



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Faculdade de Matemática

Edmilson Rodrigues Pinto

**Modelagem da variabilidade em experimentos com
misturas**

Uberlândia
2021

Edmilson Rodrigues Pinto

Modelagem da variabilidade em experimentos com misturas

Tese apresentada à Faculdade de Matemática para
promoção ao cargo de Professor Titular da
Universidade Federal de Uberlândia.

Uberlândia
2021

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

P659m Pinto, Edmilson Rodrigues, 1971-
2021 Modelagem da variabilidade em experimentos com misturas [recurso eletrônico] / Edmilson Rodrigues Pinto. - 2021.

Tese (Promoção para classe E - Professor Titular) - Universidade Federal de Uberlândia, Faculdade de Matemática.

Modo de acesso: Internet.

Disponível em: <http://doi.org/10.14393/ufu.te.2021.5536>

Inclui bibliografia.

Inclui ilustrações.

1. Estatística. I. Universidade Federal de Uberlândia. Faculdade de Matemática. III. Título.

CDU: 311.1

Glória Aparecida
Bibliotecária - CRB-6/2047

FOLHA DE APROVAÇÃO

Abaixo são apresentados os membros da Comissão Julgadora da sessão pública de defesa de tese, intitulada: “Modelagem da variabilidade em experimentos com misturas”, para promoção ao cargo de Professor Titular na área de concentração Estatística, a que se submeteu o professor **Edmilson Rodrigues Pinto**, em 27 de agosto de 2021 na Faculdade de Matemática da Universidade Federal de Uberlândia, em Uberlândia - MG.

Prof. Dr. Ednaldo Carvalho Guimarães

Presidente da Comissão Julgadora
Universidade Federal de Uberlândia - UFU

Profa. Dra. Clarice Garcia Borges Demétrio

Examinadora Externa
Universidade de São Paulo - USP

Prof. Dr. João Domingos Scalon

Examinador Externo
Universidade Federal de Lavras - UFLA

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho

Examinador Externo
Universidade Federal de Lavras - UFLA

A Ata da Defesa, assinada pelos membros da Comissão Julgadora, pode ser obtida em: https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 2997917 e o código CRC 9896170A.

Agradecimentos

À minha esposa Aurélia e aos meus filhos: Sabrina e Guilherme.

À minha mãe, pelo apoio e carinho.

Ao professor Leandro Alves Pereira, meu parceiro nas pesquisas e a quem devo muito do trabalho realizado nesta tese.

À FAPEMIG pelo apoio financeiro ao Projeto de Pesquisa APQ-03365-18, sob minha coordenação, cuja tese é fruto de parte dos resultados desse projeto.

À Faculdade de Matemática e à Universidade Federal de Uberlândia por oferecerem um ambiente agradável para o desenvolvimento desta pesquisa.

Aos professores, funcionários e colegas da Faculdade de Matemática da UFU.

Aos membros da Comissão Julgadora desta tese; professores Ednaldo, Clarice, Júlio e Scalon.

Resumo

Em experimentos industriais, controlar a variabilidade é de suma importância para garantir a qualidade dos produtos. Modelos de regressão clássicos são amplamente utilizados na indústria para experimentos com mistura; porém, quando a suposição de variância constante não é satisfeita, a construção de procedimentos que permitam minimizar a variabilidade torna-se necessária e outros métodos de modelagem estatística devem ser considerados. A abordagem considerada nesta tese utiliza a classe dos modelos lineares generalizados. Esta classe é bem geral e bastante flexível, generalizando algumas das mais importantes distribuições de probabilidade e possibilita modelar a variabilidade através da modelagem conjunta da média e da dispersão (MCMD). A MCMD fornece um método eficiente para estimar os parâmetros dos modelos conjuntos da média e dispersão; contudo o processo de seleção de variáveis não é claro e se baseia em critérios subjetivos para escolha dos termos dos modelos. Nesta tese o problema de seleção de variáveis na MCMD é resolvido. Um procedimento de seleção de variáveis, com base em teste de hipóteses e na qualidade do ajuste do modelo é proposto. Métodos de simulação são utilizados para verificar a eficiência do procedimento. O procedimento de seleção de variáveis é adaptado para o caso de experimentos com misturas e resulta num método bastante eficiente de obtenção dos modelos da média e da variância. A teoria de planejamento ótimo de experimentos é apresentada para a MCMD e aplicada a experimentos de mistura. Os resultados obtidos nesta tese através da aplicação da MCMD em experimentos com misturas são bastante animadores, indicando que a teoria promete ser de grande utilidade na modelagem e na condução de experimentos industriais envolvendo misturas.

Palavras-chave: Experimentos com misturas, seleção de variáveis, modelagem conjunta da média e dispersão, modelos lineares generalizados, planejamento ótimo de experimentos, planejamento robusto.

Códigos de Classificação Matemática por Assunto: **62JXX:** Inferência linear, regressão, **62J12:** Modelos lineares generalizados, **62KXX:** Planejamento de experimentos, **62K05:** Planejamento ótimo de experimentos, **62K25:** Planejamento robusto, **62K20:** Superfície de resposta, **62PXX:** Aplicações de estatística, **62P30:** Aplicação na engenharia e na indústria.

Abstract

In industrial experiments, controlling variability is of paramount importance to ensure product quality. Classical regression models are widely used in industry for mixture experiments; however, when the assumption of constant variance is not satisfied, the building of procedures that allow to minimize the variability becomes necessary and other methods of statistical modeling should be considered. The approach considered in this thesis uses the class of generalized linear models. This class is very general and quite flexible, generalizing some of the most important probability distributions and allowing to model variability through of the joint modeling for mean and dispersion (JMMD). The JMMD provides an efficient method for estimating the parameters of the joint models of mean and dispersion; however, the variable selection process is not clear and is based on subjective criteria for choosing the terms of the models. In this thesis the variable selection problem in JMMD is solved. A variable selection procedure, based on hypothesis testing and the goodness of fit of the model is proposed. Simulation methods are used to verify the efficiency of the procedure. The variable selection procedure is adapted for the case of experiments with mixtures and results in a very efficient method for obtaining the mean and variance models. The theory of optimal design of experiments is presented to the JMMD and applied to mixture experiments. The results obtained in this thesis through the application of JMMD in mixture experiments are very encouraging, indicating that the theory promises to be of great utility in modeling and conducting industrial experiments involving mixtures.

Keywords: Mixture experiments, variable selection, joint modeling of mean and dispersion, generalized linear models, design of experiments, optimal experimental design, robust parameter designs.

Mathematics Subject Classification Codes: **62JXX:** Linear inference, regression, **62J12:** Generalized linear models, **62KXX:** Design of experiments, **62K05:** Optimal designs, **62K25:** Robust parameter designs, **62K20:** Response surface designs, **62PXX:** Applications of statistics, **62P30:** Application in engineering and industry.

Lista de Figuras

2.1	Região simplex 2-dimensional para uma mistura com três componentes.	27
2.2	Planejamento lattice-simplex (3, 2).	29
2.3	Desenho centróide-simplex para $a = 3$	29
3.1	Desenho fatorial completo 3^2 em duas variáveis de processo cruzado com um desenho centroide simplex em três componentes de mistura	44
3.2	Gráficos de diagnóstico para dados de mistura de pão.	47
4.1	Gráficos, obtidos das simulações da distribuição Normal, quanto à relação entre tamanho da amostra e porcentagem encontrada para os modelos Ótimo e Tipo 1, para os critérios \tilde{R}_m^2 , \tilde{R}_d^2 , $EAIC$ e AIC_c , onde a linha sólida representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2) com $\lambda_n = \log n$, a linha tracejada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = \sqrt{n}$, a linha pontilhada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = 1$ e a linha tracejada e pontilhada representa o gráfico para $EAIC(AIC_c)$	63
4.2	Gráficos, obtidos nas simulações da distribuição Binomial, da relação entre tamanho da amostra e porcentagem encontrada para os modelos Ótimo e Tipo 1, para os critérios \tilde{R}_m^2 , \tilde{R}_d^2 , $EAIC$ e AIC_c , onde a linha sólida representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2) com $\lambda_n = \log n$, a linha tracejada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = \sqrt{n}$, a linha pontilhada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = 1$ e a linha tracejada e pontilhada representa o gráfico para $EAIC(AIC_c)$	66
4.3	Gráficos, obtidos nas simulações da distribuição Poisson, da relação entre tamanho da amostra e porcentagem encontrada para os modelos Ótimo e Tipo 1, para os critérios \tilde{R}_m^2 , \tilde{R}_d^2 , $EAIC$ e AIC_c , onde a linha sólida representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2) com $\lambda_n = \log n$, a linha tracejada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = \sqrt{n}$, a linha pontilhada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = 1$ e a linha tracejada e pontilhada representa o gráfico para $EAIC(AIC_c)$	68
4.4	Gráficos de diagnóstico para o modelo da média obtido dos dados de moldagem por injeção.	71
4.5	Gráficos de diagnóstico para o modelo da dispersão obtido dos dados de moldagem por injeção.	73
5.1	Gráficos de diagnóstico para o modelo da média obtido para o problema da panificação.	85
5.2	Gráficos de diagnóstico para o modelo da dispersão obtido para o problema da panificação.	85
6.1	Desenho D -ótimo Bayesiano para o modelo conjunto da média e dispersão aplicado ao experimento de mistura de farinha na situação de ruído $z_1 = 0$ e $z_2 = 0$	94

Lista de Tabelas

2.1	Resumo da MCMD para a i -ésima resposta	35
3.1	Volume de pão para os 10 tipos de farinha e as 9 combinações das variáveis de ruído	45
3.2	Coefficientes de regressão do modelo da média usando o método de MQO	46
3.3	Valores ótimos de mistura com seus respectivos valores de $Var(Y)$ e coeficiente de variação para alguns cenários envolvendo as variáveis de ruído: tempo de mistura e tempo de descanso.	50
4.1	Porcentagens de tipos de modelos encontrados em 1000 simulações de Monte Carlo do procedimento de seleção de variáveis proposto para a distribuição Normal	62
4.2	Porcentagens dos tipos de modelos encontrados em 1000 simulações de Monte Carlo do procedimento de seleção de variáveis proposto para a distribuição Binomial	65
4.3	Porcentagens dos tipos de modelos encontrados em 1000 simulações de Monte Carlo do procedimento de seleção de variáveis proposto para a distribuição de Poisson	67
4.4	Dados experimentais de moldagem por injeção	71
4.5	Coefficientes de regressão e teste de Wald para os modelos da média e dispersão obtidos dos dados de moldagem por injeção.	72
4.6	Passos do algoritmo de seleção de variáveis para a MCMD aplicado aos dados do experimento de moldagem por injeção.	74
5.1	Passos do algoritmo de seleção de variáveis para a MCMD aplicado ao problema de panificação.	83
5.2	Coefficientes de regressão e teste de Wald para os modelos da média e dispersão obtidos do problema de panificação.	84

Sumário

1	Introdução	12
1.1	Objetivos da tese	14
1.2	Organização da tese	15
2	Revisão da literatura e conceitos básicos	17
2.1	Introdução	17
2.2	Revisão da literatura	17
2.3	Conceitos básicos	26
2.3.1	Experimentos com misturas	26
2.3.2	Modelos Lineares Generalizados	31
2.3.3	Planejamento ótimo de experimentos	38
3	Variabilidade em experimentos de mistura com variáveis de ruído	42
3.1	Introdução	42
3.2	Método delta	43
3.3	Modelagem e otimização em um problema de panificação	43
3.4	Considerações finais	50
4	Procedimento de seleção de variáveis na MCMD	52
4.1	Introdução	52
4.2	Critério e procedimento para seleção de variáveis na MCMD	54
4.3	Avaliação numérica	60
4.3.1	Estudo de simulação para a distribuição Normal	61
4.3.2	Estudo de simulação para a distribuição Binomial	64
4.3.3	Estudo de simulação para a distribuição de Poisson	66
4.4	Aplicação	68
4.5	Considerações finais	71
5	MCMD aplicada a experimentos com misturas	75
5.1	Introdução	75
5.2	Processo de seleção de variáveis na MCMD com misturas	77
5.3	Aplicação ao problema de panificação	79
5.4	Considerações finais	82
6	Desenho ótimo para MCMD - aplicação em modelos de mistura	87
6.1	Introdução	87
6.2	Teoria da Equivalência	88
6.3	Teorema Geral da Equivalência para a MCMD	90
6.3.1	Critério D -ótimo Bayesiano	91

6.4	Aplicação em experimentos de mistura	92
6.5	Considerações finais	95
7	Conclusões	97
	Referências	100
A	Apêndice - Capítulo 4	107
A.1	Sequência para gerar dados normais com parâmetro de dispersão variável . . .	107
A.2	Sequência para gerar dados binomiais com parâmetros de dispersão variável .	107
A.3	Sequência para gerar dados de Poisson com parâmetros de dispersão variável .	108
B	Apêndice - Capítulo 6	109
B.1	Prova da Proposição (6.3.1)	109
B.2	Prova da Proposição (6.3.2)	109

Capítulo 1

Introdução

Produtos envolvendo misturas têm grande impacto na nossa vida cotidiana. Diariamente, do início ao fim do dia, dependemos de produtos provenientes de misturas como, por exemplo, alimentos, bebidas, cosméticos, medicamentos, vacinas, combustível, tintas, cimentos, detergentes, corantes, borracha, etc. Assim, problemas com misturas são comumente encontrados em formulações de produtos industriais envolvendo processamento de alimentos, composições químicas, fibras têxteis, drogas farmacêuticas e muitos outros.

Experimentos com mistura envolvem misturas de dois ou mais ingredientes para formar um produto final. Para tais experimentos é de interesse determinar as proporções dos componentes da mistura que conduzem a resultados ótimos em relação a alguma característica de interesse. O processo experimental envolve o planejamento e a execução do experimento, a construção do modelo de regressão, otimização, análise e interpretação dos resultados.

A relação entre a variável resposta dependente e os termos independentes (fatores, covariáveis ou variáveis explicativas) podem ser formalizadas em um modelo matemático, o qual consiste em uma tentativa de descrever matematicamente um fenômeno considerado. Nos casos em que a variável resposta apresenta incerteza, modelos estatísticos podem e devem ser considerados. Os modelos estatísticos de regressão são amplamente usados em diversas situações e são peças chave no contexto desta tese.

No caso de experimentos com misturas, o objetivo principal do pesquisador é determinar as proporções dos componentes da mistura que conduzem a resultados desejáveis em relação a alguma característica de qualidade de interesse.

Em problemas envolvendo misturas, as respostas a serem observadas e medidas estão funcionalmente relacionadas às proporções dos seus componentes, podendo também existir relação com outros fatores externos à mistura. Modelos de regressão, geralmente com respostas Normais, são comumente usados para experimentos com misturas. Essa relação de dependência proporciona a possibilidade de se usar uma abordagem, denominada metodologia de superfície de resposta, a qual é aplicada quando o interesse é otimizar uma resposta influenciada por diversos fatores. Associado ao modelo utilizado, o planejamento experimental é fundamental no processo de obtenção de um modelo eficiente e na economia de recursos.

Um dos grandes desafios da indústria é controlar a variabilidade dos produtos e, desta forma, obter produtos mais robustos, evitando retrabalho, diminuindo desperdícios e custos de produção excessivos. A variabilidade em experimentos com mistura pode acontecer devido à natureza estocástica do processo produtivo e também como consequência da ocorrência de fatores de ruído, os quais são fatores externos que não podem ser controlados pelo experimentador. Portanto, modelar e controlar a variabilidade é de suma importância.

O modelo de regressão clássico, o qual considera distribuição Normal para a variável resposta e variância constante tem sido usado sistematicamente na modelagem de experimentos envolvendo misturas, entretanto, em muitas situações essa abordagem não pode ser aplicada e outros tipos de modelos devem ser considerados.

Nesta tese, a modelagem da variabilidade em experimentos com misturas será discutida. O assunto será abordado considerando modelos lineares generalizados, que generalizam o modelo de regressão clássico e permitem que outros tipos de distribuições sejam consideradas. A metodologia utilizada será baseada na modelagem conjunta da média e da dispersão (MCMD), que permite de forma bastante eficiente modelar a variabilidade em experimentos industriais. O processo de seleção de variáveis na MCMD será abordado, tanto para o caso geral, quanto para aplicação em experimentos com misturas. A obtenção de desenhos ótimos na MCMD para misturas também será considerada.

A relevância desta tese está na proposição de um procedimento geral e eficiente de obtenção de modelos da média e da variância em experimentos com misturas. A abordagem da MCMD permite que tais modelos sejam obtidos; além disso, usando a MCMD, outros tipos de distribuição de probabilidade, além da distribuição Normal, podem ser utilizadas para a variável resposta.

A motivação para este trabalho está na possibilidade de encontrar modelos adequados para média e variância da resposta em experimentos de mistura; o que possibilitará, em situações de planejamento para a melhoria qualidade, a formulação de problemas de otimização com o objetivo de encontrar valores ótimos dos níveis de mistura, de modo a diminuir a variância e tornar a média da resposta insensível a fatores de ruído. Além disso, a obtenção de modelos adequados permitem a construção eficaz de desenhos ótimos, aumentando a eficiência e possibilitando a redução de custos.

A principal aplicação da teoria desenvolvida nesta tese é em processos industriais da engenharia da qualidade; contudo, ela pode ser aplicada a qualquer outra área que se utilize experimentos envolvendo misturas. Experimentos com misturas, modelos lineares generalizados e planejamento ótimo de experimentos constituem a base teórica para o entendimento deste trabalho.

1.1 Objetivos da tese

O principal assunto desta tese concentra-se na modelagem da variabilidade em experimentos com misturas. O problema básico com o qual estamos concentrados ocorre quando o interesse do pesquisador está não somente no valor esperado da resposta, mas também na sua variância.

Com o surgimento das ideias e dos métodos de Taguchi (TAGUCHI, 1986), na década de 1980, e a comprovada eficácia de sua metodologia, em processos industriais na área de engenharia da qualidade (PHADKE, 1989), muitos pesquisadores começaram a discutir a proposta de Taguchi e a propor métodos mais eficientes para obtenção dos seus propósitos. Esses pesquisadores passaram a sentir a necessidade de modelar não somente a média da resposta, mas também sua variância. De acordo com Nelder (NAIR, 1992), as ideias de Taguchi sobre experimentos fatoriais fracionados não são novas, mas é nova a ideia de modelar simultaneamente a média e a variância. É baseado nesta ideia que propomos modelar a variabilidade em experimentos de mistura.

Nossa abordagem para o problema em questão baseia-se na modelagem conjunta da média e da dispersão, proposta por Nelder e Lee (1991), veja Subseção 2.3.2. A abordagem da MCMD que consideraremos, utiliza como critério de otimização a quase-verossimilhança estendida ajustada, mostrada na equação na equação (2.3.13), e permite encontrar de forma eficiente modelos para média e para dispersão. Entretanto, o processo de seleção de variáveis na MCMD

não é claro e baseia-se em conhecimento subjetivo dos fatores que presumivelmente afetam os modelos da média e da dispersão, ou na utilização de gráficos normais dos efeitos para escolha desses fatores.

Nesta tese discutimos a modelagem da variabilidade em experimentos envolvendo misturas. A MCMD é utilizada como uma forma eficiente de obtenção de modelos para média e variância em problemas de misturas. Propomos um procedimento eficiente de seleção de variáveis na MCMD e adaptamos esse procedimento para o caso de experimentos com misturas. A teoria para obtenção de desenhos ótimos para a MCMD em experimentos de mistura também é apresentada.

Desta forma, os objetivos da tese são: 1) discutir o processo de variabilidade em experimentos com mistura; 2) mostrar como a MCMD pode ser usada para encontrar modelos da média e da variância em experimentos de mistura; 3) apresentar um processo eficiente de seleção de variáveis na MCMD, no caso geral e, em especial, aplicado a experimentos com misturas; 4) mostrar como a teoria de experimentos ótimos pode ser utilizada para obtenção de desenhos ótimos na MCMD aplicada a experimentos envolvendo misturas.

1.2 Organização da tese

O Capítulo 2 contém duas seções. A primeira, contendo a revisão da literatura sobre os três assuntos abordados na tese, a saber: modelagem conjunta da média e dispersão, planejamento ótimo de experimentos para modelos envolvendo variância não constante e modelagem de experimentos com misturas exibindo variabilidade. A segunda, contendo um resumo sobre os conceitos básicos, necessários para o melhor entendimento da tese. O objetivo deste capítulo é dar ao leitor um completo entendimento sobre o estado da arte e fornecer ferramentas necessárias para a melhor compreensão do texto.

Os Capítulos 3 ao 6 descrevem os principais resultados da tese. A apresentação de cada um desses capítulos é similar. Primeiro, uma breve introdução descreve o problema específico a ser tratado e menciona trabalhos relacionados ao problema. Depois se seguem seções com o desenvolvimento do assunto e métodos para resolução do problema. A penúltima seção apresenta exemplos da teoria introduzida nas seções anteriores. Finalmente, a última seção contém as considerações finais sobre o capítulo.

O capítulo 3 mostra como o método delta é utilizado para modelar a variabilidade em experimentos de mistura com variáveis de ruído. Um exemplo de aplicação na indústria de alimentos é apresentado e um processo de otimização é utilizado para obtenção dos níveis ótimos de mistura, de modo que a média da resposta seja robusta aos fatores de ruído.

Os capítulos 4 e 5 tratam da seleção de variáveis na MCMD. O capítulo 4 apresenta o procedimento para seleção de variáveis na modelagem conjunta da média e dispersão. O procedimento proposto é baseado em teste de hipóteses e na qualidade do ajuste do modelo. Um critério de ajuste do modelo, um tipo de coeficiente de determinação ajustado estendido, é proposto para a MCMD. Estudos de simulação são utilizados para verificar a eficiência do procedimento proposto. O processo de seleção de variáveis é aplicado a um exemplo real de um experimento industrial. O Capítulo 5 é uma extensão do capítulo 4 para o caso de experimentos com misturas. O procedimento de seleção de variáveis proposto no Capítulo 4 é adaptado para o caso de misturas. O exemplo introduzido no Capítulo 3 é utilizado para exemplificar o procedimento proposto.

O Capítulo 6 apresenta a metodologia para obtenção de desenhos D -ótimos em modelos conjuntos da média e da dispersão. A prova do Teorema Geral da Equivalência para a MCMD é apresentada. A teoria é aplicada ao modelo de mistura com variáveis de ruído, obtido no Capítulo 5; para o qual o desenho D -ótimo é obtido para uma combinação específica de variáveis de ruído.

O capítulo 7 traz as conclusões da tese; com um resumo de todos os resultados principais obtidos nos capítulos e sugestões para pesquisa futura. Como cada capítulo contém sua própria seção de discussão e conclusões, algumas dessas conclusões foram omitidas no capítulo final. O leitor pode referir à última seção de cada capítulo para conclusões mais específicas.

Capítulo 2

Revisão da literatura e conceitos básicos

2.1 Introdução

Este capítulo contém revisão da literatura e conceitos básicos, referentes aos temas abordados na tese.

A Seção 2.2 contém a revisão da literatura sobre os principais temas tratados nesta tese: modelagem conjunta da média e dispersão (MCMD) e variabilidade em experimentos envolvendo misturas. Desenhos ótimos para modelos abordando variabilidade, como a MCMD, também são considerados.

A Seção 2.3 trata dos conceitos básicos envolvidos na tese. O objetivo desta seção é fornecer um resumo e referências sobre temas básicos, citados nos capítulos seguintes e fundamentais para o entendimento desses capítulos.

2.2 Revisão da literatura

Nesta tese, o assunto principal é a modelagem conjunta da média e dispersão e sua aplicação em experimentos envolvendo misturas. A MCMD é muito útil para modelar variabilidade, especialmente em situações onde os experimentos são não replicados e possuem poucas observações, como aqueles utilizados em experimentos industriais. Entretanto, a seleção de variáveis na MCMD não está totalmente resolvida na literatura.

Os procedimentos usados para selecionar modelos na regressão linear clássica podem ser estendidos diretamente para a classe de modelos lineares generalizados (MLGs). Porém, a escolha do modelo envolve, além da definição da matriz do modelo, que definirá o preditor

linear, a definição da distribuição da variável resposta, que determina a função de variância, e a escolha da função de ligação. Na MCMD, devemos definir as matrizes do modelo para ambos os modelos da média e da dispersão. Para o modelo da média também é necessário definir a função de variância e a função de ligação. Para o modelo da dispersão, assume-se a distribuição Gama com função de ligação logarítmica.

A MCMD tem um enorme potencial de aplicação na ciência e na indústria e, portanto, tem atraído a atenção de muitos pesquisadores. A seguir, apresentamos alguns trabalhos relevantes nesta área.

Aitkin (1987) considerou a modelagem conjunta da média e da variância para modelos Normais, usando um modelo de regressão log-linear para a variância.

Davidian e Carroll (1987) consideraram o método da Pseudo-Verossimilhança (PV) para o ajuste do modelo na modelagem conjunta da média e dispersão. De acordo com Davidian e Carroll (1988), o método de pseudo-verossimilhança é assintoticamente equivalente à regressão ponderada sobre o quadrado dos resíduos com pesos estimados; e a iteração completa de tal regressão produz a estimativa de pseudo-verossimilhança. Para mais detalhes sobre a pseudo-verossimilhança, veja Carroll e Ruppert (1988), p. 71.

Smyth (1989) considerou uma abordagem mais geral, usando modelos lineares generalizados com parâmetro de dispersão variável, propondo a classe de modelos lineares generalizados duplos (MLGDs). Nessa abordagem, vários resultados como expressões para as equações de verossimilhança, testes assintóticos, algoritmos de estimação e suas propriedades de convergência, foram obtidos para as distribuições Normal, Inversa Gaussiana e Gama.

Nelder e Lee (1991) introduziram a MCMD para planejamento robusto, proposto por Taguchi (TAGUCHI, 1986). Eles observaram que a metodologia de Taguchi conduzia à modelagem conjunta da média e dispersão, com a utilização da classe dos modelos lineares generalizados. O ajuste da MCMD usa como critério de otimização a Quase-Verossimilhança Estendida (QVE), definida por Nelder e Pregibon (1987) e posteriormente desenvolvida por Godambe e Thompson (1989) e McCullagh e Nelder (1989). Nelder e Lee (1991) mostraram que a MCMD é geral e suficiente para a análise dos modelos de Taguchi.

A QVE, estendeu a definição original de quase-verossimilhança em modelos lineares generalizados de Wedderburn para permitir a comparação de modelos com diferentes funções de variância ou estruturas de dispersão. A Quase-Verossimilhança (QV), proposta

por Wedderburn (1974), tem estimativas de máxima quase-verossimilhança, que têm muitas propriedades em comum com as estimativas de máxima verossimilhança. O estimador de quase-verossimilhança é usado na análise de dados onde uma verossimilhança não é disponível. Lee e Nelder (1999) mostraram que os estimadores de máxima quase-verossimilhança são robustos contra uma possível especificação incorreta da verossimilhança ou dos cumulantes do modelo em relação à classe de todas as distribuições com uma determinada função de variância.

Davidian e Carroll (1988) mostraram a equivalência assintótica entre a QVE e a PV. Eles também mostraram que a QVE pode ser afetada por um viés assintótico quando a distribuição subjacente é assimétrica e fora da família exponencial.

Nelder e Lee (1992) compararam a quase-verossimilhança estendida e a pseudo-verossimilhança em estudos de simulação com amostras finitas. Eles verificaram limitações dos argumentos assintóticos nas propriedades das estimativas dos parâmetros de dispersão e concluíram que as eficiências relativas assintóticas podem ser bastante enganosas quando extrapoladas para amostras finitas. Além disso, seus estudos de simulação também mostraram que o estimador de máxima quase-verossimilhança estendida é geralmente superior ao estimador de pseudo-verossimilhança para minimizar o erro quadrático médio, veja também Nelder (2000). De acordo com Nelder e Lee (1992), para conjuntos de dados nos quais a dispersão deve ser modelada com quantidades razoavelmente pequenas de dados, como em experimentos de planejamento robusto, o melhor critério de otimização para o ajuste do modelo na modelagem conjunta da média e dispersão é a quase-verossimilhança estendida.

Verbyla (1993) considerou modelos normais heterocedásticos, para os quais suspeita-se que a variância possui uma dependência log-linear em relação às covariáveis. Verbyla discutiu diagnósticos para a máxima verossimilhança restrita e testes para a heterogeneidade da variância.

Smyth e Verbyla (1996) forneceram uma abordagem de verossimilhança condicional para a máxima verossimilhança restrita em modelos lineares generalizados. Eles mostraram que a máxima verossimilhança restrita tem interpretação como uma verossimilhança condicional, onde o condicionamento é feito sobre uma estatística suficiente apropriada, a fim de remover a dependência sobre os parâmetros de ruído. Usando a representação de verossimilhança

condicional, o conceito de máxima verossimilhança restrita é estendido para modelos lineares generalizados com parâmetro de dispersão variável e ligação canônica.

Lee e Nelder (1996) introduziram a classe de modelos lineares generalizados hierárquicos (MLGHs) como uma síntese da MCMD, generalizando as ideias de quase-verossimilhança e QVE. Para mais detalhes sobre MLGHs, veja Lee e Nelder (2001) e Lee, Nelder e Pawitan (2006).

Engel e Huele (1996) formularam uma extensão do método de superfície de resposta de Myers, Khuri e Vining (1992), usando modelos lineares generalizados Normais. Os autores propuseram um modelo linear generalizado com um parâmetro de dispersão não constante. A metodologia proposta por eles é baseada na modelagem simultânea da superfície de resposta para a média e para a variância. Na verdade, o que Engel e Huele propuseram foi apenas um caso particular da MCMD, proposta por Nelder e Lee (1991), considerando distribuição Normal para a variável resposta no modelo da média.

Dey, Gelfand e Peng (1997) consideraram uma abordagem Bayesiana para modelos lineares generalizados com sobredispersão, usando a família exponencial dupla de Efron (EFRON, 1986). Os cálculos Bayesianos foram conduzidos usando o algoritmo de Metropolis-Hastings, dentro do Amostrador de Gibbs.

Lee e Nelder (1998) introduziram um método de ajustamento para a máxima verossimilhança restrita na MCMD proposta por Nelder e Lee (1991). O ajustamento é importante quando o número de parâmetros no modelo da média é uma grande fração do tamanho da amostra.

Smyth e Verbyla (1999b) e Smyth e Verbyla (1999a) consideraram extensões da estimativa de MVR dos parâmetros da variância para MLGDs. Lee e Nelder (2000) mostraram que a família exponencial dupla, veja Efron (1986), e a QVE geram inferências idênticas.

Lee e Nelder (2000) discutiram o relacionamento entre a família exponencial dupla e a quase-verossimilhança estendida. Eles mostraram que QVE e a família exponencial dupla não-normalizada conduzem a inferências idênticas, pois suas expressões são as mesmas, a menos de uma constante.

Smyth, Huele e Verbyla (2001) compararam a máxima verossimilhança restrita, exata e aproximada, para modelos de regressão heterocedásticos, usando os métodos propostos por Lee e Nelder (1998) e Smyth e Verbyla (1999b). Eles mostraram que, para um modelo linear normal, nenhum dos dois métodos reduz à máxima verossimilhança restrita correta. Eles

também sugeriram um novo método de estimação, o qual retorna a máxima verossimilhança restrita correta e fornece uma aproximação melhorada para o erro padrão.

Cuervo e Gamerman (2001) apresentaram uma versão Bayesiana da MCMD para modelos lineares normais heterocedásticos. Os cálculos Bayesianos foram conduzidos utilizando o algoritmo de Metropolis-Hastings.

Smyth (2002) apresentou um eficiente algoritmo para a máxima verossimilhança restrita em modelos de regressão normais heterocedásticos. O algoritmo se propõe a resolver os possíveis problemas de convergência encontrados na modelagem conjunta da média e dispersão.

Lee e Nelder (2003) fizeram uma explanação sobre planejamentos robustos usando transformação para os dados, e usando modelos lineares generalizados. Eles mostraram que os modelos lineares generalizados fornecem uma extensão para a metodologia de superfície de resposta.

Pinto e Ponce de Leon (2006a) fizeram uma síntese da modelagem conjunta da média e da dispersão, proposta por Nelder e Lee, explicitando, de uma forma concisa, os principais pontos da teoria.

Bonat e Jorgensen (2016) propuseram uma estrutura geral para análise de dados multivariados não normais, denominados modelos lineares generalizados de covariância multivariada (MLGCMs), formulados para lidar com variáveis de resposta multivariadas, juntamente com uma ampla gama de estruturas de correlação temporal e espacial definidas em termos de uma função de ligação de covariância, combinada com uma matriz preditor linear envolvendo matrizes conhecidas.

Nesta tese a principal aplicação da MCMD é em planejamento robusto e o critério de estimação utilizado é a quase-verossimilhança estendida ajustada, proposta por Lee e Nelder (1998).

Em problemas de planejamento robusto, o processo de estimação gangorra da MCMD, no qual o processo de ajuste alterna-se entre os modelos da média e da dispersão (veja Seção 2.3.2), é bastante útil, pois, geralmente, como não há replicações nesses experimentos, o número de observações pode ser bastante reduzido e um processo de estimação considerando ambos os parâmetros da média e da dispersão simultaneamente poderia ser proibitivo, pois poderíamos ter mais parâmetros do que o número de observações. No procedimento de gangorra de Nelder e Lee, os modelos de média e dispersão são conhecidos (dados antes do início do procedimento)

e o que se deseja é obter estimativas de parâmetros para esses modelos fixos. Em planejamento robusto, os modelos da média e da dispersão são escolhidos usando o conhecimento *a priori* sobre os efeitos que podem ser importantes para cada modelo ou através do gráfico normal dos efeitos. Desta forma, o problema que surge é como encontrar os modelos adequados da média e da dispersão, independentemente de suposições subjetivas e com base em algum critério estatístico. A seleção de modelos para a modelagem conjunta da média e dispersão é um assunto ainda pouco estudado. Na literatura existem poucos trabalhos dedicados à seleção de variáveis e este tema ainda é um desafio. A seguir, apresentamos alguns trabalhos que tratam da seleção de variáveis, considerando algum tipo de modelagem conjunta da média e dispersão.

