

5

Análise Conjunta da Média e da Dispersão

Neste capítulo vamos apresentar métodos de identificação dos fatores que afetam a média e a variância da resposta, em experimentos fatoriais não replicados.

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

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

Dois outros tipos de fatores devem ser considerados para a análise da dispersão. Fatores de controle são aqueles possíveis de manter em um dado nível durante o processo de produção. Fatores de ruído são aqueles que oscilam em torno de um dado nível durante o processo de produção, embora possam ser controlados durante a fase experimental do projeto do processo ou produto. Os dois tipos de fatores podem produzir efeitos tanto na média quanto na dispersão. Quando se faz o ajuste no modelo não é necessário distinguir os dois tipos de fatores. Apenas na análise dos resultados, é levada em conta a distinção entre eles.

Na década de 80 foi o consultor Genichi Taguchi quem caracterizou os fatores de ruído como importantes para a análise de efeitos na dispersão. Em sua conceituação é necessário identificar os fatores de controle que afetam a dispersão da resposta e, então, o nível do fator deve ser ajustado de modo que resulte em

produtos e processos robustos às variações dos fatores de ruído. Os conceitos por ele estabelecidos são considerados de grande importância, embora as técnicas quantitativas por ele propostas tenham sido criticadas por vários autores, que apresentaram técnicas mais eficientes. Foi cunhado o termo **Projeto Robusto de Parâmetros** (*Robust Parameter Design*) para caracterizar a busca das condições ótimas no desenvolvimento de produtos e processos produtivos. Para uma descrição e análise das técnicas de Taguchi, ver Myers e Montgomery (2002), Cap.11.

Na Seção 5.1 mostramos a importância da elaboração do modelo para média. Na Seção 5.2 apresentamos um método gráfico muito simples e útil para identificar efeitos na dispersão. É o gráfico dos resíduos *versus* nível do fator de produção.

Na Seção 5.3 apresentamos o método de Box e Meyer (1986), usado para identificar fatores que afetam a dispersão. Nesse método a identificação de efeitos na dispersão é feita diretamente a partir da análise da variabilidade dos resíduos provenientes do modelo previamente elaborado para a média, sem que seja elaborado um modelo para a dispersão. Na Seção 5.4, apresentamos o método iterativo de Lee e Nelder (1998). Esses autores aplicam modelos lineares generalizados para modelar explicitamente tanto a média quanto a dispersão.

Na Seção 5.5, veremos como analisar experimentos onde são encontrados fatores de ruído. Durante a elaboração dos modelos da média e da dispersão, não é necessário fazer distinção entre fator de controle e fator de ruído. Entretanto, uma vez construído os modelos, podemos ver como os fatores de controle interagem com os de ruído.

Na Seção 5.6 apresentamos um roteiro que sintetiza todas as técnicas estatísticas indicadas para a análise da média e da dispersão.

5.1. A Relação do Modelo da Média com os Efeitos na Dispersão

Um modelo inadequado para a média em experimentos fatoriais não replicados pode acarretar a identificação de fatores espúrios afetando a dispersão. Isto porque, não havendo replicações, o modelo para a dispersão é obtido através dos resíduos provenientes do modelo da média, o que exige uma identificação correta dos fatores que afetam a média, sob pena de identificar efeitos espúrios na

dispersão ou não identificar efeitos reais. Pan (1999) e McGrath e Lin (2001) estudaram este problema aplicando os métodos de Box e Meyer (1986) e de Bergman e Hynén (1997) a casos reais.

Segundo Pan, a má identificação de efeitos na média diminui a eficiência do método usado para identificar os efeitos na dispersão, mas não afeta a sua validade. Isto quer dizer que o método ainda identifica efeitos reais na dispersão, embora produza mais erros na identificação. Por outro lado, efeitos na média não identificados, mesmo que pequenos, podem, cumulativamente, invalidar o método usado para identificar os efeitos na dispersão. Pan afirma que os efeitos na média não identificados podem causar um grande impacto no poder e na probabilidade de erro do método de identificação de efeitos na dispersão, e que a intensidade desse impacto é imprevisível se o verdadeiro modelo da média é desconhecido.

As conclusões McGrath e Lin (2001) são similares às de Pan. Eles também mostram que, em certos casos, não se pode determinar o que é real: um efeito na dispersão ou dois efeitos na média.

Nas seções 5.2 e 5.3 ilustramos, com o Exemplo 4.1, os problemas detectados por estes autores.

Pode-se supor que o contrário também é verdade: a não identificação de efeitos na dispersão pode levar a erros na identificação de efeitos na média, fazendo que a estimativa desses últimos seja aumentada ou reduzida e, em particular, que efeitos existentes possam ser considerados nulos e vice-versa.

Como foi dito, as conclusões de McGrath e Lin (2001) e de Pan (1999) referem-se aos métodos de Box e Meyer (1986) e de Bergman e Hynén (1997). Entretanto, os problemas detectados por estes autores ocorrem também com o método iterativo de Lee e Nelder (1998), como mostramos na Seção 5.4. Aplicando este método ao Exemplo 4.1, consideramos dois modelos para a média: o Modelo 5.1 e o Modelo 5.2; modelos diferentes foram produzidos para a dispersão.

5.2. Método Gráfico

O gráfico dos resíduos *versus* fator é útil para uma primeira inspeção do efeito do fator sobre a dispersão. Embora em muitos casos não forneça indicações conclusivas, em outros as fornece. A análise é feita observando o espalhamento

dos resíduos quando o nível do fator é mudado. Quando o espalhamento fica visivelmente maior com esta mudança, é indício de que a variabilidade é sensível a esse fator.

Exemplo 4.1 (continuação)

Na Seção 4.2 tínhamos o modelo

$$\hat{y} = 27,31 + 6,94A + 17,81B + 5,94AB. \quad (5.1)$$

Na Figura 5.1 o apresentamos o gráfico dos resíduos *versus* nível do fator *C*.

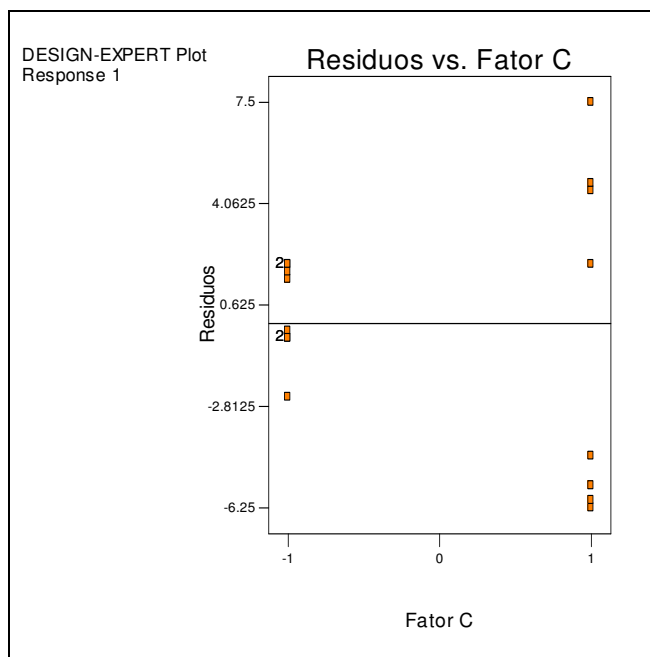


Figura 5.1 - Gráfico dos Resíduos *versus* Fator *C* - Modelo (5.1).

