

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Modelagem simultânea de média e dispersão e aplicações
na pesquisa agronômica**

Afrânio Márcio Corrêa Vieira

Tese apresentada para obtenção do título de Doutor em
Agronomia. Área de concentração: Estatística e Experi-
mentação Agronômica

**Piracicaba
2008**

Afrânio Márcio Corrêa Vieira
Estatístico

Modelagem simultânea de média e dispersão e aplicações na pesquisa agrônoma

Orientadora:

Prof^a. Dra. **CLARICE GARCIA BORGES DEMÉTRIO**

Tese apresentada para obtenção do título de Doutor em
Agronomia. Área de concentração: Estatística e Experi-
mentação Agrônoma

Piracicaba

2008

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Vieira, Afrânio Márcio Corrêa

Modelagem simultânea de média e dispersão e aplicações na pesquisa agrônômica /
Afrânio Márcio Corrêa Vieira. - - Piracicaba, 2008.
117 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 2008.
Bibliografia.

1. Análise de dados longitudinais 2. Análise de variância 3. Conforto térmico das construções
4. Cultura de tecidos 5. Delineamento experimental 6. Inferência bayesiana 7. Modelos lineares
generalizados 8. Verossimilhança I. Título

CDD 519.5
V657m

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

DEDICATÓRIA

Este trabalho é dedicado a José Afrânio Reis Vieira e
Darci Corrêa Vieira (*in memoriam*).

Em memória ao amigo
Pedro Augusto Monteiro Barbosa.

AGRADECIMENTOS

Nesta jornada de 4 anos, encontram-se obstáculos, dificuldades, perdas e momentos em que somente os queridos amigos nos dão força. Também experimentamos a felicidade de ter amigos, mestres e familiares à nossa volta, assim como o apoio de pessoas especiais. Agradeço especialmente minha orientadora, Prof^a. Dra. Clarice Demétrio que acreditou sempre em nosso trabalho, amizade, o estímulo, o companherismo e nos ajudou a alçar vôos maiores (e também por me ensinar a cozinhar!). Ao Prof. Dr. Geert Molenberghs o entusiasmo contagiante, a atenção e a dedicação na co-orientação dos trabalhos desenvolvidos durante o meu estágio de doutoramento na Universiteit Hasselt, Bélgica (Dank u wel, meneer!). Meus sinceros agradecimentos aos docentes, funcionários e pós-graduandos da Universiteit Hasselt, que tornaram agradável e produtiva a minha estadia na Bélgica. Aos docentes e funcionários do departamento de Estatística da Universidade de Brasília, o apoio, a amizade e o investimento em minha formação. Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, o apoio ao nosso trabalho, em especial, à secretária do Programa de Pós-Graduação em Estatística e Experimentação Agrônômica, Sra. Solange Sabadin, a ajuda incansável, o apoio e a gentileza de sempre. Aos professores do Departamento de Ciências Exatas da ESALQ/USP, em especial aos meus professores Dr. Carlos Tadeu Dias, Dr. Edwin Ortega e Dra. Roseli Leandro, que compartilharam conosco seus conhecimentos, amizade e exemplo. Aos professores Dr. Silvio Zocchi, Dra. Sônia Piedade, Dr. Décio Barbin, Dr. Sérgio Moraes e Dr. Vitor Ozaki, as boas conversas, amizade e ensinamentos. Ao Prof. Dr. Jarbas Miranda, a oportunidade e a amizade. Aos professores visitantes do LCE/ESALQ, Dra. Rosemary Bailey, Dr. Bryan Manly, Dr. Bill Venables, Dr. Marcos Rodrigues e Professor John Hinde, que compartilharam conosco seus ensinamentos, a cordialidade e a constante atenção com o nosso aprendizado. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) a bolsa de estudos e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) o financiamento de meu estágio de doutoramento na Bélgica. Ao Programa de Pós-graduação da ESALQ/USP, a infra-estrutura e suporte para o desenvolvimento da pesquisa. Ao Prof. Dr. Iran Silva, do núcleo de pesquisa em ambiência (NUPEA), o apoio, a oportunidade, a lealdade e a amizade incondicional. Ao meu companheiro de pesquisa, Dr. Frederico Vieira, as nossas incansáveis, intermináveis e estimulantes discussões sobre estatística, filosofia, ciência, bem-estar animal e demais assuntos aleatórios. Aos colegas do Instituto Nacional de Meteorologia, em especial ao Prof. Dr. Paulo Sérgio Lucio e Dr. Lauro For-

