto, constata-se que existe diferença significativa entre métodos que utilizam diferentes tipos de média: os métodos que estimam efeitos de dispersão por meio de médias aritméticas possuem *PIF* médio de 0,4415, inferior aos métodos que empregam médias geométricas que possuem *PIF* médio de 0,5086, apresentando, portanto, melhor desempenho.

Não é constatada, porém, diferença significativa entre métodos que usam variâncias amostrais e resíduos quadráticos (Tabela 5).

Hipótese	Soma dos quadrados (SS)	Graus de liberdade (df)	$Q^2$	Valor-p
(a)	0,000770	1	1,5401	$P > 0,05$
(b)	0,286346	1	0,2863	$p > 0,05$

Nota:  $\chi^2_{(1;0,05)} = 3,84$ .

**Tabela 4** – Análise do efeito dos contrastes na variável transformada  $\arcsen\sqrt{PIC}$ .

Hipótese	Soma dos quadrados (SS)	Graus de liberdade (df)	$Q^2$	Valor-p
(a)	0,020445	1	40,8903	$P < 0,001^*$
(b)	0,000552	1	1,1041	$P < 0,05$

Nota:  $\chi^2_{(1;0,001)} = 10,83$ ;  $\chi^2_{(1;0,05)} = 3,84$ .

**Tabela 5** – Análise do efeito dos contrastes na variável transformada  $\arcsen\sqrt{PIF}$ .

## 5 Considerações Finais

Nesta simulação, em relação às proporções de identificação correta de todos os efeitos de dispersão, não é possível detectar diferença significativa entre os métodos investigados, o que também acontece para a comparação entre métodos que usam diferentes tipos de média e entre métodos que usam diferentes tipos de medida para a variabilidade. Entretanto, constata-se que o método *R* apresenta resultado ligeiramente superior, seguido do método *BM*, o que se acentua quando existe apenas um fator com efeito de dispersão.

Em relação às proporções de identificação falsa de algum efeito de dispersão, embora não tenha sido possível detectar diferença entre os métodos investigados, nem entre métodos que usam diferentes medidas para a variabilidade, é constatado que métodos que usam média aritmética apresentam melhor resultado. O método *BM* falha menos, seguido do método *R*, ambos estimando efeitos de dispersão por meio de médias aritméticas.

Portanto, os resultados parecem indicar que, quando são usadas poucas replicações, o desempenho de um método é mais influenciado pelo tipo de média empregado na estimativa do efeito: média aritmética ou média geométrica, do que pelo tipo de medida da variabilidade: variância amostral ou resíduo quadrático.

## 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. (2000) - Variance Regression Models in Experiments with Few Replications. Quality and Reliability Engineering International, 16, p.397-404.

Bartlett, M.S.; Kendall, D.G. (1946) - The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation. *Journal of the Royal Statistics Society, Ser. B*, 8, p.128-138.

Box, G. E. P.; Meyer, R.D. (1986) - Dispersion Effects From Fractional Design. *Technometrics*, 28, p. 19-27.

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

Harvey, A. C. (1976) - Estimating Regression Models with Multiplicative Heterocedasticity. *Econometrica*, Vol. 44, N°3, p. 461-465.

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