Observamos que o espalhamento dos resíduos fica visivelmente maior com a mudança do nível. Isto é um indício de que a variabilidade aumenta quando o fator *C* muda de -1 para +1.

Na Seção 4.2 também tínhamos o modelo

$$\hat{y} = 27,31 + 6,94A + 17,81B + 5,94AB - 0,44C - 2,44G - 2,69CG \quad (5.2)$$

Na Figura 5.2 o apresentamos, para o Modelo 5.2, o gráfico dos resíduos *versus* o nível do fator *C*.

Observamos que o espalhamento dos resíduos não muda significativamente com a mudança do nível. Isto é um indício de que a variabilidade não muda quando o fator *C* muda de -1 para +1. Portanto, temos evidências de que a não inclusão de *C* e *CG* no modelo para a média resulta em identificar um efeito espúrio do fator *C* sobre a dispersão.

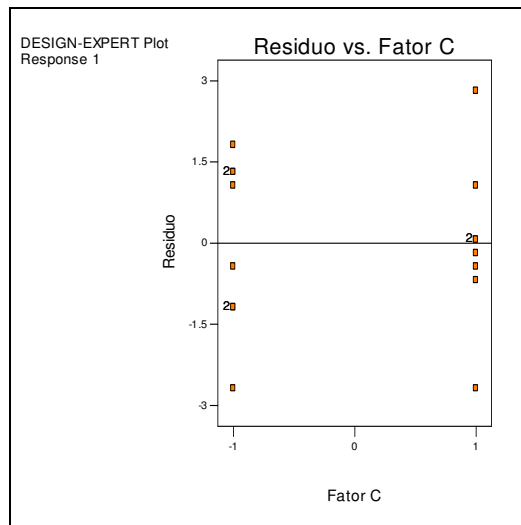


Figura 5.2 - Gráfico dos Resíduos *versus* Fator C - Modelo (5.2).

Entretanto, analisando os gráficos do Modelo 5.2, encontramos indicação de que os fatores F e G afetam a dispersão. É o que evidencia a análise do gráfico da Figura 5.3.

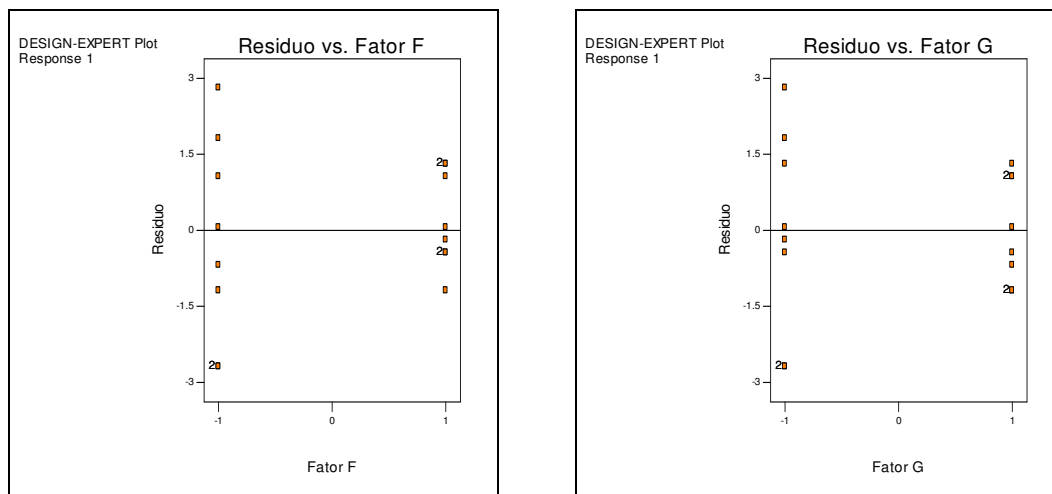


Figura 5.3 - Gráficos dos Resíduos *Versus* Fatores F e G - Modelo (5.2).

Observamos que em ambos os gráficos o espalhamento dos resíduos fica menor com a mudança do nível. Isto é um indício de que a variabilidade diminui quando os fatores G mudam do nível -1 para +1.

Evidentemente, este método gráfico não é um teste formal. Mas ele pode, em muitos casos, fornecer alguma evidência dos fatores que afetam a dispersão.

5.3. Método de Box e Meyer

Box e Meyer (1986) propuseram um método baseado na variabilidade dos resíduos provenientes do modelo ajustado para a média por mínimos quadrados.

Seja o modelo linear

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

onde

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad e \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

e a variável x_{ij} tanto pode representar um fator principal quanto uma interação entre fatores.

Este método só é aplicável a experimentos fatoriais com dois níveis, completos (2^k) ou fracionados (2^{k-p}). Portanto, as variáveis x_{ij} são geralmente codificadas como sendo igual a -1 ou +1.

Definem-se então os seguintes conjuntos M_j e P_j

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

Para estimativa dos parâmetros é aplicado o método dos mínimos quadrados (MQ).

Sejam $e_i = (y_i - \hat{y}_i)$ os resíduos, $i = 1, 2, \dots, n$, do modelo ajustado.

Para cada coluna (variável j) define-se:

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

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

Se os valores dos desvios-padrão s_{j+} e s_{j-} não são muito diferentes, tem-se a indicação de que, quando o fator j muda de um nível para outro, a variância da resposta não é afetada por esse fator. Ou seja, espera-se que a variância dos resíduos seja a mesma para os dois níveis do fator j .

Por outro lado, se um valor é muito maior do que o outro, tem-se a indicação de que, quando o fator j muda de um nível para outro, a variância da resposta é afetada por esse fator.

Exemplo 4.1 (continuação)

Na Tabela 5.1, apresentamos, para o Modelo 5.1, os conjuntos M_j e L_j e os cálculos dos desvios-padrão s_{j+} e s_{j-} ; para $j = B$ e $j = C$.

Tabela 5.1 - Cálculo dos Desvios-padrão dos Fatores B e C .

| y | \hat{y} | Resíduos | Fator B | M_B | L_B | Fator C | M_C | L_C | |
|-----|-----------|----------|-----------|-------------|-------|-----------|-------|-------|------|
| 6 | 8,50 | -2,50 | -1 | | -2,50 | -1 | | -2,50 | |
| 10 | 10,50 | -0,50 | -1 | | -0,50 | -1 | | -0,50 | |
| 32 | 32,25 | -0,25 | 1 | -0,25 | | -1 | | -0,25 | |
| 60 | 58,00 | 2,00 | 1 | 2,00 | | -1 | | 2,00 | |
| 4 | 8,50 | -4,50 | -1 | | -4,50 | 1 | -4,50 | | |
| 15 | 10,50 | 4,50 | -1 | | 4,50 | 1 | 4,50 | | |
| 26 | 32,25 | -6,25 | 1 | -6,25 | | 1 | -6,25 | | |
| 60 | 58,00 | 2,00 | 1 | 2,00 | | 1 | 2,00 | | |
| 8 | 8,50 | -0,50 | -1 | | -0,50 | -1 | | -0,50 | |
| 12 | 10,50 | 1,50 | -1 | | 1,50 | -1 | | 1,50 | |
| 34 | 32,25 | 1,75 | 1 | 1,75 | | -1 | | 1,75 | |
| 60 | 58,00 | 2,00 | 1 | 2,00 | | -1 | | 2,00 | |
| 16 | 8,50 | 7,50 | -1 | | 7,50 | 1 | 7,50 | | |
| 5 | 10,50 | -5,50 | -1 | | -5,50 | 1 | -5,50 | | |
| 37 | 32,25 | 4,75 | 1 | 4,75 | | 1 | 4,75 | | |
| 52 | 58,00 | -6,00 | 1 | -6,00 | | 1 | -6,00 | | |
| | | | | Variância | 16,11 | 19,43 | | 32,44 | 2,66 |
| | | | | Erro-padrão | 4,01 | 4,41 | | 5,70 | 1,63 |