tes, o incentivo e a oportunidade de aprender. Ao pesquisador Yann Naudet, da DevGen (Belgium) e ao Dr. Steve Powers, do Rothamsted Research Centre (UK), a oportunidade de conhecer esses importantes centros de pesquisa. Aos colegas de mestrado e doutorado da ESALQ/USP, que foram pacientes e amigos. Agradecimento especial para Giovana, Idemauro, César, Julio, Andréa, Ana Maria, Geneville, Denise Viola, Sandra, Osmar e Édila. Para Vanderly, Lucimary, Ângela, Sr. Wilson (sem café eu não sobreviveria!), Renata, Simone Werner e Simone Sartorio (sem doces a vida seria mais triste!), obrigado por terem sido tão amigos com seu colega forasteiro! Minha gratidão aos colegas do núcleo de pesquisa em ambiência (NUPEA/USP) a oportunidade de aprendizagem, a convivência e a amizade, em especial à Dra. Valéria Rodrigues, Dr. José Antônio Barbosa Filho. Que tenhamos muitos trabalhos pela frente! Aos meus amigos de Piracicaba, em especial Carolina Grando, Luiz Otávio, Osvaldo Santana, o carinho e a atenção. Aos amigos de Sampa: Maestro Roberto Anzai e a turma do ComoCanto, Eng. Rubens e família Monteiro de Arruda, minha adorada família de São Paulo, o apoio, o carinho e a presença nos momentos difíceis e nos momentos felizes; aos amigos Prof. Dr. Excelso Ruberti, Prof. Dr. Aldy Silva e Humberto Werneck, o apoio e a amizade. Aos amigos de Brasília: Rauélison Muniz, Lutero Leme, Ligiane e família Leme, que me receberam tão bem, o suporte e o carinho; ao meu eterno amigo e irmão, Dr. Antônio (Foca) Moreira, que me aturou, a presença, a amizade, a lealdade e o suporte durante minha ausência do Brasil. Aos colegas do ABN-AMRO Bank, em especial a Anderson Marcomini, Cristina Higashi e Marcos Pedote, o apoio na difícil decisão de “mudança de rota”, a amizade e o aprendizado.

To my dear friends around the world: Rudolf Sedlacek-Parker, Roberto Guglielmucci and Cary Virtue (EUA); Duncan Oliver and friends, Alain de Proost, Marc van Litsenborgh, Jan (for the wonderful dinners and friendship!), Pascal Delizée and Prof. Hugo Vandekerckhof (Belgium); Mr. Graham Steer, my British brother (UK) ; Herbert van Sluys (The Netherlands); Dr. Babis Pothoulakis (Greece); all the little messages of support that you sent to me, all the advice you gave to help me, and the comfort during these times was so important to keep me walking on this journey. Thank you all!

Ao meu pai, José Afrânio, meus queridos irmãos Frederico e Alfredo, o apoio e o amor ao compreenderem a minha ausência.

A DEUS, sua generosidade, amor incondicional, saúde concedida e proteção.

The scientist is not a person who gives the right answers, he is one who asks the right questions.

Claude Lévi-Strauss

SUMÁRIO

| | |
|---|----|
| RESUMO | 9 |
| ABSTRACT | 10 |
| LISTA DE FIGURAS | 11 |
| LISTA DE TABELAS | 12 |
| 1 INTRODUÇÃO | 13 |
| 2 DESENVOLVIMENTO | 16 |
| 2.1 Revisão metodológica | 16 |
| 2.1.1 Técnicas de Taguchi e algumas alternativas | 16 |
| 2.1.1.1 Filosofia de Taguchi e razão sinal-ruído | 16 |
| 2.1.1.2 Transformação de dados | 18 |
| 2.1.2 Modelos lineares generalizados e extensões | 20 |
| 2.1.2.1 Algoritmo de estimação de parâmetros | 21 |
| 2.1.2.2 Medidas de discrepância em MLG | 22 |
| 2.1.2.3 Quase-verossimilhança | 23 |
| 2.1.2.4 Quase-verossimilhança estendida | 23 |
| 2.1.3 Modelagem simultânea de média e dispersão | 24 |
| 2.1.3.1 Método de Engel | 24 |
| 2.1.3.2 MLGS - Modelos lineares generalizados simultâneos | 25 |
| 2.1.3.3 Estratégia de Análise | 27 |
| 2.1.4 Modelos com efeitos fixo e aleatório | 29 |
| 2.1.4.1 Modelo linear geral misto | 29 |
| 2.1.4.2 Famílias de modelos para dados longitudinais | 32 |
| 2.1.4.3 Modelo linear generalizado misto | 33 |
| 2.1.4.4 Algoritmos para otimização de funções | 36 |
| 2.1.4.5 Algoritmos para integração numérica | 39 |
| 2.2 Avaliação da mortalidade de frangos de corte na condição de espera pré-abate utilizando modelos lineares generalizado duplos | 44 |
| 2.2.1 Introdução | 44 |
| 2.2.2 Materiais e métodos | 45 |