Wang e Zhang (2009) estudaram o problema de seleção de variáveis para modelos conjuntos de média e dispersão e propuseram uma extensão do critério de informação de Akaike baseado na quase-verossimilhança estendida, denominado critério de informação de Akaike estendido (EAIC).

Wu e Li (2012) propuseram um procedimento para selecionar simultaneamente variáveis dos modelos de média e dispersão, considerando como critério de estimação a máxima verossimilhança penalizada e considerando a distribuição Normal Inversa para o modelo da média. Ainda com base no critério de estimação de máxima verossimilhança penalizada, Wu, Zhang e Xu (2012) propuseram um método para selecionar simultaneamente variáveis do modelo da média e da dispersão, considerando a distribuição lognormal para o modelo da média.

Charalambous, Pan e Tranmer (2014) consideraram a seleção das variáveis em modelos conjuntos de média e variância para os modelos multiníveis via verossimilhança penalizada, onde para o modelo da média foi utilizado um modelo linear generalizado misto e para os componentes da variância foi utilizado um modelo linear generalizado. Charalambous, Pan e Tranmer (2015) estenderam a seleção de variáveis na modelagem conjunta da média e variância via verossimilhança penalizada, proposta por Charalambous, Pan e Tranmer (2014), para modelos lineares generalizados hierárquicos.

Antoniadis et al. (2016) consideraram a estimação e seleção de variáveis para modelos conjuntos da média e da dispersão em modelos próprios de dispersão. O procedimento de seleção proposto foi investigado por meio de um estudo de simulação considerando um grande número de covariáveis. Para mais detalhes sobre modelos próprios de dispersão, veja Jorgensen (1987, 1992, 1997).

Bayer e Cribari Neto (2017), consideraram critérios para seleção de modelos na regressão beta com parâmetro de dispersão variável. Eles utilizaram um modelo consistindo de dois submodelos, um para média e outro para a dispersão, ambos com distribuição beta para variável resposta. O critério de estimação utilizado foi o da máxima verossimilhança.

Wu, Li e Tao (2020), consideraram um processo de estimação e seleção de variáveis para mistura de modelos conjuntos da média e da variância. Eles utilizaram a metodologia proposta para analisar dados normais heterocedásticos, provenientes de uma população heterogênea. O problema de seleção de variáveis também foi considerado para o modelo proposto.

Modelagem e planejamento experimental são dois processos complementares dentro de um círculo virtuoso, especialmente em processos industriais. Neste sentido, o processo experimental envolve o planejamento e a implementação do experimento, a construção do modelo de regressão, otimização, análise e interpretação dos resultados.

As técnicas de planejamento de experimento são utilizadas para construção de um processo experimental; ou seja, para definir quais dados, em que quantidade e em que condições devem ser coletados durante um determinado experimento, buscando, basicamente, satisfazer dois grandes objetivos: a maior precisão estatística possível na resposta e o menor custo. É, portanto, uma ferramenta de extrema importância para a indústria, pois seu emprego permite resultados confiáveis economizando tempo e dinheiro, fatores fundamentais em tempos de concorrência acirrada entre as empresas.

A Teoria Clássica de Planejamento de Experimento (veja, por exemplo, Cox e Reid (2000)) originada na primeira metade do século XX, contribuiu para engrandecer o papel da Estatística no contexto da metodologia científica. Na outra metade desse século, graças aos avanços dos métodos matemáticos, surgiu a Teoria de Planejamento Ótimo de Experimentos (veja Subseção 2.3.3), que ganhou destaque, com o rápido desenvolvimento dos recursos computacionais, devido às diversas possibilidades de aplicação em várias áreas da ciência e da indústria. O planejamento (delineamento, desenho) ótimo de experimento é uma poderosa e flexível ferramenta para gerar desenhos experimentais eficientes, possibilitando a diminuição do número de ensaios experimentais e com isso reduzindo custos e diminuindo o tempo de execução do experimento. A obtenção do desenho ótimo está associada ao modelo considerado, ou seja, dado um modelo o desenho ótimo pode ser obtido.

Desta forma, no contexto da modelagem conjunta da média e da dispersão desenhos ótimos se fazem necessários e têm adquirido grande importância nos processos industriais,

especialmente em experimentos envolvendo misturas. A seguir apresentamos alguns trabalhos relacionados a obtenção de desenhos ótimos em modelos envolvendo variabilidade.

Dasgupta, Mukhopadhyay e Studden (1992) consideraram o problema de encontrar desenhos ótimos, com uma função de eficiência proporcional ao inverso da variância da resposta, em modelos lineares Normais heterocedásticos, nos quais a variância da resposta é uma função exponencial ou potência de sua média.

Burridge e Sebastiani (1994) obtiveram desenhos D -ótimos locais para modelos lineares generalizados com variância proporcional ao quadrado da média. Eles consideraram funções de ligação, como sendo uma função potência da média.

Atkinson e Cook (1995), assumindo uma resposta Normal, mostraram como obter experimentos D -ótimos em modelos de regressão com erros heterocedásticos. Eles consideraram uma classe de modelos na qual a variância do erro e a média da resposta são funções paramétricas de, possivelmente, diferentes conjuntos de variáveis explicativas.

Wong (1995) encontrou condições sob as quais, para modelos heterocedásticos, os desenhos D e G -ótimos permanecem equivalentes. Wong concluiu que os desenhos D e G -ótimos raramente são equivalentes para modelos heterocedásticos, contudo eles vêm a ser, aproximadamente equivalentes, quando a função eficiência é limitada por um valor suficientemente distante de zero.

Vining e Schaub (1996) propuseram uma abordagem semi-Bayesiana para modelos Normais heterocedásticos, na qual o critério de otimalidade é maximizar o determinante da matriz de informação esperada.

Atkinson e Cook (1997) obtiveram desenhos D -ótimos Bayesianos para os casos nos quais, após uma transformação na resposta, um modelo linear é obtido. A transformação de Box e Cox permitiu encontrar a condição de linearidade; e usando a suposição de normalidade, as técnicas obtidas em Atkinson e Cook (1995) puderam ser aplicadas para encontrar desenhos ótimos. Uma excelente revisão sobre desenhos ótimos Bayesianos é dada por Chaloner e Verdinelli (1995).

Dette e Wong (1998) consideraram o problema de encontrar desenhos ótimos no contexto de modelos de regressão polinomial, onde a variância da resposta depende exponencialmente da variável independente, sendo que esta dependência não é exatamente conhecida. O critério de otimalidade utilizado foi o critério D -ótimo Bayesiano, sobre um intervalo compacto, com

o número de pontos suporte restrito a ser o grau do polinômio mais um. Desenhos ótimos Bayesianos foram encontrados analiticamente para uma ampla classe de *prioris*.

Brown e Wong (2000) consideraram o problema de encontrar um desenho ótimo para estimar a superfície de resposta de um experimento sobre uma região compacta arbitrária, quando existe heterocedasticidade no modelo. O critério de otimalidade foi o de minimizar a máxima variância da superfície de resposta, estimada sobre a região compacta. Eles propuseram um algoritmo minmax para encontrar desenhos ótimos para modelos de regressão linear simples com erros heterocedásticos.

Tack, Goss e Wandebroek (2002), considerando um modelo heterocedástico Normal, verificaram o impacto do conhecimento da *priori* na obtenção de desenhos ótimos Bayesianos em comparação com os desenhos clássicos, baseados na suposição de erros homocedásticos.

Pinto e Ponce de Leon (2004) estenderam os resultados de Atkinson e Cook (1995) para a classe mais geral de modelos lineares generalizados, com a utilização da MCMD, onde a suposição do conhecimento da verossimilhança não é mais necessária, bastando somente o conhecimento da relação existente entre a média e a variância. O conceito de quase-verossimilhança estendida foi usado para obtenção da matriz de informação de Fisher. Desenhos ótimos locais e Bayesianos foram considerados. Pinto e Ponce de Leon (2007) consideraram a obtenção de desenhos D_S -ótimos para a MCMD, onde o interesse está em apenas um conjunto de parâmetros, tanto no modelo da média, quanto no modelo da dispersão. Pinto e Ponce de Leon (2014) provaram a validade do Teorema Geral da Equivalência para modelos conjuntos da média e dispersão.

A modelagem da variabilidade em experimentos envolvendo misturas é considerada somente em situações onde existem variáveis de ruído (veja Seção 3.1). Em todas as abordagens consideradas para a modelagem da variância, foram utilizadas distribuição Normal para a variável resposta e parâmetro de dispersão constante, ou seja, modelo linear normal homocedástico para a variável resposta; assim, a variância não constante se deveu apenas à inclusão de variáveis de ruído.

A modelagem estatística da variância em experimentos de mistura com variáveis de ruído foi considerada em Steiner e Hamada (1997), que propuseram um modelo combinado de mistura e variáveis de ruído. Eles consideraram um problema de otimização para minimizar uma função de perda quadrática envolvendo a média e variância da resposta.

Goldfarb, Borrer e Montgomery (2003) propuseram uma forma de modelar a variância em experimentos de mistura com variáveis de ruído com a utilização do método delta (CASELLA; BERGER, 2002, p. 240), o qual emprega uma aproximação de série de Taylor de primeira ordem do modelo de regressão em um vetor de variáveis de ruído.

Granato, Calado e Pinto (2020) apresentaram uma aplicação didática do método delta para modelar a variância em experimentos de mistura com variáveis de ruído, na engenharia de alimentos.

2.3 Conceitos básicos

Nesta seção daremos um breve resumo sobre os principais assuntos tratados nesta tese. Na Subseção 2.3.1, discutimos os principais conceitos sobre experimentos com mistura, incluindo modelagem, desenhos de Scheffé e experimentos de mistura com variáveis de processo; na Subseção 2.3.2, a teoria dos modelos lineares generalizados é abordada, especialmente a modelagem conjunta da média e da dispersão, que é um conceito fundamental nesta tese. Finalmente, na Subseção 2.3.3 apresentamos uma visão geral da teoria de planejamento ótimo de experimentos, sob o enfoque da teoria aproximada de Kiefer (KIEFER, 1974), e sua relação com experimentos de misturas.

2.3.1 Experimentos com misturas

Um experimento de mistura envolve proporções de mistura de dois ou mais componentes para fazer diferentes composições de um produto final. Em experimentos envolvendo misturas, o objetivo principal é determinar as proporções dos componentes da mistura que levam a resultados desejáveis com relação a alguma propriedade de interesse. As proporções dos componentes de mistura x_i estão sujeitas às seguintes restrições

$$0 \leq x_i \leq 1 \quad i = 1, 2, \dots, a \quad \text{e} \quad \sum_{i=1}^a x_i = 1, \quad (2.3.1)$$

onde a é o número de componentes envolvidos no experimento de mistura. Conseqüentemente, o espaço de design ou região experimental é um simplex $(a - 1)$ -dimensional. Uma região experimental simplex n -dimensional é um polítopo n -dimensional determinado pelas restrições de mistura. Por exemplo, um simplex 1-dimensional é um segmento de linha; um simplex 2-dimensional é um triângulo (incluindo

seu interior); um simplex 3-dimensional é um tetraedro (veja Boyd e Vandenberghe (2004) para um tratamento mais abrangente). A Figura 2.1 mostra a região experimental simplex 2-dimensional para uma mistura com três componentes.

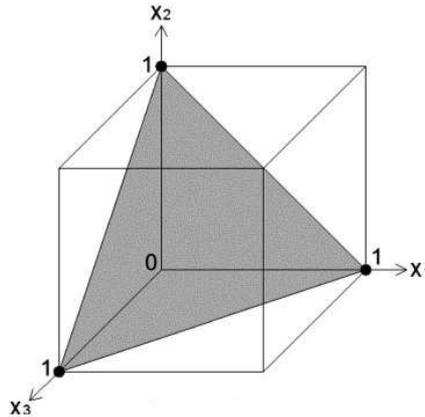


Figura 2.1: Região simplex 2-dimensional para uma mistura com três componentes.

A região experimental é parte de um simplex n -dimensional se houver outras condições nas proporções de mistura, como $L_i \leq x_i \leq U_i$ para $i = 1, 2, \dots, a - 1$, com a proporção x_a tomando o valor que compõem a mistura. L_i e U_i são, respectivamente, os limites inferior e superior da mistura.

Quando a existência de limites inferior e/ou superior não afeta o formato da região de mistura, ou seja, ela ainda é uma região simplex, é possível definir um novo conjunto de componentes que assumem os valores de 0 a 1 sobre a região viável. Esses novos componentes são chamados de pseudocomponentes (veja Cornell (2002) para mais detalhes). Myers e Montgomery (2002) recomendam o uso de pseudocomponentes para reduzir a multicolinearidade entre as covariáveis.

Modelos e desenhos padrão

De acordo com Khuri (2005), considere o modelo linear

$$y = X\beta + \epsilon, \quad (2.3.2)$$

onde $y^t = (y_1, \dots, y_n)$ é um vetor resposta, X é uma matriz de planejamento de ordem $n \times p$ e posto (*rank*) $p \leq n$, β é um vetor de parâmetros desconhecidos, e $\epsilon^t = (\epsilon_1, \dots, \epsilon_n)$ é um vetor

de erros aleatórios com média zero e matriz de variância-covariância $\sigma^2 \mathbf{I}$, onde $\sigma^2 = \text{Var}(\epsilon_i)$ para $i = 1, \dots, n$ e \mathbf{I} é a matriz identidade. O valor da resposta em algum ponto $\mathbf{x}^t = (x_1, \dots, x_a)$ na região de interesse \mathcal{X} é dado por $y(\mathbf{x}) = \zeta(\mathbf{x}, \boldsymbol{\beta}) + \epsilon$, onde $\zeta(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{f}^t(\mathbf{x})\boldsymbol{\beta}$ é um modelo polinomial de determinado grau nas variáveis de controle x_1, \dots, x_a . O vetor $\mathbf{f}(\mathbf{x})$ é avaliado no ponto \mathbf{x} e representa a linha da matriz \mathbf{X} .

Os modelos comumente usados em experimentos com mistura são os chamados modelos de Scheffé. Scheffé (1958) propôs modelos polinomiais canônicos para experimentos de mistura, considerando uma reparametrização do modelo polinomial, baseada na restrição de mistura. O modelo polinomial de primeira ordem de Scheffé é dado por

$$\zeta(\mathbf{x}, \boldsymbol{\beta}) = \sum_{i=1}^a \beta_i x_i, \quad (2.3.3)$$

que não contém explicitamente o termo constante. O modelo de segunda ordem (modelo quadrático) de Scheffé é dado por

$$\zeta(\mathbf{x}, \boldsymbol{\beta}) = Q(\mathbf{x}, \boldsymbol{\beta}) = \sum_{i=1}^a \beta_i x_i + \sum_{i=1}^{a-1} \sum_{j=i+1}^a \beta_{ij} x_i x_j. \quad (2.3.4)$$

Para modelos de ordem superior, a reparametrização não conduz a modelos simples; por exemplo, para o modelo de terceira ordem (modelo cúbico), o modelo polinomial canônico de Scheffé é dado por

$$\zeta(\mathbf{x}, \boldsymbol{\beta}) = Q(\mathbf{x}, \boldsymbol{\beta}) + \sum_{i=1}^{a-1} \sum_{j=i+1}^a \beta_{i-j} x_i x_j (x_i - x_j) + \sum_{i=1}^{a-2} \sum_{j=i+1}^{a-1} \sum_{k=j+1}^a \beta_{ijk} x_i x_j x_k. \quad (2.3.5)$$

Para estimação dos parâmetros do modelo, Scheffé propôs os planejamentos em rede simplex (a, m) , que são caracterizados por arranjos simétricos de pontos dentro da região experimental. O número de parâmetros no modelo polinomial canônico é exatamente igual ao número de pontos escolhidos dentro da região experimental. O planejamento em rede simplex (a, m) consiste em $\frac{(a+m-1)!}{m!(a-1)!}$ pontos experimentais. Cada componente da mistura leva $m + 1$ valores igualmente espaçados entre 0 e 1, ou seja, $x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, \frac{m}{m} = 1$ para $i = 1, \dots, a$. Todas as misturas envolvendo essas proporções são usadas no planejamento experimental. Por exemplo, um planejamento em rede simplex $(3, 2)$ consiste em $\frac{(3+2-1)!}{2!(3-1)!} = 6$ pontos experimentais. Cada componente da mistura x_i pode assumir $2 + 1 = 3$ valores possíveis, ou seja, $x_i = 0, \frac{1}{2}, 1$; que resulta nos seguintes pontos experimentais $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$,

$(1/2, 1/2, 0)$, $(0, 1/2, 1/2)$, $(1/2, 0, 1/2)$. O planejamento lattice-simplex $(3, 2)$ é mostrado na Figura 2.2.

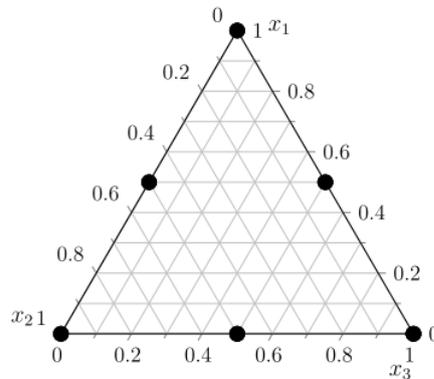


Figura 2.2: Planejamento lattice-simplex $(3, 2)$.

Scheffé (1963) definiu um caso especial de um modelo polinomial canônico dado por

$$\zeta(\mathbf{x}, \boldsymbol{\beta}) = \sum_{i=1}^a \beta_i x_i + \sum_{i=1}^{a-1} \sum_{j=i+1}^a \beta_{ij} x_i x_j + \dots + \beta_{1, \dots, a} x_1 \dots x_a. \quad (2.3.6)$$

Novamente, para a estimativa dos parâmetros do modelo polinomial canônico Scheffé propôs um planejamento experimental denominado centróide-simplex, que consiste em $2^a - 1$ pontos experimentais. Os pontos do desenho experimental centróide-simplex são formados por a permutações de misturas puras $(1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$, $\frac{q!}{2!(q-2)!}$ permutações de misturas binárias $(1/2, 1/2, 0, \dots, 0), \dots, (0, \dots, 0, 1/2, 1/2)$, e o centróide global, a mistura dos a componentes $(1/a, 1/a, \dots, 1/a)$. Por exemplo, para $a = 3$, o centróide-simplex consiste em sete pontos, dados por: $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, $(1/2, 1/2, 0)$, $(1/2, 0, 1/2)$, $(0, 1/2, 1/2)$ e $(1/3, 1/3, 1/3)$. O centróide-simplex para $a = 3$ é mostrado na Figura 2.3.

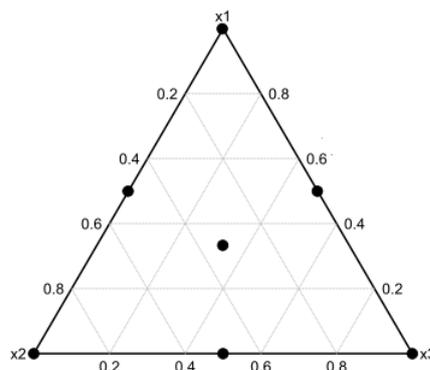


Figura 2.3: Desenho centróide-simplex para $a = 3$.

Alternativamente aos modelos canônicos de Scheffé, também podemos usar modelos de mistura com variável de folga. Devido às restrições de mistura (2.3.1), as proporções dos componentes não são independentes; assim, conhecendo as proporções dos primeiros $a - 1$ componentes, podemos determinar a proporção do componente restante. Desta forma, os modelos de mistura com a variável de folga podem ser obtidos a partir dos modelos canônicos de Scheffé, substituindo a variável de folga escolhida, digamos x_l , por $1 - \sum_{i \neq l}^a x_i$. O objetivo deste processo é produzir modelos de mistura que dependem de $a - 1$ variáveis independentes. Por exemplo, ao definir x_a como a variável de folga, o modelo de mistura com a variável de folga, obtido do modelo quadrático de Scheffé (2.3.4), é dado por

$$\psi(\mathbf{x}, \boldsymbol{\alpha}) = \alpha_0 + \sum_{i=1}^{a-1} \alpha_i x_i + \sum_{i=1}^{a-1} \alpha_{ii} x_i^2 + \sum_{i=1}^{a-2} \sum_{j=i+1}^{a-1} \alpha_{ij} x_i x_j. \quad (2.3.7)$$

Note que, comparando os coeficientes dos modelos (2.3.4) e (2.3.7), temos as seguintes equivalências $\alpha_0 = \beta_a$, $\alpha_i = \beta_i - \beta_a + \beta_{ia}$, $\alpha_{ii} = -\beta_{ia}$ e $\alpha_{ij} = \beta_{ij} - (\beta_{ia} + \beta_{ja})$.

A utilização de modelos com variável de folga, bem como a escolha de qual componente deve ser designado como variável de folga, não está claramente apresentada na literatura. Uma discussão sobre os prós e contras do uso da variável de folga é dada por Cornell (2000); veja também Smith (2005). Estudos comparativos entre modelos de misturas com variável de folga e modelos de Scheffé são dados em Cornell (2002), Cornell e Gorman (2003) e Khuri (2005).

Experimentos de mistura com variáveis de processo

Problemas de experimentos de mistura com variáveis de processo surgem quando no experimento de mistura a característica de interesse é função das proporções dos ingredientes e de outros fatores que não constituem parte alguma da mistura, como calor ou tempo, por exemplo. Esses fatores, que dependem das condições do processo, são chamados de variáveis de processo. Desta forma, a variável de resposta depende não apenas da proporção dos componentes da mistura, mas também das variáveis de processo.

Usualmente, em problemas de experimentos com mistura incluindo variáveis de processo, o objetivo é determinar as proporções dos componentes da mistura juntamente com as situações de condições de processo, a fim de satisfazer alguma propriedade ou característica de interesse envolvendo a variável resposta.

Assim, se em adição aos a componentes de mistura $\mathbf{x}^t = (x_1, \dots, x_a)$ existirem r variáveis de processo $\mathbf{z}^t = (z_1, \dots, z_r)$, podemos considerar modelos aditivos como $\eta(\mathbf{x}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \zeta(\mathbf{x}, \boldsymbol{\beta}) +$

$\vartheta(z, \boldsymbol{\gamma})$ ou modelos completos de produtos cruzados do tipo $\eta(\boldsymbol{x}, z, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \zeta(\boldsymbol{x}, \boldsymbol{\beta})\vartheta(z, \boldsymbol{\gamma})$ ou combinações desses modelos como $\eta(\boldsymbol{x}, z, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \zeta(\boldsymbol{x}, \boldsymbol{\beta}) + \nu(\boldsymbol{x}, z, \boldsymbol{\beta}, \boldsymbol{\gamma})$, onde $\vartheta(z, \boldsymbol{\gamma})$ representa o modelo para as variáveis de processo e $\nu(\boldsymbol{x}, z, \boldsymbol{\beta}, \boldsymbol{\gamma})$ compreende produto de termos em $\zeta(\boldsymbol{x}, \boldsymbol{\beta})$ e $\vartheta(z, \boldsymbol{\gamma})$.

Em geral, a metodologia usada para construir experimentos de misturas envolvendo variáveis de processo é uma combinação de dois planejamentos, sendo um para os componentes da mistura e outro sendo fatorial ou fatorial fracionário para as variáveis de processo. Por exemplo, a Figura 3.1 no problema de fabricação de pão, discutido na Seção 3.3, mostra um projeto fatorial completo 3^2 em duas variáveis de processo, z_1 e z_2 , cruzadas com um desenho simplex-centroide de terceira ordem em três componentes da mistura, x_1 , x_2 e x_3 .

Uma apresentação abrangente de vários aspectos da modelagem estatística e planejamento experimental envolvendo experimentos de mistura com variáveis de processo é fornecida por Cornell (2002). O livro de Smith (2005) também é uma excelente referência.

Piepel (2004) faz um levantamento bibliográfico sobre experimentos de mistura ao longo de um período de 50 anos, desde 1955 a 2004. Uma extensa bibliografia sobre artigos envolvendo experimentos de mistura com variáveis de processo, bem como vários outros tipos de experimentos de mistura é fornecida por Piepel e Cornell (2001).

2.3.2 Modelos Lineares Generalizados

A classe dos modelos lineares generalizados (MLGs) foi desenvolvida por Nelder e Wedderburn (1972) e se baseia na família exponencial uniparamétrica, que possui propriedades interessantes para estimação, testes de hipóteses e outros problemas de inferência. Alguns dos mais importantes modelos de probabilidade pertencem a esta classe, como Normal, Poisson, Binomial, Multinomial, Gama, Binomial Negativa e Inversa Gaussiana. Antes do advento dos MLGs, esses modelos e as técnicas específicas para ajustá-los eram tratados separadamente como se fossem métodos estatísticos distintos. A teoria de MLG unificou tais modelos em uma mesma estrutura. Esta unificação é baseada nas propriedades da família exponencial de distribuições de probabilidade e tem como principal resultado o fato de que todo MLG pode ser ajustado da mesma forma, através do método iterativo de mínimos quadrados ponderados.

O modelo linear generalizado é definido por uma distribuição de probabilidade, membro da família exponencial de distribuições para a variável resposta; um conjunto de covariáveis

independentes, descrevendo a estrutura linear do modelo e uma função de ligação entre a média da variável resposta e a estrutura linear.

Os modelos lineares generalizados estendem a classe dos modelos lineares. Eles permitem que aleatoriedade venha de uma classe de distribuições ao invés de somente a distribuição Normal. Um MLG gera uma escala aditiva para as variáveis explicativas proveniente da estrutura do componente aleatório. A escala na qual os efeitos das covariáveis são assumidos aditivos é relacionada à média da distribuição do componente aleatório por uma função de ligação. Usando MLGs não há necessidade de transformar os dados para obter aditividade. Muitas características dos modelos lineares clássicos são imediatamente estendidas por MLGs. Essas características incluem a estrutura do preditor linear, a tabela ANOVA e as ideias de diagnósticos de modelos por meio de análise de resíduos, efeito de alavanca, influência, e assim por diante. Além disso, um algoritmo simples, uma versão do Algoritmo Iterativo de Mínimos Quadrados Ponderados, ajusta todos os MLGs.

Sejam y_1, \dots, y_n , n observações independentes da variável resposta Y , e sejam x_1, \dots, x_k , k covariáveis, cujos efeitos combinam aditivamente para produzir um preditor linear $\eta = \mathbf{f}^t(\mathbf{x})\boldsymbol{\beta} = \sum_{j=1}^p f_j(\mathbf{x})\beta_j$, onde $\boldsymbol{\beta}^t = (\beta_1, \dots, \beta_p)$ representa o vetor de parâmetros desconhecidos do modelo e $\mathbf{f}^t(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_p(\mathbf{x}))$ denota o vetor de funções das covariáveis.

De acordo com McCullagh e Nelder (1989), um MLG é caracterizado por três componentes:

1. Componente aleatório, cuja esperança é denotada por $E(Y) = \mu$;
2. Componente sistemático $\eta = \mathbf{f}^t(\mathbf{x})\boldsymbol{\beta}$
3. A ligação entre os componentes aleatório e sistemático é dada por $\eta = \varphi(\mu)$, onde $\varphi(\cdot)$ é qualquer função monótona diferenciável.

A função $\varphi(\cdot)$ é conhecida como função de ligação do MLG e define a escala na qual os efeitos das covariáveis são assumidos serem aditivos. A grande vantagem proveniente da teoria de MLG é que ela permite a representação de qualquer membro da família exponencial de distribuições de probabilidade, contínua ou discreta, na estrutura acima.

Seja a função de distribuição na família exponencial, para uma observação y_i , dada por:

$$p_{Y_i}(y_i, \boldsymbol{\theta}, \phi) = \exp \left\{ \frac{y_i \boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\phi)} + c(y_i, \phi) \right\}, \quad (2.3.8)$$

onde θ é o parâmetro canônico, $a(\phi)$ tem a forma de $\frac{\phi}{m}$, ϕ é o parâmetro de dispersão e m , o peso a priori. O logaritmo da verossimilhança para um conjunto de y_1, \dots, y_n observações independentes é

$$l(\theta, \phi; \mathbf{y}) = \sum_{i=1}^n \ln p_{Y_i}(y_i; \theta, \phi) = \sum_{i=1}^n \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi). \quad (2.3.9)$$

Para os MLGs $E(Y) = \mu = \frac{\partial b(\theta)}{\partial \theta} = b'(\theta)$ e que $V(\mu) = \frac{\partial \mu}{\partial \theta} = b''(\theta)$, onde $V(\mu)$ é a função de variância. Uma propriedade importante de todos os MLGs é a forma da variância, isto é, $Var(Y) = \phi V(\mu)$, ou seja, a variância se divide em duas partes: ϕ , independente da média e $V(\mu)$, que descreve como a variância varia com a média.

Para uma observação y_i e para ϕ fixado, as equações de máxima verossimilhança para β_j , com $j = 1, \dots, p$ são independentes de ϕ e dadas por:

$$\sum_{j=1}^p w_i (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i} f_j(\mathbf{x}_i) = 0, \quad (2.3.10)$$

onde $w_i = \left[\left(\frac{\partial \eta_i}{\partial \mu_i} \right)^2 V(\mu_i) \right]^{-1}$ é o peso para a i -ésima observação. Essas equações podem ser ajustadas pelo método iterativo de mínimos quadrados ponderados, usando uma variável dependente R , com valores $r_i = \eta_i + (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i}$ com peso w_i .

Uma medida de discrepância para os MLGs, generalizando a soma de quadrados dos resíduos para o modelo normal, é o desvio (*deviance*), D , dado por:

$$D = -2\phi[l(\hat{\boldsymbol{\mu}}, \phi; \mathbf{y}) - l(\mathbf{y}, \phi; \mathbf{y})], \quad (2.3.11)$$

onde $l(\hat{\boldsymbol{\mu}}, \phi; \mathbf{y})$, expresso em função de $\hat{\boldsymbol{\mu}}$, é o máximo da função logarítmica da verossimilhança, com média $\boldsymbol{\mu}$ e parâmetro de dispersão ϕ fixo; e $l(\mathbf{y}, \phi; \mathbf{y})$ é o máximo do logaritmo da função de verossimilhança para o modelo completo, isto é, com n parâmetros, onde n é o número de observações.

Para uma visão mais abrangente sobre a teoria dos modelos lineares generalizados veja McCullagh e Nelder (1989).

Modelagem conjunta da média e dispersão

De acordo com Nelder e Lee (1991), a modelagem conjunta da média e dispersão consiste em encontrar modelos conjuntos para a média e para a dispersão em um MLG. Em sua abordagem,

usando a quase-verossimilhança estendida, dois modelos lineares generalizados interligados são necessários, um para a média e outro para a dispersão.

Sejam Y_1, \dots, Y_n n variáveis aleatórias independentes com a mesma distribuição de probabilidade, representando as variáveis respostas dependentes, cujos valores observados são dados por y_1, \dots, y_n . Para a i -ésima resposta Y_i é assumido somente serem conhecidos $E(Y_i) = \mu_i$ e $Var(Y_i) = \phi_i V(\mu_i)$, onde ϕ_i é o parâmetro de dispersão e $V(\cdot)$ é a função de variância nos MLGs. Uma vez que a modelagem conjunta de média e dispersão é baseada na quase-verossimilhança estendida, o conhecimento completo da distribuição de probabilidade da variável de resposta não é necessário. Os modelos da média e da dispersão são construídos da seguinte maneira.

Suponha que $\mathbf{t}^t = (t_1, \dots, t_s)$ e $\mathbf{u}^t = (u_1, \dots, u_r)$ sejam os vetores das variáveis independentes que afetam os modelos da média e da dispersão, respectivamente. As variáveis independentes para o modelo de dispersão são comumente, mas não necessariamente, um subconjunto das variáveis independentes para o modelo média.

Seja φ a função de ligação para o modelo da média, isto é, para a i -ésima resposta, $\eta_i = \varphi(\mu_i) = \mathbf{f}^t(\mathbf{t}_i)\boldsymbol{\beta}$ com $\mathbf{f}^t(\mathbf{t}_i) = (f_1(\mathbf{t}_i), \dots, f_p(\mathbf{t}_i))$ onde $f_j(\mathbf{t}_i)$, for $j = 1, \dots, p$, é uma função conhecida de \mathbf{t}_i e $\boldsymbol{\beta}$ é um vetor $p \times 1$ de parâmetros desconhecidos.

Para o modelo de dispersão é utilizado como variável de resposta o componente de desvio, que, para cada observação y_i , é dado por

$$d_i = 2 \int_{\mu_i}^{y_i} \frac{y_i - l}{V(l)} dl, \quad (2.3.12)$$

veja McCullagh e Nelder (1989), p. 360.

Seguindo Lee e Nelder (1998), para o modelo de dispersão, estamos assumindo um modelo Gama com uma função de ligação logarítmica, ou seja, para a resposta i -ésima resposta, $\zeta_i = \log(\phi_i) = \mathbf{g}^t(\mathbf{u}_i)\boldsymbol{\gamma}$, com $\mathbf{g}^t(\mathbf{u}_i) = (g_1(\mathbf{u}_i), \dots, g_q(\mathbf{u}_i))$, onde $g_j(\mathbf{u}_i)$, para $j = 1, \dots, q$, é uma função conhecida de \mathbf{u}_i e $\boldsymbol{\gamma}$ é um vetor $q \times 1$ de parâmetros desconhecidos. Também definimos $\mathbf{T} = [\mathbf{f}(\mathbf{t}_1), \dots, \mathbf{f}(\mathbf{t}_n)]^t$ a matriz experimental $n \times p$ para o modelo da média e $\mathbf{U} = [\mathbf{g}(\mathbf{u}_1), \dots, \mathbf{g}(\mathbf{u}_n)]^t$ a matriz experimental $n \times q$ para o modelo da dispersão.