Para o fator B temos

$$s_{B+} = 4,01 \text{ e } s_{B-} = 4,41,$$

enquanto que, para o fator C temos

$$s_{C+} = 5,70 \text{ e } s_{C-} = 1,63.$$

É evidente que a diferença é bem maior no caso do fator C . Entretanto, não sabemos se esta diferença é significativa. Para uma melhor verificação da magnitude relativa dos efeitos sobre a dispersão, Box e Meyer (1986) definiram a estatística

$$F_j = \ln \frac{s_{j+}^2}{s_{j-}^2} = \ln(s_{j+}^2) - \ln(s_{j-}^2) \quad j = 1, 2, \dots, r \quad (5.3)$$

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

Os autores do método propõem a construção de um gráfico de pontos da estatística F_j ao longo de uma reta. Isto, porém, não é um teste formal.

Na Tabela 5.2 apresentamos a matriz \mathbf{X} e os valores da estatística F_j , para cada variável.

Tabela 5.2 - Estatística F_j do Modelo (5.1).

| A | B | AB | C | AC | BC | E | D | AD | BD | ABD | CD | G | F | DE |
|-------|-------|------|-----|-------|-------|-------|------|------|-------|------|------|------|------|------|
| -1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | 1 | -1 | -1 | 1 |
| 1 | -1 | -1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 |
| -1 | 1 | -1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | 1 | -1 | 1 | -1 |
| 1 | 1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 |
| -1 | -1 | 1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 |
| 1 | -1 | -1 | 1 | 1 | -1 | -1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 |
| -1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| -1 | -1 | 1 | -1 | 1 | 1 | -1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 | -1 |
| 1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 | -1 | 1 | 1 |
| -1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 | -1 | 1 | -1 | -1 | 1 | -1 | 1 |
| 1 | 1 | 1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 | -1 |
| -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 |
| 1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 |
| -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| -0,38 | -0,18 | 0,11 | 2,5 | -0,41 | -0,24 | -0,03 | 0,51 | 0,42 | -0,18 | 0,52 | 0,51 | 0,23 | -0,3 | 0,72 |

Na Figura 5.4, apresentamos o gráfico de pontos proposto por Box e Meyer, com os valores de F_j , correspondentes aos valores da Tabela 5.1.

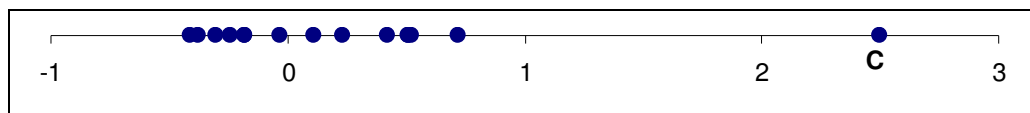


Figura 5.4 - Gráfico com os valores de F_j .

Observamos que o valor de F_j para o fator C se destaca dos demais.

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

Na Figura 5.5 apresentamos o gráfico de probabilidade normal de F_j , construído na planilha *Excel*, onde, efetivamente, o ponto correspondente ao fator C está afastado da reta em torno da qual se alinham os demais pontos.

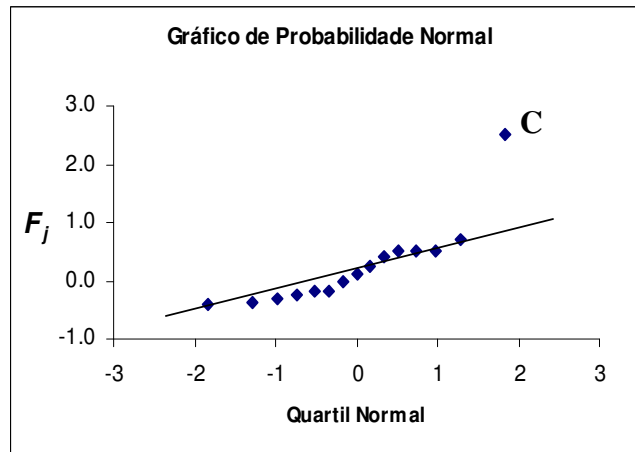


Figura 5.5 - Gráfico de Probabilidade Normal de F_j . – Modelo (5.1).

Agora vamos usar o Modelo 5.2. Na tabela abaixo apresentamos os valores da estatística F_j , construída na planilha *Excel*.

| F | G | B | AD | BD | ABD | A | D | C | AC | CD | BC | DE | E | AB |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|------|------|------|
| -1,50 | -1,31 | -0,38 | -0,38 | -0,36 | -0,30 | -0,22 | -0,22 | -0,06 | -0,02 | 0,13 | 0,25 | 0,33 | 0,39 | 1,80 |

Verificamos que o valor de F_j para o fator C diminuiu. Na Figura 5.6 apresentamos o gráfico de probabilidade normal de F_j para o Modelo 5.2, construído na planilha *Excel*. Observando o gráfico, não temos mais indicação de que a dispersão seja afetada pelo fator C (o que confirma a indicação do gráfico da Fig. 5.2), mas sim pelos fatores F e G (o que confirma a indicação dos gráficos da Figura 5.3) e pela interação AB .

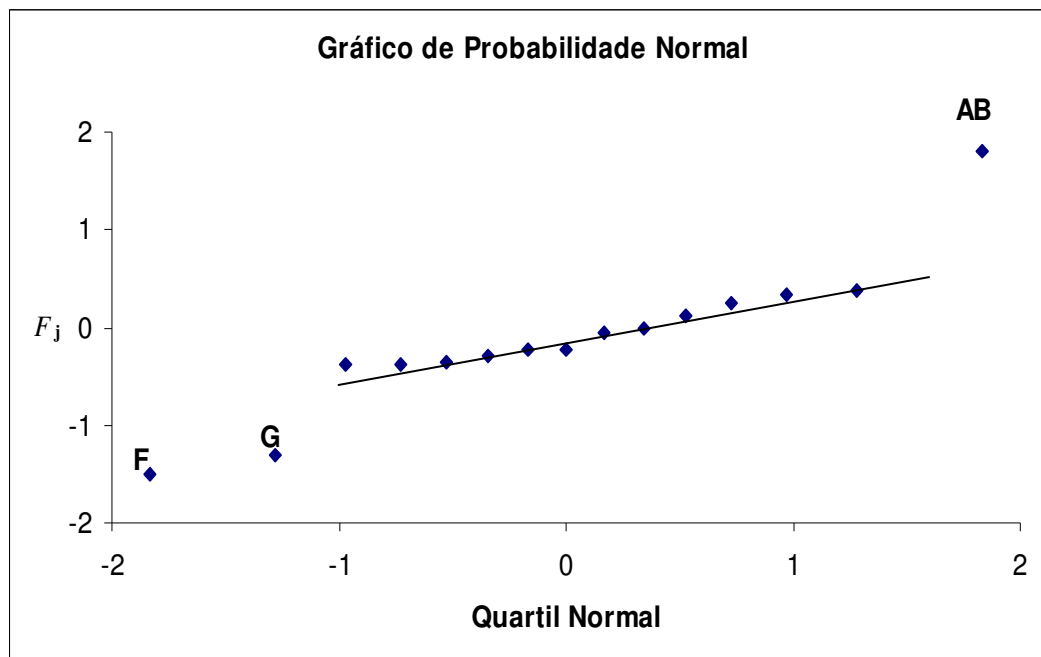


Figura 5.6 - Gráfico de Probabilidade Normal de F_j . – Modelo (5.2).