| | |
|--|-----|
| | 8 |
| 2.2.2.1 Materiais | 45 |
| 2.2.2.2 Métodos | 47 |
| 2.2.3 Resultados e discussão | 50 |
| 2.2.4 Conclusões | 54 |
| 2.3 Modelo linear generalizado duplo: uma perspectiva Bayesiana | 55 |
| 2.3.1 Introdução | 55 |
| 2.3.2 Modelagem simultânea de média e dispersão | 56 |
| 2.3.3 Modelo linear generalizado duplo Bayesiano | 59 |
| 2.3.3.1 Modelo para medidas com distribuição normal | 59 |
| 2.3.3.2 Modelo para proporções com superdispersão | 60 |
| 2.3.3.3 Aplicação: experimento de cultura de tecidos | 67 |
| 2.3.4 Conclusões | 70 |
| 2.4 Modelos para dados binários longitudinais com superdispersão | 74 |
| 2.4.1 Introdução | 74 |
| 2.4.2 Desenvolvimento | 75 |
| 2.4.2.1 Um estudo clínico em dermatologia | 75 |
| 2.4.2.2 Modelos lineares generalizados e extensões | 76 |
| 2.4.2.2.1 Modelos lineares generalizados | 76 |
| 2.4.2.2.2 Modelos para superdispersão | 78 |
| 2.4.2.2.3 Modelos com efeitos fixos e aleatórios | 79 |
| 2.4.2.2.4 Modelos combinando superdispersão e estrutura longitudinal | 81 |
| 2.4.2.3 Estimação | 82 |
| 2.4.2.3.1 Modelos lineares generalizados | 82 |
| 2.4.2.3.2 Modelos lineares generalizados mistos | 83 |
| 2.4.2.3.3 Modelos combinados | 84 |
| 2.4.2.4 Aplicação e resultados | 93 |
| 2.4.3 Considerações finais | 97 |
| 3 COMENTÁRIOS FINAIS | 99 |
| REFERÊNCIAS | 101 |
| ANEXOS | 108 |

RESUMO

Modelagem simultânea de média e dispersão e aplicações na pesquisa agrônômica

Diversos delineamentos experimentais que são aplicados correntemente tomam como base experimentos agrônômicos. Esses dados experimentais são, geralmente, analisados usando-se modelos que consideram uma variância residual constante (ou homogênea), como pressuposto inicial. Entretanto, esta pressuposição mostra-se relativamente forte quando se está diante de situações para as quais fatores ambientais ou externos exercem considerável influência nas medidas experimentais. Neste trabalho, são estudados modelos para a média e a variância, simultaneamente, com a variância estruturada de duas formas: (i) por meio de um preditor linear, que permite incorporar variáveis externas e fatores de ruído e (ii) por meio de efeitos aleatórios, que permitem acomodar tanto o efeito longitudinal quanto o efeito de superdispersão, no caso de medidas binárias repetidas no tempo. A classe de modelos lineares generalizados duplos (MLGD) foi aplicada a um estudo observacional que consistiu em medir a mortalidade de frangos de corte no fim da condição de espera pré-abate. Nesse problema, é forte a evidência de que alguns fatores influenciam a variabilidade, e consequentemente, diminuem a precisão das análises inferenciais. Outro problema agrônômico relevante, associado à horticultura, são os experimentos de cultura de tecidos vegetais, em que o número de explantes que regeneram são contados. Como esse tipo de experimento apresenta um grande número de parâmetros a serem estimados, comparado ao tamanho da amostra, os modelos existente podem gerar estimativas questionáveis ou até levar a conclusões errôneas, uma vez que são baseados em grandes amostras para se fazer inferência estatística. Foi proposto um modelo linear generalizados duplo, para os dados de proporções, de uma perspectiva Bayesiana, visando a análise estatística sob pequenas amostras e a incorporação do conhecimento especialista no processo de estimação dos parâmetros. Um problema clínico, que envolve dados binários medidos repetidamente no tempo é apresentado e são propostos dois modelos que acomodam o efeito da superdispersão e a dependência longitudinal das medidas, utilizando-se efeitos aleatórios. Foram obtidos resultados satisfatórios nos três problemas estudados. Os MLGD permitiram identificar os fatores associados à mortalidade das aves de corte, o que permitirá minimizar perdas e habilitar os processos de manejo, transporte e abate aos critérios de bem-estar animal e exigências da comunidade europeia. O MLGD Bayesiano permitiu identificar o genótipo associado ao efeito de superdispersão, aumentando a precisão da inferência de seleção de variedades. Dois modelos combinados foram propostos logit-normal-Bernoulli-beta e o probit-normal-Bernoulli-beta, que acomodaram satisfatoriamente a superdispersão e a dependência longitudinal das medidas binárias. Esses resultados reforçam a importância de se modelar a média e a variância conjuntamente, o que aumenta a precisão na pesquisa agrônômica, tanto em estudos experimentais quanto em estudos observacionais.