O processo de ajuste para a MCMD usa como critério de otimização a quase-verossimilhança estendida, introduzida por Nelder e Pregibon (1987), ver McCullagh e Nelder (1989), p. 349. Nesta tese usamos a quase-verossimilhança estendida ajustada, introduzida por Lee e Nelder (1998), dada por

$$Q^+(\boldsymbol{\mu}, \boldsymbol{\phi}; \mathbf{y}) = \sum_{i=1}^n -\frac{1}{2} \left(\frac{d_i^*}{\phi_i} + \log\{2\pi\phi_i V(y_i)\} \right) \quad (2.3.13)$$

onde $d_i^* = \frac{d_i}{1-h_i}$ é o componente do desvio padronizado e h_i é o i -ésimo elemento da diagonal da matriz $\mathbf{H} = \mathbf{W}^{\frac{1}{2}} \mathbf{T} (\mathbf{T}^t \mathbf{W} \mathbf{T})^{-1} \mathbf{T} \mathbf{W}^{\frac{1}{2}}$, sendo \mathbf{W} , a matriz $n \times n$ dos pesos para os MLGs, uma matriz diagonal com elementos dados por $w_i = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{\phi_i V(\mu_i)}$ com $i = 1, \dots, n$.

A Tabela 2.1 dá um resumo da modelagem conjunta de média e dispersão. Da Tabela 2.1, podemos observar que o componente de desvio padronizado do modelo para a média torna-se a resposta para o modelo de dispersão, e o inverso dos valores ajustados para o modelo da dispersão fornece os pesos *a priori* para o modelo da média.

Tabela 2.1: Resumo da MCMD para a i -ésima resposta

Componente	Modelo da média	Modelo da dispersão [†]
Variável resposta	y_i	d_i^*
Média	μ_i	ϕ_i
Variância	$\phi_i V(\mu_i)$	$2\phi_i^2$
Função de ligação	$\eta_i = \varphi(\mu_i)$	$\zeta_i = \log(\phi_i)$
Preditor linear	$\eta_i = \mathbf{f}^t(\mathbf{t}_i) \boldsymbol{\beta}$	$\zeta_i = \mathbf{g}^t(\mathbf{u}_i) \boldsymbol{\gamma}$
Componente do desvio	$d_i = 2 \int_{\mu_i}^{y_i} \frac{y_i - l}{V(l)} dl$	$2 \left\{ -\log \left(\frac{d_i^*}{\phi_i} \right) + \frac{(d_i^* - \phi_i)}{\phi_i} \right\}$
peso <i>a priori</i>	$\frac{1}{\phi_i}$	$(1 - h_i)/2$

[†] Para o modelo da dispersão nós estamos assumindo um modelo Gama com função de ligação logarítmica

O algoritmo para estimação de parâmetros é uma extensão do algoritmo padrão dos MLGs, no qual o modelo para a média é ajustado assumindo que os valores ajustados para a dispersão são conhecidos e que o modelo para dispersão é ajustado usando os valores ajustados para o modelo da média. O processo de ajuste alterna-se entre os modelos de média e da dispersão até que uma convergência seja alcançada. o Algoritmo 2.1 explica o processo de estimação de parâmetros para a modelagem conjunta de média e dispersão.

Algoritmo 2.1: Pseudo algoritmo para a modelagem conjunta da média e da dispersão

1. Modelo da média - Considerando uma distribuição de probabilidade para a variável de resposta, um preditor linear e uma função de ligação, calcule $\boldsymbol{\beta}^t = (\beta_1, \dots, \beta_p)$ usando o algoritmo para ajustar modelos lineares generalizados (ver McCullagh e Nelder (1989), p. 40).

2. Com o valor encontrado para β no passo 1, calcule $\mathbf{d}^* = (d_1^*, \dots, d_n^*)^t$.
3. Modelo da dispersão - Considerando \mathbf{d}^* como resposta para o modelo da dispersão e assumindo um MLG Gama com função de ligação logarítmica, calcule $\boldsymbol{\gamma}^t = (\gamma_1, \dots, \gamma_q)$ usando o algoritmo para ajustar modelos lineares generalizados.
4. Com o valor encontrado para $\boldsymbol{\gamma}$ no passo 3, calcule $\boldsymbol{\phi}^t = (\phi_1, \dots, \phi_n)$.
5. Com o valor encontrado para \mathbf{d}^* no passo 2 e o valor encontrado para $\boldsymbol{\phi}$ no passo 4, calcule $Q^+(\boldsymbol{\mu}, \boldsymbol{\phi}; \boldsymbol{\gamma})$ e use esta medida para verificar a convergência do modelo conjunto na próxima iteração, ou seja, comparando seu valor com o valor obtido na iteração seguinte. Se a convergência for alcançada, pare. Os atuais β e $\boldsymbol{\gamma}$ são os parâmetros finais para o modelo conjunto. Caso contrário, com o valor de $\boldsymbol{\phi}$, atualizado no passo 4, calcule os novos pesos w_1, \dots, w_n para o modelo da média e vá para o passo 1.

No passo 5 do Algoritmo 2.1, os novos pesos para o modelo da média na próxima iteração do Algoritmo 2.1 são $w_i = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{1}{\phi_i V(\mu_i)}$ para $i = 1, \dots, n$. Observe que o parâmetro de dispersão influencia os pesos para o modelo da média apenas a partir da segunda iteração do algoritmo. Para maiores detalhes sobre o algoritmo para a modelagem conjunta da média e dispersão, veja Pinto e Ponce de Leon (2006a).

Existem pacotes estatísticos que realizam a modelagem conjunta da média e da dispersão. No entanto, o Algoritmo 2.1 pode ser facilmente implementado no software R (R CORE TEAM, 2019) usando a função de estimação de parâmetros para os modelos lineares generalizados.

Modelos da média e da variância

Embora a MCMD possa ser aplicada em diversas áreas, uma de suas principais aplicações é em estatística industrial, especialmente em planejamento robusto, uma ferramenta muito útil para melhorar a qualidade em experimentos industriais, onde os dados apresentam grande variabilidade e os experimentos são, em geral, experimentos fatoriais fracionários não replicados com poucos ensaios experimentais. Geralmente, em problemas de planejamento robusto, existem alguns fatores que não podem ser controlados, tais fatores são considerados como fatores de ruído.

De acordo com Engel e Huele (1996), ao contrário do planejamento experimental clássico, que trata de fatores que são fixados tanto durante o experimento quanto sob as condições

de processamento, os chamados fatores de planejamento, no planejamento robusto existem também fatores de ruído, introduzidos por Taguchi (1986). Fatores de ruído são fixados durante o experimento, mas variam aleatoriamente fora do contexto experimental.

O objetivo nos planejamentos robustos é reduzir a variação nos produtos/processos, mantendo a média em um determinado valor alvo. Em planejamento robusto a identificação dos fatores que afetam a média e a variância das características de qualidade é essencial para a otimização de produtos e processos de produção. Para mais detalhes sobre planejamento robusto, veja Phadke (1989). Um painel de discussão que fornece um excelente resumo sobre planejamento robusto com várias ideias e comentários úteis sobre sua aplicação é fornecido por Nair (1992).

Desta forma, em planejamento robusto modelos que permitem modelar a variabilidade são de fundamental importância. A MCMD é um mecanismo muito útil para modelagem da variabilidade, pois ela permite que a variância seja modelada de forma muito eficaz, mesmo sem replicação no experimento. O modelo de variância pode ser obtido a partir do modelo de dispersão de uma forma muito simples.

Se no experimento não existirem fatores de ruído, após obter o modelo conjunto final para média e para a dispersão, podemos obter os modelos para $E(Y)$ e $Var(Y)$ da seguinte forma. Da Tabela 2.1 podemos ver que o modelo da média é obtido diretamente por $E(Y) = \mu = \varphi^{-1}(\eta)$, enquanto o modelo para variância é dado por $Var(Y) = \exp(\zeta)V(\mu)$.

Para o caso em que o experimento envolve fatores de ruído, considerados variáveis aleatórias, os resultados do experimento são obtidos condicionados às variáveis de ruído e os modelos da média e da variância, anteriormente obtidos, passam a ser condicionados às variáveis de ruído e dados por $E(Y|\mathbf{N})$ e $Var(Y|\mathbf{N})$, onde \mathbf{N} é um vetor de variáveis aleatórias de ruído.

Desta forma, podemos obter os modelos incondicionais às variáveis de ruído para $E(Y)$ e $Var(Y)$ usando as expressões $E(Y) = E(E(Y|\mathbf{N}))$ e $Var(Y) = Var(E(Y|\mathbf{N})) + E(Var(Y|\mathbf{N}))$. Neste caso, assume-se que a distribuição de probabilidade ou, pelo menos, a média e a variância de \mathbf{N} são conhecidas ou podem ser estimadas.

Vale ressaltar que, em um problema de planejamento robusto, conhecendo os modelos de média e da variância, podemos minimizar a variabilidade encontrando as configurações ótimas para as variáveis explicativas que afetam o modelo de variância e então ajustar a média para seu valor alvo, encontrando configurações apropriadas para as variáveis explicativas que afetam o

modelo da média. Para o caso de experimentos de mistura, esse processo leva a um problema de otimização, no qual a restrição de mistura deve ser considerada, além de outras restrições nas variáveis de mistura, se houver.

2.3.3 Planejamento ótimo de experimentos

A teoria de planejamento ótimo de experimentos é uma poderosa ferramenta para obtenção de planejamentos experimentais eficientes e constitui-se em uma rara mistura entre as áreas de álgebra linear, análise convexa e estatística. Seja um conjunto de m covariáveis, denominadas variáveis experimentais, reunidas em um vetor $\mathbf{x}^t = (x_1, x_2, \dots, x_m)$, onde $x_j \in \mathbb{R}$; $j = 1, \dots, m$. Suponha que os níveis das variáveis experimentais, para os quais as respostas serão observadas, possam ser escolhidos pelo investigador, com a restrição de que as escolhas sejam realizadas no interior de uma região pré-determinada, bem definida, um conjunto compacto $\mathcal{X} \subset \mathbb{R}^m$, denominada região experimental. Considere o conjunto de vetores $\{\mathbf{x}_i, i = 1, \dots, n\}$ representando uma possível escolha de n unidades experimentais, nas quais N respostas serão observadas. O desenho do experimento ou delineamento experimental pode ser representado por

$$\xi_N = \left\{ \begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \\ p_1 & p_2 & \dots & p_n \end{array} \right\} \quad (2.3.14)$$

com $p_i = r_i/N$ e $N = \sum_{i=1}^n r_i$, onde r_i é o número de replicações feitas na unidade experimental \mathbf{x}_i , $i = 1, \dots, n$. Na representação na equação (2.3.14) as unidades experimentais $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ são os pontos de suporte do experimento e p_1, p_2, \dots, p_n são os pesos associados aos pontos de suporte. O desenho exato ξ_N pode ser visto como uma medida de probabilidade discreta em \mathcal{X} , com $\sum_{i=1}^n p_i = 1$. De acordo com Kiefer e Wolfowitz (1960), a classe de todas as possíveis distribuições discretas, mostrada em (2.3.14), pode ser estendida para uma classe mais abrangente de distribuições de probabilidade, definida nos conjuntos de Borel da região experimental \mathcal{X} . Desta forma, o desenho experimental aproximado será denotado por ξ e, neste caso, deve-se ter $\int_{\mathcal{X}} d\xi(x) = 1$.

Seja $p(y|\mathbf{x}, \theta)$ a função de probabilidade ou de densidade de probabilidade da variável resposta Y , onde $\mathbf{x} \in \mathcal{X}$ e $\theta \in \Theta$, em que $\theta^t = (\theta_1, \dots, \theta_p)$ é o parâmetro de interesse e Θ é o espaço paramétrico. Assim, considerando n observações independentes da variável resposta

Y , a matriz de informação de Fisher, tomada no ponto \mathbf{x} é dada por $\mathbf{I}(\mathbf{x}, \boldsymbol{\theta}) = E_Y \left\{ \frac{-\partial^2 \log p(y|\mathbf{x}, \boldsymbol{\theta})}{\partial \theta \partial \theta^t} \right\}$. A matriz experimental ou a matriz de informação de Fisher esperada, $\mathbf{M}(\boldsymbol{\theta}|\xi)$ por observação, obtida como a esperança da informação de Fisher com respeito à medida de planejamento ξ sobre \mathcal{X} é dada por, $\mathbf{M}(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} \mathbf{I}(\boldsymbol{\theta}|\xi) d\xi(\mathbf{x})$.

Defina Ξ como o conjunto de todas as medidas de probabilidade sobre \mathcal{X} e considere o problema de planejamento ótimo consistindo em encontrar uma medida ξ^* em Ξ que maximize uma função critério $\psi(\mathbf{M}(\xi))$ ou, em notação simplificada, $\psi(\xi)$. A função critério $\psi(\xi) : \Xi \rightarrow \mathbb{R}$. $\psi(\xi)$ é uma função crescente no seguinte sentido, se $\mathbf{M}(\xi_1) - \mathbf{M}(\xi_2)$ é uma matriz não negativa definida, então $\psi(\xi_1) \geq \psi(\xi_2)$. Como mostrado em Fedorov e Hackal (1997), o conjunto Ξ é convexo. Desta forma, considerando uma função côncava ψ o problema de obtenção do desenho ótimo se reduz a um problema de otimização convexa, isto é, $\xi^* = \arg \max_{\xi \in \Xi} \psi(\xi)$.

Na teoria aproximada de planejamento ótimo de experimentos, a derivada direcional Fréchet tem um papel fundamental para obtenção de desenhos ótimos. Seja $\psi : \Xi \rightarrow \mathbb{R}$, define-se como derivada direcional de Fréchet da função ψ em ξ_1 na direção de ξ_2 , quando o limite existe, como

$$F_\psi(\xi_1, \xi_2) = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \{ \psi [(1 - \epsilon)\xi_1 + \epsilon\xi_2] - \psi(\xi_1) \}, \quad (2.3.15)$$

sendo F_ψ chamada de função derivada. A função ψ será dita diferenciável se satisfizer a equação (2.3.15).

De acordo com Fedorov e Hackal (1997), a função derivada $F_\psi(\xi^*, \xi)$ pode ser apresentada como $F_\psi(\xi^*, \xi) = \int_{\mathcal{X}} F_\psi(\xi^*, \xi_x) d\xi(\mathbf{x})$, onde ξ_x é o desenho experimental com um único ponto \mathbf{x} . A função $\phi_\psi(\xi^*, \xi_x) = F_\psi(\xi^*, \xi_x) + C$, onde C é uma constante, é denominada função de sensibilidade e tem um papel importante na teoria de planejamento ótimo.

A teoria da convexidade em planejamento ótimo de experimentos está muito bem apresentada nos livros de Fedorov e Hackal (1997) e Silvey (1980).

Em modelos não lineares, em especial nos modelos lineares generalizados, a matriz experimental depende dos parâmetros do modelo o que torna o processo de busca pelo desenho ótimo mais complexo. Várias abordagens para solucionar esse problema foram introduzidas na literatura, entre elas, desenhos ótimos locais, *minmax*, sequenciais e Bayesianos. Nesta tese será considerada somente a abordagem Bayesiana, como descrita em Chaloner e Larntz (1989) ou em Atkinson, Donev e Tobias (2002).

O critério que será considerado nesta tese é o critério da D -otimalidade, que é utilizado para minimizar a covariância dos estimadores dos parâmetros do modelo.

A função critério D -ótimo Bayesiana é dada por $\psi_\theta(\xi) = E_\theta\{\log \det[\mathbf{M}(\theta|\xi)]\}$, onde E_θ se refere a esperança em relação à distribuição a priori sobre θ e $\log \det[\mathbf{M}(\theta|\xi)]$ representa o logaritmo natural do determinante da matriz experimental. Desta forma, o problema de planejamento D -ótimo Bayesiano consiste em encontrar uma medida de planejamento, ξ^* , na classe de medidas de planejamento Ξ , que maximiza $\psi_\theta(\xi)$.

O Teorema Geral da Equivalência (TGE), proposto por Kiefer e Wolfowitz (1960), tem um papel fundamental na obtenção de desenhos ótimos aproximados, pois permite verificar se determinado desenho encontrado é, de fato, ótimo; facilitando a busca computacional e possibilitando a utilização eficiente de um critério de parada para o procedimento computacional de busca pelo ótimo. Uma extensão do TGE para o caso Bayesiano foi dada por Chaloner e Larntz (1989), com base nos resultados obtidos por Whittle (1973).

A derivada de Fréchet para a classe dos MLG é dada por $F_{\psi_\theta}(\xi, \xi_x) = E_\theta[\text{tr}(\mathbf{M}_x \mathbf{M}^{-1})] - p$, onde $\mathbf{M}_x = \mathbf{M}(\theta|\xi_x)$ é a matriz de planejamento associada ao desenho ξ_x , o desenho que possui apenas o ponto \mathbf{x} , e p é o número de parâmetros do MLG, veja Pinto e Ponce de Leon (2006b). Desta forma, a função de sensibilidade é dada por $\phi_{\psi_\theta}(\xi, \xi_x) = E_\theta[\text{tr}(\mathbf{M}_x \mathbf{M}^{-1})]$. Para os MLGs, $\mathbf{M}_x = w(\theta, \mathbf{x})\mathbf{f}^t(\mathbf{x})\mathbf{f}(\mathbf{x})$, onde $w(\theta, \mathbf{x}) = \left(\frac{\partial \mu}{\partial \eta}\right)^2 \frac{1}{\phi V(\mu)}$. Desta forma a função de sensibilidade é dada por $\phi_{\psi_\theta}(\xi, \xi_x) = E_\theta[w(\theta, \mathbf{x})\mathbf{f}^t(\mathbf{x})\mathbf{M}^{-1}(\theta|\xi)\mathbf{f}(\mathbf{x})]$. Para o desenho discreto dado em (2.3.14), $\phi_{\psi_\theta}(\xi, \xi_x) = E_\theta[w(\theta, \mathbf{x})\mathbf{f}^t(\mathbf{x})(\mathbf{X}^t \mathbf{W} \mathbf{P} \mathbf{X})^{-1}\mathbf{f}(\mathbf{x})]$, onde $\mathbf{P} = \text{Diag}\{p_1, \dots, p_n\}$ é a matriz diagonal contendo os pesos do desenho. O Teorema Geral da Equivalência, dado por Chaloner e Larntz (1989), é apresentado a seguir

Teorema 2.3.1 (TGE). *Considere Ξ a classe de todas as medidas de probabilidade em um conjunto compacto \mathcal{X} . Seja a função critério $\psi_\theta = E_\theta[\log \det(\mathbf{M}(\theta|\xi))]$, então um desenho ψ_θ -ótimo, ξ^* , pode ser equivalentemente caracterizado por alguma das três seguintes condições.*

- a) o desenho ξ^* maximiza ψ_θ sobre Ξ
- b) ξ^* minimiza o $\sup_{\mathbf{x} \in \mathcal{X}} F_{\psi_\theta}(\xi, \xi_x)$
- c) $\sup_{\mathbf{x} \in \mathcal{X}} \phi_\theta(\xi, \xi_x) = p$,

onde p é o número de parâmetros do modelo.

Desenhos ótimos para experimentos com misturas

Em experimentos envolvendo misturas, a construção de planejamentos experimentais ótimos é de fundamental importância, principalmente quando os desenhos tradicionais, usados para situações padrão, não puderem ser aplicados, por exemplo, quando tivermos experimentos de mistura com restrição, onde a região experimental é uma parte da região simplex.

Estudos sobre a relação entre alguns desenhos de Scheffé e desenhos D -ótimos foram apresentados na literatura. Kiefer (1961) provou que os desenhos simplex, usados por Scheffé para os modelos polinomiais de primeira e segunda ordem, são D -ótimos. Atkinson, Donev e Tobias (2002) discutiram a D -otimalidade do planejamento experimental simplex-centroide para modelos de terceira ordem e verificaram que a eficiência do desenho simplex-centroide em relação ao desenho D -ótimo é próxima de 1. Donev (1989) estabeleceu as condições para a D -otimalidade em experimentos de mistura com fatores qualitativos.

A teoria de planejamento ótimo de experimentos em experimentos com misturas está bem apresentada no livro de Atkinson, Donev e Tobias (2002). Outra referência é Lawson e Willden (2016), que discutem planejamento ótimo e análise de experimentos com mistura usando o *software* R (R CORE TEAM, 2019).

Capítulo 3

Variabilidade em experimentos de mistura com variáveis de ruído

3.1 Introdução

Quando algumas variáveis de processo são incontroláveis ou difíceis de serem controladas, elas devem ser tratadas como variáveis de ruído. Variáveis de ruído são consideradas variáveis aleatórias com uma distribuição de probabilidade supostamente conhecida. Assim, a presença de variáveis de ruído no modelo significa que a variância não pode mais ser considerada constante. Uma maneira de controlar a variabilidade presente no modelo é modelando a variância.

A modelagem estatística da variância em experimentos de mistura com variáveis de ruído foi considerada em Steiner e Hamada (1997), que propuseram um modelo combinado de mistura com variáveis de ruído, construíram e resolveram um problema de otimização para minimizar uma função de perda quadrática, levando em consideração a média e a variância da resposta. Outra abordagem para modelar a variância foi proposta por Goldfarb, Borror e Montgomery (2003) usando o método delta, que emprega uma aproximação de série de Taylor de primeira ordem do modelo de regressão em um vetor de variáveis de ruído.

O método delta para experimentos com misturas é apresentado com detalhes na Seção 3.2, na Seção 3.3 é apresentada uma aplicação da modelagem da variabilidade num problema da indústria de alimentos, adicionalmente um processo de otimização para escolha da melhor combinação dos ingredientes de mistura também é apresentado. Considerações finais são dadas na Seção 3.4.

3.2 Método delta

O método delta é uma técnica bem conhecida, baseada em expansões da série de Taylor, para encontrar aproximações para a média e variância de funções de variáveis aleatórias. Considerando a aproximação de Taylor de primeira ordem, o método delta é dado a seguir.

Sejam Z_1, \dots, Z_r variáveis aleatórias com médias μ_1, \dots, μ_r e matriz de variância-covariância Σ_z . Defina $Z^t = (Z_1, \dots, Z_r)$ e $\mu^t = (\mu_1, \dots, \mu_r)$. Para nosso problema de experimentos com misturas, nós também consideramos x como o vetor de componentes de mistura. Suponha que exista uma função diferenciável $\eta(x, Z, \beta, \gamma)$ para a qual nós queremos uma estimativa aproximada da média e da variância $\eta'(x, \mu, \beta, \gamma) = \left. \frac{\partial}{\partial z} \eta(x, z, \beta, \gamma) \right|_{z=\mu}$. A expansão de série de Taylor de primeira ordem de η em relação a μ é dada por

$$\eta(x, z, \beta, \gamma) = \eta(x, \mu, \beta, \gamma) + (z - \mu)^t \eta'(x, \mu, \beta, \gamma) + \mathcal{R}, \quad (3.2.1)$$

onde \mathcal{R} é o resto da aproximação e será ignorado. Assim, como $\eta(x, Z, \beta, \gamma)$ é uma variável aleatória, pois é função de Z que é uma variável aleatória, nós podemos obter

$$E [\eta(x, Z, \beta, \gamma)] \approx \eta(x, \mu, \beta, \gamma) \quad (3.2.2)$$

e

$$Var [\eta(x, Z, \beta, \gamma)] \approx [\eta'(x, \mu, \beta, \gamma)]^t \Sigma_z \eta'(x, \mu, \beta, \gamma). \quad (3.2.3)$$

Para um tratamento abrangente sobre o método delta, veja Casella e Berger (2002), p. 240.

3.3 Modelagem e otimização em um problema de panificação

O problema de panificação, originalmente apresentado por Faergestad e Naes (1997), de acordo com Naes, Faergestad e Cornell (1998), consistiu de um experimento com três ingredientes de mistura e duas variáveis de ruído, e teve como objetivo investigar e avaliar a qualidade final da farinha, composta por diferentes misturas de farinha de trigo, para produção de pães.

Além disso, de acordo com Naes, Faergestad e Cornell (1998), Faergestad e Naes (1997) consideraram três tipos de farinha de trigo: duas noruegueses, *Tjalve* (x_1) e *Folke* (x_2) e uma americana, *Hard Red Spring* (x_3), que foram consideradas como variáveis de controle, e dois tipos de variáveis de processo: tempo de mistura (z_1) e tempo de descanso da massa (z_2), consideradas como variáveis de ruído. A variável resposta foi considerada como o volume do pão após assado, com valor-alvo de 530 ml. As misturas de farinha foram consideradas ingredientes de mistura com $x_1 + x_2 + x_3 = 1$ e com restrições $0,25 \leq x_1 \leq 1,0$; $0 \leq x_2 \leq 0,75$ e $0 \leq x_3 \leq 0,75$, onde x_1 , x_2 e x_3 são as proporções das farinhas Tjalve, Folke e Hard Red Spring, respectivamente. Para as variáveis de ruído, foram consideradas três situações para o tempo de mistura: 5, 15 e 25 minutos e também três situações para o tempo de descanso: 35, 47,5 e 60 minutos.

Um desenho fatorial completo 3^2 foi usado para as variáveis de ruído e as 10 corridas experimentais, correspondente a um planejamento em rede simplex (*simplex lattice design*), foram replicadas em cada uma das nove combinações dos tempos de mistura (tempo em que os ingredientes são misturados) e de descanso (tempo de fermentação da massa), de modo que o planejamento completo envolveu 90 corridas experimentais conforme mostrado na Figura 3.1. Consideramos as variáveis de ruído codificadas como $z_1 = (\text{tempo de mistura} - 15)/10$ and $z_2 = (\text{tempo de descanso} - 47.5)/12.5$, desta forma, as variáveis de ruído foram codificadas como -1, 0, 1 de acordo com seus valores mínimo, máximo e médio.

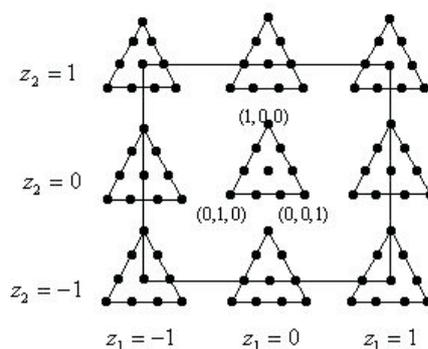


Figura 3.1: Desenho fatorial completo 3^2 em duas variáveis de processo cruzado com um desenho centroeide simplex em três componentes de mistura

Os volumes registrados para os 10 tipos de farinha e as 9 combinações das variáveis de ruído são reproduzidos na Tabela 3.1. Detalhes adicionais sobre a forma como o experimento foi conduzido são fornecidos por Naes, Faergestad e Cornell (1998), e uma descrição mais

detalhada dos aspectos práticos do estudo é fornecida por Faergestad e Naes (1997), de acordo com Naes, Faergestad e Cornell (1998).

Tabela 3.1: Volume de pão para os 10 tipos de farinha e as 9 combinações das variáveis de ruído

n ^o	Ingred. de mistura			Fatores de ruído								
				z ₁			z ₂					
	x ₁	x ₂	x ₃	-1	0	1	-1	0	1	-1	0	1
1	0,25	0,75	0,00	378,89	396,67	392,22	445,56	452,22	487,78	457,22	500,56	472,78
2	0,50	0,50	0,00	388,89	423,33	416,11	460,00	488,89	475,78	472,78	478,00	506,11
3	0,75	0,25	0,00	426,11	483,33	389,44	474,44	514,44	462,78	506,67	591,67	522,22
4	1,00	0,00	0,00	386,11	459,11	423,33	458,33	506,11	514,44	545,56	522,22	551,11
5	0,25	0,50	0,25	417,78	437,22	444,56	484,44	490,00	495,00	497,78	531,11	577,78
6	0,50	0,25	0,25	389,44	447,22	415,00	490,89	528,89	507,78	517,78	567,22	538,33
7	0,75	0,00	0,25	448,33	459,44	455,56	436,00	535,00	552,22	507,44	578,89	590,00
8	0,25	0,25	0,50	413,89	485,56	462,22	483,89	529,44	540,00	565,00	598,89	580,56
9	0,50	0,00	0,50	415,56	514,44	437,78	493,89	583,33	578,89	524,44	694,44	640,00
10	0,25	0,00	0,75	432,78	498,33	517,22	474,44	568,33	579,44	541,11	638,89	638,89

Naes, Faergestad e Cornell (1998) realizaram um estudo utilizando a metodologia de planejamento robusto para o problema de panificação a fim de investigar as relações subjacentes entre a variável de resposta volume de pão e as variáveis de mistura e de ruído, comparando três técnicas para analisar o volume de pão, isto é, o erro quadrático médio, a análise de variância e a abordagem de regressão, onde todos os fatores, os três componentes de mistura e as duas variáveis de ruído, são modelados simultaneamente. Na análise realizada por eles, o modelo cruzado completo para três ingredientes de mistura e duas variáveis de ruído foi tomado como o modelo inicial. No entanto, um argumento detalhado foi apresentado a fim de reduzir o número de parâmetros, removendo alguns dos termos de mistura de segunda e terceira ordens antes de realizar o cruzamento dos modelos de mistura e de variáveis de ruído. Desta forma, o modelo inicial reduzido ficou com 28 termos.

A análise aqui apresentada começa com o mesmo modelo inicial reduzido, envolvendo 28 termos, considerado por Naes, Faergestad e Cornell (1998) e dados na equação (3.3.1).

$$\begin{aligned}
\eta(\mathbf{x}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = & \beta_1^0 x_1 + \beta_2^0 x_2 + \beta_3^0 x_3 + \beta_{12}^0 x_1 x_2 + \beta_{13}^0 x_1 x_3 + \gamma_{12}^0 x_1 x_2 (x_1 - x_2) + \\
& \gamma_{13}^0 x_1 x_3 (x_1 - x_3) + \{\beta_1^1 x_1 + \beta_2^1 x_2 + \beta_3^1 x_3 + \beta_{12}^1 x_1 x_2 + \beta_{13}^1 x_1 x_3 + \\
& \gamma_{12}^1 x_1 x_2 (x_1 - x_2) + \gamma_{13}^1 x_1 x_3 (x_1 - x_3)\} z_1 + \{\beta_1^2 x_1 + \beta_2^2 x_2 + \\
& \beta_3^2 x_3 + \beta_{12}^2 x_1 x_2 + \beta_{13}^2 x_1 x_3 + \gamma_{12}^2 x_1 x_2 (x_1 - x_2) + \\
& \gamma_{13}^2 x_1 x_3 (x_1 - x_3)\} z_2 + \{\beta_1^{11} x_1 + \beta_2^{11} x_2 + \beta_3^{11} x_3 + \beta_{12}^{11} x_1 x_2 + \\
& \beta_{13}^{11} x_1 x_3 + \gamma_{12}^{11} x_1 x_2 (x_1 - x_2) + \gamma_{13}^{11} x_1 x_3 (x_1 - x_3)\} z_1^2
\end{aligned} \tag{3.3.1}$$

Inicialmente, um modelo de regressão linear para a média foi ajustado pelo método de Mínimos Quadrados Ordinários (MQO). Após a aplicação do método *stepwise-backward*, que consiste em adicionar ou eliminar variáveis de acordo com sua significância no modelo, foi possível eliminar os termos do modelo que realmente não influenciavam a variável resposta e obter um modelo parcimonioso com os 18 termos mostrados na Tabela 3.3.

Tabela 3.2: Coeficientes de regressão do modelo da média usando o método de MQO

Termos	Estimativa	Erro padrão	Valor t	Valor-p
x_1	484,624	6,363	76,161	0,0000
x_2	474,875	13,369	35,521	0,0000
x_3	436,381	64,837	6,730	0,0000
$x_1 x_3$	468,313	164,234	2,851	0,0057
$x_1 x_2 (x_1 - x_2)$	375,341	94,623	3,997	0,0002
$x_1 x_3 (x_1 - x_3)$	-403,031	199,679	-2,018	0,0473
$x_1 z_1$	16,768	5,452	3,076	0,0029
$x_3 z_1$	51,876	8,406	6,171	0,0000
$x_1 x_2 (x_1 - x_2) z_1$	-144,553	60,706	-2,381	0,0199
$x_1 z_2$	54,933	6,703	8,195	0,0000
$x_2 z_2$	42,504	8,470	5,018	0,0000
$x_1 x_3 z_2$	188,762	25,167	7,500	0,0000
$x_1 x_3 (x_1 - x_3) z_2$	-202,822	61,681	-3,288	0,0016
$x_2 z_1^2$	-52,644	14,972	-3,516	0,0008
$x_3 z_1^2$	164,077	79,249	2,070	0,0420
$x_1 x_3 z_1^2$	-600,046	199,173	-3,013	0,0036
$x_1 x_2 (x_1 - x_2) z_1^2$	-440,721	109,730	-4,016	0,0001
$x_1 x_3 (x_1 - x_3) z_1^2$	525,480	244,486	2,149	0,0349

A qualidade do ajuste do modelo encontrado é avaliada usando diferentes tipos de gráficos de diagnóstico mostrados na Figura 3.2. Pela observação dos gráficos é possível concluir que não há pontos discrepantes e que o modelo se ajusta bem aos dados.