Agora temos um dilema. Temos um modelo simples para a média (5.1) com um fator de dispersão (C) ou modelo mais complexo para a média (5.2), com dois fatores (F e G) e uma interação (AB) que afetam a dispersão. Na Figura 5.7, apresentamos o gráfico dos resíduos *versus* nível da interação AB .

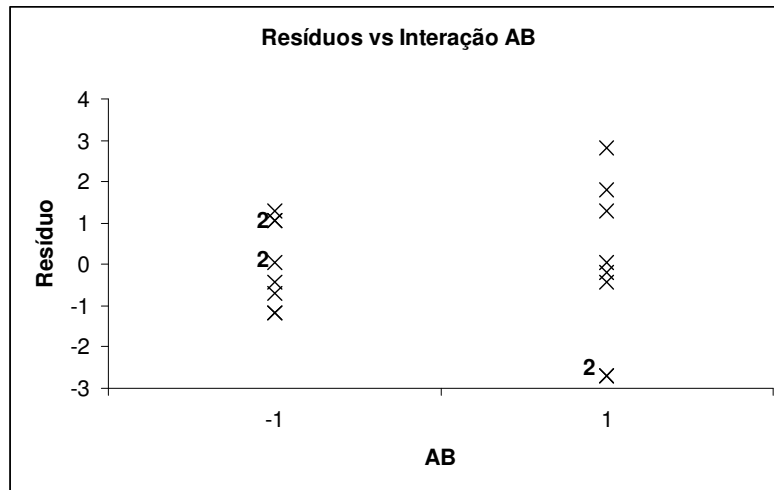


Figura 5.7 - Gráfico dos Resíduos *versus* o Nível da Interação AB .

Os experimentos fracionados (como no Exemplo 4.1) são usados nos estágios iniciais da investigação. Se optarmos pelo modelo mais simples (5.1), estaremos eliminando fatores potencialmente importantes, que poderiam ser identificados em experimentos posteriores.

Cabe lembrar que, devido à estrutura do experimento, a interação AB está confundida com CE e FG , e a interação CG está confundida com AD e EF . Caso não haja conhecimentos práticos sobre a existência ou não dessas interações, há necessidade de um outro experimento para eliminar as ambigüidades.

Para maiores comentários sobre o método de Box e Meyer, ver Brenneman e Nair (2001) e Myers e Montgomery (2002).

Brenneman e Nair (2001) analisaram, além do método de Box e Meyer (1986), os métodos de Bergman e Hynén (1997) e Brenneman e Nair (2001). Nesses métodos, a identificação de efeitos na dispersão é feita diretamente a partir da análise da variabilidade dos resíduos provenientes do modelo previamente elaborado para a média, sem que seja elaborado um modelo para a dispersão. Eles são aplicáveis apenas em experimentos fatoriais com dois níveis, completos (2^k) ou fracionados (2^{k-p}) e só consideram, para a modelagem da média, os modelos lineares com o método dos mínimos quadrados.

Brenneman e Nair (2001), concluíram, com resultados de estudos de simulação, que todos esses métodos apresentam algum viés, que depende do modelo ajustado para a média e do número de efeitos ativos na dispersão. Essas simulações evidenciaram que, mesmo que o modelo para a média seja o correto, esses métodos podem identificar efeitos espúrios na dispersão.

5.4. Métodos Iterativos – Modelagem Conjunta da Média e Variância

Métodos iterativos para identificação de efeitos na média e na dispersão foram desenvolvidos por McCullagh e Nelder (1989), Engel e Huele (1996) e Lee e Nelder (1998). Trata-se de métodos que utilizam os modelos lineares generalizados (MLG), sendo que McCullagh e Nelder (1989) e Lee e Nelder (1998) aplicam MLG para modelar explicitamente tanto a média quanto a dispersão, enquanto Engel e Huele aplicam MLG só para modelar a dispersão. Para a modelagem da média, Engel e Huele assumem um modelo cuja função de ligação é a identidade, sendo portanto um caso particular dos modelos MLG propostos por McCullagh e Nelder (1989) e Lee e Nelder (1998). Além disso, Engel e Huele não consideram o uso da técnica de máxima verossimilhança restrita (MVR) (*restricted maximum likelihood*), utilizada por McCullagh e Nelder (1989) e Lee e Nelder (1998), que é um ajuste feito no modelo da média, necessário em experimentos fatoriais altamente fracionados.

Brenneman e Nair (2001) analisaram também os métodos iterativos, concluindo pela superioridade destes, especialmente o método de Lee e Nelder com MVR, em relação aos métodos de Box e Meyer (1986), Bergman e Hynén (1997) e Brenneman e Nair (2001).

5.4.1. Método de Lee e Nelder (1998)

Lee e Nelder (1998) apresentam um método de análise dos fatores que produzem efeitos na média e na dispersão. Eles propõem usar a quase-verossimilhança estendida para obter o modelo da média, e MLG para o modelo da dispersão.

A seguir descrevemos o modelo para a média e, em seguida, o modelo para a dispersão. Uma exposição detalhada dos aspectos teóricos deste método pode ser vista em McCullagh e Nelder (1989), Cap. 10.

5.4.1.1 Modelo para a Média

Considere que a resposta y_i tem média $E(y_i) = \mu_i$ e variância $\text{var}(y_i) = \phi_i V(\mu_i)$, sendo ϕ_i o parâmetro de dispersão e $V(\mu_i)$ a função de variância. Observe que o parâmetro de dispersão é constante para os membros da família exponencial e agora é considerado diferente para cada combinação dos níveis dos fatores. Isto porque, quando se constrói apenas o modelo para a média não há condição de se fazer uma estimativa de ϕ_i para cada uma destas combinações. Com a modelagem da dispersão isto será possível, como será visto.

O modelo proposto para a média é

$$\eta_i = g(\mu_i) = \mathbf{x}_i' \boldsymbol{\beta},$$

e a média e a variância de y_i são

$$E(y_i) = \mu_i \quad \text{e} \quad \text{var}(y_i) = \phi_i V(\mu_i)$$

Em muitos casos, a escolha de um dos membros da família exponencial pode ser inadequada, quer pela própria inadequação da distribuição de probabilidade, quer pela impropriedade da função de variância. Para lidar com distribuições não pertencentes à família exponencial foi então proposta a função de quase-verossimilhança, vista na Seção 3.6.