Palavras-chave: Modelos lineares generalizados; Heterocedasticidade; Quase-verossimilhança estendida; Análise Bayesiana de dados; Bem-estar animal; Cultura de tecidos; Modelos mistos; Análise de dados longitudinais

ABSTRACT

Joint modeling of mean and dispersion and applications to agricultural research

Several experimental designs that are currently applied are based on agricultural experiments. These experimental data are, usually, analysed with statistical models that assume constant residual variance (or homogeneous), as basic assumption. However, this assumption shows hard to stand for, when environmental or external factors exert strong influence over the measurements. In this work, we study the joint modelling for the mean and the variance, the latter being structured on two ways: (i) through a linear predictor, which allows the incorporation of external variables and/or noise factors and (ii) by the use of random effects, that accommodate jointly the possible overdispersion effect and the dependence of longitudinal data in the case of binary measurements taken over time. The class of double generalized linear models (DGLM) was applied to an observational study where the poultry mortality was measured in the preslaughter operations. With this situation, it can be observed that there is a strong influence from some environmental factors over the variability observed, and consequently, this reduces the precision of the inferential analysis. Another relevant agricultural problem, related to horticulture, is the tissue culture experiments, where the number of regenerated explants is counted. Usually, this kind of experiment use a large number of parameters to be estimated, when compared with the sample size. The current frequentist models are based on large samples for statistical inference and, under this experimental condition, can generate unreliable estimates or even lead to erroneous conclusions. A double generalized linear model was proposed to analyse proportion data, under the Bayesian perspective, which can be applied to small samples and can incorporate expert knowledge into the parameter estimation process. One clinical research, that measured binary data repeatedly through the time is presented and two models are proposed to fit the overdispersion effect and the dependence of longitudinal measurements, using random effects. It was obtained satisfactory results under these three problems studied. the DGLM allowed to identify factors associated with the poultry mortality, that will allow to minimize loss and improve the process, since the catching until lairage on slaughterhouse, agreeing with animal welfare criteria and the European community rules. The Bayesian DGLM allowed to identify the genotype associated with the overdispersion effect, increasing the precision on the inference about varieties selection. Two combined models were proposed, a logit-normal-Bernoulli-beta and a probit-normal-Bernoulli-beta, which have both addressed the overdispersion effect and the longitudinal dependence of the binary measurements. These results reinforce the importance to modelling mean and dispersion jointly, as a way to increase the precision of agricultural experimentation, be it on experimental studies or observational studies.

Keywords: Generalized linear models; Heterocedasticity; Extended quasi-likelihood; Bayesian data analysis; Animal welfare; Tissue culture; Mixed models; Longitudinal data analysis

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1 - Fluxograma das variáveis respostas envolvidas nas operações pré-abate | 46 |
| Figura 2 - Número de aves mortas estimado em função da interação entre a distância granja-abatedouro e o tempo de espera | 53 |
| Figura 3 - Modelo gráfico acíclico direcionado para o modelo linear generalizado duplo binomial Bayesiano | 62 |
| Figura 4 - Média do número de explantes regenerados em função do meio de cultura e do tipo de Explante | 68 |
| Figura 5 - Desvio padrão do número de explantes regenerados em função do meio de cultura e do tipo de Explante | 69 |
| Figura 6 - Média do número de explantes regenerados em função do meio de cultura e do tipo de Explante para o modelo 5 | 71 |
| Figura 7 - Desvio-padrão do número de explantes regenerados em função do meio de cultura e do tipo de Explante para o modelo 5 | 72 |
| Figura 8 - Proporção de severidade de TDO para os tratamentos estudados | 76 |
| Figura 9 - Tendência populacional esperada para o percentual de severidade dos dois tratamentos, com base nos parâmetros da Tabela 8 | 97 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1 - Comparação entre o modelo linear generalizado e Quase-Poisson | 51 |
| Tabela 2 - Modelo linear generalizado duplo (MLGD), para a média e dispersão | 51 |
| Tabela 3 - Número de explantes (y_i) de maçã que regeneraram, dentre 25, considerando 16 placas de Petri | 70 |
| Tabela 4 - Medidas de qualidade de ajuste e complexidade dos modelos MLGD Binomial Bayesiano ajustados | 71 |
| Tabela 5 - Estatísticas sumárias das amostras das densidades marginais dos parâmetros do modelo 5, calculadas com base em 4.200 observações | 72 |
| Tabela 6 - Resultados dos ajustes dos modelos MLG, beta-binomial, MLGM e o modelo combinado, assumindo função de ligação logit | 94 |
| Tabela 7 - Resultados dos ajustes dos modelos MLG, beta-binomial, MLGM e o modelo combinado, assumindo função de ligação probit | 95 |
| Tabela 8 - Resultados do ajuste do modelo combinado probit-normal-Bernoulli-beta | 96 |

1 INTRODUÇÃO

O papel principal da Estatística Experimental na pesquisa científica é garantir que evidências factuais baseadas em experimentos, relacionadas com hipóteses pré-estabelecidas, sejam obtidas livres de viés ou influências externas, que comprometam as conclusões do estudo. A pesquisa agrônômica foi o berço da estatística experimental no início do século XX e diversos delineamentos experimentais que são aplicados no dia-a-dia, baseados em métodos conhecidos de longa data, tomam como base experimentos agrônômicos. Estes métodos assumem que os dados experimentais serão analisados através de modelos que consideram uma variância residual constante (ou homogênea), como pressuposto inicial. Entretanto, essa pressuposição mostra-se, relativamente, forte quando se está diante de situações para as quais fatores ambientais ou externos exercem considerável influência nas medidas experimentais.