Na Figura 3.2 O gráfico superior esquerdo representa os resíduos padronizados em relação à ordem das observações. Como nenhum padrão é encontrado nesse gráfico, podemos assumir que os dados são independentes. O gráfico superior direito representa as distâncias de Cook em relação à ordem das observações. Como todos os valores da distância de Cook não excedem o valor do quantil $F_{(0,5,18,72)} \approx 0,97$, não há evidência de valores discrepantes. O gráfico central esquerdo exibe o histograma dos resíduos, mostrando que a hipótese de normalidade não deve ser rejeitada. O gráfico central direito representa os resíduos padronizados em relação ao preditor linear. O gráfico não indica evidência de heteroscedasticidade ou não linearidade. O gráfico inferior esquerdo exibe o gráfico dos quantis dos resíduos (*QQ plot*), reforçando a hipótese de normalidade dos resíduos. O gráfico inferior direito apresenta os valores preditos versus valores observados, mostrando que o modelo está bem ajustado aos dados.

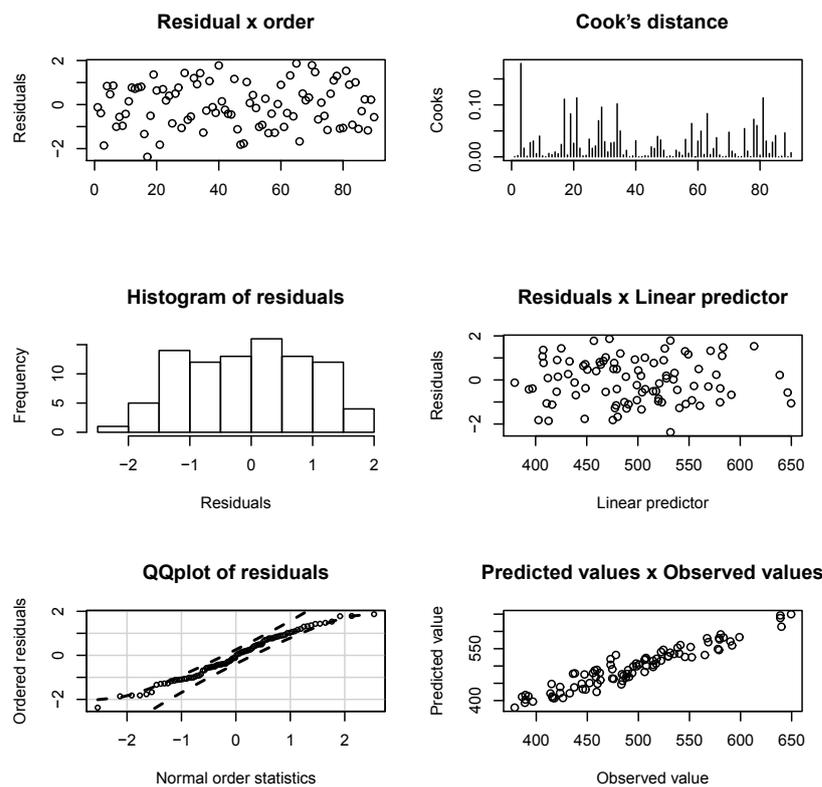


Figura 3.2: Gráficos de diagnóstico para dados de mistura de pão.

Para facilitar a análise, o modelo encontrado será escrito como

$$\hat{\eta}(x, z, \hat{\beta}, \hat{y}) = c_1 + c_2 z_1 + c_3 z_2 + c_4 z_1^2, \quad (3.3.2)$$

onde $c_1 = 484,62x_1 + 474,88x_2 + 436,38x_3 + 468,31x_1x_3 + 375,34x_1x_2(x_1 - x_2) - 403,03x_1x_3(x_1 - x_3)$, $c_2 = 16,77x_1 + 51,88x_3 - 144,55x_1x_2(x_1 - x_2)$, $c_3 = 54,93x_1 + 42,50x_2 + 188,76x_1x_3 - 202,82x_1x_3(x_1 - x_3)$ e $c_4 = -52,64x_2 + 164,08x_3 - 600,05x_1x_3 - 440,72x_1x_2(x_1 - x_2) + 525,48x_1x_3(x_1 - x_3)$. Note que, c_1 , c_2 , c_3 e c_4 são constantes com respeito a z .

Como as variáveis de processo são consideradas variáveis de ruído, elas influenciam a variabilidade do modelo. Desta forma, a modelagem estatística de variância deve ser considerada. Assim, supondo que as variáveis de ruído não são correlacionadas e que também não haja correlação entre as variáveis de ruído e os erros do modelo, a média e a variância do modelo podem ser obtidas pelo método delta da seguinte forma.

Considere $E(Z_i) = \mu_i$ e $Var(Z_i) = \sigma_i^2$, para $i = 1, 2$. Note que, desde que Z_1 and Z_2 são variáveis contínuas, poderia ter sido considerado $Z_i \sim N(\mu_i, \sigma_i^2)$, para $i = 1, 2$, mas isso não é necessário, pois somente precisamos conhecer a média e a variância das variáveis de ruído. Da equação (3.3.2) é possível obter $[\eta'(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\gamma})]^t = (c_2 + 2c_4\mu_1, c_3)$. Desta forma, como $Y = \eta(\mathbf{x}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}) + \epsilon$, onde $\epsilon \sim N(0, \sigma^2)$, $\Sigma_z = \text{Diag}(\sigma_1^2, \sigma_2^2)$, usando as equações (3.2.2) e (3.2.3), as expressões para a média e para a variância são aproximadas por

$$E(Y) \approx c_1 + c_2\mu_1 + c_3\mu_2 + c_4\mu_1^2, \quad (3.3.3)$$

e

$$Var(Y) \approx (c_2 + 2c_4\mu_1)^2\sigma_1^2 + c_3^2\sigma_2^2 + \sigma^2. \quad (3.3.4)$$

Se for assumido que $E(\mathbf{Z}) = \mathbf{0}$, então $E(Y) = c_1$ e $Var(Y) = c_2^2\sigma_1^2 + c_3^2\sigma_2^2 + \sigma^2$. O erro quadrático médio para o modelo ajustado é usado como uma estimativa de σ^2 . Para nosso modelo $\hat{\sigma}^2 = 58.36$.

Seguindo as ideias de Taguchi para a melhoria da qualidade, veja Taguchi (1986), após encontrarmos as equações para $E(Y)$ e $Var(Y)$ e desde que o valor alvo para a variável resposta (volume do pão após assar) é 530 ml, o problema de minimização, mostrado na equação (3.3.5) deve ser resolvido.

$$\begin{array}{l}
 \text{Min } \text{Var}(Y) \\
 \text{Sujeito a } \left\{ \begin{array}{l}
 E(Y) = 530.0 \\
 x_1 + x_2 + x_3 = 1.0 \\
 0.25 \leq x_1 \leq 1.0 \\
 0.0 \leq x_2 \leq 0.75 \\
 0.0 \leq x_3 \leq 0.75
 \end{array} \right. \quad (3.3.5)
 \end{array}$$

onde $E(Y)$ e $\text{Var}(Y)$ são funções de $x_1, x_2, x_3, \mu_1, \mu_2, \sigma_1^2$ e σ_2^2 , sendo que μ_1, μ_2, σ_1^2 e σ_2^2 são fixos e conhecidos. As expressões para $E(Y)$ e $\text{Var}(Y)$ são dadas nas equações (3.3.3) e (3.3.4), respectivamente.

O problema de otimização foi resolvido considerando alguns cenários envolvendo os valores da média e da variância para as variáveis aleatórias tempo de mistura e tempo de descanso. A Tabela 3.3 mostra a combinação ótima para a mistura, sua variância estimada e seu coeficiente de variação, dado por $\sqrt{\text{Var}(Y)}/530$, para cada cenário envolvendo as variáveis de ruído. In Table 3.3, de acordo com a codificação previamente assumida para as variáveis de ruído, os valores da média e da variância de Z_1 and Z_2 foram obtidos por $\mu_1 = (\mu_m - 15)/10$, $\sigma_1^2 = \sigma_m^2/100$, $\mu_2 = (\mu_p - 47,5)/12,5$ e $\sigma_2^2 = \sigma_p^2/156,25$, onde μ_m and σ_m^2 representam, respectivamente, a média e a variância da variável aleatória tempo de mistura e, da mesma forma, μ_p e σ_p^2 representam, nesta ordem, a média e a variância da variável aleatória tempo de descanso.

A partir dos cenários considerados na Tabela 3.3, é possível observar que alguns pontos ótimos foram robustos para alguns cenários e que o coeficiente de variação em todos os casos foi inferior a 12%, indicando uma pequena variabilidade na resposta. Vale ressaltar que os cenários foram considerados apenas como uma forma ilustrativa de mostrar como o problema de otimização poderia ser conduzido. Os cenários poderiam ser escolhidos de forma adequada se houvesse algum estudo prévio sobre o comportamento das variáveis de ruído no processo produtivo. A decisão sobre qual composição de mistura seria ideal, ou seja, robusta aos fatores de ruído, depende do conhecimento sobre o comportamento das variáveis de ruído e, como essa informação não é conhecida, não é possível avaliar quais cenários deveriam ser considerados para a escolha da composição ideal.

Tabela 3.3: Valores ótimos de mistura com seus respectivos valores de $Var(Y)$ e coeficiente de variação para alguns cenários envolvendo as variáveis de ruído: tempo de mistura e tempo de descanso.

Variáveis de ruído não codificadas		Variáveis de ruído codificadas		Ótimo (x_1, x_2, x_3)	$Var(Y)$ estimada	Coef. of variação
Tempo de mistura [†] (μ_m, σ_m^2)	Tempo de descanso [‡] (μ_p, σ_p^2)	Z_1 (μ_1, σ_1^2)	Z_2 (μ_2, σ_2^2)			
(10, 0; 6, 25)	(47, 50; 9, 766)	(-0, 50; 0, 0625)	(0, 00; 0, 0625)	(0,25;0,10;0,65)	604,91	0,046
(12, 5; 6, 25)	(44, 375; 9, 766)	(-0, 25; 0, 0625)	(-0, 25; 0, 0625)	(0,25;0,10;0,65)	480,79	0,041
(12, 5; 6, 25)	(53, 75; 9, 766)	(-0, 25; 0, 0625)	(0, 50; 0, 0625)	(0,27;0,00;0,73)	643,55	0,048
(12, 5; 25, 0)	(44, 375; 39, 063)	(-0, 25; 0, 25)	(-0, 25; 0, 25)	(0,25;0,10;0,65)	1748,08	0,078
(12, 5; 56, 25)	(44, 375; 87, 891)	(-0, 25; 0, 5625)	(-0, 25; 0, 5625)	(0,25;0,10;0,65)	3860,23	0,117
(15, 0; 6, 25)	(53, 75; 9, 766)	(0, 00; 0, 0625)	(0, 50; 0, 0625)	(0,61;0,36;0,03)	216,34	0,028
(15, 0; 6, 25)	(41, 25; 9, 766)	(0, 00; 0, 0625)	(-0, 50; 0, 0625)	(0,44;0,10;0,46)	396,81	0,038
(15, 0; 6, 25)	(47, 50; 9, 766)	(0, 00; 0, 0625)	(0, 00; 0, 0625)	(0,62;0,17;0,21)	260,33	0,030
(15, 0; 25, 0)	(47, 50; 39, 063)	(0, 00; 0, 250)	(0, 00; 0, 250)	(0,62;0,17;0,21)	866,23	0,056
(15, 0; 56, 25)	(47, 50; 87, 891)	(0, 00; 0, 5625)	(0, 00; 0, 5625)	(0,62;0,17;0,21)	1876,07	0,082
(20, 0; 6, 25)	(47, 50; 9, 766)	(0, 50; 0, 0625)	(0, 00; 0, 0625)	(0,25;0,30;0,45)	234,32	0,029
(20, 0; 6, 25)	(53, 75; 9, 766)	(0, 50; 0, 0625)	(0, 50; 0, 0625)	(0,26;0,43;0,31)	211,24	0,027
(20, 0; 25, 0)	(53, 75; 39, 063)	(0, 50; 0, 250)	(0, 50; 0, 250)	(0,26;0,43;0,31)	669,87	0,049
(20, 0; 56, 25)	(53, 75; 87, 891)	(0, 50; 0, 5625)	(0, 50; 0, 5625)	(0,26;0,43;0,31)	1434,25	0,071

[†]Tempo de mistura é uma variável aleatória com média μ_m e variância σ_m^2 .

[‡]Tempo de descanso é uma variável aleatória com média μ_p e variância σ_p^2 .

3.4 Considerações finais

Neste capítulo, a modelagem da variabilidade em experimentos com mistura envolvendo variáveis de ruído foi considerada utilizando o método delta para obtenção dos modelos da média e da variância.

A obtenção de modelos para a média e a variância da resposta possibilitou a construção de um processo de otimização com o objetivo de encontrar uma combinação ótima dos ingredientes da mistura, a fim de tornar robusta a média da variável resposta em relação às condições de ruído.

No exemplo da panificação, que teve como objetivo investigar e avaliar a qualidade final da farinha composta por diferentes misturas de farinha de trigo para produção de pão, a combinação ótima para a mistura de farinha deveria ser obtida de modo que fosse robusta aos tempos de mistura e de descanso, considerados como variáveis de ruído.

Em nossa análise foram considerados vários cenários envolvendo as variáveis de ruído, mostrados na Tabela 3.3. Obtivemos combinações ótimas de mistura que são robustas às condições de ruído, ou seja, para as combinações de misturas encontradas espera-se que o pão produzido tenha um volume médio de 530 ml, independentemente das situações de ruído a que a mistura seja exposta.

Vale também mencionar que em nossa abordagem para o problema considerado a variância da resposta, condicionada às variáveis de ruído, foi tomada como constante, isto é, $Var(Y|Z_1, Z_2) = \sigma^2$, ou seja, assumimos dispersão constante. Contudo, em algumas situações tal suposição não poderia ser considerada. Esse problema será abordado com detalhes nesta tese, a metodologia utilizada será a da MCMD. No próximo capítulo uma metodologia eficiente para obtenção dos modelos da média e da variância na MCMD será apresentada.

Capítulo 4

Procedimento de seleção de variáveis na MCMD

4.1 Introdução

Na análise de regressão, selecionar um modelo adequado é de fundamental importância, mas não é uma tarefa simples. Na literatura sobre modelos de regressão clássicos, existem vários procedimentos para seleção de modelos, embora nenhum deles seja completamente eficiente. Os critérios mais populares para selecionar modelos são: coeficiente de determinação (R^2), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC), Mallows C_p , que é uma variante de AIC , métodos de *forward*, *backward* e *stepwise*, além de outros métodos que usam computação intensiva. Para mais detalhes sobre esses procedimentos, veja, por exemplo, Neter et al. (1996) e Miller (2002).

O processo de seleção de variáveis que será considerado neste capítulo diz respeito à modelagem conjunta da média e dispersão, que é uma extensão dos modelos lineares generalizados, introduzidos por Nelder e Wedderburn (1972).

Os MLGs assumem uma distribuição na família exponencial para a variável resposta e são mais gerais do que os modelos lineares normais. Nos MLGs pode ser acomodada uma relação de média-variância apropriada para os dados e uma escala apropriada pode ser escolhida para modelar a média, na qual a ação das covariáveis é aproximadamente linear. Em modelos lineares generalizados, o foco está na modelagem e estimação da estrutura da média dos dados, enquanto trata o parâmetro de dispersão como uma constante, uma vez que os MLGs permitem automaticamente a dependência da variância em relação à média por meio de

suposições sobre a distribuição da variável resposta. No entanto, existem situações em que os dados observados podem apresentar maior variabilidade do que aquela que está implícita na relação média-variância e, portanto, a perda de eficiência na estimativa dos parâmetros do modelo da média, usando modelos de dispersão constante quando a dispersão é variável, pode ser substancial Antoniadis et al. (2016).

A modelagem conjunta da média e dispersão é um método geral que supera essa dificuldade, permitindo a modelagem da dispersão em função das covariáveis, por meio da utilização de dois modelos lineares generalizados interligados.

Da mesma forma como apresentada na Seção 2.3.2, a abordagem considerada para a MCMD utiliza a quase verossimilhança estendida (QVE) ajustada, conforme proposto por Lee e Nelder (1998). Embora a MCMD possa ser aplicada em diversas áreas, uma de suas principais aplicações e, portanto, o foco deste capítulo, é em planejamento robusto (PR), uma ferramenta muito útil para melhorar a qualidade em experimentos industriais, ver Phadke (1989) e Nair (1992). Segundo Nelder e Lee (1992), para conjuntos de dados em que a dispersão deve ser modelada com quantidades de dados razoavelmente pequenas, como em planejamento robusto, o melhor critério de otimização para ajuste do modelo na MCMD é a QVE.

Em planejamento robusto, modelar e controlar a variância é de suma importância, porém esta tarefa é dificultada pelo fato de que os dados experimentais são, em geral, obtidos de experimentos fatoriais fracionários não replicados com poucas execuções experimentais. A MCMD permite que a dispersão seja modelada de forma muito eficaz em problemas de planejamento robusto, mesmo sem replicação, e o modelo da variância pode ser obtido a partir do modelo de dispersão de uma forma muito simples. Dada a importância da modelagem da média e da variabilidade nos processos industriais, se estivermos considerando a MCMD, a questão que se coloca é como encontrar o melhor modelo para a média e o melhor modelo para a dispersão.

Os métodos usados para selecionar as variáveis na MCMD aplicados em planejamento robusto não são muito claros e baseiam-se principalmente em análise gráfica, ver, por exemplo, Engel e Huele (1996). Por outro lado, para realizar um processo formal de seleção de variáveis, teríamos que considerar um grande número de modelos possíveis, o que torna o processo bastante trabalhoso. Assim, um método claro, confiável e eficiente de seleção de variáveis é de fundamental importância.

O objetivo deste capítulo é propor uma metodologia consistente que permita a seleção de variáveis para os modelos de média e dispersão na MCMD. Propomos um procedimento de seleção de variáveis, com base em testes de hipóteses e na qualidade do ajuste do modelo. Um critério de verificação da qualidade do ajuste é utilizado, a cada iteração do processo de seleção, como um filtro para a escolha dos termos que serão avaliados por um teste de hipótese. Três tipos de critérios foram considerados para verificar a qualidade do ajuste do modelo em nosso procedimento de seleção de variáveis. Os critérios utilizados foram: o critério de informação de Akaike estendido, o critério de informação de Akaike corrigido e um novo critério, específico para a MCMD, um tipo de coeficiente de determinação ajustado estendido.

A seleção das variáveis para modelos conjuntos da média e dispersão conjunta foi considerada por Wang e Zhang (2009), onde foi proposta uma extensão do critério de informação de Akaike baseado em QVE; Wu e Li (2012), para a distribuição normal inversa; Wu, Zhang e Xu (2012), para distribuição lognormal; Charalambous, Pan e Tranmer (2014), para modelos multinível; Charalambous, Pan e Tranmer (2015), para modelos lineares generalizados hierárquicos; Antoniadis et al. (2016), para modelos próprios de dispersão (*proper dispersion models*); Bayer e Cribari Neto (2017), para modelos de regressão beta e Wu, Li e Tao (2020), para mistura de modelos de regressão.

Na Seção 4.2, são propostos um critério e um procedimento para seleção de variáveis na MCMD. Na Seção 4.3, estudos de simulação são realizados para avaliar o procedimento e critério propostos. Uma aplicação a dados reais de um experimento industrial é apresentada na Seção 4.4. Considerações finais são dadas na Seção 4.5.

4.2 Critério e procedimento para seleção de variáveis na MCMD

Para nosso procedimento de seleção de variáveis na modelagem conjunta da média e dispersão, consideramos dois critérios alternativos para verificar a qualidade do ajuste do modelo. O primeiro critério, proposto por nós, é uma combinação dos critérios propostos por Hu e Shao (2008) e Zhang (2017) para seleção de variáveis em modelos lineares generalizados. Hu e Shao (2008) propuseram um critério de qualidade de ajuste dado por

$$R_{HS}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (n - \lambda_n r)}{\sum_{i=1}^n (y_i - \bar{y})^2 / (n - 1)}, \quad (4.2.1)$$

onde $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, $\hat{\mu}_i = \varphi^{-1}(\hat{\eta}_i)$, n é o número de observações, r é o número de parâmetros estimados dos modelo e λ_n é uma função de n , satisfazendo $\lambda_n \rightarrow \infty$ e $\lambda_n/n \rightarrow 0$ quando o tamanho da amostra $n \rightarrow \infty$. As funções λ_n consideradas por eles nos estudos de simulação foram \sqrt{n} e $\log n$.

O critério ajustado proposto por Zhang (2017) para verificar a qualidade do ajuste do modelo é dado por

$$R_Z^2 = 1 - \frac{\sum_{i=1}^n d_{V_i}(y_i, \hat{\mu}_i)/(n-r)}{\sum_{i=1}^n d_{V_i}(y_i, \bar{y})/(n-1)}, \quad (4.2.2)$$

onde $d_V(a, b) = \left[\int_a^b \sqrt{1 + [V'(t)]^2} dt \right]^2$ com $V'(t) = \frac{dV(t)}{dt}$ e $V(t)$ é uma função contínua e derivável em (a, b) . A função $d_V(a, b)$ é o quadrado do comprimento de arco da função de variância entre $V(a)$ e $V(b)$. Na equação (4.2.2), $\sum_{i=1}^n d_{V_i}(y_i, \hat{\mu}_i)$ representa a variação na resposta não explicada pelo modelo e $\sum_{i=1}^n d_{V_i}(y_i, \bar{y})$ representa a variação total na resposta.

De acordo com Zhang (2017), $d_V(a, b)$ pode diferir dramaticamente da distância Euclidiana $(a - b)^2$ quando a função de variância subjacente é não linear. Por exemplo, em um caso particular, mas bastante geral, de função de variância quadrática, isto é, $V(\mu) = c_2\mu^2 + c_1\mu + c_0$, onde c_0 , c_1 e c_2 são constantes conhecidas e $c_2 \neq 0$, $d_V(a, b) = \frac{1}{16c_2^2} \left\{ \log \left(\frac{v(b)}{v(a)} \right) + v(b) - v(a) \right\}^2$, onde $v(t) = V'(t) + \sqrt{1 + [V'(t)]^2}$. Se $c_2 = 0$, a função de variância é linear e $d_V(a, b) = (1 + c_1^2)(a - b)^2$. Consequentemente, para o modelo Normal, onde $V(\mu) = 1$, $d_V(a, b) = (a - b)^2$; para o modelo de Poisson, onde $V(\mu) = \mu$, $d_V(a, b) = 2(a - b)^2$.

O critério proposto por Zhang (2017) é bem definido, desde que para o modelo subjacente, como quase-modelos, sejam especificados a média e a função de variância. Desta forma, R_Z^2 é aplicável aos modelos de quase verossimilhança, pois não depende da especificação completa da função de verossimilhança, necessitando apenas o conhecimento da média e da função de variância.

Desta forma, a extensão do critério proposto por Zhang (2017) para modelos conjuntos da média e dispersão é imediata, porém, no modelo da média, o parâmetro de dispersão deve ser levado em consideração para o cálculo do critério. Na verdade, como no processo iterativo para a MCMD, a variância para a i -ésima observação é dada por $Var(Y_i) = \phi_i V(\mu_i)$, onde ϕ_i é conhecido, dado na iteração anterior. Podemos pensar em ϕ_i como uma medida de peso para $V(\mu_i)$.

Assim, usando a ideia considerada por Zhang (2017) para medir a mudança de variação da variável resposta, associada à função de penalidade proposta por Hu e Shao (2008) e

considerando o parâmetro de dispersão como peso para a função de variância, propomos, para o processo de modelagem conjunta da média e dispersão, os seguintes critérios de qualidade de ajuste. Para o modelo da média, propomos o seguinte critério

$$\tilde{R}_m^2 = 1 - \frac{\sum_{i=1}^n \tilde{d}_{V_i}(y_i, \hat{\mu}_i)/(n - \lambda_n p)}{\sum_{i=1}^n \tilde{d}_{V_i}(y_i, \bar{y})/(n - 1)} \quad (4.2.3)$$

e para o modelo da dispersão o critério proposto é

$$\tilde{R}_d^2 = 1 - \frac{\sum_{i=1}^n d_{V_i}(d_i^*, \hat{\phi}_i)/(n - \lambda_n q)}{\sum_{i=1}^n d_{V_i}(d_i^*, \bar{d}^*)/(n - 1)}, \quad (4.2.4)$$

onde $\tilde{d}_V(a, b) = \left[\int_a^b \sqrt{1 + \phi^2 [V'(t)]^2} dt \right]^2$ é calculado para o modelo da média da mesma forma como proposto por Zhang (2017), mas considerando o parâmetro de dispersão como peso para a função de variância; $d_V(a, b)$, calculado para o modelo da dispersão, é o mesmo que aqueles dado na equação (4.2.2); n é o número de observações, λ_n é uma função de n , como proposta por Hu e Shao (2008), isto é, $\lambda_n \rightarrow \infty$ e $\lambda_n/n \rightarrow 0$ quando o tamanho da amostra $n \rightarrow \infty$, p é o número de parâmetros do modelo da média, q é o número de parâmetros no modelo da dispersão, $\hat{\mu}_i = \varphi^{-1}(\hat{\eta}_i)$, $\hat{\phi}_i = \exp(\hat{\xi}_i)$, $\bar{y} = (1/n) \sum_{i=1}^n y_i$ e $\bar{d}^* = (1/n) \sum_{i=1}^n d_i^*$, com y_i e d_i^* dados na Tabela 2.1. Na equação (4.2.4), podemos notar que se $\lambda_n = 1$ temos o critério proposto por Zhang (2017).

Note também que, como no critério proposto por Zhang (2017), $\sum_{i=1}^n \tilde{d}_{V_i}(y_i, \hat{\mu}_i)$ e $\sum_{i=1}^n d_{V_i}(d_i^*, \hat{\phi}_i)$ representam a variação na resposta não explicada pelo modelo e $\sum_{i=1}^n \tilde{d}_{V_i}(y_i, \bar{y})$ e $\sum_{i=1}^n d_{V_i}(d_i^*, \bar{d}^*)$ representam a variação total na resposta. Assim, para ambos os critérios \tilde{R}_m^2 e \tilde{R}_d^2 , quanto maior o valor, melhor.

Vale ressaltar também que, dependendo dos modelos considerados, incluindo seu número de parâmetros, o tamanho da amostra e a escolha da função λ_n , os critérios \tilde{R}_m^2 e \tilde{R}_d^2 podem ter valores fora do intervalo $[0, 1]$.

Na equação (4.2.4), desde que estamos considerando o modelo Gama para a dispersão, $d_V(a, b) = \frac{1}{16} \left\{ \log \left(\frac{2b + \sqrt{1+4b^2}}{2a + \sqrt{1+4a^2}} \right) + 2b\sqrt{1+4b^2} - 2a\sqrt{1+4a^2} \right\}^2$. Para os modelos da média que serão considerados neste capítulo os valores de $\tilde{d}_V(a, b)$, usados na equação (4.2.3), são calculados por: $\tilde{d}_V(a, b) = (b - a)^2$ para o modelo do tipo Normal; $\tilde{d}_V(a, b) = (1 + \phi^2)(b - a)^2$ para o modelo do tipo Poisson e para o modelo do tipo Binomial, $\tilde{d}_V(a, b) = \frac{1}{16\phi^2} \left\{ \log \left(\frac{\beta + \sqrt{1+\beta^2}}{\alpha + \sqrt{1+\alpha^2}} \right) + \beta\sqrt{1+\beta^2} - \alpha\sqrt{1+\alpha^2} \right\}^2$, onde $\alpha = \phi(1 - 2a)$ e $\beta = \phi(1 - 2b)$. Note que no caso do modelo Normal nosso critério é o mesmo critério proposto por Hu e Shao (2008).

O segundo critério, alternativo ao primeiro, considerado para o modelo da média em nosso procedimento de seleção de variáveis na MCMD, é o critério de informação de Akaike estendido (*EAIC*), proposto por Wang e Zhang (2009), dado por

$$EAIC = -2Q^+(\boldsymbol{\mu}, \boldsymbol{\phi}; \mathbf{y}) + F(\kappa, n), \quad (4.2.5)$$

onde $\kappa = p + q$ é a soma do número de parâmetros dos modelos da média e da dispersão e $F(\kappa, n)$ é uma função de penalidade que depende de κ e n . Em nossos estudos de simulação consideramos $F(\kappa, n) = \frac{2\kappa n}{n-\kappa-1}$. Tal função de penalidade, originalmente proposta por Sugiura (1978) como uma correção de viés ao critério de informação de Akaike no contexto de modelos de regressão linear, foi estendida para modelos de regressão não linear e modelos de séries temporais autoregressivos por Hurvich e Tsai (1989). Para o modelo de dispersão, como estamos considerando a verossimilhança conhecida, o critério utilizado será o critério de informação de Akaike corrigido (*AIC_c*), dado por $AIC_c = -2 \ln \hat{L} + \frac{2qn}{nq-1}$, onde \hat{L} é o valor máximo da função de verossimilhança para o modelo de dispersão Gama. Para os critérios *EAIC* e *AIC_c*, quanto menor o valor, melhor.

O Algoritmo 4.1 mostra como os termos do modelo considerado (média ou dispersão) são selecionados para encontrar o modelo ótimo.

Algoritmo 4.1: Pseudo algoritmo para seleção de variáveis no modelo da média (dispersão)

Na MCMD, para um dado e fixo modelo da dispersão (média), escolha os termos do modelo de média (dispersão), considerando a seguinte sequência de passos.

1. Seja \mathcal{V} o conjunto de termos que serão usados no processo de seleção de variáveis e considere $\vartheta = \theta_0$ o preditor linear do modelo inicial.
2. Ajuste modelos com os preditores lineares $\vartheta_j = \vartheta + \theta_j v_j$, para todo $v_j \in \mathcal{V}$.
3. Para cada modelo ajustado no passo 2, calcule a medida de seleção apropriada (\tilde{R}_m^2 , \tilde{R}_d^2 , *EAIC*, ou *AIC_c*) para verificar a qualidade do ajuste e escolha o modelo com preditor linear, digamos $\vartheta_k = \vartheta + \theta_k v_k$, que tenha o melhor ajuste. Avalie o valor da medida de seleção encontrada na iteração atual com o valor da medida de seleção encontrada na iteração anterior. Se o valor da medida de seleção atual for melhor do que o valor anterior, vá para o passo 4, caso contrário, vá para o passo 5.

4. Para os modelos com preditores lineares encaixados ϑ e ϑ_k , aplique um teste apropriado para verificar a significância da adição de v_k no preditor linear ϑ . Se v_k for significativo, remova o termo v_k de \mathcal{V} , faça $\vartheta = \vartheta_k$ e volte para o passo 2. Caso contrário, o procedimento termina e ϑ é o modelo final escolhido.
5. Avalie, da mesma forma como no passo 4, a significância da adição de v_k ao preditor linear ϑ . Se v_k for significativo, faça $\vartheta = \vartheta_k$. Pare o procedimento e ϑ é o modelo final escolhido.

No passo 4 do Algoritmo 4.1, no caso do modelo de dispersão, para o qual estamos assumindo um MLG com distribuição Gama, o teste para comparar dois modelos encaixados é a análise usual do desvio para MLGs. Neste caso, se H_a e H_b são duas hipóteses encaixadas de dimensão $a < b$, ou seja, $\zeta_a < \zeta_b$, então, sob H_a , a mudança no desvio, dada por $\frac{1}{2} \{D(\mathbf{d}^*, \boldsymbol{\phi}_a) - D(\mathbf{d}^*, \boldsymbol{\phi}_b)\} = \frac{1}{2} \left\{ \sum_{i=1}^n d_{d_i}^a - \sum_{i=1}^n d_{d_i}^b \right\}$, tem uma distribuição assintótica χ_{b-a}^2 , onde $D(\mathbf{d}^*, \boldsymbol{\phi}_k)$ é o desvio (*deviance*) para o modelo com média ϕ_k e $d_{d_i}^k = 2 \left\{ -\log \left(\frac{d_i^*}{\phi_{k_i}} \right) + \frac{(d_i^* - \phi_{k_i})}{\phi_{k_i}} \right\}$ é o componente do desvio, como mostrado na Tabela 2.1. No caso do modelo da média, onde $\boldsymbol{\phi}$ é dado, o teste para comparar dois modelos encaixados é obtido pelo teste de razão de quase-verossimilhanças aplicado à quase-verossimilhança estendida. A justificativa para usar o teste de razão de quase-verossimilhanças no modelo da média é dada pelo fato de que a quase-verossimilhança estendida é uma aproximação do ponto de sela para o logaritmo da verossimilhança, especialmente quando existe uma distribuição da família exponencial com uma função de variância determinada, veja Nelder e Lee (1991) e McCullagh e Nelder (1989), p. 350. Assim, da mesma forma como foi considerado anteriormente no caso do modelo da dispersão, se H_c e H_d são duas hipóteses encaixadas de dimensão $c < d$, isto é, $\eta_c < \eta_d$, então, sob H_c , a mudança no quase desvio estendido, dada por $-2 \left\{ Q^+(\boldsymbol{\mu}_c, \boldsymbol{\phi}; \mathbf{y}) - Q^+(\boldsymbol{\mu}_d, \boldsymbol{\phi}; \mathbf{y}) \right\} = \sum_{i=1}^n \frac{1}{\phi_i} d_{\mu_{c_i}}^* - \sum_{i=1}^n \frac{1}{\phi_i} d_{\mu_{d_i}}^*$, tem uma distribuição assintótica χ_{d-c}^2 , onde $d_{\mu_{k_i}}^*$ é o componente do desvio padronizado em relação a μ_{k_i} , dado na Tabela 2.1. Contudo, este resultado é verdadeiro quando ϕ_i é o valor correto, o que não é o caso em nosso processo de aprendizagem iterativo para a construção dos modelos conjuntos da média e da dispersão. Desta forma, para o nosso procedimento, devido à incerteza na estimação de ϕ_i , nós assumimos que ϕ_i seja substituído por $\tau \phi_i$, onde τ é uma constante desconhecida. Podemos dizer que τ é uma compensação por não sermos capazes de estimar ϕ_i de forma exata. Consequentemente, o teste de hipótese considerado no nosso procedimento de seleção

no modelo da média, que supera a dependência do parâmetro desconhecido τ , é o teste F e não é difícil de verificar que

$$\frac{\left(\sum_{i=1}^n \frac{1}{\phi_i} d_{\mu_{c_i}}^* - \sum_{i=1}^n \frac{1}{\phi_i} d_{\mu_{d_i}}^*\right) / (d - c)}{\left(\sum_{i=1}^n \frac{1}{\phi_i} d_{\mu_{d_i}}^*\right) / (n - d)} \quad (4.2.6)$$

tem uma distribuição assintótica $F_{d-c, n-d}$. Para uma compreensão abrangente da relação entre o teste F e o teste da razão de verossimilhanças, considerando a aproximação do ponto de sela, consulte Jorgensen (1992), p. 89.