Com o parâmetro de dispersão sendo diferente para cada resposta (y_i), a função de quase-verossimilhança não é suficiente para a modelagem da média. Para este modelo, Lee e Nelder (1998) propõem então a função de quase-verossimilhança estendida (*extended quasi-likelihood*), vista na Seção 3.7.

$$-2Q^+ = \sum_{i=1}^n \left\{ \frac{d_i}{\phi_i} + \ln[2\pi\phi_i V(y_i)] \right\} \quad (5.4)$$

onde d_i é o quase-deviance, já visto e dado por

$$d_i = 2 \int_{\mu_i}^{y_i} \frac{y_i - t}{V(t)} dt$$

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

A maximização da função de quase-verossimilhança estendida (QVE) com relação a β será obtida com os estimadores de quase-verossimilhança (QV), com pesos $1/\phi_i$, cujas equações-escore são

$$\frac{\partial Q^+}{\partial \beta} = \sum_{i=1}^n \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \beta} = 0$$

ou ainda:

$$\sum_{i=1}^n \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \mathbf{x}_i = 0 \quad (5.5)$$

Ou seja, as equações de quase-verossimilhança estendida (EQL) são as mesmas de quase-verossimilhança (QL) com pesos conhecidos $1/\phi_i$, como foi visto na Seção 3.7. Por conseguinte, para resolver essas equações, utiliza-se a mesma metodologia usada em QV. Na primeira iteração da modelagem da média, ϕ_i é considerado igual a um. Na segunda iteração já são obtidas as suas estimativas a partir da modelagem da dispersão, como veremos mais adiante.

Para a família exponencial, a relação da função de variância com a média pode ser representada por uma função de potência: $V(\mu) = \mu^t$.

No Capítulo 3 vimos que para $t = 0, 1, 2$ e 3 , as funções de quase-verossimilhança correspondem, respectivamente, às funções de log-verossimilhança (a menos das constantes) das distribuições normal, de Poisson, gama e normal inversa. O mesmo acontece com a distribuição binomial quando a função de variância é $V(\mu) = \mu(1 - \mu)$.

Nestes casos, QV e MLG produzem as mesmas estimativas dos coeficientes, portanto, pode-se usar o mesmo algoritmo de estimativa dos coeficientes, visto na Seção 3.3.1. Entretanto, a estimativa do parâmetro de dispersão em QV é feita com estatística generalizada de Pearson (McCullagh e Nelder (1989), pág. 328), e não com a *deviance*.

Isto quer dizer que, nestes casos, podemos usar o mesmo algoritmo para resolver as equações de MVE (5.5). A única diferença é na equação (3.5) que passa a ser

$$w_{ii}^{(m)} = \frac{1}{\phi_i V(y_i)} \left(\left(\frac{\partial \mu_i}{\partial \eta_i} \right)^{(m)} \right)^2 \quad i = 1, 2, \dots, n.$$

Na verdade, a utilização da QVE é um artifício para os casos em que a função de variância não “explica” completamente a variabilidade da resposta, quando então, o parâmetro de dispersão não é constante para cada resposta (como é em MLG), dependendo dos fatores x_{ij} . Um modelo para dispersão é então construído para estabelecer esta relação de dependência.

Máxima verossimilhança restrita.

Em experimentos altamente fracionados, nos quais o número de dados é pequeno em relação ao número de parâmetros do modelo para a média, os estimadores dos parâmetros do modelo de dispersão podem ter vieses acentuados, devido à escassez de graus de liberdade para modelagem da dispersão (Lee e Nelder, 1998). Neste caso, procedimentos de ajuste devem ser adotados para o modelo da média. Lee e Nelder citam vários tipos de ajuste e propõem uso da técnica de máxima verossimilhança restrita (MVR) com a maximização da seguinte função de verossimilhança

$$-2Q_B = \sum_{i=1}^n \left\{ \frac{d_i^*}{\phi_i} + \ln[2\pi\phi_i V(y_i)] \right\} \quad (5.6)$$

onde

$$d_i^* = d_i / (1 - h_i)$$

e h_i , $i = 1, 2, \dots, n$ são os elementos da diagonal da matriz chapéu (*hat matrix*) \mathbf{H} , que é a matriz de projeção dos valores ajustados sobre os valores observados para a média, ou seja: $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$.

Então, o ajuste consiste apenas em substituir d_i por d_i^* , na função de quase-verossimilhança estendida.

5.4.1.2. Modelo para a Dispersão

O modelo proposto para a dispersão é $\xi_i = h(\phi_i) = \mathbf{z}'_i \boldsymbol{\gamma}$.

Na modelagem da dispersão, a variável de resposta é o resíduo *deviance* d_i . Para representar d_i Lee e Nelder (1998) recomendam a distribuição gama com parâmetro de dispersão igual a 2. Portanto, é assumido que:

$$E(d_i) = \phi_i \quad \text{e} \quad \text{var}(d_i) = 2\phi_i^2.$$

Os vetores \mathbf{x} , do modelo da média, e \mathbf{z} podem ter elementos comuns. Ou seja, fatores que afetam a média podem afetar ou não a dispersão e vice-versa.

Lee e Nelder (1998) afirmam que a escolha usual da função de ligação é

$$h(\phi_i) = \ln(\phi_i),$$

ou seja, eles consideram que os efeitos dos fatores sobre a dispersão são multiplicativos. Na Tabela 5.3 é apresentado um resumo dos modelos e componentes para a média e para a dispersão.

Tabela 5.3 - Modelos e Componentes para a Média e a Dispersão.

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

5.4.1.3. Procedimento Iterativo.

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

Exemplo 4.1 (continuação)

Neste exemplo foi usado um experimento fatorial 2^{7-3} com sete fatores A-G sem replicações. Se consideramos o Modelo 5.1 como o adequado, temos que os fatores A e B afetam a média e temos indicação de que o fator C afeta a dispersão. Tratamos então os outros fatores como sendo inertes e podemos considerar um experimento 2^3 com duas replicações. Na Tabela 5.4 apresentamos este experimento.

Tabela 5.4 - Experimento 2^3 com Duas Replicações.

| A | B | C | y_{i1} | y_{i2} | $ y_{i1} - y_{i2} $ |
|----|----|----|----------|----------|---------------------|
| -1 | -1 | -1 | 6 | 8 | 2 |
| 1 | -1 | -1 | 10 | 12 | 2 |
| -1 | 1 | -1 | 32 | 34 | 2 |
| 1 | 1 | -1 | 60 | 60 | 0 |
| -1 | -1 | 1 | 4 | 16 | 12 |
| 1 | -1 | 1 | 15 | 5 | 10 |
| -1 | 1 | 1 | 26 | 37 | 11 |
| 1 | 1 | 1 | 60 | 52 | 8 |

Sem nenhuma análise formal, observamos que as diferenças entre os valores de y é muito menor no nível do fator C é -1. Isto indica que C tem um efeito considerável na dispersão.

Para uma análise formal, vamos considerar os modelos para a média e para a dispersão

$$y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB + \varepsilon,$$

onde $E(\varepsilon) = 0$ e

$$\text{var}(\varepsilon) = \phi = \exp(\gamma_0 + \gamma_1 A + \gamma_2 B + \gamma_3 C)$$

onde ϕ tem distribuição gama.