Um problema abordado naquela época era o aspecto da heterogeneidade inerente aos problemas experimentais, cujo controle local era feito através da aleatorização na alocação de tratamentos dentro de blocos. Na mesma época em que se propunham novas formas de se planejarem experimentos, Fisher (1922) propôs o *Princípio da Verossimilhança* que se tornou a base da estatística inferencial moderna. Entretanto, somente na década de 60 que a abordagem da modelagem explícita da média e da variância, simultaneamente, começou a ser explorada. Uma abordagem baseada no método da máxima verossimilhança foi apresentada por Harvey (1976), descrita como uma estrutura geral para um modelo de regressão para respostas com distribuição normal e uma variância estruturada na forma de um preditor linear.

Na década de 80, o Eng. Genichi Taguchi abordava o problema da heterocedasticidade dentro da experimentação industrial, lançando o novo paradigma para a *gestão da qualidade*, denominado *planejamento robusto*, que visava otimizar processos industriais e reduzir a variabilidade, simultaneamente. Naquela época, os EUA e alguns países da Europa começavam a despertar para o desenvolvimento da qualidade de processos, produtos e serviços. Apesar da contribuição valiosa e relevante de Taguchi para a Engenharia da Qualidade, alguns de seus métodos foram criticados e bastante discutidos, trazendo à tona, novamente, a questão de se modelar explicitamente a variância conjuntamente com a média. Uma nota de revisão de Pregibon (1984) retoma o problema de modelagem dentro do “framework” dos modelos lineares generalizados (MLG). O modelo proposto por Smyth (1989), denominado *modelo linear generalizado duplo (MLGD)*, e o *modelo linear ge-*

neralizado simultâneo (MLGS), proposto por Nelder e Lee (1991) visam à modelagem de efeitos dos fatores de *controle* e dos fatores de *variação* ou *ruído* (também denominados como efeitos de locação e dispersão, respectivamente).

Apesar da evolução dessa classe de modelos, poucas são as aplicações que utilizam desse tipo de modelagem estatística na experimentação agrônômica. Essas metodologias têm potencial de utilização não só na experimentação industrial mas também na pesquisa agrônômica, em estudos observacionais retrospectivos, em biotecnologia e nas ciências biológicas em geral, cujos objetos de estudo estão inerentemente sujeitos a grande variabilidade. Um exemplo de estudo observacional desse tipo consiste em medir a temperatura retal e o número de frangos de corte mortos no fim da condição de espera pré-abate. Identificar e estudar os fatores que influenciam essas medidas é de grande importância para minimizar perdas e habilitar os processos de manejo, transporte e abate aos critérios de bem-estar animal e exigências da comunidade europeia. No entanto, é forte a evidência de que alguns fatores influenciam a variabilidade, e conseqüentemente, diminuem a precisão das análises inferenciais.

Outro problema agrônômico relevante, associado à horticultura, são os experimentos de cultura de tecidos vegetais, em que o número de explantes que regeneraram são contados. Na modelagem estatística para esse tipo de experimento, o número de parâmetros a serem estimados é, relativamente, grande quando comparado ao tamanho da amostra. Sob esse tipo de situação, mesmo os modelos existentes, que são baseados em grandes amostras para se fazer inferência estatística, podem gerar estimativas questionáveis ou até levar a conclusões errôneas.

Outro problema potencial ocorre quando medidas binárias são registradas, repetidamente, ao longo do tempo, e apresentam uma variabilidade gerada por fatores não controláveis. Uma situação agrônômica que caracteriza esse problema seria um estudo comportamental de suínos em uma granja, em que o número de animais é grande e muitos deles pertencem à mesma leitegada, além de cada animal ter seu comportamento classificado como ‘estressado’ ou não ‘estressado’, com base na pressão sonora emitida pelos animais em diferentes horários do dia.

Esses e outros problemas motivam a pesquisa de métodos que modelem simultaneamente a média e a dispersão, aplicados à pesquisa agrônômica. Dessa forma, os objetivos desse trabalho são:

- i. revisar e aplicar os modelos lineares generalizados simultâneos (MLGS) e os modelos lineares generalizados duplos (MLGD) a problemas agronômicos;
- ii. apresentar uma proposta Bayesiana para o modelo MLGD/MLGS para análise de dados na forma de proporções;
- iii. apresentar propostas para modelar dados binários longitudinais com superdispersão.