Nossa estratégia de seleção, mostrada no Algoritmo 4.2, consiste em um esquema de três etapas em que o modelo da dispersão selecionado é usado para selecionar o melhor modelo da média e vice-versa. Em nosso esquema de seleção, usamos um procedimento recursivo que somente é finalizado quando o valor usado como uma medida de qualidade de ajuste para o modelo da média atinge um valor ótimo, por exemplo, quando \tilde{R}_m^2 deixar de aumentar ou quando $EAIC$ parar de diminuir.

Algoritmo 4.2: Pseudo algoritmo para seleção de variáveis na MCMD

1. Assumindo dispersão constante, use o Algoritmo 4.1 para encontrar os termos do atual melhor modelo da média e, para esse modelo, calcule o valor de \tilde{R}_m^2 ($EAIC$).
2. Assumindo que o selecionado modelo da média atual é adequado, use o Algoritmo 4.1 para encontrar os termos do melhor modelo atual da dispersão.
3. Assumindo que o selecionado modelo da dispersão atual é adequado, use o Algoritmo 4.1 para encontrar os termos do melhor modelo da média atual e para este modelo calcule um novo valor de \tilde{R}_m^2 ($EAIC$). Se o valor atualizado de \tilde{R}_m^2 diminuir ($EAIC$ aumentar) ou for igual ao valor obtido anteriormente, pare e os modelos encontrados anteriormente para média e dispersão são escolhidos. Caso contrário, volte para o passo 2.

O esquema de seleção proposto neste capítulo consiste em um método que, além de fornecer um processo claro e objetivo de seleção de variáveis, visa reduzir custos computacionais inerentes à grande complexidade envolvida na seleção de variáveis para modelos conjuntos de média e dispersão. Na verdade, em nosso processo de seleção de variáveis, se assumirmos que m regressores são considerados para cada modelo no Algoritmo 4.1, temos $\sum_{j=0}^{k-1} \binom{m-j}{1} + 1 = \frac{k}{2}(2m - k + 1) + 1$ possibilidades de modelos conjuntos a serem

avaliados, onde $k \leq m$ representa o número de iterações do algoritmo. Em nosso processo, regido pelo Algoritmo 4.2, o Algoritmo 4.1 é usado duas vezes, exceto na primeira iteração, onde é usado três vezes. Portanto, se considerarmos l iterações do Algoritmo 4.2, teríamos que verificar $(2l + 1)\binom{k}{2}(2m - k + 1) + 1$ modelos conjuntos. Sem usar nosso processo, teríamos que avaliar $(\sum_{j=0}^m \binom{m}{j}) \times (\sum_{j=0}^m \binom{m}{j}) = 2^m \times 2^m = 2^{2m}$ diferentes modelos conjuntos. Como ilustração desse resultado, considere $l = m = k = 10$. Em nosso procedimento, devemos avaliar $21 \times 56 = 1176$ modelos conjuntos. Sem usar nosso procedimento, teríamos que avaliar $2^{20} = 1048576$ modelos conjuntos.

4.3 Avaliação numérica

Para avaliar a acurácia do procedimento de seleção de variáveis, foram realizadas simulações de Monte Carlo para verificar a porcentagem de acertos. O objetivo principal é avaliar como o esquema de seleção de variáveis se comporta para diferentes tamanhos de amostra e como os critérios considerados afetam o esquema proposto.

No processo de avaliação dos modelos selecionados, foram consideradas três situações possíveis: categoria 1 (modelo Tipo 1): quando o modelo selecionado não contém todos os termos do modelo simulado; categoria 2 (modelo Tipo 2): quando o modelo selecionado contém todos os termos do modelo simulado mais outros termos e modelo correto (Ótimo): quando o modelo selecionado contém exatamente os termos do modelo simulado. Todas as simulações foram realizadas no software estatístico R (R CORE TEAM, 2019).

As simulações foram realizadas para modelos da média em que a variável resposta foi considerada seguindo a distribuição Normal, Binomial ou Poisson. Para o modelo da dispersão, a variável resposta foi considerada como tendo uma distribuição Gama. Os preditores lineares, escolhidos para o processo de geração de dados, foram:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (4.3.1)$$

e

$$\zeta = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 z_3, \quad (4.3.2)$$

onde (4.3.1) foi o predito linear usado para o modelo da média e (4.3.2) foi o preditor linear utilizado para o modelo da dispersão.

Para as simulações, foram considerados 16 diferentes cenários, que se originaram da combinação de quatro possibilidades para os critérios de bondade de ajuste com quatro diferentes tamanhos de amostra. Os tamanhos de amostra usados nas simulações foram 25, 50, 100 e 150. Os critérios considerados foram o critério de informação Akaike estendido, o critério de informação Akaike corrigido e três variantes do critério por nós proposto; uma usando $\lambda_n = 1$, outra com $\lambda_n = \log n$ e outra com $\lambda_n = \sqrt{n}$. O critério de informação de Akaike estendido foi usado no modelo da média e o critério de informação de Akaike corrigido foi usado no modelo da dispersão. Para cada cenário considerado foram realizadas 1000 replicações de Monte Carlo. Os valores de $\mathbf{x}^t = (x_1, x_2, x_3)$ e $\mathbf{z}^t = (z_1, z_2, z_3)$ foram obtidos aleatoriamente de uma distribuição uniforme entre -1 e 1 .

Em relação ao nível de significância, considerado para os testes de hipóteses no passo 4 do Algoritmo 4.1, entendemos que um nível de 0,05 é muito estrito e pode excluir variáveis importantes do modelo. Lee e Koval (1997) examinaram a questão do nível de significância em procedimentos *stepwise* para o caso de regressão logística e sugeriram que o nível de significância adotado deveria estar entre 0,15 e 0,20. Com base nas discussões apresentadas por Lee e Koval (1997) e nas referências consideradas no artigo publicado por eles, optamos por utilizar um valor intermediário entre 0,05 e 0,15, ou seja, um nível de significância de 0,10. Em todos os estudos simulados realizados, $i = 1, \dots, n$ refere-se à i -ésima resposta observada e n é o tamanho da amostra. Os estudos de simulação são mostrados a seguir.

4.3.1 Estudo de simulação para a distribuição Normal

Para o modelo da média com distribuição de probabilidade Normal ou tipo Normal, isto é, com $V(\mu_i) = 1$, o processo de simulação dos dados deve assegurar que $E(Y_i) = \mu_i$ e $Var(Y_i) = \phi_i$. Tais resultados podem ser obtidos diretamente tomando $Y_i \sim Normal(\mu_i, \phi_i)$ ou considerando $S_i \sim \phi_i \chi_1^2$, onde χ_1^2 representa a distribuição Qui-quadrado com 1 grau de liberdade, e tomando $Y_i = \pm \sqrt{S_i} + \mu_i$.

Os valores escolhidos para o vetor de parâmetros $(\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3)$ usados para executar as simulações foram $(4, 15, 13, 0, 0.3, 0, 3, 0)$, onde o valor zero significa que o parâmetro não foi usado na simulação. As respostas, para cada simulação de Monte Carlo, foram geradas considerando função de ligação logarítmica para o modelo da dispersão e função de ligação identidade para o modelo da média, de acordo com o Apêndice A.1.

A Tabela 4.1 apresenta a porcentagem dos modelos Ótimos, Tipo 1 e Tipo 2, obtidos a partir de 1000 simulações de Monte Carlo do procedimento proposto de seleção de variáveis, conforme previamente estabelecido no início da Seção 4.3.

Analisando os percentuais para os modelos da média, referindo-se ao modelo Ótimo na Tabela 4.1, podemos ver que as simulações realizadas com tamanho de amostra 25 já apresentam altas taxas de sucesso, mas com índices consideráveis de modelos Tipo 1. Quando aumentamos o tamanho da amostra, podemos observar o rápido aumento nas porcentagens para o modelo Ótimo e uma queda acentuada nas porcentagens para o modelo Tipo 1. Para os modelos de dispersão, há uma menor taxa do modelo Ótimo, mas uma taxa aceitável é alcançada para as amostras de tamanho 100 e 150. Com relação aos modelos Tipo 2, notamos que tanto para os modelos da média quanto para os modelos da dispersão o esquema de seleção de variáveis apresentou percentuais baixos e continuaram com percentuais baixos, mesmo aumentando o tamanho da amostra.

Tabela 4.1: Porcentagens de tipos de modelos encontrados em 1000 simulações de Monte Carlo do procedimento de seleção de variáveis proposto para a distribuição Normal

Modelo	Critério	Ótimo				Tipo 2				Tipo 1			
		Tamanho de amostra				Tamanho de amostra				Tamanho de amostra			
		25	50	100	150	25	50	100	150	25	50	100	150
Média	$\tilde{R}_m^2(\log n)$	71,0	83,5	92,0	95,1	4,9	5,1	5,5	4,8	24,1	11,4	2,5	0,1
	$\tilde{R}_m^2(\sqrt{n})$	76,6	88,5	92,0	98,9	1,0	0,2	5,5	0,1	22,4	11,3	2,5	1,0
	$\tilde{R}_m^2(1)$	69,2	81,9	90,4	90,4	5,8	7,0	5,9	8,4	25,0	11,1	3,7	1,2
	$E A I C$	66,6	88,5	91,7	94,7	6,7	0,2	6,3	5,0	26,7	11,3	2,0	0,3
Dispersão	$\tilde{R}_d^2(\log n)$	30,9	53,4	62,8	68,8	0,8	2,1	4,5	4,0	68,3	44,5	32,7	27,2
	$\tilde{R}_d^2(\sqrt{n})$	32,0	53,8	62,9	67,5	1,2	3,5	4,4	3,4	66,8	42,7	32,7	29,1
	$\tilde{R}_m^2(1)$	30,9	52,1	63,8	67,9	1,8	3,6	2,7	1,5	67,3	44,3	33,5	30,6
	$A I C_c$	36,1	58,1	71,9	78,2	1,6	5,6	7,0	4,0	62,3	36,3	21,1	17,8

Os resultados apresentados na Tabela 4.1 mostram que o procedimento proposto é eficiente para encontrar o modelo ótimo a partir de tamanho de amostra 150, mas também é aceitável com tamanho de amostra menor. Quando comparamos os percentuais encontrados para os modelos da média e da dispersão, verificamos que tanto nos modelos da média quanto nos modelos da dispersão o procedimento de seleção de variáveis é eficiente, porém para o modelo

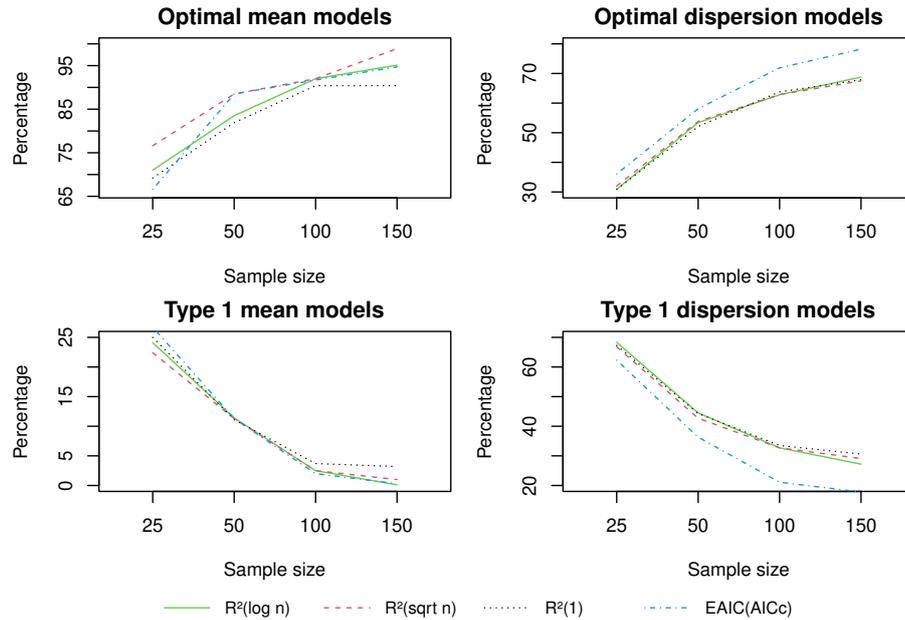


Figura 4.1: Gráficos, obtidos das simulações da distribuição Normal, quanto à relação entre tamanho da amostra e porcentagem encontrada para os modelos Ótimo e Tipo 1, para os critérios \tilde{R}_m^2 , \tilde{R}_d^2 , $EAIC$ e AIC_c , onde a linha sólida representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2) com $\lambda_n = \log n$, a linha tracejada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = \sqrt{n}$, a linha pontilhada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = 1$ e a linha tracejada e pontilhada representa o gráfico para $EAIC(AIC_c)$.

da dispersão a eficiência é menor que a do modelo da média, principalmente para pequenas amostras. Para os modelos da dispersão, tivemos percentual máximo de modelo Ótimo igual a 78,2%, considerando o critério AIC_c ; enquanto para os modelos da média, o percentual máximo foi de 98,9%, considerando o critério \tilde{R}_m^2 com $\lambda_n = \sqrt{n}$.

Ao comparar os critérios \tilde{R}_m^2 (\tilde{R}_d^2) e $EAIC(AIC_c)$, usados como medidas de seleção, observamos que, para os modelos da média, o critério \tilde{R}_m^2 com $\lambda_n = \sqrt{n}$ foi melhor e para os modelos da dispersão, o critério AIC_c teve melhor desempenho. Essas conclusões foram observadas para todos os tamanhos de amostra.

Os gráficos exibidos na Figura 4.1 para os percentuais apresentados na Tabela 4.1, mostram apenas os resultados para os modelos Ótimo e Tipo 1 e confirmam visualmente parte das conclusões obtidas anteriormente. A linha sólida mostra a porcentagem de acertos do procedimento proposto usando o critério \tilde{R}_m^2 (\tilde{R}_d^2) com $\lambda_n = \log n$, a linha tracejada mostra a porcentagem de acertos usando o critério \tilde{R}_m^2 (\tilde{R}_d^2) com $\lambda_n = \sqrt{n}$, a linha pontilhada mostra a porcentagem de acertos usando o critério \tilde{R}_m^2 (\tilde{R}_d^2) com $\lambda_n = 1$ e a linha tracejada e pontilhada apresenta a porcentagem de acertos usando o critério $EAIC(AIC_c)$. Os gráficos representam

bem a eficiência do procedimento proposto, pois mostram o crescimento das porcentagens dos modelos Ótimos e a diminuição considerável das porcentagens dos modelos Tipo 1. No que diz respeito à comparação dos critérios, os gráficos referentes ao modelo da média possuem as linhas do gráfico muito próximas, com uma vantagem para o critério \tilde{R}_m^2 com $\lambda_n = \sqrt{n}$. Nos gráficos referentes ao modelo de dispersão, o critério AIC_c é representado pela reta mais afastada das demais, indicando uma melhor eficiência que o critério \tilde{R}_d^2 , independente da função λ_n considerada.

4.3.2 Estudo de simulação para a distribuição Binomial

Para o modelo da média com distribuição de probabilidade Binomial com índice m ou tipo Binomial, isto é, com $V(\mu_i) = \mu_i(1 - \mu_i/m)$, o processo de simulação dos dados deve assegurar que $E(Y_i) = \mu_i$ e $Var(Y_i) = \phi_i\mu_i(1 - \mu_i/m)$, garantindo, assim, sobredispersão (*overdispersion*). De acordo com McCullagh e Nelder (1989), sobredispersão ocorre quando a variância da resposta excede a variância nominal do modelo teórico. Para a distribuição Binomial a variância nominal é $\mu(1 - \mu/m)$. A sobredispersão pode surgir de várias maneiras, mas o mecanismo mais simples e comum é o agrupamento (*clustering*) na população. O processo de geração dos dados é adaptado de McCullagh e Nelder (1989), p. 125.

Os agrupamentos (*clusters*) geralmente variam em tamanho, mas, da mesma forma como considerado por McCullagh e Nelder (1989), também assumimos que o tamanho do *cluster*, k , é fixo e que os m indivíduos amostrados são provenientes de $\ell = m/k$ *clusters*. Consideramos o número total de respondentes positivos, representado pela variável aleatória condicional $Y|\Pi = S_1 + \dots + S_\ell$, onde S_j , o número de respondentes positivos no *cluster* j , com $j = 1, \dots, \ell$, são variáveis aleatórias independentes com a mesma distribuição que a variável aleatória $S \sim Binomial(k, \pi)$. Também assumimos que a variável aleatória Π tem uma distribuição Beta. Observe que, para dado π_i , $Y_i|\pi_i \sim Binomial(m, \pi_i)$. Desta forma, para $\Pi_i \sim Beta(a_i, b_i)$, podemos obter $E(Y_i) = m\lambda_i$ e $Var(Y_i) = m\lambda_i(1 - \lambda_i)[1 + (m - 1)\delta_i]$, onde $\lambda_i = \frac{a_i}{a_i + b_i}$ e $\delta_i = \frac{1}{a_i + b_i + 1}$. A variável aleatória Y_i tem a distribuição Beta-Binomial, veja McCullagh e Nelder (1989), p. 140. Assim, tomando $\mu_i = m\lambda_i$ e $\phi_i = 1 + (m - 1)\delta_i$ temos o resultado desejado.

Para a distribuição binomial, os valores escolhidos para o vetor de parâmetro $(\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3)$ usados para realizar simulações foram $(0, 2; 0, 6; 0, 8; 1; 0; 0; 2, 5)$, onde o valor zero significa que o parâmetro não foi usado na simulação. Também consideramos $k = 5$ e $m = 10$. Para cada simulação de Monte Carlo, as respostas foram geradas considerando a

função de ligação logarítmica para o modelo de dispersão e a função logística para o modelo da média, de acordo com o Apêndice A.2.

Tabela 4.2: Porcentagens dos tipos de modelos encontrados em 1000 simulações de Monte Carlo do procedimento de seleção de variáveis proposto para a distribuição Binomial

Modelo	Critério	Ótimo				Tipo 2				Tipo 1			
		Tamanho de amostra				Tamanho de amostra				Tamanho de amostra			
		25	50	100	150	25	50	100	150	25	50	100	150
Média	$\tilde{R}_m^2(\log n)$	33,9	63,1	66,1	88,8	0,4	2,1	2,0	1,4	65,7	34,8	31,9	9,8
	$\tilde{R}_m^2(\sqrt{n})$	13,4	51,8	82,2	92,4	0,4	0,2	0,3	0,3	86,2	48,0	17,5	7,3
	$\tilde{R}_m^2(1)$	35,3	69,0	80,8	87,7	2,9	3,4	4,7	4,6	61,8	27,6	14,5	7,7
	$EAIC$	53,9	88,6	94,4	96,2	2,4	4,3	5,3	0,0	43,7	7,1	0,3	3,8
Dispersão	$\tilde{R}_d^2(\log n)$	13,8	34,6	38,4	80,8	0,2	1,4	1,3	6,5	86,0	64,0	60,3	12,7
	$\tilde{R}_d^2(\sqrt{n})$	11,9	33,8	66,2	83,6	0,4	1,8	4,3	4,9	87,7	64,4	29,5	11,5
	$\tilde{R}_d^2(1)$	11,2	33,4	65,5	79,4	0,8	2,3	4,1	6,4	88	64,3	30,4	14,5
	AIC_c	17,3	43,5	69,7	83,6	0,4	3,1	4,1	5,5	81,9	53,4	26,2	10,9

No caso Binomial, as simulações tiveram uma baixa taxa de sucesso para os modelos Ótimos para tamanho de amostra de 25, porém houve um rápido crescimento dessa taxa, atingindo valores excelentes para tamanho de amostra de 150. Esse comportamento foi observado tanto para o modelo da média quanto para o modelo de dispersão. Ao contrário das simulações para o caso Normal, o critério $EAIC$ apresentou melhores resultados para o modelo da média, porém observa-se que os critérios convergem para uma similaridade quando o tamanho da amostra é aumentado. Para amostras menores, observamos uma grande diferença entre os critérios, indicando que neste caso o critério $EAIC$ deve ser escolhido, mas para grandes amostras ambos os critérios podem ser escolhidos. Para o modelo da dispersão, o critério AIC_c foi ligeiramente melhor do que os outros critérios, embora em alguns casos os valores foram próximos. Esses resultados mostram que a escolha do critério deve ser feita com cautela.

Os gráficos exibidos na Figura 4.2, para percentuais apresentados na Tabela 4.2, mostram o que foi discutido anteriormente, onde podemos ver as linhas, que são divergentes para pequenas amostras, convergindo para um ponto comum. A linha tracejada e pontilhada está acima das demais nos gráficos dos modelos ótimos e abaixo das outras linhas nos gráficos

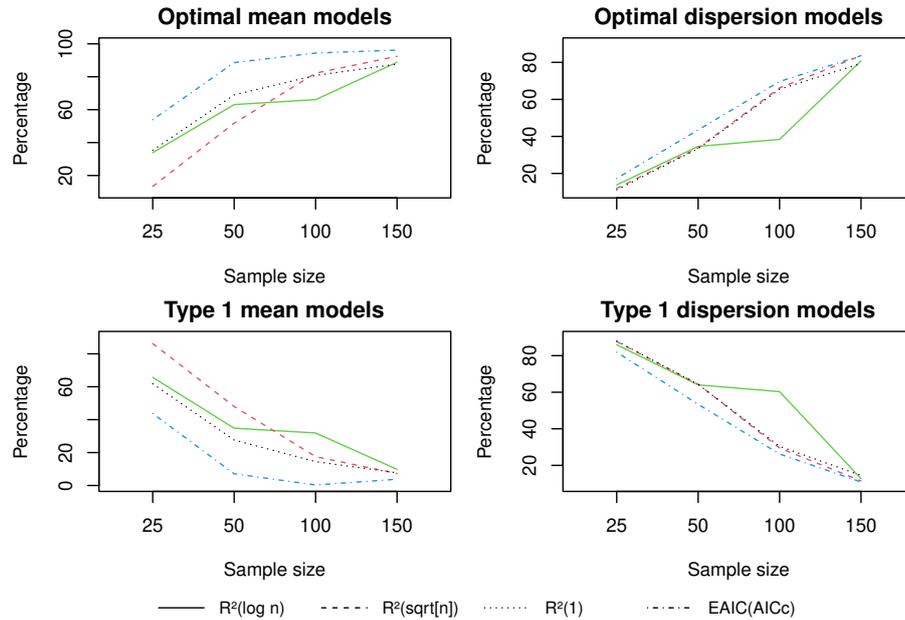


Figura 4.2: Gráficos, obtidos nas simulações da distribuição Binomial, da relação entre tamanho da amostra e porcentagem encontrada para os modelos Ótimo e Tipo 1, para os critérios \tilde{R}_m^2 , \tilde{R}_d^2 , $EAIC$ e AIC_c , onde a linha sólida representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2) com $\lambda_n = \log n$, a linha tracejada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = \sqrt{n}$, a linha pontilhada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = 1$ e a linha tracejada e pontilhada representa o gráfico para $EAIC(AIC_c)$.

dos modelos Tipo 1, indicando o melhor resultado para o critério $EAIC$ (AIC_c) para pequenas amostras; e a aproximação das linhas em tamanho da amostra igual a 150, indicando que os critérios convergem para similaridade.

4.3.3 Estudo de simulação para a distribuição de Poisson

Para o modelo da média com distribuição de probabilidade Poisson ou tipo Poisson, ou seja, com $V(\mu_i) = \mu_i$, o processo de simulação de dados deve garantir que $E(Y_i) = \mu_i$ e $Var(Y_i) = \phi_i \mu_i$. A geração de dados foi adaptada de McCullagh e Nelder (1989), p. 198. Considere a variável aleatória $Y|N = S_1 + \dots + S_N$, onde S_j são variáveis aleatórias independentes, distribuídas identicamente e N é uma variável aleatória com distribuição de Poisson, independente de S_j , $\forall j = 1, \dots, N$. Desta forma, podemos obter $E(Y_i) = E(N_i)E(S_i)$ e $Var(Y_i) = E(N_i)E(S_i^2)$. Observe que haverá sobredispersão se $Var(Y_i) > E(Y_i)$, ou seja, se $E(S_i^2) > E(S_i)$. Assim, para obter o resultado desejado, basta tomar $N_i \sim Poisson(\mu_i/\rho_i)$ e $S_i \sim Poisson(\rho_i)$ com $\rho_i = \phi_i - 1 > 0$.

Os valores escolhidos para o vetor de parâmetros $(\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3)$, usados para realizar simulações, foram $(1, 5; 3; 2; 0; 0, 2; 0; 3; 0)$, onde o valor zero significa que o parâmetro

não foi utilizado na simulação. Para cada simulação de Monte Carlo, as respostas foram geradas considerando a função de ligação logarítmica para os modelos da média e da dispersão, de acordo com o Apêndice A.3.

Tabela 4.3: Porcentagens dos tipos de modelos encontrados em 1000 simulações de Monte Carlo do procedimento de seleção de variáveis proposto para a distribuição de Poisson

Modelo	Critério	Ótimo				Tipo 2				Tipo 1			
		Tamanho de amostra				Tamanho de amostra				Tamanho de amostra			
		25	50	100	150	25	50	100	150	25	50	100	150
Média	$\tilde{R}_m^2(\log n)$	88,3	90,5	91,1	92,4	11,3	9,3	8,8	7,5	0,4	0,2	0,1	0,1
	$\tilde{R}_m^2(\sqrt{n})$	92,6	92,6	93,3	95,3	7,00	6,8	6,5	4,6	0,4	0,6	0,2	0,1
	$\tilde{R}_m^2(1)$	89,5	90,7	91,3	90,8	10,4	9,2	8,7	9,2	0,1	0,1	0,1	0,0
	$E A I C$	85,9	91,8	91,3	92,4	6,7	6,6	8,6	7,5	7,4	1,6	0,1	0,1
Dispersão	$\tilde{R}_d^2(\log n)$	10,6	27,3	47,3	58,3	0,4	3,4	9,5	13,4	89,0	69,3	43,2	28,3
	$\tilde{R}_d^2(\sqrt{n})$	13,5	27,2	51,6	59,8	0,3	2,3	7,7	13,1	86,2	70,5	40,7	27,1
	$\tilde{R}_d^2(1)$	11,2	29,7	46,0	57,6	1,0	3,2	9,2	15,6	87,8	67,1	44,8	28,8
	$A I C_c$	51,7	64,9	78,8	84,0	2,4	3,1	3,1	4,7	51,7	32,0	18,1	11,3

A tabela 4.3 mostra que, para os modelos da média, o procedimento de seleção de variáveis continua eficiente para a distribuição de Poisson e temos até maior eficiência para amostras menores em relação às distribuições anteriores. O critério \tilde{R}_m^2 com $\lambda_n = \sqrt{n}$ teve uma ligeira vantagem sobre os outros critérios.

Para os modelos da dispersão, o critério $A I C_c$ foi melhor, com uma alta taxa de modelos ótimos para grandes amostras. O critério \tilde{R}_d^2 teve uma melhora na taxa ótima dos modelos com o aumento do tamanho da amostra, chegando perto de 60%. Observe também que os modelos do Tipo 1 mostraram taxas altas para amostras pequenas; no entanto, houve uma diminuição acelerada dessa taxa quando o tamanho da amostra foi aumentado.

Nos gráficos exibidos na Figura 4.3, podemos observar linhas que estão acima das demais, ou seja, para os modelos da média observamos a vantagem do critério \tilde{R}_m^2 com $\lambda_n = \sqrt{n}$ (linha tracejada) e para os modelos da dispersão observamos a vantagem do critério $A I C_c$ (linha com traço-ponto).

Para modelos do tipo 1, o critério $A I C_c$ foi melhor para modelos da dispersão, com a linha traço-ponto sempre abaixo das demais. Para os modelos da média, o critério \tilde{R}_m^2 , para qualquer

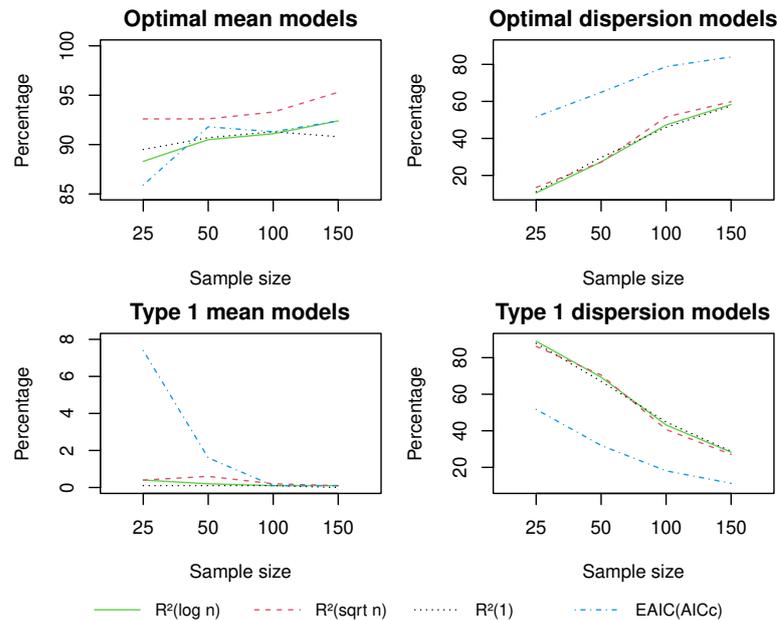


Figura 4.3: Gráficos, obtidos nas simulações da distribuição Poisson, da relação entre tamanho da amostra e porcentagem encontrada para os modelos Ótimo e Tipo 1, para os critérios \tilde{R}_m^2 , \tilde{R}_d^2 , $EAIC$ e AIC_c , onde a linha sólida representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2) com $\lambda_n = \log n$, a linha tracejada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = \sqrt{n}$, a linha pontilhada representa o gráfico para \tilde{R}_m^2 (\tilde{R}_d^2), com $\lambda_n = 1$ e a linha tracejada e pontilhada representa o gráfico para $EAIC(AIC_c)$.

valor de λ_n , foi melhor do que $EAIC$ para pequenas amostras, no entanto, os critérios tornam-se equivalentes à medida que o tamanho da amostra aumenta.

4.4 Aplicação

O procedimento proposto de seleção de variáveis foi aplicado a dados de um experimento de moldagem por injeção. O experimento foi realizado para estudar a influência de sete fatores controláveis e três fatores de ruído sobre a média e a variação na porcentagem de encolhimento de produtos feitos por moldagem por injeção.

O objetivo do experimento foi determinar as configurações dos parâmetros do processo de forma que a porcentagem de encolhimento ficasse próxima do valor alvo e fosse robusta contra variações ambientais. As respostas foram porcentagens de encolhimento de produtos feitos por moldagem por injeção, o planejamento experimental para sete fatores controláveis (A = tempo do ciclo, B = temperatura do molde, C = espessura da cavidade, D = pressão de retenção, E = velocidade de injeção, F = tempo de retenção e G = tamanho do porta) foi um fatorial

fracionário 2^{7-4} . Em cada configuração dos fatores controláveis, quatro observações foram obtidas a partir de um fatorial fracionário 2^{3-1} com três fatores de ruído (M = porcentagem de moagem, N = teor de umidade e O = temperatura ambiente). Os dados experimentais da moldagem por injeção são apresentados na Tabela 4.4. Engel e Huele (1996) fornecem uma boa discussão sobre a origem desses dados.

O conjunto de dados considerado é bastante conhecido na literatura de experimentos industriais e foi analisado por diversos autores como Engel (1992), Engel e Huele (1996) e Lee e Nelder (1998). Embora os artigos de Engel e Huele (1996) e Lee e Nelder (1998) tenham usado a MCMD, o processo de seleção de variáveis não foi discutido. Em nossa abordagem, além de obter estimativas dos parâmetros dos modelos da média e da dispersão por meio da MCMD, também construímos um procedimento investigativo consistente para verificar a adequação do modelo conjunto final encontrado. Para a aplicação do Algoritmo 4.2, consideramos $V(\mu) = 1$ e função de ligação identidade para o modelo da média. Para o modelo de dispersão consideramos a distribuição Gama com função de ligação logarítmica. Os modelos conjuntos que consideramos para média e para a dispersão foram os mesmos propostos por Engel e Huele (1996) e Lee e Nelder (1998).

A medida de seleção considerada no Algoritmo 4.1 para o modelo da média foi o \tilde{R}_m^2 com $\lambda_n = \sqrt{n}$ e para o modelo da dispersão a medida de seleção foi o AIC_c . Conforme examinado na Seção 4.3, o nível de significância para os testes de hipóteses realizados foi de 0,10.