Como agora temos replicações, podemos estimar a variância de cada resposta e usá-la para fazer o ajuste de um modelo independentemente do modelo da média. Na Tabela 5.5 apresentamos a estimativa da variância de cada resposta para cada combinação dos fatores A e B e C.

Tabela 5.5 - Estimativa da Variância da Resposta.

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

A estimativa da variância resposta 4 é zero. Para este valor o valor da função de log-verossimilhança da distribuição gama é infinito, o que torna inviável a estimativa dos coeficientes do modelo. Entretanto podemos atribuir um valor para está resposta, digamos 0.01 e ajustar o modelo. Na Tabela 5.6 apresentamos as estimativas dos coeficientes para o modelo da dispersão, fornecidos pelo software *ARC*.

Tabela 5.6 - Coeficientes e Erro-Padrão do Modelo para a Dispersão.

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

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

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

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

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

Outra maneira de modelar a média e a variância é aplicando o procedimento de Lee e Nelder (1998), descrito na Seção 5.4.1. Nesse caso, para dar início ao procedimento iterativo, ajusta-se o modelo da média sem considerar as replicações,. Após a convergência, obtém-se um modelo para a média e para a dispersão de *y*.

Para os cálculos do modelo da média usamos a planilha *Excel* (apenas as operações com matrizes) e para o modelo da dispersão usamos o software *Arc*.

Adotamos a técnica de máxima verossimilhança restrita, uma vez que temos 6 coeficientes para estimar e apenas 16 observações. Houve convergência após 4 iterações. Na Tabela 5.7 apresentamos as estimativas dos coeficientes para o modelo da média.

Tabela 5.7 - Coeficientes e Erro-Padrão do Modelo da Média.

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

Na Tabela 5.8 apresentamos as estimativas dos coeficientes para o modelo da dispersão.

Tabela 5.8 - Coeficientes e Erro-Padrão do Modelo da Dispersão.

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

Os modelos para a média e dispersão são, portanto:

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

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

Os modelos da média obtidos pelos dois procedimentos são praticamente iguais e devem produzir as mesmas estimativas da resposta.

Os modelos da dispersão obtidos pelos dois procedimentos estão de acordo com a indicação de que o fator *C* afeta a dispersão. Pode haver casos em que haja desacordo entre os dois modelos na indicação do fator de dispersão. Nesses casos, uma explicação possível é que as medidas de variabilidade dos dois procedimentos medem variabilidades de diferentes tipos ou de fontes diferentes (McCullagh e Nelder 1989, pág. 369).

Para otimizar o processo é desejável minimizar a contração média \hat{y} do molde, bem como a sua dispersão. Com os modelos acima (obtidos por qualquer um dos dois procedimentos) podemos minimizar a média com $A = -1$ e $B = -1$, e minimizar a variância com $C = -1$.

Entretanto, para estimar o efeito do fator C sobre a dispersão, o modelo da dispersão ajustado com a variância das observações replicadas é mais confiável porque o valor esperado da variável de resposta deste modelo é a própria variância da resposta enquanto a estimativa de ϕ a partir dos resíduos do modelo da média é enviesada (Lee e Nelder 1988, pág. 96). Portanto, o efeito de diminuir o fator C de +1 para -1 é reduzir a variância em $\exp(1,79) = 5,99$.

Continuando a análise da média e da dispersão no Exemplo 4.2, vamos considerar o Modelo 5.2.

$$y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB + \beta_4 C + \beta_5 G + \beta_6 CG + \varepsilon$$

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

$$\text{var}(\varepsilon) = \phi = \exp(\gamma_0 + \gamma_1 F + \gamma_2 G + \gamma_3 AB).$$

Houve convergência após 6 iterações. Na Tabela 5.9 apresentamos as estimativas dos coeficientes para o modelo da média.

Tabela 5.9 - Coeficientes e Erro-Padrão do Modelo da Média.

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

Na Tabela 5.10 apresentamos as estimativas dos coeficientes para o modelo da dispersão.

Tabela 5.10 - Coeficientes e Erro-Padrão do Modelo da Dispersão.

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

Observamos que o fator G não deve ser considerado significativo porque o quociente entre a estimativa e o erro padrão tem módulo menor do que 1.

O modelo para a dispersão passa a ser

$$\phi = \exp(\gamma_0 + \gamma_1 F + \gamma_2 AB)$$

Repetindo o processo, após 5 iterações temos seguintes modelos para a média e a dispersão

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

$$\phi = \exp(0,1675 - 0,80F + 0,83AB)$$

O efeito de aumentar o fator F de -1 para +1 é reduzir a variância em $\exp(-0,8001) = 0,45$.

Se A e B tiverem sinais contrários a redução na variância será $\exp(-0,8295) = 0,44$, em relação a quando A e B tiverem sinais iguais.

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

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

Minimizar $0,8295AB$

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

No Exemplo 4.1 vimos que dois diferentes modelos para a média produziram diferentes modelos para a dispersão. Portanto, no método iterativo de Lee e Nelder (1998), um modelo inadequado para a média pode acarretar a identificação de fatores espúrios afetando a dispersão.

5.5. Análise da Média e da Dispersão em Presença de Fatores de Ruído

Cabe lembrar que fatores de controle são aqueles possíveis de manter em um dado nível durante o processo de produção. Fatores de ruído são aqueles que oscilam em torno de um dado nível (previamente ajustado) durante o processo de produção, embora possam ser controlados durante a fase experimental do projeto do processo ou produto.

Como foi dito no início deste Capítulo, durante a elaboração dos modelos da média e da dispersão, não é necessário fazer distinção entre fator de controle e fator de ruído. Entretanto, uma vez construídos os modelos — com eles buscam-se as condições ótimas de operação —, esses fatores são tratados de maneira diversa. É necessário identificar os fatores de controle que afetam a dispersão da resposta — através das interações com os fatores de ruído — e, então, o nível do fator de controle deve ser ajustado de modo que resulte em produtos e processos robustos às variações dos fatores de ruído.

5.5.1. Fatores de Ruído

Vamos considerar inicialmente o caso em que não há fatores de dispersão. Vejamos um exemplo simples com um fator de controle e um fator de ruído.

Na Seção 4.2 vimos um experimento de moldagem, o Exemplo 4.1. Um dos modelos ajustados foi

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

Neste experimento o fator A é a temperatura do molde. Suponha que esta temperatura pode ser controlada em condições experimentais, entretanto, é difícil de controlar nas condições de operação, quando oscila em torno de um valor previamente ajustado. O fator B permanece fixo no ponto ajustado. Então, B é um fator de controle e A é um fator de ruído. É necessário identificar como B afeta a dispersão da resposta através da sua interação AB com o fator B .

Para melhor distinguir os tipos de fatores nas equações, os fatores de ruído vão ser denominados variáveis z e os fatores de controle, variáveis x . Então, $A = z_1$ e $B = x_1$, e temos modelo:

$$y(x_1, z_1) = 27,31 + 6,94z_1 + 17,81x_1 + 5,94z_1x_1 + \varepsilon$$

onde $E(\varepsilon) = 0$ e $\text{var}(\varepsilon) = \sigma^2$, que foi estimada quando o modelo foi ajustado: $\hat{\sigma}^2 = 20,73$.