O problema da mortalidade de frangos de corte na condição pré-abate é estudado utilizando os MLGD no capítulo 2.2. A análise dos dados de um experimento de cultura de tecidos é apresentado na seção 2.3, sendo que o processo de estimação é feito do ponto de vista Bayesiano. Dois modelos para problema da análise de dados binários longitudinais e com variação extra (ou *superdispersão*) são propostos na seção 2.4, exemplificados não em um problema agronômico mas em um problema clínico, mas que se estende a outras situações agronômicas. No capítulo a seguir, será apresentada uma revisão dos principais métodos utilizados nesse trabalho.

2 DESENVOLVIMENTO

2.1 Revisão metodológica

Na seção 2.1.1, descrevem-se as primeiras abordagens de Taguchi para o problema da modelagem da média e da variância, simultaneamente, e alternativas baseadas em transformações de dados. Na seção 2.1.2, são descritos os modelos lineares generalizados (MLG), o algoritmo de estimação, suas extensões e medidas de discrepâncias para avaliação de modelos ajustados. Na seção 2.1.3, são descritas algumas abordagens para modelagem simultânea de média e dispersão, em especial a que foi proposta por Nelder e Lee (1991). A revisão dos modelos lineares mistos e suas extensões estão apresentadas na seção 2.1.4. Alguns algoritmos para otimização, utilizados em conjunto com o método da máxima verossimilhança, são apresentados na seção 2.1.4.4 e alguns algoritmos para integração numérica são revisados na seção 2.1.4.5.

2.1.1 Técnicas de Taguchi e algumas alternativas

2.1.1.1 Filosofia de Taguchi e razão sinal-ruído

Em 1980, o Eng. Genichi Taguchi apresentou uma abordagem dos diferentes delineamentos experimentais robustos com três objetivos, a saber:

1. projeto de produtos e processos que fosse robusto às condições ambientais;
2. projeto e desenvolvimento de produtos e processos que fossem robustos a componentes de variação (fatores que influenciam na variação do parâmetro de interesse);
3. redução de variação em torno do valor nominal (valor-alvo).

Segundo Montgomery (2004), o termo “robusto” pode ser entendido como

a característica de um produto ou processo que é consistente com o valor-alvo estabelecido e que é relativamente insensível a fatores de ruído ou que são difíceis de controlar no dia-a-dia.

Um ponto-chave na “Filosofia de Taguchi” (expressão que o próprio Taguchi rejeita mas que ficou amplamente difundida) é a *redução da variabilidade*. Considerando que uma característica

de engenharia de processo ou produto tem um valor nominal, a meta é reduzir a variabilidade dessa característica em torno dele. Ele propôs uma *função-perda* para mensurar o distanciamento do parâmetro de interesse do valor nominal, através do raciocínio que se segue (TAGUCHI, 1985). Considera-se que a perda $L \mapsto \mathfrak{R}^+ : y \mapsto L(y)$ é igual a zero quando a característica de qualidade y tem distância do valor nominal T igual a zero, ou seja, coincide com o valor nominal T e portanto, $L(T) = 0$. Quando isso ocorre, $L(y)$ é zero e mínimo, e, portanto, a derivada de primeira ordem de L é $L'(T) = 0$. Uma aproximação da função-perda L , em torno do ponto $y = T$ pode ser obtida utilizando-se até o terceiro termo da expansão em séries de Taylor,

$$L(y) \doteq L(T) + \frac{L'(T)}{1!}(y - T) + \frac{L''(T)}{2!}(y - T)^2$$

sendo que o termo constante e o termo linear tornam-se iguais a zero e supondo que os termos iguais ou superiores à terceira ordem são desprezíveis. Tem-se, então, a aproximação quadrática da função perda dada por

$$L(y) \approx k(y - T)^2, \quad (1)$$

sendo que T é o valor nominal, y é o valor do parâmetro em estudo e k é uma constante. A partir dessa função, Taguchi propôs as seguintes Razões Sinal-Ruído (*Signal-to-Noise Ratios*):

$$\begin{aligned} SN_L &= -10 \log \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{y^2} \right), \\ SN_S &= -10 \log \left(\frac{1}{n} \sum_{i=1}^n y^2 \right) e \\ SN_T &= 10 \log \frac{\bar{y}^2}{s^2} \end{aligned}$$

sendo que os índices em SN significam ‘Maior é Melhor’ (L), ‘Menor é Melhor’ (S) e ‘No Alvo é Melhor’ (T). Considerando a razão sinal-ruído como variável resposta, em que \bar{y} e s^2 representam, respectivamente, a média e a variância dos tratamentos gerados por delineamentos fatoriais sobre os fatores de controle e de ruído, pode-se conduzir uma análise de dados e/ou uma análise estatística. Algumas críticas a essa abordagem foram apresentadas na literatura estatística. Box (1988) reescreve SN_T como