A Tabela 4.6 apresenta os resultados de todas as etapas do processo de seleção de variáveis (Algoritmo 4.2) tomadas para obter o modelo conjunto final para a média e para dispersão. Em cada iteração do Algoritmo 4.2, o Algoritmo 4.1 foi usado para encontrar os termos dos modelos da média e da dispersão. Os termos adicionais aos modelos atuais, que aparecem em sequência na Tabela 4.6, são aqueles que, para todos os termos não pertencentes ao modelo atual, possuem o maior valor de \tilde{R}_m^2 , no caso do modelo da média, e o menor valor de AIC_c , no caso do modelo da dispersão. Para estes termos, candidatos a entrar no modelo atual, foram utilizados os testes de hipótese Qui-quadrado, no caso do modelo da dispersão, e o teste F , no caso do modelo da média, para verificar se, de fato, o termo escolhido deve ser adicionado ao modelo atual. As medidas D e D^* foram usadas como estatísticas de teste. Os valores de D , D^* , χ^2 , F e \tilde{R}_m^2 , com $\lambda_n = \sqrt{n}$ e com $\lambda_n = 1$, são mostrados para cada etapa do Algoritmo 4.1, onde D é o desvio para o modelo Gama, $D^* = \sum_{i=1}^n \frac{1}{\phi_i} d_i^*$ (ver Algoritmo 4.1), χ^2 , a estatística de teste para o teste de Qui-quadrado no modelo de dispersão, é a diferença de desvio para

modelos encaixados Gama e F é a estatística de teste, dada pela equação (4.2.6), para o teste F no modelo da média. Vale ressaltar que a medida \tilde{R}_m^2 com $\lambda_n = 1$, que não foi utilizada como medida de seleção, foi apresentada na Tabela 4.6 apenas para mostrar o impacto no valor de \tilde{R}_m^2 , quando o número de termos no modelo aumenta, considerando a penalidade com $\lambda_n = \sqrt{n}$ e com $\lambda_n = 1$. Em cada iteração, os modelos finais da média e da dispersão são mostrados em negrito.

Seguindo o Algoritmo 4.2, como o valor de \tilde{R}_m^2 para o modelo da média, obtido na terceira iteração, foi inferior ao obtido na segunda iteração, o procedimento foi interrompido. Desta forma, foram realizadas três iterações do Algoritmo 4.2, obtendo ao final, na segunda iteração, um modelo para a média com $\tilde{R}_m^2(\sqrt{n})$ igual a 0,803 e $\tilde{R}_m^2(1)$ igual a 0,973.

Todos os modelos da média encontrados nas três iterações do Algoritmo 4.2 tiveram os mesmos termos (ver Tabela 4.6), diferindo apenas no valor das estimativas dos efeitos. A tabela 4.5 apresenta as estimativas finais, com seus desvios-padrão e testes de Wald. Observe que todas as estimativas foram significativas ao nível de 1%, exceto os termos E , N e C no modelo da média, que entraram no modelo somente após o término do procedimento de seleção de variáveis, devido aos termos CN e EN terem sido significativos.

Os gráficos de diagnósticos para o modelo da média, apresentados na Figura 4.4, mostram que o modelo obtido é adequado, pois não apresentou evidências em violar as premissas de independência, variância constante e normalidade dos resíduos, também não há evidência de inadequação das funções de ligação e variância ou pontos de influência.

Para a dispersão, os modelos obtidos em cada iteração do Algoritmo 4.2 foram diferentes. Isso se deve ao fato de que a variável de resposta, d^* , não é a mesma em cada iteração do Algoritmo 4.2. O modelo final da dispersão teve um valor de AIC_c igual a $-74,2471$ (ver Tabela 4.6) e mostrou efeitos altamente significativos pelo teste de Wald (ver Tabela 4.5). Os gráficos de diagnóstico para o modelo de dispersão são apresentados na Figura 4.5.

Tabela 4.4: Dados experimentais de moldagem por injeção

ensaio	Fatores de controle							Fatores de ruído			
	A	B	C	D	E	F	G	M	N	O	
1	-1	-1	-1	-1	-1	-1	-1	2,2	2,1	2,3	2,3
2	-1	-1	-1	1	1	1	1	2,5	0,3	2,7	0,3
3	-1	1	1	-1	-1	1	1	0,5	3,1	0,4	2,8
4	-1	1	1	1	1	-1	-1	2,0	1,9	1,8	2,0
5	1	-1	1	-1	1	-1	1	3,0	3,1	3,0	3,0
6	1	-1	1	1	-1	1	-1	2,1	4,2	1,0	3,1
7	1	1	-1	-1	1	1	-1	4,0	1,9	4,6	2,2
8	1	1	-1	1	-1	-1	1	2,0	1,9	1,9	1,8

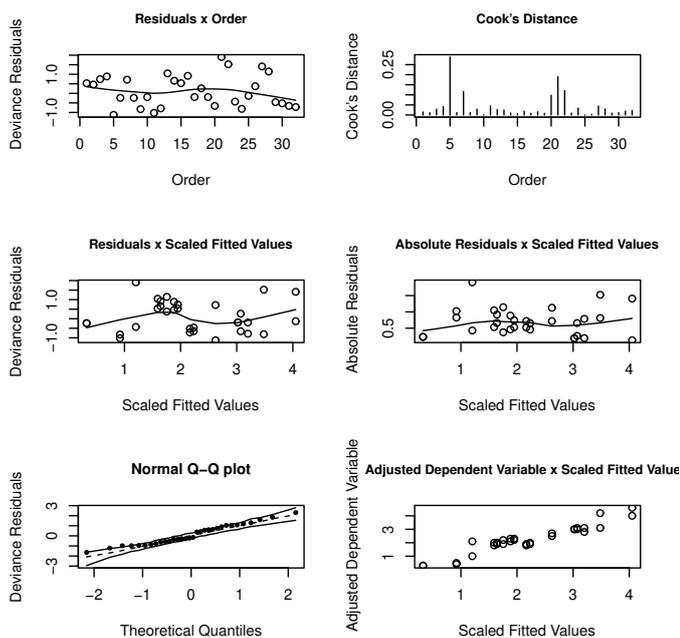


Figura 4.4: Gráficos de diagnóstico para o modelo da média obtido dos dados de moldagem por injeção.

4.5 Considerações finais

O capítulo apresentou um procedimento para seleção de variáveis na modelagem conjunta da média e dispersão com aplicação em experimentos de projeto robusto. O procedimento proposto fornece um mecanismo iterativo para seleção de variáveis na MCMD. O processo

Tabela 4.5: Coeficientes de regressão e teste de Wald para os modelos da média e dispersão obtidos dos dados de moldagem por injeção.

Modelo da média				
Termos	Estimativa	Erro padrão	Valor t	$Pr(> t)$
1	2,24903	0,03322	67,693	0,0000
A	0,42802	0,06575	6,509	0,0000
C	0,07172	0,05802	1,236	0,2279
N	-0,00433	0,05994	-0,072	0,9430
D	-0,28639	0,06407	-4,470	0,0001
E	0,06528	0,05971	1,093	0,2851
CN	0,58684	0,03322	17,663	0,0000
EN	-0,55727	0,05994	-9,297	0,0000
Modelo da dispersão				
Termos	Estimativa	Erro padrão	Valor t	$Pr(> t)$
1	-2,2973	0,1754	-13,097	0,0000
E	-0,8670	0,1754	-4,942	0,0000
B	0,6773	0,1754	3,861	0,0006
G	-0,6015	0,1754	-3,429	0,0019

aproveita a estrutura de gangorra (vai e vem) do algoritmo para estimar os parâmetros na MCMD, onde os pesos do modelo de dispersão são incorporados ao modelo da média. Para os modelos da média e da dispersão, uma medida de seleção é utilizada para escolher, em cada iteração do procedimento, as variáveis que contribuem para o melhor ajuste do modelo. Testes de hipóteses apropriados são usados para confirmar a escolha dessas variáveis ou para encerrar o procedimento. À medida que os pesos do modelo da dispersão são incorporados ao modelo da média, o ajuste do modelo da média torna-se melhor. O modelo conjunto final obtido é aquele que melhor se ajusta ao modelo da média. Dessa forma, o modelo da dispersão é constituído pelos termos que fornecem os melhores pesos para o modelo da média.

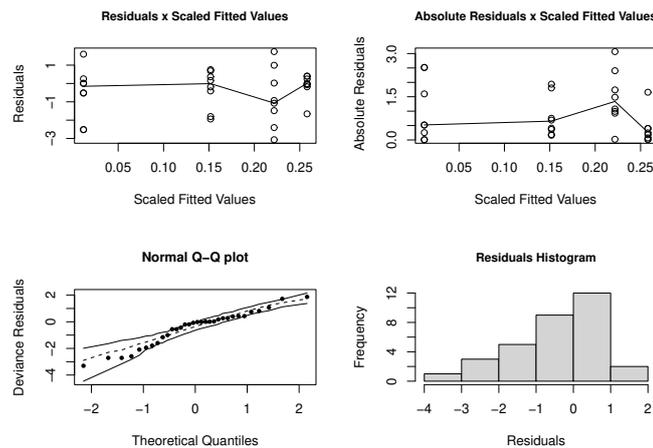


Figura 4.5: Gráficos de diagnóstico para o modelo da dispersão obtido dos dados de moldagem por injeção.

Para cada um dos modelos da média e da dispersão, duas medidas alternativas de seleção foram utilizadas nos estudos de simulação. As medidas de seleção \tilde{R}_m^2 e $EAIC$ foram usadas no modelo da média e \tilde{R}_d^2 e AIC_c foram usadas no modelo da dispersão. Cada um dos critérios \tilde{R}_m^2 e \tilde{R}_d^2 foram usados considerando as variações baseadas nas escolhas de λ_n igual a \sqrt{n} , $\log n$ e 1.

Nos estudos de simulação, realizados para as distribuições Normal, Poisson e Binomial, o critério \tilde{R}_m^2 com $\lambda_n = \sqrt{n}$ foi melhor para os casos Normal e Poisson, enquanto o $EAIC$ foi melhor para o caso Binomial. O AIC_c foi melhor em todos os casos considerados.

A eficiência do procedimento proposto depende do critério utilizado como medida de seleção e do tamanho da amostra. Como era de se esperar, a eficiência cresce com o aumento do tamanho da amostra. As simulações mostraram que, para uma amostra grande, o procedimento proposto reduz consideravelmente o risco de se encontrar modelos ruins e garante a obtenção de modelos satisfatórios para média e para dispersão. O procedimento também mostrou-se razoavelmente bom para pequenas amostras.

Na aplicação, discutida na Seção 4.4, o procedimento de seleção das variáveis foi explicado passo a passo e os resultados obtidos para os modelos da média e da dispersão foram bastante satisfatórios, indicando a eficácia da metodologia proposta e corroborando os resultados encontrados pelas simulações.

No próximo capítulo a modelagem conjunta da média e dispersão será aplicada em experimentos envolvendo misturas. Entretanto, neste caso, o procedimento de seleção de variáveis deverá ser adaptado para satisfazer a restrição de mistura.

Tabela 4.6: Passos do algoritmo de seleção de variáveis para a MCMD aplicado aos dados do experimento de moldagem por injeção.

Iteração	Modelo da dispersão	AIC_c	D^\ddagger	Valor χ_c^2 *	$Pr(> \chi_c^2)$	
1	1	-	-	-	-	
	Modelo da média	$\tilde{R}_m^2(\sqrt{n})$	$\tilde{R}_m^2(1)$	D^* ‡	Valor F_c ° $Pr(> F_c)$	
	1	-	-	34,6839	-	-
	1+CN	-0,0060	0,3063	24,0587	13,2491	0,0010
	1+CN+EN	0,2232	0,5974	13,9628	20,9687	0,0001
	1+CN+EN+A	0,3234	0,7735	7,8557	21,7672	0,0001
	1+CN+EN+A+D	-0,0782	0,8516	5,1467	14,2120	0,0008
Iteração	Modelo da dispersão	AIC_c	D^\ddagger	Valor χ_c^2 *	$Pr(> \chi_c^2)$	
2	1	-	80,8899	-	-	
	1+E	-64,1778	72,0353	4,4273	0,0354	
	1+E+B	-68,4785	61,7544	5,1404	0,0234	
	1+E+B+G	-74,2471	50,8136	5,4704	0,0193	
	1+E+B+G+D	-77,5280	44,3609	3,2263	0,0725	
	Modelo da média	$\tilde{R}_m^2(\sqrt{n})$	$\tilde{R}_m^2(1)$	D^* ‡	Valor F_c ° $Pr(> F_c)$	
	1	-	-	1098,102	-	-
	1+A	0,133	0,402	699,342	17,106	0,0003
	1+A+CN	0,378	0,678	404,023	21,197	0,0001
	1+A+CN+EN	0,841	0,947	52,454	187,666	0,0000
	1+A+CN+EN+D	0,803	0,973	26,624	26,195	0,0000
Iteração	Modelo da dispersão	AIC_c	D^\ddagger	Valor χ_c^2 *	$Pr(> \chi_c^2)$	
3	1	-	75,4049	-	-	
	1+D	42,5318	65,4232	4,9908	0,0255	
	1+D+F	41,9160	61,3397	2,0417	0,1530	
	Modelo da média	$\tilde{R}_m^2(\sqrt{n})$	$\tilde{R}_m^2(1)$	D^* ‡	Valor F_c ° $Pr(> F_c)$	
	1	-	-	59,6112	-	-
	1+CN	-0,0078	0,3051	41,8419	12,7404	0,0012
	1+CN+EN	0,2266	0,5992	24,3629	20,8059	0,0001
	1+CN+EN+A+D	0,2260	0,8935	6,2942	13,4942	0,0010

‡ D é o desvio para o modelo Gama e $D^* = \sum_{i=1}^n \frac{1}{\phi_i} d_i^*$.* χ_c^2 é o valor da estatística Qui-quadrado, calculado a partir da diferença dos desvios para dois modelos Gama encaixados° F_c é o valor da estatística de teste F , calculado pela equação (4.2.6).

Capítulo 5

MCMD aplicada a experimentos com misturas

5.1 Introdução

Experimentos com misturas envolvem mistura ou combinação de dois ou mais ingredientes para formar um produto final. Para este tipo de experimento é de interesse determinar as proporções dos componentes da mistura que levam a resultados desejáveis no que diz respeito a alguma característica de qualidade de interesse. Veja Seção 2.3.1 para maiores detalhes.

Modelos de regressão clássicos são geralmente usados em experimentos de mistura, os quais assumem as suposições de normalidade dos erros, variância constante e linearidade dos efeitos sistemáticos. No entanto, podem surgir situações em que tais suposições não possam ser totalmente satisfeitas. De acordo com Nelder e Lee (1991), para muitos conjuntos de dados as suposições subjacentes aos modelos lineares clássicos são insatisfatórias e as transformações da variável resposta não podem necessariamente satisfazer todas as suposições necessárias aos modelos de regressão clássicos, especialmente normalidade.

Modelos lineares generalizados permitem a análise de dados para os quais as suposições do modelo de regressão clássico não são satisfeitas e fornecem uma estrutura geral para modelagem da variabilidade por meio da modelagem conjunta de média e dispersão, veja Seção 2.3.2. A MCMD foi introduzida por Nelder e Lee (1991) como uma alternativa aos métodos de Taguchi em experimentos de melhoria de qualidade e fornece uma excelente metodologia para encontrar e verificar o ajuste dos modelos encontrados com uma sólida base estatística.

Nos modelos lineares generalizados, a variância da resposta é igual ao parâmetro de dispersão multiplicado por uma função da média, chamada de função de variância. Apenas no caso da distribuição normal, onde a função de variância é a função de identidade, o parâmetro de dispersão coincide com a variância da resposta.

Em experimentos de mistura, a modelagem de variância é considerada apenas em situações onde existem variáveis de ruído. Variáveis de ruído são variáveis de processo que não podem ser controladas, como umidade ou temperatura. Variáveis de processo são fatores em um experimento que não formam nenhuma parte da mistura, mas cujos níveis, quando alterados, podem afetar os resultados finais da mistura, veja Subseção 2.3.1.

A modelagem estatística da variância em experimentos de mistura com variáveis de ruído foi considerada por Steiner e Hamada (1997), que propuseram um modelo combinado de mistura com variáveis de processo e ruído, construíram e resolveram, um problema de otimização para minimizar uma função de perda quadrática, levando em consideração ambos os média e variância da resposta. Outra abordagem para modelar a variância em experimentos de mistura com variáveis de ruído é devida a Goldfarb, Borror e Montgomery (2003) usando o método delta, que emprega uma aproximação de série de Taylor de primeira ordem do modelo de regressão em um vetor de variáveis de ruído, veja Seção 3.2. Uma aplicação do Método Delta para modelar a variabilidade em experimentos de mistura com variáveis de ruído em engenharia de alimentos também é apresentada por Granato, Calado e Pinto (2020) e mostrada na Seção 3.3. No entanto, em todas as abordagens apresentadas para modelagem a variância, são assumidas as suposições de normalidade dos erros e parâmetro de dispersão constante. Desta forma, o modelo da média foi considerado normal homocedástico e a modelagem da variância se deu apenas em função das variáveis de ruído, presentes no modelo.

A abordagem que consideramos neste capítulo é bastante geral porque, ao considerar a classe de modelos lineares generalizados, distribuições diferentes da distribuição normal podem ser consideradas. Além disso, considerando um modelo para a dispersão, o processo de modelagem da variabilidade permite modelar a variância não apenas em situações onde existem variáveis de ruído.

O objetivo deste capítulo é mostrar como a modelagem conjunta de média e dispersão pode ser aplicada para modelar a variabilidade em experimentos de mistura. O Algoritmo 2.1 é o

mesmo para estimação de parâmetros no caso de experimentos com mistura, entretanto no processo de seleção de variáveis algumas adaptações deverão ser realizadas.

Na Seção 5.2, apresentamos o processo de seleção de variáveis para a MCMD adaptado para experimentos com misturas. Na Seção 5.3, aplicamos a MCMD ao problema de panificação, que envolveu três ingredientes de mistura e duas variáveis de processo/ruído, conforme apresentado na Seção 3.3. Considerações finais são dadas na Seção 5.4.

5.2 Processo de seleção de variáveis na MCMD com misturas

O processo de seleção de variáveis considerado para a MCMD em experimentos com misturas é baseado no procedimento proposto no Capítulo 4. O esquema de seleção proposto consiste em um método que visa reduzir os custos computacionais inerentes à grande complexidade envolvida na seleção de variáveis para modelos conjuntos de média e dispersão.

O procedimento de seleção de variáveis proposto no Capítulo 4 é baseado em testes de hipóteses e na qualidade do ajuste do modelo. No procedimento proposto no Capítulo 4, um critério de verificação da qualidade do ajuste é utilizado, a cada iteração do processo de seleção, como filtro para a escolha dos termos que serão avaliados por um teste de hipótese. A estratégia de seleção, proposta, apresentada no Algoritmo 4.2, consiste em um esquema de três etapas em que o modelo da dispersão selecionado é utilizado para selecionar o melhor modelo da média e vice-versa. O esquema de seleção utiliza um procedimento recursivo que só é finalizado quando o valor do critério, considerado como medida de seleção para o modelo da média, deixa de melhorar.

Nesta seção, o processo de seleção de variáveis, apresentado na Seção 4.2, é adaptado para o caso de experimentos envolvendo misturas. O Algoritmo 4.2 não sofrerá mudança, porém o Algoritmo 4.1 deverá ser adaptado, pois, devido à restrição de mistura (2.3.1), o termo inicial no processo de busca pelo modelo ótimo não necessariamente poderá ser considerado como constante.

No Algoritmo 4.1, o termo inicial no preditor linear é o termo constante, porém no caso de experimentos envolvendo misturas, na maioria dos casos, isso pode não ser verdade. Para verificação do termo de inicialização no preditor linear, deveremos realizar um teste de hipótese para saber se o termo constante existe ou não no preditor linear. Assim,

precisaremos verificar se os parâmetros relacionados aos termos lineares (efeitos principais) são iguais; Se forem iguais, significa que o modelo inicial deverá conter somente o termo constante; caso contrário, o modelo inicial não terá o termo constante e deverá conter todos os termos de efeitos principais de mistura. Assim, se tivermos a variáveis de mistura dadas por, x_1, \dots, x_a , o preditor linear contendo os efeitos principais será $\eta = \beta_1 x_1 + \dots + \beta_a x_a$; se $\beta_j = \beta_0$ para todo $j = 1, \dots, a$, então $\eta = \beta_0 (\sum_{j=1}^a x_j) = \beta_0$. Desta forma, se aceitarmos que $\beta_j = \beta_0$ para todo $j = 1, \dots, a$, então o preditor linear inicial será $\eta = \beta_0$ e não haverá termos envolvendo os efeitos principais x_1, \dots, x_a . Se rejeitarmos que $\beta_j = \beta_0$ para algum j , então o preditor linear inicial conterá todos os efeitos principais, ou seja, $\eta = \beta_1 x_1 + \dots + \beta_a x_a$, veja Marquardt e Snee (1974) para maiores detalhes.

O teste de hipóteses para verificação da existência de termo constante no preditor linear será realizado como segue. Queremos testar se os parâmetros relacionados aos efeitos principais, x_1, \dots, x_a são iguais ou diferentes de um valor constante, digamos, β_0 ; para isso, consideramos as hipóteses $H_0 : \beta_i = \beta_0$ para todo $i = 1, \dots, a$, versus $H_1 : \beta_i \neq \beta_0$ para algum i , com $i = 1, \dots, a$. Observe que sob H_0 , o preditor η_0 , considerando somente os efeitos principais, pode ser escrito como $\eta_0 = \beta_0$; sob H_1 , o preditor linear é dado por $\eta_1 = \sum_{i=1}^a \beta_i x_i$. Note ainda que, devido à restrição de mistura 2.3.1 e considerando, por exemplo, x_a como variável de folga (veja Seção 2.3.1), o preditor linear η_1 pode ser escrito como $\eta_1^* = \beta_a + (\beta_1 - \beta_a)x_1 + \dots + (\beta_{(a-1)} - \beta_a)x_{(a-1)} = \beta_0^* + \beta_1^* x_1 + \dots + \beta_{(a-1)}^* x_{(a-1)}$. Assim, $\eta_0 \subset \eta_1^*$, onde, η_1^* corresponde ao preditor η_0 acrescido dos termos $x_1, \dots, x_{(a-1)}$. Consequentemente, como agora temos dois modelos encaixados, os testes de hipóteses utilizados no Algoritmo 4.1 podem ser aplicados. Note que a escolha da variável de folga é arbitrária, pois não afeta o funcionamento do teste de hipótese. O Algoritmo 5.1, que faz a adaptação do Algoritmo 4.1 para o caso de experimentos com misturas, é mostrado a seguir.

Algoritmo 5.1: Pseudo algoritmo para seleção de variáveis no modelo de mistura da média (dispersão)

Na MCMD, para um determinado modelo da dispersão (média) fixo, escolha os termos do modelo da média (dispersão), considerando a seguinte sequência de passos.

1. Sejam x_1, \dots, x_a os termos referentes aos componentes da mistura, satisfazendo as restrições da mistura, de acordo com a equação (2.3.1), e considere um possível modelo para inicialização com preditor linear $\vartheta_0 = \theta_1 x_1 + \dots + \theta_a x_a$. Faça um teste de hipótese

para verificar se $H_0 : \theta_i = \theta_0, \forall i = 1, \dots, a$. Se H_0 for aceita, faça o preditor linear inicial $\vartheta = \theta_0$, caso contrário, $\vartheta = \vartheta_0$.

2. Seja \mathcal{V} o conjunto de termos que serão utilizados no processo de seleção de variáveis e, de acordo com o passo 1, considere ϑ o preditor linear do modelo inicial. Note que os termos no conjunto \mathcal{V} também dependerão do teste realizado no passo 1. Se H_0 tiver sido aceita, o conjunto \mathcal{V} não poderá conter os termos x_1, \dots, x_a . Caso contrário, \mathcal{V} não deverá conter o termo constante.
3. Ajuste os modelos com preditores lineares $\vartheta_j = \vartheta + \theta_j v_j$, para todo $v_j \in \mathcal{V}$.
4. Para cada modelo ajustado no passo 3, calcule a medida de seleção apropriada para verificar a qualidade do ajuste e escolha o modelo com preditor linear, digamos $\vartheta_k = \vartheta + \theta_k v_k$, que tenha o melhor ajuste. Avalie o valor da medida de seleção encontrada na iteração atual com o valor da medida de seleção encontrada na iteração anterior. Se o valor da medida de seleção atual for melhor do que o valor anterior, vá para o passo 5, caso contrário, vá para o passo 6.
5. Para aqueles modelos com preditores lineares encaixados ϑ e ϑ_k , aplique um teste apropriado para verificar a significância da adição de v_k no preditor linear ϑ . Se v_k for significativo, remova o termo v_k de \mathcal{V} , faça $\vartheta = \vartheta_k$ e volte para o passo 3. Caso contrário, o procedimento termina e ϑ é o modelo final escolhido.
6. Avalie, da mesma forma como no passo 5, a significância da adição de v_k ao preditor linear ϑ . Se v_k é significativo, faça $\vartheta = \vartheta_k$. Pare o procedimento e ϑ é o modelo final escolhido.

5.3 Aplicação ao problema de panificação

Nesta seção, a modelagem conjunta da média e dispersão será aplicada ao problema da panificação apresentado na Seção 3.3. O problema consistiu de um experimento com três ingredientes de mistura e duas variáveis de ruído, e teve como objetivo investigar e avaliar a qualidade final da farinha, composta por diferentes misturas de farinha de trigo, para produção de pães.

Os ingredientes de mistura consistiram de três tipos de farinha de trigo: duas noruegueses, *Tjalve* (x_1) e *Folke* (x_2) e uma americana, *Hard Red Spring* (x_3), que foram

consideradas como variáveis de controle, e dois tipos de variáveis de processo: tempo de mistura (z_1) e tempo de descanso da massa (z_2), consideradas como variáveis de ruído. A variável resposta foi considerada como o volume do pão após assado, com valor-alvo de 530 ml. As restrições nos ingredientes de mistura foram $0,25 \leq x_1 \leq 1,0$; $0 \leq x_2 \leq 0,75$ e $0 \leq x_3 \leq 0,75$. Para as variáveis de ruído, foram consideradas três situações para o tempo de mistura: 5, 15 e 25 minutos e três situações para o tempo de descanso da massa: 35, 47,5 e 60 minutos. O desenho experimental é mostrado na Figura 3.1 e o conjunto de dados na Tabela 3.1.

Para a aplicação do processo de seleção de variáveis, consideramos $V(\mu) = 1$ e função de ligação identidade para o modelo da média. Para o modelo de dispersão consideramos a distribuição Gama com função de ligação logarítmica. Os termos considerados em ambos os modelos da média e da dispersão consistiram dos termos do modelo cúbico nos componentes de mistura cruzado com o modelo quadrático nas variáveis de ruído.

Da mesma forma como fizemos na Seção 4.4, a medida de seleção considerada no Algoritmo 5.1 para o modelo da média foi o \tilde{R}_m^2 com $\lambda_n = \sqrt{n}$ e para o modelo da dispersão a medida de seleção foi o AIC_c . A Tabela 5.1 apresenta os resultados de todas as etapas do processo de seleção de variáveis (Algoritmo 4.2) tomadas para obter o modelo conjunto final para a média e para dispersão. O teste inicial, apresentado no início de cada iteração, tanto para o modelo da média quanto para o modelo da dispersão, diz respeito ao teste de hipótese para verificar se o modelo deveria ter o termo constante ou não; desta forma, a hipótese nula é que os parâmetros dos termos dos efeitos principais são iguais a um valor constante.

Em cada iteração do Algoritmo 4.2, o Algoritmo 5.1 foi usado para encontrar os termos dos modelos da média e da dispersão. Os termos adicionais aos modelos atuais, que aparecem em sequência na Tabela 5.1, são aqueles que, para todos os termos não pertencentes ao modelo atual, possuem o maior valor de \tilde{R}_m^2 , no caso do modelo da média, e o menor valor de AIC_c , no caso do modelo da dispersão. Para estes termos, candidatos a entrar no modelo atual, foram utilizados os testes de hipótese Qui-quadrado, no caso do modelo da dispersão, e o teste F , no caso do modelo da média, para verificar se, de fato, o termo escolhido deve ser adicionado ao modelo atual. As medidas D e D^* foram usadas como estatísticas de teste. Os valores de D , D^* , χ^2 , F e \tilde{R}_m^2 , com $\lambda_n = \sqrt{n}$ e com $\lambda_n = 1$, são mostrados para cada etapa do Algoritmo 5.1, onde D é o desvio para o modelo Gama, $D^* = \sum_{i=1}^n \frac{1}{\phi_i} d_i^*$ (ver Algoritmo 5.1), χ^2 , a estatística de teste para o teste de Qui-quadrado no modelo da dispersão, é a diferença de desvio para

modelos encaixados Gama e F é a estatística de teste, dada pela equação (4.2.6), para o teste F no modelo da média. Vale ressaltar que a medida \tilde{R}_m^2 com $\lambda_n = 1$, que não foi utilizada como medida de seleção, foi apresentada na Tabela 5.1 apenas para mostrar o impacto no valor de \tilde{R}_m^2 , quando o número de termos no modelo aumenta, considerando a penalidade com $\lambda_n = \sqrt{n}$ e com $\lambda_n = 1$. Em cada iteração, os modelos finais da média e da dispersão são mostrados em negrito.

Seguindo o Algoritmo 4.2, como o valor de \tilde{R}_m^2 para o modelo da média, obtido na terceira iteração, foi inferior ao obtido na segunda iteração, o procedimento foi interrompido. Desta forma, foram realizadas três iterações do Algoritmo 4.2, obtendo ao final, na segunda iteração, um modelo para a média com $\tilde{R}_m^2(\sqrt{n})$ igual a 0,9913 e $\tilde{R}_m^2(1)$ igual a 0,9975.

Todos os modelos da média encontrados nas três iterações do Algoritmo 4.2 tiveram os mesmos termos (ver Tabela 5.1), diferindo apenas no valor das estimativas dos efeitos. A tabela 5.2 apresenta as estimativas finais, com seus desvios-padrão e testes de Wald. Observe que todas as estimativas foram significativas ao nível de 1%, exceto o termo x_2x_3 , no modelo da dispersão, que foi significativo ao nível de 5%.

Os gráficos de diagnósticos para o modelo da média, apresentados na Figura 5.1, mostram que o modelo obtido é adequado, pois não apresentou evidências em violar as premissas de independência, variância constante e normalidade dos resíduos, também não há evidência de inadequação das funções de ligação e variância ou pontos de influência.

Para a dispersão, os modelos obtidos em cada iteração do Algoritmo 4.2 foram diferentes. Isso se deve ao fato de que a variável de resposta, d^* , não ser a mesma em cada iteração do Algoritmo 4.2. O modelo final da dispersão teve um valor de AIC_c igual a 1321,135 (ver Tabela 5.1) e mostrou efeitos significativos pelo teste de Wald (ver Tabela 5.2). Os gráficos de diagnóstico para o modelo de dispersão são apresentados na Figura 5.2.

Os modelos da média e da variância incondicionais às variáveis de ruído podem ser obtidos a partir dos modelos condicionais da média e da dispersão, veja Seção 2.3.2. Os modelos para $\mu = E(Y|Z_1, Z_2)$ e $\log(\phi)$, cujas estimativas dos parâmetros são exibidas na Tabela 5.2, são dados pelas equações (5.3.1) e (5.3.2), respectivamente.

$$E(Y|Z_1, Z_2) = 488,961x_1 + 432,21x_2 + 574,124x_3 + 174,216x_1x_3Z_1 + (56,621x_1 + 35,904x_2 + 79,146x_3)Z_2 \quad (5.3.1)$$

$$\log(\phi) = 6,9984x_1 + 5,94x_2 + 7,325x_3 - 7,9662x_2x_3 \quad (5.3.2)$$

Desta forma, considerando Z_1 e Z_2 variáveis aleatórias independentes com $E(Z_i) = \mu_i$ e $Var(Z_i) = \sigma_i^2$, para $i = 1, 2$, o modelo da média incondicional às variáveis de ruído é obtido como $E(Y) = E[E(Y|Z_1, Z_2)]$ e mostrado na equação (5.3.3).

$$\begin{aligned} E(Y) = & 488,961x_1 + 432,21x_2 + 574,124x_3 + 174,216x_1x_3\mu_1 + \\ & (56,621x_1 + 35,904x_2 + 79,146x_3)\mu_2. \end{aligned} \quad (5.3.3)$$

O modelo da variância incondicional às variáveis de ruído é obtido como $Var(Y) = E[Var(Y|Z_1, Z_2)] + Var[E(Y|Z_1, Z_2)]$. Como $Var(Y|Z_1, Z_2) = \phi V(\mu)$ e como supomos que $V(\mu) = 1$ e conhecemos o modelo para $\log(\phi)$, mostrado na equação (5.3.2), temos que $Var(Y|Z_1, Z_2) = \exp\{6,9984x_1 + 5,94x_2 + 7,325x_3 - 7,9662x_2x_3\}$. Desta forma, como $Var[E(Y|Z_1, Z_2)] = (174,216x_1x_3)^2\sigma_1^2 + (56,621x_1 + 35,904x_2 + 79,146x_3)^2\sigma_2^2$ e $E[Var(Y|Z_1, Z_2)] = \exp\{6,9984x_1 + 5,94x_2 + 7,325x_3 - 7,9662x_2x_3\}$ obtemos que

$$\begin{aligned} Var(Y) = & \exp\{6,9984x_1 + 5,94x_2 + 7,325x_3 - 7,9662x_2x_3\} + \\ & (174,216x_1x_3)^2\sigma_1^2 + (56,621x_1 + 35,904x_2 + 79,146x_3)^2\sigma_2^2. \end{aligned} \quad (5.3.4)$$

Observe que agora poderíamos realizar o mesmo processo de otimização exibido em (3.3.5), porém isso não será feito, visto que o procedimento já foi apresentado no Capítulo 3.

5.4 Considerações finais

Neste capítulo, a modelagem conjunta da média e da dispersão foi adaptada para ser utilizada em problemas envolvendo misturas. Uma aplicação em um problema de mistura com variáveis de ruído foi apresentada.