No experimento (com variáveis codificadas), a média de z_1 é zero. A variância do fator A é conhecida (σ_A^2). No experimento, os níveis mais alto (+1) e mais baixo (-1) de z_1 foram fixados em $\pm \sigma_A$. Então, a variância de z_1 é igual a 1.

Como $E(z_1) = 0$, valor esperado da resposta y é

$$E[y(x_1, z_1)] = 27,31 + 17,81x_1.$$

Aplicando o operador variância à resposta y têm-se

$$\begin{aligned} \text{var}[y(x_1, z_1)] &= \text{var}[27,31 + 6,94z_1 + 17,81x_1 + 5,94z_1x_1 + \varepsilon] \\ &= \text{var}[(6,94 + 5,94x_1)z_1] + \sigma^2 \\ &= (6,94 + 5,94x_1)^2 \sigma_z^2 + 20,73 \\ &= (6,94 + 5,94x_1)^2 + 20,73 \end{aligned}$$

Portanto, para minimizar a variância de y fazemos $x_1 = -1$.

Exemplo 5.1

Montgomery (2001), pág. 282, apresenta um experimento com quatro fatores, mostrado na Tabela 5.11. A resposta é o rendimento de um processo.

Tabela 5.11 - Experimento do Exemplo 5.1.

| $z_1 = A$ | $x_1 = B$ | $x_2 = C$ | $x_3 = D$ | y |
|-----------|-----------|-----------|-----------|-----|
| -1 | -1 | -1 | -1 | 12 |
| 1 | -1 | -1 | -1 | 18 |
| -1 | 1 | -1 | -1 | 13 |
| 1 | 1 | -1 | -1 | 16 |
| -1 | -1 | 1 | -1 | 17 |
| 1 | -1 | 1 | -1 | 15 |
| -1 | 1 | 1 | -1 | 20 |
| 1 | 1 | 1 | -1 | 15 |
| -1 | -1 | -1 | 1 | 10 |
| 1 | -1 | -1 | 1 | 25 |
| -1 | 1 | -1 | 1 | 13 |
| 1 | 1 | -1 | 1 | 24 |
| -1 | -1 | 1 | 1 | 19 |
| 1 | -1 | 1 | 1 | 21 |
| -1 | 1 | 1 | 1 | 17 |
| 1 | 1 | 1 | 1 | 23 |

No experimento original todos os fatores são de controle. Entretanto, para fins deste exemplo, vamos considerar que apenas B , C e D são fatores de controle e que A é um fator de ruído. Para análise deste experimento usamos o software *Design Expert*.

O modelo ajustado é

$$y(\mathbf{x}, z_1) = 17,38 + 2,25z_1 + 1,00x_2 + 1,63x_3 - 2,13z_1x_2 + 2,00z_1x_3 + \varepsilon$$

e a estimativa da variância do erro é $\hat{\sigma}^2 = 1,625$.

O valor esperado da resposta y é $E[y(\mathbf{x}, z_1)] = 17,38 + 1,00x_2 + 1,63x_3$.

A variância da resposta é

$$\begin{aligned} \text{var}[y(\mathbf{x}, z_1)] &= \text{var}[17,38 + 2,25z_1 + 1,00x_2 + 1,63x_3 - 2,13z_1x_2 + 2,00z_1x_3 + \varepsilon] \\ &= \text{var}[(2,25z_1 - 2,13z_1x_2 + 2,00z_1x_3)] + \sigma^2 \end{aligned}$$

$$\begin{aligned} \text{var}[y(\mathbf{x}, z_1)] &= \sigma_z^2 (2,25 - 2,13x_2 + 2,00x_3)^2 + 1,625 \\ &= 6,69 - 9,57x_2 + 9,00x_3 + 4,54x_2^2 + 4,00x_3^2 + 8,52x_2x_3 \end{aligned}$$

Lembrando que, com as variáveis codificadas, $\sigma^2 = 1$.

Para achar as condições de variância mínima usamos a opção POE (propagation of error) do *Design Expert*. Na Figura 5.7 apresentamos o gráfico POE, do desvio padrão da estimativa da resposta y versus variáveis x_2 e x_3 .

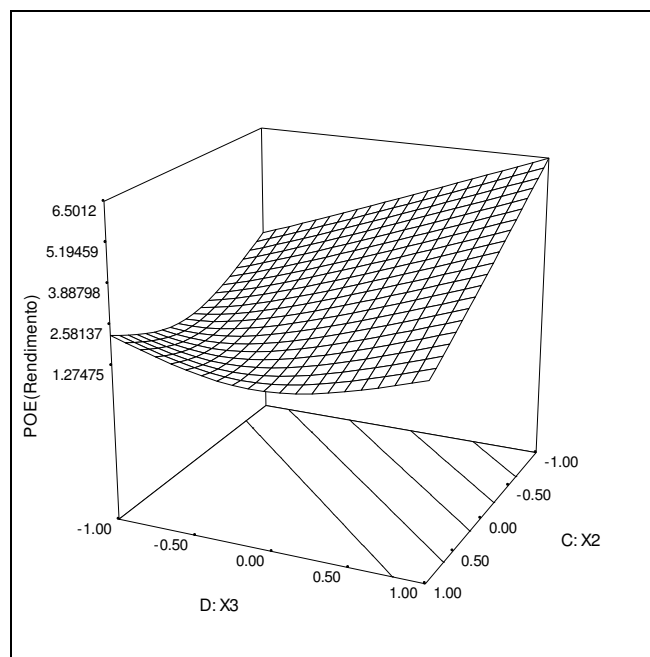


Figura 5.8 – Gráfico de POE versus x_2 e x_3 .

Observamos que a variância mínima se dá ao longo de uma reta, e é aproximadamente de 1,28. isto pode ser melhor observado no gráfico da Figura 5.8.

5.5.2. Fatores de Dispersão e Fatores de Ruído

Vamos ver um exemplo em que os fatores de ruído e de dispersão afetam simultaneamente a dispersão: o de dispersão atuando diretamente e o de ruído, através de interações.

Exemplo 4.1 (continuação)

Na Seção 5.4 tínhamos os seguintes modelos para a média e dispersão:

$$y = 27,71 + 7,68A + 18,67B + 5,77AB + \varepsilon$$

$$E(\varepsilon) = 0 \quad \text{e} \quad \text{var}(\varepsilon) = \phi = \exp(1,95 + 1,57C).$$

Neste experimento o fator A é a temperatura do molde. Suponha que B é um fator de controle e A é um fator de ruído. O fator C foi identificado como sendo um fator de dispersão.

É necessário identificar como B afeta a dispersão da resposta através da interação AB com o fator B . Façamos $A = z_1$, $B = x_1$ e $C = x_2$.

Então:

$$y(x_1, z_1) = 27,71 + 7,68z_1 + 18,67x_1 + 5,77z_1x_1 + \varepsilon$$

$$\text{var}(\varepsilon) = \phi(x_2) = \exp(1,95 + 1,57x_2).$$

Agora temos que a variabilidade de y é afetada por duas fontes: o polinômio de regressão (devido à variabilidade de z_1) e a variância de ε (devido à influência de x_2).