$$SN_T = 10 \log \frac{\bar{y}^2}{s^2} = -10 \log \gamma^2,$$

sendo γ o coeficiente de variação s/\bar{y} . Logo, SN_T é proporcional a $\log(\bar{y})$ e essa última é mais eficiente para exprimir as fontes de variação associadas aos fatores de ruído. Schmidt e Boudot¹ (1989, apud MONTGOMERY, 2004) conduziram um estudo baseado em simulação Monte Carlo em que avaliaram a efetividade das razões sinal-ruído na detecção de efeitos de dispersão. Concluíram que SN_L e SN_S são ineficazes para detectar efeitos de dispersão significativos, apesar de detectarem os efeitos que influenciam a média (o que pode ser feito modelando-se diretamente os efeitos dos fatores em estudo sobre a média). Outra crítica recai sobre modelar uma função do coeficiente de variação, que não tem propriedades distribucionais probabilísticas evidentes, tornando complexa a inferência estatística, quando a análise sobre \bar{y} e s^2 é mais intuitiva, natural e mais simples. Um ponto enfatizado por Nelder e Lee (1991) é que quando se utiliza SN , considera-se $\log(y)$ como a única transformação capaz de atender aos critérios de **parcimônia** (modelos mais simples são preferíveis aos mais complexos) e **separação** (eliminação da dependência da média e parâmetro de dispersão), o que não é verdade. Essa transformação somente faz sentido se $\log(Y) \sim N(\mu, \sigma^2)$. Além disso, trata-se de uma transformação de dados não selecionada com base nos dados e sim de forma arbitrária e prévia à análise dos dados. Isto vai contra a todo o raciocínio estatístico e ao procedimento científico vigente. Maiores discussões sobre as propriedades e os problemas relacionados à utilização de razões sinal-ruído estão descritas em León et al. (1987). A partir desse artigo, a comunidade estatística buscou estudar propostas consistentes e abrangentes para a análise desse tipo de experimento, cujo grande interesse é modelar o efeito de dispersão, conjuntamente, com a modelagem dos efeitos que afetam uma característica de qualidade.

2.1.1.2 Transformação de dados

Considere o modelo de regressão linear múltipla

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

sendo que \mathbf{Y} é o vetor representando a variável resposta, \mathbf{X} a matriz de planejamento do experimento, $\boldsymbol{\beta}$ o vetor de parâmetros representando os efeitos dos fatores em estudo que serão estimados e $\boldsymbol{\epsilon}$ o vetor de erros aleatórios. Esse tipo de modelo é largamente utilizado em estudos de otimiza-

¹SCHMIDT, S.R., BOUDOT, J.R. A Monte Carlo simulation study comparing effectiveness of Signal-to-Noise ratios and other methods for identifying dispersion effects. In: The Rocky Mountain Quality Conference, 1989.

ção de processos industriais e desenvolvimento de novos produtos. Quase sempre são assumidas as seguintes suposições: (i) $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$; (ii) σ^2 constante e (iii) $\text{Cov}(\epsilon_i, \epsilon_j) = 0, \forall i \neq j$.

Quando alguma dessas suposições não são atendidas (o que acontece em muitas situações reais na análise estatística de dados), uma abordagem seria selecionar uma transformação que atenderia aproximadamente às duas primeiras suposições descritas. Uma família importante de transformações do tipo y^λ é aquela proposta por Box e Cox (1964), dada por

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}}, & \lambda \neq 0 \\ \dot{y} \ln y, & \lambda = 0 \end{cases}$$

sendo que λ é a potência de transformação que estabiliza a variância e/ou normaliza, aproximadamente, a distribuição e \dot{y} é a média geométrica de y . Entretanto, acontece, freqüentemente, que a transformação de dados selecionada atende a um critério ou a outro e, dificilmente, a ambos os critérios de parcimônia e separação. Box (1988) sugere a utilização do *Lambda Plot*, priorizando o critério de parcimônia. Esse gráfico consiste em estabelecer uma grade de valores para λ e para cada valor ajustar um modelo de regressão e tomar o valor da estatística t de cada um dos efeitos. Assim, uma linha é plotada para cada efeito estudado e pode-se notar que transformações inadequadas podem gerar efeitos de interações falsamente significativas e efeitos de dispersão inflacionados. Entretanto, Shoemaker et al. (1988) pontuaram que o critério de separação é mais importante, por ser uma condição necessária para um procedimento de ajuste em dois passos (modelagem da dispersão para, então, modelagem dos efeitos da média).

Uma desvantagem do método proposto por Box é que se trata de uma ferramenta exploratória, uma vez que a média e a dispersão são ajustadas separadamente e baseadas em estatísticas sumárias (\bar{y} e s). Mesmo assim, o *Lambda Plot* pode ser utilizado em estudos do tipo exploratório como em delineamentos fatoriais e fatoriais fracionados. Outros métodos gráficos para seleção de transformação na presença de fatores de ruído são o *Mean-Variance Plot* (NAIR; PREGIBON, 1988) e o *Log-log Regression* (LOGOTHETIS, 1990).