No problema da panificação, apresentado na Seção 3.3, o objetivo foi obter modelos para a média e variância da resposta, incondicionais às variáveis de ruído. No Capítulo 3, os modelos foram obtidos considerando uma aproximação pelo método delta; porém o método somente

Tabela 5.1: Passos do algoritmo de seleção de variáveis para a MCMD aplicado ao problema de panificação.

Iteração	Modelo da dispersão	AIC_c	D^{\ddagger}	Valor χ_c^2 *	$Pr(> \chi_c^2)$	
1	1	-	-	-	-	
	Modelo da média	$\tilde{R}_m^2(\sqrt{n})$	$\tilde{R}_m^2(1)$	$D^* \ddagger$	Valor F_c $^\circ$ $Pr(> F_c)$	
	Teste inicial	-	-	-	-	0,0000
	$x_1 + x_2 + x_3$	0,9810	0,9865	305938,54	-	-
	$x_1 + x_2 + x_3 + x_1 z_2$	0,9893	0,9935	148184,67	91,55	0,0000
	$x_1 + x_2 + x_3 + x_1 z_2 + x_3 z_2$	0,9901	0,9951	113114,00	26,35	0,0000
	$x_1 + x_2 + x_3 + x_1 z_2 + x_3 z_2 + x_1 x_3 z_1$	0,9911	0,9965	80267,00	34,37	0,0000
$x_1 + x_2 + x_3 + x_1 z_2 + x_3 z_2 + x_1 x_3 z_1 + x_2 z_2$	0,9888	0,9968	72773,67	8,55	0,0045	
Iteração	Modelo da dispersão	AIC_c	D^{\ddagger}	Valor χ_c^2 *	$Pr(> \chi_c^2)$	
2	Teste inicial	-	-	-	0,0330	
	$x_1 + x_2 + x_3$	1323,527	268,68	-	-	
	$x_1 + x_2 + x_3 + x_2 x_3$	1321,135	259,14	4,77	0,03	
	$x_1 + x_2 + x_3 + x_2 x_3 + x_1 x_3$	1321,544	255,76	1,69	0,19	
	Modelo da média	$\tilde{R}_m^2(\sqrt{n})$	$\tilde{R}_m^2(1)$	$D^* \ddagger$	Valor F_c $^\circ$ $Pr(> F_c)$	
	Teste inicial	-	-	-	-	0,0000
	$x_1 + x_2 + x_3$	0,9831	0,9880	436,08	-	-
	$x_1 + x_2 + x_3 + x_1 z_2$	0,9911	0,9946	197,54	103,85	0,0000
	$x_1 + x_2 + x_3 + x_1 z_2 + x_3 z_2$	0,9923	0,9962	139,80	35,11	0,0000
	$x_1 + x_2 + x_3 + x_1 z_2 + x_3 z_2 + x_1 x_3 z_1$	0,9927	0,9971	104,44	28,44	0,0000
$x_1 + x_2 + x_3 + x_1 z_2 + x_3 z_2 + x_1 x_3 z_1 + x_2 z_2$	0,9913	0,9975	90,16	13,15	0,0005	
Iteração	Modelo da dispersão	AIC_c	D^{\ddagger}	Valor χ_c^2 *	$Pr(> \chi_c^2)$	
3	Teste inicial	-	-	-	0,9946	
	1	128,2646	268,80	-	-	
	$1 + x_1 x_3$	130,0634	268,36	0,22	0,64	
	Modelo da média	$\tilde{R}_m^2(\sqrt{n})$	$\tilde{R}_m^2(1)$	$D^* \ddagger$	Valor F_c $^\circ$ $Pr(> F_c)$	
	Teste inicial	-	-	-	-	0,0000
	$x_1 + x_2 + x_3$	0,9810	0,9865	305402,98	-	-
	$x_1 + x_2 + x_3 + x_1 z_2$	0,9893	0,9935	147925,26	91,55	0,0000
	$x_1 + x_2 + x_3 + x_1 z_2 + x_3 z_2$	0,9901	0,9951	112915,99	26,35	0,0000
	$x_1 + x_2 + x_3 + x_1 z_2 + x_3 z_2 + x_1 x_3 z_1$	0,9911	0,9965	80126,49	34,37	0,0000
	$x_1 + x_2 + x_3 + x_1 z_2 + x_3 z_2 + x_1 x_3 z_1 + x_2 z_2$	0,9888	0,9968	72646,28	8,55	0,0045

$\ddagger D$ é o desvio para o modelo Gama e $D^* = \sum_{i=1}^n \frac{1}{\phi_i} d_i^*$.

* χ_c^2 é o valor da estatística Qui-quadrado, calculado a partir da diferença dos desvios para dois modelos Gama encaixados

$^\circ F_c$ é o valor da estatística de teste F , calculado pela equação (4.2.6).

Tabela 5.2: Coeficientes de regressão e teste de Wald para os modelos da média e dispersão obtidos do problema de panificação.

Modelo da média				
Termos	Estimativa	Erro padrão	Valor t	$Pr(> t)$
x_1	488,961	7,263	67,323	0,0000
x_2	432,210	7,791	55,472	0,0000
x_3	574,124	9,675	59,340	0,0000
$x_1 z_2$	56,621	8,895	6,365	0,0000
$x_3 z_2$	79,146	11,850	6,679	0,0000
$x_2 z_2$	35,904	9,543	3,762	0,0003
$x_1 x_3 z_1$	174,216	29,706	5,865	0,0000
Modelo da dispersão				
Termos	Estimativa	Erro padrão	Valor t	$Pr(> t)$
x_1	6,9984	0,3439	20,352	0,0000
x_2	5,9400	0,5607	10,594	0,0000
x_3	7,3250	0,5607	13,064	0,0000
$x_2 x_3$	-7,9662	3,4523	-2,307	0,0234

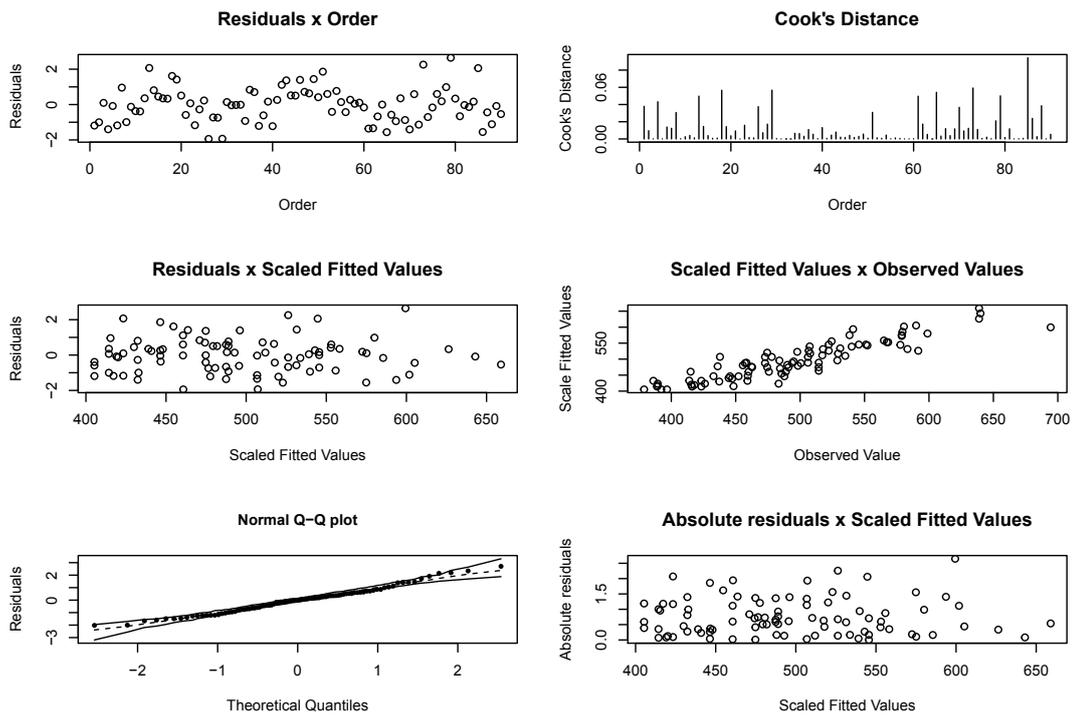


Figura 5.1: Gráficos de diagnóstico para o modelo da média obtido para o problema da panificação.

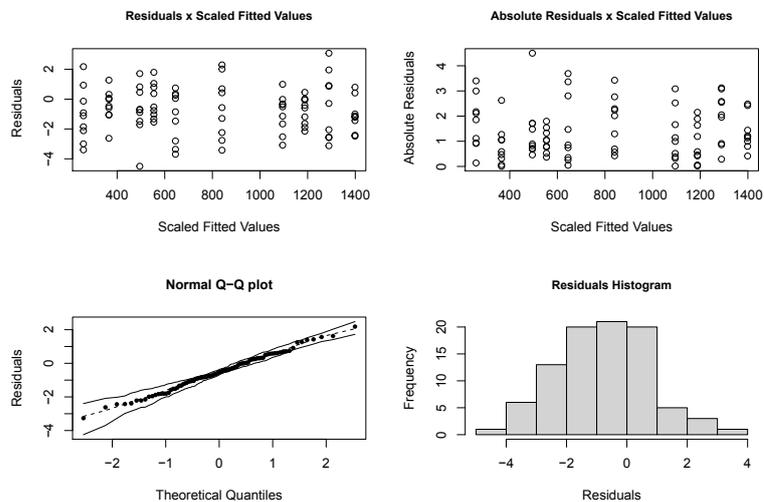


Figura 5.2: Gráficos de diagnóstico para o modelo da dispersão obtido para o problema da panificação.

pode ser empregado se tivermos variáveis de ruído. Na abordagem utilizando a MCMD, se o processo apresentar variabilidade, os modelos da média e da variância podem ser obtidos mesmo sem a presença de fatores de ruído. Além disso, para o problema da panificação, os modelos da média e da variância, incondicionais às variáveis de ruído, dados nas equações

(5.3.3) e (5.3.4), são mais parcimoniosos do que aqueles encontrados usando o método delta, mostrados nas equações (3.3.3) e (3.3.4).

Entretanto, não se pode fazer comparações entre as duas abordagens; ambas fornecem aproximações para a média e a variância incondicional às variáveis de ruído. Contudo, é importante enfatizar que quando usamos a MCMD outras distribuições, que não apenas a distribuição Normal, poderiam ser consideradas para o modelo da média, por exemplo, distribuições para contagens ou proporções. Porém, no caso de um modelo para a média ser do tipo Poisson, com $V(\mu) = \mu$, a complexidade do modelo da variância aumenta, pois $Var(Y|Z) = V(\mu)\phi = \mu \exp\{\mathbf{g}^t(\mathbf{u})\boldsymbol{\gamma}\}$.

A construção de modelos de regressão e a obtenção de planejamentos experimentais são procedimentos fortemente relacionados, onde um forma a base para o outro e vice-versa. Em experimentos envolvendo misturas, a obtenção de planejamentos experimentais ótimos é muito utilizada, especialmente em processos industriais. No próximo capítulo apresentaremos uma metodologia que permite a obtenção de desenho ótimo para modelos conjuntos da média e dispersão aplicados a experimentos com mistura.

Capítulo 6

Desenho ótimo para MCMD - aplicação em modelos de mistura

6.1 Introdução

Em muitos problemas reais, não apenas a média, mas também, no caso mais geral, a dispersão depende de um conjunto de variáveis. Assim, a análise de tais problemas requer a modelagem conjunta de média e dispersão. A MCMD consiste em dois modelos lineares generalizados interligados, um para a média e outro para a dispersão. Modelos lineares generalizados conjuntos da média e dispersão foram introduzidas por Nelder e Lee (1991) como uma alternativa aos métodos de Taguchi em experimentos de melhoria de qualidade. Exemplos adicionais aparecem em Lee e Nelder (1998) e Lee e Nelder (2003).

Assim como a modelagem, a experimentação desempenha um papel importante no método científico. Modelagem e experimentação são processos que estão fortemente vinculados. A estimativa adequada dos parâmetros do modelo depende de um experimento bem planejado. Ao conduzir um experimento, existem muitas questões de planejamento a serem consideradas, incluindo decidir quais tratamentos estudar, quais fatores controlar, quais aspectos de um experimento devem ser aleatorizados, quantas unidades experimentais são necessárias, quantas observações devem ser alocadas para cada tratamento, quais níveis de outras variáveis de entrada ou de controle devem ser usados. Limitações devido a custos, tempo, ética ou recursos humanos, tamanhos de amostra são, geralmente, restritas, portanto, o uso eficiente dos recursos disponíveis é crítico. O objetivo do planejamento experimental ideal é melhorar a inferência estatística sobre as quantidades de interesse, selecionando de

forma otimizada os valores dos fatores de planejamento, sob o controle do investigador, dentro, é claro, das limitações dos recursos disponíveis (CLYDE, 2001).

Quando consideramos a obtenção de planejamentos ótimos para modelos lineares generalizados, nos deparamos com o problema da dependência do valores dos parâmetros a serem estimados. Uma maneira de superar esse problema é considerar desenhos experimentais Bayesianos. A abordagem Bayesiana para o desenho experimental fornece uma maneira de incorporar informações *a priori* no processo de planejamento. Abordagens Bayesianas para planejamento ótimo de experimentos, quando o modelo estatístico é não linear, foram revistas por Atkinson, Donev e Tobias (2002) e por Chaloner e Verdinelli (1995). Na teoria de desenhos ótimos, o Teorema de Equivalência Geral (TEG), proposto por Kiefer e Wolfowitz (1960) e Kiefer (1974), desempenha um papel importante. A maioria dos algoritmos para obtenção de desenhos ótimos fazem uso do TEG, uma vez que ele fornece uma forma de verificar se um determinado planejamento é, de fato, ótimo. Sem considerar o TEG, a obtenção de desenhos ótimos, *D*-ótimo em especial, se reduz a um problema difícil de otimização, com múltiplos máximos e nenhum mecanismo para avaliar quando o ótimo foi obtido (ATKINSON; COOK, 1995). Uma extensão de TEG para *D*-otimalidade Bayesiana é dada por Chaloner e Larntz (1989) e, para um caso mais geral, por Firth e Hinde (1997), ambos utilizando os resultados obtidos por Whittle (1973).

Desenhos ótimos Bayesianos para a MCMD foram introduzidos por Pinto e Ponce de Leon (2004) e Pinto e Ponce de Leon (2007); a prova do Teorema Geral da Equivalência foi apresentada por Pinto e Ponce de Leon (2014). Os resultados apresentados neste capítulo são baseados em Pinto e Ponce de Leon (2014).

Na Seção 6.2, apresentamos a Teoria da Equivalência para desenhos ótimos em modelos não lineares e o Teorema de Whittle. Na Seção 6.3 apresentamos a versão Bayesiana do critério da *D*-otimalidade para a MCMD e sua versão do TEG. Na Seção 6.4 fazemos uma aplicação a experimentos envolvendo misturas. Considerações finais são dadas na Seção 6.5.

6.2 Teoria da Equivalência

Defina Ξ como o conjunto de todas as medidas de probabilidade na região de planejamento denotada por \mathcal{X} . Considere o problema de planejamento de encontrar uma medida em Ξ que maximiza uma função critério ψ_θ , onde o subscrito denota dependência do valor do parâmetro

θ . A informação total ou a matriz de informação Fisher esperada $\mathbf{M}(\theta|\xi)$ é obtida como a matriz de informação de Fisher esperada por observação, com relação à medida de planejamento ξ em \mathcal{X} , ou seja,

$$\mathbf{M}(\theta|\xi) = \int_{\mathcal{X}} \mathbf{I}(\mathbf{u}|\theta) d\xi(\mathbf{u}) \quad (6.2.1)$$

Para duas medidas ξ_1 and ξ_2 em Ξ , a derivada de Fréchet de ψ_θ em $\mathbf{M}_1 = \mathbf{M}(\theta|\xi_1)$ na direção de $\mathbf{M}_2 = \mathbf{M}(\theta|\xi_2)$ é definida, quando o limite existe, por

$$F_{\psi_\theta}(\mathbf{M}_1, \mathbf{M}_2) = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \{ \psi_\theta [(1 - \varepsilon)\mathbf{M}_1 + \varepsilon\mathbf{M}_2] - \psi_\theta(\mathbf{M}_1) \} \quad (6.2.2)$$

com $\psi_\theta : \mathcal{M} \rightarrow \mathbb{R}$, onde $\mathcal{M} = \{ \mathbf{M}(\theta|\xi) : \xi \in \Xi, \theta \in \Theta \}$ e Θ é o espaço paramétrico. Denotamos ξ_u como uma medida que coloca toda massa de probabilidade em um único ponto \mathbf{u} em \mathcal{X} , e adicionalmente definimos $D_{\psi_\theta}(\mathbf{u}, \theta, \xi) = F_{\psi_\theta}[\mathbf{M}(\theta|\xi), \mathbf{M}(\theta|\xi_u)]$.

O teorema proposto por Whittle (1973) para desenhos ótimos em modelos lineares e usado por Chaloner e Larntz (1989) para obtenção de desenhos ótimos em modelos não lineares, forma a base para verificar se desenhos específicos são ótimos ou não, e é reproduzido abaixo em nossa notação.

Teorema 6.2.1 (Whittle, 1973). *Se ψ_θ é côncava, então um desenho ψ_θ -ótimo ξ^* pode ser equivalentemente caracterizado por alguma das três condições seguintes.*

- (i) ξ^* maximiza ψ_θ ;
- (ii) ξ^* minimiza $\sup_{\mathbf{u} \in \mathcal{X}} D_{\psi_\theta}(\mathbf{u}, \theta, \xi)$;
- (iii) $\sup_{\mathbf{u} \in \mathcal{X}} D_{\psi_\theta}(\mathbf{u}, \theta, \xi^*) = 0$.

Como apontado por Chaloner e Larntz (1989), a prova do Teorema de Whittle, para problemas de planejamento linear, pode ser aplicada de forma geral a problemas não lineares sob as seguintes suposições adicionais, isto é, \mathcal{X} deve ser um conjunto compacto, as derivadas devem existir e serem contínuas em $\mathbf{u} \in \mathcal{X}$, deve existir pelo menos uma medida em Ξ para a qual ψ_θ seja finita, e ψ_θ deve ser tal que se $\xi_i \rightarrow \xi$ em convergência fraca então $\psi_\theta(\mathbf{M}(\theta|\xi_i)) \rightarrow \psi_\theta(\mathbf{M}(\theta|\xi))$.

6.3 Teorema Geral da Equivalência para a MCMD

Seja \mathbf{x} um vetor representando o conjunto de fatores que presumivelmente influenciam a média da variável resposta e seja \mathbf{z} um vetor representando o conjunto de fatores que presumivelmente influenciam a dispersão. Seja \mathbf{u} um vetor contendo os fatores ocorrendo em \mathbf{x} e em \mathbf{z} . O vetor \mathbf{z} pode conter alguns ou todos os elementos de \mathbf{x} , bem como outros elementos.

Para a modelagem conjunta da média e da dispersão, seja y_u a resposta a ser observada na unidade experimental $\mathbf{u} \in \mathcal{X}$. Suponha que a distribuição de Y seja desconhecida, no entanto, assume-se que $E(Y_u) = \mu_u$ e $Var(Y_u) = \phi_u V(\mu_u)$, onde ϕ_u é o parâmetro de dispersão e $V(\cdot)$ é a função de variância. Seja φ uma função de ligação para o modelo da média, ou seja, $\eta_u = \varphi(\mu_u) = \mathbf{f}^t(\mathbf{x}_u)\boldsymbol{\beta}$, com $\mathbf{f}^t(\mathbf{x}_u) = [f_1(\mathbf{x}_u), \dots, f_p(\mathbf{x}_u)]$, onde $f_j(\mathbf{x}_u)$, para $j = 1, \dots, p$, é uma função conhecida de \mathbf{x}_u e $\boldsymbol{\beta}$ é um vetor $p \times 1$ de parâmetros desconhecidos; para o modelo de dispersão, seja ζ a função de ligação, geralmente log, ou seja, $\tau_u = \zeta(\phi_u) = \mathbf{g}^t(\mathbf{z}_u)\boldsymbol{\gamma}$, com $\mathbf{g}^t(\mathbf{z}_u) = [g_1(\mathbf{z}_u), \dots, g_q(\mathbf{z}_u)]$, onde $g_j(\mathbf{z}_u)$, para $j = 1, \dots, q$, é uma função conhecida de \mathbf{z}_u e $\boldsymbol{\gamma}$ é um vetor $q \times 1$ de parâmetros desconhecidos. Geralmente é considerado que $f_1(\mathbf{x}_u) = g_1(\mathbf{z}_u) = 1$; porém, para o caso de experimentos com misturas o termo constante não é colocado no modelo devido à restrição de mistura (2.3.1).

Usando a quase-verossimilhança estendida como um critério de estimação e considerando o modelo da dispersão como Gama, a matriz de quase informação de Fisher por observação é dada por

$$I_C(\mathbf{u}|\boldsymbol{\theta}) = \begin{bmatrix} w_u \mathbf{f}(\mathbf{x}_u) \mathbf{f}^t(\mathbf{x}_u) & \mathbf{0} \\ \mathbf{0} & v_u \mathbf{g}(\mathbf{z}_u) \mathbf{g}^t(\mathbf{z}_u) \end{bmatrix} \quad (6.3.1)$$

onde $\boldsymbol{\theta}^t = (\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t)$, $w_u = \left(\frac{\partial \mu_u}{\partial \eta_u}\right)^2 \frac{1}{\phi_u V(\mu_u)}$ e $v_u = \left(\frac{\partial \phi_u}{\partial \tau_u}\right)^2 \frac{1}{2\phi_u^2}$.

Para um desenho $\xi \in \xi$, a quase matriz de informação para a MCMD é dada por

$$\mathbf{M}_C(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} \mathbf{I}_C(\mathbf{u}|\boldsymbol{\theta}) d\xi(\mathbf{u}) = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}, \quad (6.3.2)$$

onde $\mathbf{C} = \mathbf{C}(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} w_u \mathbf{f}(\mathbf{x}_u) \mathbf{f}^t(\mathbf{x}_u) d\xi(\mathbf{u})$ e $\mathbf{D} = \mathbf{D}(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} v_u \mathbf{g}(\mathbf{z}_u) \mathbf{g}^t(\mathbf{z}_u) d\xi(\mathbf{u})$.

Observe que a matriz de quase informação $\mathbf{M}_C(\boldsymbol{\theta}|\xi)$ depende do valor do parâmetro $\boldsymbol{\theta}$. Neste caso, usaremos a abordagem Bayesiana descrita por Chaloner e Larntz (1989) e Atkinson, Donev e Tobias (2002) para modelos não lineares.

6.3.1 Critério D -ótimo Bayesiano

O critério D -ótimo Bayesiano para a MCMD é definido como

$$\psi_{\theta}(\mathbf{M}_C(\boldsymbol{\theta}|\xi)) = E_{\theta}[\log |\mathbf{M}_C(\boldsymbol{\theta}|\xi)|], \quad (6.3.3)$$

com $\psi_{\theta} : \mathcal{M} \rightarrow \mathbb{R}$, onde $\mathcal{M} = \{\mathbf{M}_C(\boldsymbol{\theta}|\xi) : \xi \in \Xi, \boldsymbol{\theta} \in \Theta\}$, E_{θ} refere-se à esperança com respeito à distribuição à priori em $\boldsymbol{\theta}$ e $\log |\mathbf{M}_C(\boldsymbol{\theta}|\xi)|$ representa o logaritmo natural do determinante da matriz de quase informação de Fisher. Desta forma, o problema de planejamento D -ótimo Bayesiano consiste em encontrar uma media ξ^* , de uma classe específica de medidas de planejamento Ξ , que maximiza ψ .

Proposição 6.3.1. *A função critério Bayesiana $\psi_{\theta} = E_{\theta}[\ln |\mathbf{M}_C(\boldsymbol{\theta}|\xi)|]$ é estritamente côncava em \mathcal{M}^+ , o subconjunto de \mathcal{M} onde ψ_{θ} é finita.*

Seja a matriz de quase informação, associada ao desenho ξ_u , que colocada toda massa de probabilidade em um único ponto \mathbf{u} em \mathcal{X} , denotada por $\mathbf{M}_C(\boldsymbol{\theta}|\xi_u)$.

Proposição 6.3.2. *A derivada de Fréchet da função critério $\psi_{\theta} = E_{\theta}[\ln |\mathbf{M}_C(\boldsymbol{\theta}|\xi)|]$ em $\mathbf{M}_C(\boldsymbol{\theta}|\xi)$ na direção de $\mathbf{M}_C(\boldsymbol{\theta}|\xi_u)$ é dada por*

$$D_{\psi_{\theta}}(\mathbf{u}, \xi) = E_{\theta}[\text{tr}(\mathbf{M}_{C_u} \mathbf{M}_C^{-1})] - t \quad (6.3.4)$$

onde $\mathbf{M}_C = \mathbf{M}_C(\boldsymbol{\theta}|\xi)$, $\mathbf{M}_{C_u} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_u)$, t é o número de parâmetros no modelo conjunto e $\text{tr}(\cdot)$ representa o operador traço.

As provas das proposições 6.3.1 e 6.3.2 foram dadas por Pinto e Ponce de Leon (2014) e também são mostradas nos Apêndices B.1 e B.2.

Em (6.3.4) $\mathbf{M}_{C_u} = \mathbf{I}_C(\mathbf{u}|\boldsymbol{\theta})$ é dado em (6.3.1), desta forma

$$D_{\psi_\theta}(\mathbf{u}, \xi) = E_\theta[\mathbf{w}_u \mathbf{f}^t(\mathbf{x}_u) \mathbf{C}^{-1} \mathbf{f}(\mathbf{x}_u) + \mathbf{v}_u \mathbf{g}^t(\mathbf{z}_u) \mathbf{D}^{-1} \mathbf{g}(\mathbf{z}_u)] - t \quad (6.3.5)$$

Vamos definir a função de sensibilidade como $d_{\psi_\theta}(\mathbf{u}, \xi) = E_\theta[\text{tr}(\mathbf{M}_{C_u} \mathbf{M}_C^{-1})]$; desta forma, a função de sensibilidade no ponto \mathbf{u} é dada por

$$d_{\psi_\theta}(\mathbf{u}, \xi) = E_\theta[\mathbf{w}_u \mathbf{f}^t(\mathbf{x}_u) \mathbf{C}^{-1} \mathbf{f}(\mathbf{x}_u) + \mathbf{v}_u \mathbf{g}^t(\mathbf{z}_u) \mathbf{D}^{-1} \mathbf{g}(\mathbf{z}_u)] \quad (6.3.6)$$

Teorema 6.3.1 (Pinto e Ponce de Leon, 2014). *Considere Ξ a classe de todas as medidas de probabilidade sobre \mathcal{X} . Seja $\psi_\theta = E_\theta[\ln |\mathbf{M}_C(\boldsymbol{\theta}|\xi)|]$ o critério D-ótimo Bayesiano para a MCMD. Então um desenho ψ_θ -ótimo ξ^* pode ser equivalentemente caracterizado por alguma das três seguintes condições.*

- a) O desenho ξ^* maximiza ψ_θ em Ξ
- b) ξ^* minimiza $\sup_{\mathbf{u} \in \mathcal{X}} D_{\psi_\theta}(\mathbf{u}, \xi)$
- c) $\sup_{\mathbf{u} \in \mathcal{X}} d_{\psi_\theta}(\mathbf{u}, \xi) = t$

onde $D_{\psi_\theta}(\mathbf{u}, \xi)$ e $d_{\psi_\theta}(\mathbf{u}, \xi)$ são dados em (6.3.5) e (6.3.6), respectivamente e t é o número de parâmetros nos modelos da média e da dispersão.

Como $\psi_\theta = E_\theta[\ln |\mathbf{M}_C(\boldsymbol{\theta}|\xi)|]$ é uma função côncava, pela Proposição 6.3.1, a prova do Teorema 6.3.1 segue diretamente de Chaloner e Larntz (1989), usando o Teorema 6.2.1.

Para obtenção do desenho ótimo em cálculos computacionais são utilizados desenhos discretos, como representado em (2.3.14). As expressões de $\mathbf{M}_C(\boldsymbol{\theta}|\xi)$ e $d_{\psi_\theta}(\mathbf{u}, \xi)$, considerando desenhos discretos, são dadas em Pinto e Ponce de Leon (2006b).

6.4 Aplicação em experimentos de mistura

Nesta seção vamos aplicar a teoria desenvolvida neste capítulo a um problema de mistura com restrição para o qual conhecemos os modelos conjuntos da média e da dispersão. O problema considerado é o de encontrar o desenho ótimo para uma das situações de ruído do problema

de mistura de farinha para produção de pão, introduzido na Seção 3.3. Os modelos conjuntos da média e da dispersão foram obtidos no Capítulo 5.

Nosso objetivo agora é, escolhida uma determinada situação de ruído, encontrar um desenho ótimo para uma próxima experimentação. Para esta situação específica, o problema se reduz ao problema de encontrar um desenho ótimo para um experimento de mistura com restrição. As restrições para cada uma das covariáveis de mistura são $0,25 \leq x_1 \leq 1,0$; $0 \leq x_2 \leq 0,75$ e $0 \leq x_3 \leq 0,75$.

De acordo com Pinto (2005), que também analisou esse problema, para cada situação de ruído, pontos (z_1, z_2) , com $z_1, z_2 \in \{-1, 0, 1\}$, mostrados na Figura 3.1, a substituição dos valores de z_1 e z_2 nas equações (5.3.1) e (5.3.2) definem modelos diferentes para a média e para a dispersão, isto é, para cada situação de ruído, temos um par de modelos diferentes. Vamos considerar apenas uma dessas situações. Optamos pela situação em que $z_1 = 0$ e $z_2 = 0$, ou seja, a situação central na Figura 3.1. Neste caso, o preditor linear para o modelo da média é dado por

$$\eta = 488,96x_1 + 432,21x_2 + 574,12x_3 \quad (6.4.1)$$

(7,26) (7,79) (9,68)

e, para o modelo da dispersão, o preditor linear é

$$\zeta = 6,99x_1 + 5,94x_2 + 7,33x_3 - 7,97x_2x_3. \quad (6.4.2)$$

(0,34) (0,56) (0,56) (3,45)

Nas equações (6.4.1) e (6.4.2), os valores entre parênteses representam os respectivos valores estimados dos desvios-padrão dos estimadores dos parâmetros, conforme Tabela 5.2. Os valores estimados dos desvios-padrão dos parâmetros foram usados para construir *prioris* uniformes discretas em torno das estimativas dos parâmetros. Os valores dos pontos das *prioris* foram escolhidos de modo a formarem um hipercubo em torno dos valores estimados dos parâmetros $\beta^t = (488,96; 432,21; 574,12)$ e $\gamma^t = (6,99; 5,94; 7,33; -7,97)$, mostrados nas equações (6.4.1) e (6.4.2). Note que os pontos das *prioris* são vetores de dimensão 3, no caso da *priori* para β e 4, no caso da *priori* para γ . Os vetores de parâmetros usados na construção das *prioris* uniformes foram construídos somando e subtraindo os valores dos desvios-padrão

aos valores estimados dos parâmetros. Os vetores β e γ também fizeram parte da *priori*. Desta forma, a *priori* uniforme discreta para β teve com 9 pontos e aquela para γ teve 17 pontos. Para maiores detalhes sobre construção de *prioris* discretas, veja Pinto e Ponce de Leon (2006b), p. 52.

Conforme considerado no Capítulo 5, para o modelo da média foi utilizado $V(\mu) = 1$ e função de ligação identidade; para o modelo da dispersão foi utilizado um modelo Gama com função de ligação logarítmica.

O desenho D -ótimo Bayesiano, encontrado para a situação em que os fatores de ruído são $z_1 = 0$ e $z_2 = 0$, foi:

$$\xi^* = \left\{ \begin{array}{ccccc} (0, 25; 0, 75; 0, 0) & (0, 25; 0, 0; 0, 75) & (1, 0; 0, 0; 0, 0) & (0, 51; 0, 0; 0, 49) & (0, 25; 0, 28; 0, 47) \\ 0, 29 & 0, 25 & 0, 30 & 0, 10 & 0, 06 \end{array} \right\} \quad (6.4.3)$$

A Figura 6.1 mostra ilustra os pontos do desenho ótimo (6.4.3) na região região experimental simplex.

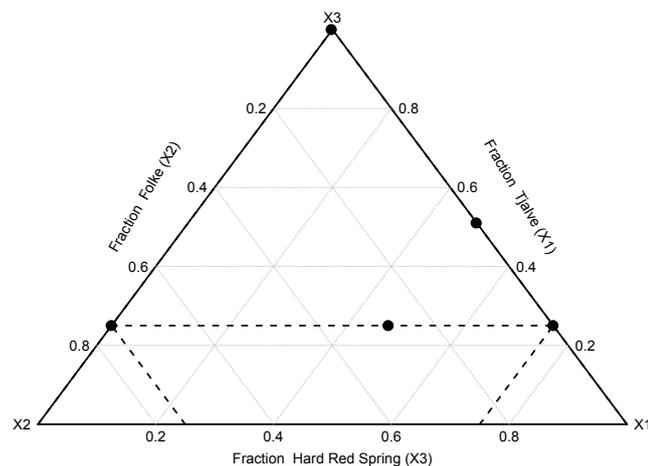


Figura 6.1: Desenho D -ótimo Bayesiano para o modelo conjunto da média e dispersão aplicado ao experimento de mistura de farinha na situação de ruído $z_1 = 0$ e $z_2 = 0$.

6.5 Considerações finais

Em situações onde não se pode assumir variância constante, a MCMD permite obter, de forma bastante eficiente, modelos para média e dispersão, possibilitando que se construam modelos conjuntos para média e variância. A obtenção de planejamento ótimo de experimento para modelos conjuntos da média e da dispersão é grande interesse na indústria, pois promove maior eficiência confiabilidade nos resultados obtidos e permite a redução do número de pontos a serem considerados no experimento, resultando em economia de tempo e dinheiro. A teoria de planejamento ótimo de experimentos apresentada por Pinto (2005) e formalizada por Pinto e Ponce de Leon (2014), permite encontrar desenhos ótimos para MCMD, em especial, modelos de experimentos com misturas.