O valor esperado de y condicionado a um valor de z_1 é

$$E[y(x_1 | z_1)] = 27,71 + 7,68z_1 + 18,67x_1 + 5,77z_1x_1.$$

Como $E(z_1) = 0$, o valor esperado de y é

$$E(y) = E\{E[y(x_1 | z_1)]\} = 27,71 + 18,67x_1.$$

Aplicando o operador variância á resposta y têm-se

$$\begin{aligned} \text{var}(y) &= \text{var}(27,71 + 7,68z_1 + 18,67x_1 + 5,77z_1x_1) + \text{var}(\varepsilon) \\ &= \text{var}[(7,68 + 5,77x_1)z_1] + \exp(1,95 + 1,57x_2) \\ &= (7,68 + 5,77x_1)^2 \sigma_z^2 + \exp(1,95 + 1,57x_2) \\ &= (7,68 + 5,77x_1)^2 + \exp(1,95 + 1,57x_2) \end{aligned}$$

Para otimizar o processo é desejável minimizar $E(y)$ e $\text{var}(y)$. Ambos objetivos são alcançados com $x_1 = -1$ e $x_2 = -1$.

Neste exemplo, a distribuição de y é normal e a função de ligação é a identidade. Pode-se generalizar este procedimento com a estrutura dos modelos lineares generalizados, o que veremos a seguir.

5.5.3 Procedimento Geral - Fatores de Dispersão e Fatores de Ruído

Sejam $x_j, j = 1, 2, \dots, q$, as variáveis de controle e $z_k, k = 1, 2, \dots, m$, as variáveis de ruído.

$$\text{Então: } \mathbf{x}'_i = (x_{i1}, \dots, x_{iq}) \text{ e } \mathbf{z}'_i = (z_{i1}, \dots, z_{im})$$

Seja $\mathbf{w}'_i, i = 1, 2, \dots, n$, a i -ésima linha da matriz das variáveis de regressão, onde cada elemento de \mathbf{w}'_i pode ser:

- O nível de uma variável de controle (x_{ij}); ou de uma variável de ruído (z_{ik})
- O quadrado do nível de uma variável de controle (x_{ij}^2), ou de uma variável de ruído (z_{ik}^2)
- O produto entre os níveis das variáveis de controle ($x_{ij} x_{il}$), ou das variáveis de ruído ($z_{ik} z_{ir}$)
- O produto entre os níveis das variáveis de controle e de ruído ($x_{ik} z_{ir}$).

Então, modelo para a média é

$$\eta_i = g[\mu_i(\mathbf{x}_i, \mathbf{z}_i)] = \mathbf{w}'_i \boldsymbol{\beta} \text{ ou } \mu_i(\mathbf{x}_i, \mathbf{z}_i) = g^{-1}(\mathbf{w}'_i \boldsymbol{\beta}),$$

o modelo para a dispersão é

$$\phi_i = \exp(\mathbf{w}'_i \boldsymbol{\gamma}),$$

o valor esperado condicional de y dados os valores das variáveis de ruído é então

$$E(y_i | \mathbf{z}) = \mu_i(\mathbf{x}_i | \mathbf{z}_i)$$

e a média e a variância de y_i são

$$\begin{aligned} E(y_i) &= E[E(y_i | \mathbf{z})] \\ \text{var}(y_i) &= \text{var}[E(y_i | \mathbf{z})] + E[\text{var}(y_i | \mathbf{z})] \end{aligned} \quad (5.7)$$

Mas, $\text{var}(y_i | \mathbf{z}) = \phi_i V(\mu_i)$, então:

$$\text{var}(y_i) = \text{var}[E(y_i | \mathbf{z})] + E[\phi_i V(\mu_i)]$$

cuja variável de resposta é d_i , com distribuição gama.

5.6. Roteiro para Analisar a Média e a Variância

Eis um roteiro sintético do procedimento geral, seja para a análise da média (Seção 5.6.1) seja para a análise conjunta da média e da dispersão (Seção 5.6.2).

5.6.1. Análise da Média

Se há evidências de que os três pressupostos — normalidade, aditividade e variância constante — possam ser considerados, usar o modelo linear e a resposta com distribuição normal; caso contrário usar os modelos linear generalizado.

I. Modelo Linear –Resposta com Distribuição Normal

1. Estimativa dos Parâmetros – Seção 2.1
2. Teste de Significância dos Parâmetros – Seção 2.2
3. Verificação da Adequação do Modelo – Seção 2.3

Se o modelo linear não for considerado adequado, usar a resposta transformada.

II. Modelo Linear – Resposta Transformada com Distribuição Normal

1. Definir a Transformação da Resposta – Seção 4.3.1
2. Estimativa dos Parâmetros – Seção 2.1
3. Teste de Significância dos Parâmetros – Seção 2.2
4. Verificação da Adequação do Modelo – Seção 2.3

Se o modelo linear com a resposta transformada não for considerado adequado, usar os modelos lineares generalizados.

III. Modelos Lineares Generalizados

1. Escolha da Distribuição de Probabilidade e Função de Ligação para a Média – Seção 4.4.1.

Se houve estágios anteriores de modelagem da resposta pode haver indicações para a função de variância e para a função de ligação.

- Indicação de que a variância cresce com o aumento da média – Seção 2.3.1.3.
- Indicação para a função de ligação – Seção 4.1.

Se há indicação de que a variância cresce com o aumento da média, usar as distribuições gama, normal inversa e a função de quase-verossimilhança com a função de variância igual a média; e usar a função de ligação indicada.

2. Estimativa dos Parâmetros – Seção 3.3.
3. Teste de Significância dos Parâmetros – Seção 3.4.
4. Verificação da Adequação do Modelo – Seção 3.5.

5.6.2. Análise Conjunta da Média e da Dispersão

1. Escolha da Distribuição de Probabilidade e Função de Ligação para a Média – Seção 4.4.1.

Se houve estágios anteriores de modelagem da resposta pode haver indicações para a função de variância e para a função de ligação.

- Indicação de que a variância cresce com o aumento da média – Seção 2.3.1.3.
- Indicação para a função de ligação – Seção 4.1.

2. Modelo para a Média – Seção 5.4.1.1.

3. Modelo para a dispersão – Seção 5.4.1.2.

4. Procedimento Iterativo para Estimar os Parâmetros – Seção 5.4.1.3.

5. Teste de Significância dos Parâmetros:

– Seção 3.7.3 para o modelo da média

– Seção 3.4 para o modelo da dispersão.

6. Verificação da Adequação do Modelo:

– Seção 3.7.4 para o modelo da média

– Seção 3.5 para o modelo da dispersão.

7. Análise na Presença de Fatores de Ruído

– Seção 5.5.1 para quando só há fatores de ruído

– Seção 5.5.2 para quando há fatores de dispersão e fatores de ruído

Caso não seja encontrado um modelo adequado para a média, podem-se usar os modelos **não lineares** generalizados. Esses tipos de modelos foram propostos por Lane (1996), citado por Weisberg (2004). Neles tem-se a mesma estrutura dos modelos lineares generalizados, com exceção do polinômio de regressão, que não é linear. Neste caso, é necessário definir este polinômio, ou seja, deve ser conhecida a relação funcional entre os fatores e a variável de resposta. Weisberg (2004) descreve como aplicar modelos não lineares generalizados usando a mais recente versão do software *ARC*.