Em experimentos com misturas, desenhos ótimos são de fundamental importância, especialmente em situações onde, além da restrição de mistura, existem também restrições nas variáveis de mistura. A teoria de planejamento ótimo de experimentos é muito versátil para situações onde não temos uma situação padrão (veja Seção 2.3.1). Neste capítulo foi apresentada a teoria para obtenção de desenho ótimo para a MCMD. A teoria desenvolvida se aplica normalmente a problemas de mistura, desde que sejam levadas em conta as condições de mistura (2.3.1) e as restrições nas covariáveis, se houver.

Na aplicação a um problema de mistura com restrições, considerada na Seção 6.4, o desenho ótimo foi obtido apenas para uma situação de combinação de variáveis de ruído, onde $z_1 = z_2 = 0$, porém o mesmo procedimento também poderia ser realizado para obtenção de desenhos ótimos para todas as outras combinações dessas variáveis, mostradas na Figura 3.1. Nestes casos, os modelos da média e da dispersão serão diferentes para cada combinação, porém o procedimento de procura pelo ótimo será o mesmo. Note também que em cada uma dessas combinações o vetor de *prioris*, que depende do número de parâmetros dos modelos, também sofrerá mudança. Desta forma, vale ressaltar que os desenhos ótimos que serão obtidos estarão condicionados às situações de ruído (desenho fatorial nas variáveis de ruído, conforme mostrado na Figura 3.1), onde as combinações dos fatores de ruído são fixas e conhecidas. Assim, dada uma combinação de condições de ruído, o desenho ótimo é obtido para esta combinação. Entretanto, a teoria de planejamento ótimo para MCMD, apresentada na Seção 6.3, é geral e pode ser utilizada considerando que os valores das variáveis de ruído não necessitem ser fixos, podendo variar, por exemplo, no intervalo $(-1, 1)$. Neste caso, o

processo de busca pelo desenho ótimo será mais complexo e acarretará maior esforço computacional.

Capítulo 7

Conclusões

Nesta tese, discutimos a modelagem da variabilidade em experimentos com misturas. A modelagem conjunta da média e dispersão foi utilizada para obter modelos para a média e para variância nos experimentos considerados. Foi proposto um processo eficiente de seleção de variáveis para a MCMD, que foi adaptado e utilizado para obtenção de modelos da média e da variância em experimentos com misturas. A teoria para obtenção de planejamento ótimo na MCMD foi apresentada e aplicada para obtenção do desenho *D*-ótimo para uma combinação de variáveis de ruído em um modelo de mistura obtido pela MCMD.

O método delta, apresentado no Capítulo 3, fornece uma aproximação para o modelo da média e para o modelo da variância em experimentos de mistura com variáveis de ruído, porém a aproximação pode não ser boa se a dispersão não for constante.

A aplicação da modelagem conjunta da média e dispersão em problemas de misturas é um resultado novo, embora a teoria da MCMD já exista há algum tempo, não existe aplicação alguma em problemas envolvendo misturas. Nesta tese o procedimento da MCMD foi explicado com detalhes, tanto do ponto de vista da estimação, quanto do processo de seleção de variáveis. O procedimento proposto nos Capítulos 4 e 5, permite verificar se o modelo considerado para a variável resposta tem dispersão constante ou não, além disso, fornece, como resultado final, os modelos ótimos obtidos para a média e para dispersão.

O processo de seleção de variáveis na MCMD, apresentado no Capítulo 4, é geral e permite a obtenção dos modelos da média e da dispersão de uma forma bastante eficiente. O procedimento de seleção de variáveis pode ser facilmente implementado no *software* R (R CORE TEAM, 2019). O procedimento de seleção de variáveis utiliza o mesmo processo gangorra da MCMD, veja Subseção 2.3.2, o que torna sequencial a procura pelo modelo ótimo

da média, isto é, em cada iteração do processo o ajuste do modelo vai melhorando. O modelo da dispersão, por ter como resposta o desvio do modelo da média, varia para cada iteração do procedimento. Uma função no *software* R, que realiza esse processo de procura pelo modelo conjunto ótimo, está em fase final de elaboração e deverá ser disponibilizada em breve.

No Capítulo 5, o processo de seleção de variáveis na MCMD foi modificado para satisfazer a restrição de mistura e o procedimento foi aplicado ao exemplo considerado no Capítulo 3. Note que o processo identificou dispersão variável e, portanto a aproximação considerada no Capítulo 3, utilizando o método delta, não é adequada. Os modelos obtidos para a média e para a variância foram mais simples, porém agora temos um modelo para a dispersão e não apenas uma estimativa para a variância, como considerado no Capítulo 3.

Em processos industriais, modelagem e experimentação são procedimentos complementares. A obtenção de desenhos ótimos em problemas com misturas é de fundamental importância, particularmente quando existem restrições nos componentes da mistura. A teoria apresentada no Capítulo 6 permite que desenhos ótimos sejam obtidos para a MCMD, em especial, para experimentos envolvendo misturas. O processo considerado para obtenção do desenho *D*-ótimo utilizou a metodologia Bayesiana, considerando *prioris* uniformes discretas para ambos os parâmetros dos modelos da média e da dispersão. O desenho ótimo foi obtido somente para a situação de ruído $z_1 = z_2 = 0$, contudo poderíamos ter obtido desenhos ótimos para cada uma das combinações de ruído mostradas na Figura 3.1. Entretanto, o procedimento de procura pelo desenho ótimo, não necessitaria de que as situações de ruído fossem fixas. O processo de busca pelo ótimo poderia incluir também situações de ruído, porém essa abordagem ainda está sob estudo e deverá ser apresentada em trabalhos futuros.

A utilização da quase-verossimilhança estendida na MCMD, permite que os modelos, bem como os desenhos experimentais ótimos, sejam obtidos sem o conhecimento completo da verossimilhança, necessitando somente do conhecimento da relação entre a média e a variância da resposta. Além disso, a MCMD permite que a dispersão seja modelada de forma muito eficaz em problemas de planejamento robusto, mesmo sem replicação, e o modelo da variância pode ser obtido de uma forma simples a partir do modelo de dispersão.

Vale ressaltar também que os desenhos ótimos obtidos no Capítulo 6, denominados Bayesianos, não são soluções típicas de problemas Bayesianos tradicionais, desde que nenhuma distribuição a *posteriori* é envolvida. Contudo, estes desenhos podem ser obtidos na

forma tradicional Bayesiana, através da abordagem da Teoria da Decisão Bayesiana aplicada na obtenção de planejamento ótimo de experimentos. Excelentes referências sobre o assunto são Lindley (1956, 1972) e Chaloner e Verdinelli (1995). Um procedimento para encontrar desenhos ótimos usando métodos numéricos de otimização estocástica, via simulação de Monte Carlo é dado por Müller e Parmigiani (1995). Para maiores detalhes sobre desenhos ótimos Bayesianos veja Atkinson, Donev e Tobias (2002) e Pinto e Ponce de Leon (2006b).

Os resultados obtidos nesta tese através da aplicação da modelagem conjunta da média e dispersão em experimentos com misturas foram bastante animadores, indicando que a teoria promete ser de grande utilidade na modelagem e na condução de experimentos industriais envolvendo misturas. Contudo, ainda existe pouco entrosamento entre os pesquisadores acadêmicos e os profissionais da indústria. Para que a aplicação das teorias apresentadas nesta tese não fiquem restritas somente ao meio acadêmico e se tornem parte importante no planejamento industrial, seria necessário a elaboração de textos, abordando o assunto de forma didática, e o desenvolvimento de *softwares* adequados. Neste sentido, com o intuito de tentar preencher a lacuna existente entre academia e indústria, será elaborado um livro didático sobre modelagem e otimização de experimentos com misturas, abordando desde os conceitos básicos até os assuntos discutidos nesta tese. O livro, que também conterà apoio computacional, com a utilização do *software* R, já se encontra em fase de preparação e acreditamos que em breve deverá ser publicado.

Referências

- AITKIN, M. Modelling variance heterogeneity in normal regression using GLIM. **Journal of the Royal Statistical Society - Series C - Applied Statistics**, v. 36, p. 332–339, 1987.
- ANTONIADIS, A.; GIJBELS, I.; LACROIX, S. L.; POGGI, J. M. Joint estimation and variable selection for mean and dispersion in proper dispersion models. **Electronic Journal of Statistics**, v. 10, p. 1630–1676, 2016.
- ATKINSON, A. C.; COOK, R. D. D-optimal designs for heteroscedastic linear models. **Journal of The American Statistical Association**, v. 90, p. 204–212, 1995.
- ATKINSON, A. C.; COOK, R. D. Designing for a response transformation parameter. **Journal of the Royal Statistical Society, serie B**, v. 59, p. 111–124, 1997.
- ATKINSON, A. C.; DONEV, A. N.; TOBIAS, R. D. **Optimum experimental designs, with SAS**. New York: Oxford University Press, 2002.
- BAYER, F. M.; CRIBARI NETO, F. Model selection criteria in beta regression with varying dispersion. **Communications in Statistics - Simulation and Computation**, v. 46, p. 729–746, 2017.
- BONAT, W. H.; JORGENSEN, B. Multivariate covariance generalized linear models. **Journal of the Royal Statistical Society - Series C - Applied Statistics**, v. 65, p. 649–675, 2016.
- BOYD, S.; VANDENBERGHE, L. **Convex Optimization**. Cambridge: Cambridge University Press, 2004.
- BROWN, D.; WONG, W. K. An algorithmic construction of optimal minimax designs for heteroscedastic linear models. **Journal of Statistical Planning and Inference**, v. 85, p. 103–114, 2000.
- BURRIDGE, J.; SEBASTIANI, P. D-optimal designs for generalized linear models with variance proportional to square of the mean. **Biometrika**, v. 81, p. 295–304, 1994.
- CARROLL, R. J.; RUPPERT, D. **Transformation and weighting in regression**. New York: Chapman & Hall, 1988.
- CASELLA, G.; BERGER, R. L. **Statistical Inference**. 2. ed. New York: Duxbury - Thomson Learning, 2002.
- CHALONER, K.; LARNTZ, K. Optimal Bayesian design applied to logistic regression experiments. **Journal of the Statistical Planning and Inference**, v. 21, p. 191–208, 1989.

- CHALONER, K.; VERDINELLI, I. Bayesian experimental design: a review. **Statistical Science**, v. 10, p. 273–304, 1995.
- CHARALAMBOUS, C.; PAN, J.; TRANMER, M. Variable selection in joint mean and dispersion models via double penalized likelihood. **Sankhyā: The Indian Journal of Statistics**, v. 76, p. 276–304, 2014.
- CHARALAMBOUS, C.; PAN, J.; TRANMER, M. Variable selection in joint modelling of the mean and variance for hierarchical data. **Statistical Modelling**, v. 15, p. 24–50, 2015.
- CLYDE, M. A. Experimental Design: A Bayesian Perspective. **International Encyclopedia of the Social & Behavioral Sciences**, p. 5075–5081, 2001.
- CORNELL, J. A. **Experiments with mixtures: designs, models, and the analysis of mixture data**. 3. ed. New York: Wiley, 2002.
- CORNELL, J. A. Fitting a slack-variable model to mixture data: some questions raised. **Journal of Quality Technology**, v. 32, p. 133–147, 2000.
- CORNELL, J. A.; GORMAN, J. W. Two new mixture models: living with collinearity but removing its influence. **Journal of Quality Technology**, v. 35, p. 78–88, 2003.
- COX, D. R.; REID, N. **The theory of the design of experiments**. Boca Raton: Chapman & Hall, 2000.
- CUERVO, E.; GAMERMAN, D. Bayesian modeling of variance heterogeneity in normal regression models. **Brazilian Journal of Probability and Statistics**, v. 14, p. 207–221, 2001.
- DASGUPTA, A.; MUKHOPADHAY, S.; STUDDEN, W. J. Compromise designs in heteroscedastic linear models. **Journal of Statistical Planning and Inference**, v. 32, p. 363–384, 1992.
- DAVIDIAN, M.; CARROLL, R. J. A note on extended quasi-likelihood. **Journal of the Royal Statistical Society, Series B**, v. 50, p. 74–82, 1988.
- DAVIDIAN, M.; CARROLL, R. J. Variance function estimation. **Journal of the American Statistical Association**, v. 82, p. 1079–1091, 1987.
- DETTE, H.; WONG, W. K. Bayesian D -optimal designs on a fixed number of designs points for heteroscedastic polynomial models. **Biometrika**, v. 85, p. 869–882, 1998.
- DEY, K. D.; GELFAND, A. E.; PENG, F. Overdispersed generalized linear models. **Journal of Statistical Planning and Inference**, v. 64, p. 93–107, 1997.
- DONEV, A. N. Design of experiments with both mixture and qualitative factors. **Journal of the Royal Statistical Society B**, v. 50, p. 297–302, 1989.
- EFRON, B. Double exponential families and their use in generalized linear regression. **Journal of the American Statistical Association**, v. 81, p. 709–721, 1986.
- ENGEL, J. Modeling variation in industrial experiments. **Applied Statistics**, v. 41, p. 579–593, 1992.

- ENGEL, J.; HUELE, A. F. A generalized linear modeling approach to robust Design. **Technometrics**, v. 38, p. 365–373, 1996.
- FAERGESTAD, E. M.; NAES, T. **Evaluation of baking quality of wheat flours: small scale straight dough baking test of heart bread with variable mixing time and proofing time**. As, Norway, 1997.
- FEDOROV, V. V. **Theory of Optimal Experiments**. New York: Academic Press, 1972.
- FEDOROV, V. V.; HACKAL, P. **Model-Oriented Design of Experiments**. New York: Springer, 1997. (Lecture Notes in Statistics 125).
- FIRTH, D.; HINDE, J. P. On Bayesian D-optimum criteria and the Equivalence Theorem in non-linear models. **Journal of the Royal Statistical Society, Series B**, v. 59, p. 793–797, 1997.
- GODAMBE, V. P.; THOMPSON, M. E. An extension of quasi-likelihood estimation. **Journal of Statistical Planning and Inference**, v. 22, p. 137–152, 1989.
- GOLDFARB, H.; BORROR, C.; MONTGOMERY, D. Mixture-process variable experiments with noise variables. **Journal of Quality Technology**, v. 35, p. 393–405, 2003.
- GRANATO, D.; CALADO, V.; PINTO, E. R. Mathematical and Statistical Applications in Food Engineering. In: SEVDA, S.; SINGH, A. (Ed.). **Optimization of food processes using mixture experiments - some applications**. 1. ed. Boca Raton: CRC Press, 2020. cap. 3, p. 21–35.
- HU, B.; SHAO, J. Generalized linear model selection using R^2 . **Journal of Statistical Planning and Inference**, v. 138, p. 3705–3712, 2008.
- HURVICH, C. M.; TSAI, C. L. Regression and time series model selection in small samples. **Biometrika**, v. 76, p. 297–307, 1989.
- JORGENSEN, B. Exponential dispersion models (with discussion). **Journal of the Royal Statistical Society, Series B**, v. 49, p. 127–162, 1987.
- JORGENSEN, B. **The theory of dispersion models**. New York: Chapman & Hall, 1997.
- JORGENSEN, B. **The theory of exponential dispersion models and analysis of deviance**. Rio de Janeiro: IMPA, 1992.
- KHURI, A. I. Slack Variable Models Versus Scheffé's Mixture Models. **Journal of Applied Statistics**, v. 32, p. 887–908, 2005.
- KIEFER, J. General Equivalence Theorem for optimum designs (approximate theory). **Annals of Statistics**, v. 2, p. 849–879, 1974.
- KIEFER, J. Optimum designs in regression problems II. **Annals of Mathematical Statistics**, v. 32, p. 298–325, 1961.
- KIEFER, J.; WOLFOWITZ, J. The Equivalence of two extremum problems. **Canadian Journal of Statistics**, v. 12, p. 363–366, 1960.

- LAWSON, J.; WILLDEN, C. Mixture Experiments in R using mixexp. **Journal of Statistical Software**, v. 72, p. 1–20, 2016.
- LEE, K. I.; KOVAL, J. J. Determination of the best significance level in forward stepwise logistic regression. **Communications in Statistics - Simulation and Computation**, v. 26, p. 559–575, 1997.
- LEE, Y.; NELDER, J. A. Generalized linear models for analysis of quality improvement experiments. **The Canadian Journal of Statistics**, v. 26, p. 95–105, 1998.
- LEE, Y.; NELDER, J. A. Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. **Biometrika**, v. 88, p. 987–1006, 2001.
- LEE, Y.; NELDER, J. A. Hierarchical generalized linear models. **Journal of the Royal Statistical Society, Series B**, v. 58, p. 619–678, 1996.
- LEE, Y.; NELDER, J. A. Robust design via generalized linear models. **Journal of Quality Technology**, v. 35, p. 2–12, 2003.
- LEE, Y.; NELDER, J. A. The relationship between double-exponential families and extended quasi-likelihood families, with application to modelling Geissler's human Sex ratio data. **Applied Statistics**, v. 49, p. 413–419, 2000.
- LEE, Y.; NELDER, J. A. The robustness of the quasilielihood estimator. **The Canadian Journal of Statistics**, v. 27, p. 321–327, 1999.
- LEE, Y.; NELDER, J. A.; PAWITAN, Y. **Generalized linear models with random effects: unified analysis via h-likelihood**. Boca Raton: Chapman & Hall, 2006.
- LINDLEY, D. V. **Bayesian statistics: a review**. 1. ed. Philadelphia: Society for Industrial e Applied Mathematics - SIAM, 1972.
- LINDLEY, D. V. On a measure of information provided by an experiment. **The Annals of Mathematical Statistics**, v. 27, p. 986–1005, 1956.
- MARQUARDT, D. W.; SNEE, R. D. Test Statistics for Mixture Models. **Technometrics**, v. 16, p. 533–537, 1974.
- MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. 2. ed. London: Chapman & Hall, 1989.
- MILLER, A. **Subset Selection in Regression**. 2. ed. London: Chapman & Hall, 2002.
- MÜLLER, P.; PARMIGIANI, G. Optimal design by curve fitting of Monte Carlo experiments. **Journal of the American Statistical Association**, v. 90, p. 1322–1330, 1995.
- MYERS, R. H.; KHURI, A. I.; VINING, G. Response surface alternatives to the Taguchi robust parameter design approach. **The American Statistician**, v. 42, p. 131–139, 1992.
- MYERS, R. H.; MONTGOMERY, D. C. **Response surface methodology: process and product optimization using designed experiments**. 2. ed. New York: John Wiley & Sons, 2002.

- NAES, T.; FAERGESTAD, E. M.; CORNELL, J. A. A comparison of methods for analyzing data from a three component mixture experiment in the presence of variation created by two process variables. **Chemometrics and Intelligence Laboratory Systems**, v. 41, p. 221–235, 1998.
- NAIR, V. N. Taguchi's parameter design: a panel discussion. **Technometrics**, v. 34, p. 127–161, 1992.
- NELDER, J. A. Quasi-likelihood and pseudo-likelihood are not the same thing. **Journal of Applied Statistics**, v. 27, p. 1007–1011, 2000.
- NELDER, J. A.; LEE, Y. Generalized linear models for the analysis of Taguchi-type experiments. **Applied Stochastic Models and Data Analysis**, v. 7, p. 107–120, 1991.
- NELDER, J. A.; LEE, Y. Likelihood, quasi-likelihood and pseudolikelihood: some comparisons. **Journal of the Royal Statistical Society, Series B**, v. 54, p. 273–284, 1992.
- NELDER, J. A.; PREGIBON, D. An extended quasi-likelihood function. **Biometrika**, v. 74, p. 221–232, 1987.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society, Series A**, v. 135, p. 370–384, 1972.
- NETER, J.; KUTNER, M. H.; WASSERMAN, W.; NACHTSHEIM, C. J. **Applied linear regression models**. 3. ed. Boston: McGraw-Hill, 1996.
- PHADKE, M. S. **Quality engineering using robust design**. New York: Prentice-Hall, 1989.
- PIEPEL, G. F. 50 Years of mixture experiment research: 1955-2004. In: KHURI, A. I. (Ed.). **Response Surface Methodology and Related Topics**. Singapore: World Scientific, 2004. p. 283–327.
- PIEPEL, G. F.; CORNELL, J. A. **A catalogue of mixture experiment examples**. Richland, WA, 2001.
- PINTO, E. R. **Planejamento de experimentos ótimos para modelos lineares generalizados com parâmetro de dispersão variando**. 2005. Tese (Doutorado) – Universidade Federal do Rio de Janeiro.
- PINTO, E. R.; PONCE DE LEON, A. Bayesian D_s -optimum designs for generalized linear models with varying dispersion parameter. In: LÓPEZ-FIDALGO, J.; RODRÍGUEZ-DÍAS, J. M.; TORSNEY, B. (Ed.). **mODa8-Advances in Model Oriented Data Analysis - Contributions to Statistics**. New York: Springer, 2007. p. 181–188.
- PINTO, E. R.; PONCE DE LEON, A. Bayesian D -optimum designs for generalized linear models with varying dispersion parameter. In: DI BUCCHIANICO, A.; LÄUTER, H.; WYNN, H. P. (Ed.). **mODa7-Advances in Model Oriented Data Analysis - Contributions to Statistics**. New York: Springer, 2004. p. 127–134.
- PINTO, E. R.; PONCE DE LEON, A. Modelagem conjunta da média e dispersão de Nelder e Lee como alternativa aos métodos de Taguchi. **Pesquisa Operacional**, v. 26, p. 203–224, 2006.

- PINTO, E. R.; PONCE DE LEON, A. On Bayesian D -optimal design criteria and the General Equivalence Theorem in joint generalized linear models for the mean and dispersion. **Brazilian Journal of Probability and Statistics**, v. 28, p. 483–491, 2014.
- PINTO, E. R.; PONCE DE LEON, A. **Planejamento ótimo de experimentos**. São Paulo: Associação Brasileira de Estatística - ABE, 2006.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<http://www.R-project.org/>>.
- SCHEFFÉ, H. Experiments with Mixture. **Journal of the Royal Statistical Society B**, v. 20, p. 344–366, 1958.
- SCHEFFÉ, H. Simplex-centroid designs for experiments with mixtures. **Journal of the Royal Statistical Society B**, v. 25, p. 235–263, 1963.
- SILVEY, S. D. **Optimal design**. London: Chapman & Hall, 1980.
- SMITH, W. F. **Experimental design for formulation**. New York: Society for Industrial e Applied Mathematics - SIAM, 2005.
- SMYTH, G. K. An efficient algorithm for REML in heteroscedastic regression. **Journal of Graphical and Computational Statistics**, v. 11, p. 15–27, 2002.
- SMYTH, G. K. Generalized linear models with varying dispersion. **Journal of The Royal Statistical Society – Series B**, v. 51, p. 47–60, 1989.
- SMYTH, G. K.; HUELE, A. F.; VERBYLA, A. P. Exact and approximate REML for heteroscedastic regression. **Statistical Modelling**, v. 1, p. 161–175, 2001.
- SMYTH, G. K.; VERBYLA, A. P. A conditional likelihood approach to REML in generalized linear models. **Journal of The Royal Statistical Society – Series B**, v. 58, p. 565–572, 1996.
- SMYTH, G. K.; VERBYLA, A. P. Adjusted likelihood methods for modeling dispersion in generalized linear models. **Environmetrics**, v. 10, p. 696–709, 1999.
- SMYTH, G. K.; VERBYLA, A. P. Double generalized linear models: approximate REML and diagnostics. In: FRIEDL, H.; BERGHOLD, A.; KAUERMANN, G. (Ed.). **Statistical Modelling: Proceedings of the 14th International Workshop on Statistical Modelling (IWSM14)**. Graz, Austria: Technical University, jul. 1999. p. 66–88.
- STEINER, S. H.; HAMADA, M. Making mixtures robust to noise and mixing measurement errors. **Journal of Quality Technology**, v. 29, p. 441–450, 1997.
- SUGIURA, N. Further analysis of the data by Akaike's information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, v. 7, p. 13–26, 1978.
- TACK, L.; GOSS, P.; WANDEBROEK, M. Efficient Bayesian designs under heteroscedasticity. **Journal of Statistical planning and Inference**, v. 104, p. 469–483, 2002.
- TAGUCHI, G. **Introduction to Quality Engineering**. New York: Unipub-Kraus International Publications, White Plains, 1986.

- VERBYLA, A. P. Modelling variance heterogeneity: residual maximum likelihood and diagnostics. **Journal of the Royal Statistical Society – Series B**, v. 55, p. 493–508, 1993.
- VINING, G.; SCHAUB, D. Experimental designs for estimating both mean and variance functions. **Journal of Quality Technology**, v. 28, p. 135–147, 1996.
- WANG, D.; ZHANG, Z. Variable selection in joint generalized linear models. **Chinese Journal of Applied Probability and Statistics**, v. 25, p. 245–256, 2009.
- WEDDERBURN, R. W. M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. **Biometrika**, v. 61, p. 439–447, 1974.
- WHITTLE, P. Some general points in the theory of optimal experimental design. **Journal of The Royal Statistical Society, Series B**, v. 35, p. 123–130, 1973.
- WONG, W. K. On the equivalence of D and G -optimal designs in heteroscedastic models. **Statistics & Probability Letters**, v. 25, p. 317–321, 1995.
- WU, L.; LI, H. Variable selection for joint mean and dispersion models of the inverse Gaussian distribution. **Metrika**, v. 75, p. 795–808, 2012.
- WU, L.; LI, S.; TAO, Y. Estimation and variable selection for mixture of joint mean and variance models. **Communications in Statistics – Theory and Methods**, p. 1–18, 2020.
- WU, L.; ZHANG, Z.; XU, D. Variable selection for joint mean and dispersion models of the lognormal distribution. **Hacettepe Journal of Mathematics and Statistics**, v. 41, p. 307–320, 2012.
- ZHANG, D. A coefficient of determination for generalized linear models. **The American Statistician**, v. 71, p. 310–316, 2017.

Apêndice A

Apêndice - Capítulo 4

A.1 Sequência para gerar dados normais com parâmetro de dispersão variável

Como especificado na Seção 4.3, dados os vetores $\mathbf{x}_i^t = (x_{1i}, x_{2i}, x_{3i})$, $\mathbf{z}_i^t = (z_{1i}, z_{2i}, z_{3i})$, $\boldsymbol{\beta}^t = (\beta_0, \beta_1, \beta_2, \beta_3)$ e $\boldsymbol{\gamma}^t = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$, para $i = 1, \dots, n$,

- i - calcule $\phi_i = \exp(\gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \gamma_3 z_{3i})$;
- ii - desde que nós estamos considerando $S_i \sim \phi_i \chi_1^2$, gere s_i através de uma distribuição Gama com parâmetro de forma $\frac{1}{2}$ e parâmetro de escala $\frac{1}{2\phi_i}$;
- iii - calcule $\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$ e faça $\mu_i = \eta_i$;
- iv - calcule $y_i = \pm \sqrt{s_i} + \mu_i$, em que o sinal \pm é determinado aleatoriamente.

Observe que, na sequência anterior, também poderíamos ter gerado y_i diretamente de $Y_i \sim \text{Normal}(\mu_i, \phi_i)$.

A.2 Sequência para gerar dados binomiais com parâmetros de dispersão variável

Como especificado na Seção 4.3, dados os valores de k , m e dos vetores $\mathbf{x}_i^t = (x_{1i}, x_{2i}, x_{3i})$, $\mathbf{z}_i^t = (z_{1i}, z_{2i}, z_{3i})$, $\boldsymbol{\beta}^t = (\beta_0, \beta_1, \beta_2, \beta_3)$ e $\boldsymbol{\gamma}^t = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$, para $i = 1, \dots, n$,

- i - calcule $\phi_i = \exp(\gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \gamma_3 z_{3i})$;

- ii - calcule $\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$;
- iii - calcule $\lambda_i = \frac{\exp\{\eta_i\}}{1 + \exp\{\eta_i\}}$ e $\delta_i = \frac{\phi_i - 1}{m - 1}$;
- iv - calcule $a_i = \lambda_i \left(\frac{1 - \delta_i}{\delta_i} \right)$ e $b_i = (1 - \lambda_i) \frac{(1 - \delta_i)}{\delta_i}$;
- v - gere π_i através de uma distribuição Beta com parâmetros de forma a_i e b_i , ou seja $\Pi_i \sim \text{Beta}(a_i, b_i)$;
- vi - gere s_1, \dots, s_ℓ valores de uma distribuição Binomial com parâmetros k e π_i , ou seja, gere os valores s_j , para $j = 1, \dots, \ell$, de $S \sim \text{Binomial}(k, \pi_i)$, onde $\ell = m/k$;
- vii - calcule $y_i = \sum_{j=1}^{\ell} s_j$.

Observe que, conforme discutido na Subseção 4.3.2, na sequência anterior também poderíamos ter gerado y_i diretamente de $Y_i | \pi_i \sim \text{Binomial}(m, \pi_i)$.

A.3 Sequência para gerar dados de Poisson com parâmetros de dispersão variável

Como especificado na Seção 4.3, dados os vetores $\mathbf{x}_i^t = (x_{1i}, x_{2i}, x_{3i})$, $\mathbf{z}_i^t = (z_{1i}, z_{2i}, z_{3i})$, $\boldsymbol{\beta}^t = (\beta_0, \beta_1, \beta_2, \beta_3)$ e $\boldsymbol{\gamma}^t = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$, para $i = 1, \dots, n$,

- i - calcule $\phi_i = \exp(\gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \gamma_3 z_{3i})$;
- ii - calcule $\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$ e faça $\mu_i = \exp(\eta_i)$;
- iii - calcule $\rho_i = \phi_i - 1$. Se $\rho_i > 0$, gere n_i através de uma distribuição de Poisson com parâmetro μ_i / ρ_i ;
- iv - gere s_1, \dots, s_{n_i} valores através de uma distribuição de Poisson com parâmetro ρ_i ;
- v - calcule $y_i = \sum_{j=1}^{n_i} s_j$.

Apêndice B

Apêndice - Capítulo 6

B.1 Prova da Proposição (6.3.1)

Sejam $M_{C_1} = M_C(\theta|\xi_1)$ e $M_{C_2} = M_C(\theta|\xi_2)$ duas matrizes de informação com base nos desenhos ξ_1 and ξ_2 , respectivamente. Seja α um número real tal que $0 < \alpha < 1$. Então, para um dado θ e com $\psi_\theta[M_C(\theta|\xi)] = E_\theta[\ln |M_C(\theta|\xi)|]$, temos que

$$\psi_\theta[\alpha M_{C_1} + (1 - \alpha)M_{C_2}] = E_\theta[\ln |\alpha M_{C_1} + (1 - \alpha)M_{C_2}|] > E_\theta[\ln |M_{C_1}|^\alpha + \ln |M_{C_2}|^{(1-\alpha)}] = \alpha E_\theta \ln |M_{C_1}| + (1 - \alpha)E_\theta \ln |M_{C_2}| = \alpha \psi_\theta(M_{C_1}) + (1 - \alpha)\psi_\theta(M_{C_2}). \quad \square$$

B.2 Prova da Proposição (6.3.2)

A derivada de Fréchet da função ψ_θ em $M_C = M_C(\theta|\xi)$ na direção de $M_{C_u} = M_C(\theta|\xi_u)$ é dada por $F_\psi(M_C, M_{C_u}) = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \{ \psi_\theta [(1 - \varepsilon)M_C + \varepsilon M_{C_u}] - \psi_\theta(M_C) \}$.

Nossa prova é baseada em Silvey (1980), page 21. O critério Bayesiano para a D -optimalidade é $\psi_\theta = E_\theta[\ln |M_C(\theta|\xi)|]$. Nós começamos a prova calculando $F_\psi(M_C, M_{C_u})$ para uma matriz de informação não singular M_C . Seguindo Silvey (1980), calculamos a derivada de Fréchet via a derivada de Gâteaux, dada por $G_{\psi_\theta}(M_C, M_{C_u}) = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \{ \psi_\theta [M_C + \varepsilon M_{C_u}] - \psi_\theta(M_C) \}$. Então,

$$\begin{aligned} G_{\psi_\theta}(M_C, M_{C_u}) &= \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \{ E_\theta \ln |(M_C + \varepsilon M_{C_u})| - E_\theta \ln |M_C| \} = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \{ E_\theta \ln [(M_C + \varepsilon M_{C_u}) |M_C|^{-1}] \} \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \{ E_\theta \ln [(I + \varepsilon M_{C_u} M_C^{-1})] \} = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \{ E_\theta \ln [1 + \varepsilon \operatorname{tr}(M_{C_u} M_C^{-1}) + O(\varepsilon^2)] \} = \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} E_\theta \{ \varepsilon \operatorname{tr}(M_{C_u} M_C^{-1}) + O(\varepsilon^2) \} = E_\theta \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \{ \varepsilon \operatorname{tr}(M_{C_u} M_C^{-1}) + O(\varepsilon^2) \} = E_\theta [\operatorname{tr}(M_{C_u} M_C^{-1})]. \end{aligned}$$

A matriz I é a matriz identidade $t \times t$. É conhecido que (veja Silvey (1980))

$$\begin{aligned} F_{\psi_\theta}(M_C, M_{C_u}) &= G_{\psi_\theta}(M_C, M_{C_u} - M_C) = E_\theta \{ \operatorname{tr}[(M_{C_u} - M_C) M_C^{-1}] \} = E_\theta \{ \operatorname{tr}[(M_{C_u} M_C^{-1} - I)] \} = \\ &= E_\theta \{ \operatorname{tr}[M_{C_u} M_C^{-1}] \} - t. \quad \square \end{aligned}$$

Propriedades sobre determinantes podem ser encontradas em Fedorov (1972).