2.1.2 Modelos lineares generalizados e extensões

Os modelos lineares tiveram seu grande *upgrade* com a publicação do relevante artigo de Nelder e Wedderburn (1972) (extensamente estudado e descrito em McCullagh; Nelder, 1989), no qual a nova classe de Modelos Lineares Generalizados (MLG) faz a unificação de várias técnicas e metodologias estatísticas em uma única moldura. Considerando uma amostra com n observações independente e identicamente distribuídas, \mathbf{X} uma matriz com $p + 1$ colunas que representam p variáveis explicativas e \mathbf{y} um vetor de observações amostrado de Y , são definidos os 3 componentes de um MLG:

1. Y tem distribuição probabilística como membro da **Família Exponencial** de distribuições, com uma função de probabilidade ou função densidade de probabilidade (para variáveis aleatórias discretas e contínuas, respectivamente)

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{1}{a(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\},$$

com média

$$E(Y_i) = \mu_i = b'(\theta_i)$$

e

$$\text{Var}(Y_i) = a_i(\phi) b''(\theta_i) = a_i(\phi) V(\mu_i) = a_i(\phi) V_i,$$

sendo θ_i o parâmetro canônico, $a_i(\phi) = \phi/w_i$, ϕ o parâmetro de dispersão, w_i um peso “a priori” e V_i a *função de variância* dada por $V_i = d\mu_i/d\theta_i$.

Algumas distribuições-membros da Família Exponencial são a Normal, Poisson, Binomial, Gama, Gaussiana Inversa, Binomial Negativa (com o parâmetro k fixado), algumas distribuições truncadas, etc.

2. As covariáveis e os fatores em \mathbf{X} estão relacionados de forma aditiva no **preditor linear**

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta};$$

Todas as configurações admitidas para os bem conhecidos Modelos Lineares podem ser aqui utilizados.

3. Uma **Função de Ligação** monotônica e diferenciável $g(\cdot)$, que liga o preditor linear $\boldsymbol{\eta}$ à média de Y , e, portanto,

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \Rightarrow \mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}).$$

2.1.2.1 Algoritmo de estimação de parâmetros

O método de estimação para o vetor parâmetros $\boldsymbol{\beta}$, proposto por Nelder e Wedderburn (DEMÉTRIO, 2002), é o método da máxima verossimilhança. Utilizando o logaritmo da função de verossimilhança da família exponencial,

$$\begin{aligned} l(\boldsymbol{\theta}|\mathbf{y}) &= \log \left[\prod_{i=1}^n f(y_i; \theta_i, \phi) \right] = \sum_{i=1}^n l(\theta_i; y_i) \\ &= \sum_{i=1}^n \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\}; \end{aligned}$$

o j -ésimo componente do vetor escore pode ser descrito como

$$U_j = \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^n \frac{dl_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^n \frac{1}{a_i(\phi)} (y_i - \mu_i) \frac{1}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} \quad j = 1, \dots, p. \quad (2)$$

Nota-se que somente μ_i and $V(\mu_i)$ estão sendo utilizadas no processo de estimação, não sendo necessária a forma funcional completa da família exponencial. Para obter estimativas de Máxima Verossimilhança de $\boldsymbol{\beta}$, o seguinte algoritmo (*Iterative Reweighted Least Square - IRLS*) deve ser executado, considerando a m -ésima iteração, até que um critério de convergência seja atingido:

Passo 1:

$$\eta_i^{(m)} = \sum_{j=1}^p x_{ij} \beta_j^{(m)}; \quad \mu_i^{(m)} = g^{-1}(\eta_i^{(m)});$$

Passo 2:

$$q_i^{(m)} = \eta_i^{(m)} + (y_i - \mu_i^{(m)}) g'(\mu_i^{(m)}); \quad \mathbf{Q}^{(m)} = [q_1^{(m)}, q_2^{(m)}, \dots, q_n^{(m)}]^T$$

$$W_i^{(m)} = \frac{w_i}{V(\mu_i^{(m)}) [g'(\mu_i^{(m)})]^2}; \quad \mathbf{W}^{(m)} = \text{diag} \{ W_i^{(m)} \};$$

Passo 3:

$$\boldsymbol{\beta}^{m+1} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{Q}^{(m)}$$

Passo 4: Se o ciclo iterativo convergir, $\tilde{\beta} = \beta^{(m)}$. Caso contrário, vá para o Passo 1 utilizando $\beta^{(m)}$ no passo $m + 1$.

2.1.2.2 Medidas de discrepância em MLG

Como medida de discrepância entre o valor observado na amostra e aquele predito pelo modelo estimado, Nelder e Wedderburn (1972) propuseram a seguinte medida de desvio:

$$\frac{D}{\phi} = -2 \sum_{i=1}^n \{l(\hat{\mu}_i, y_i) - l(y_i, y_i)\},$$

sendo $l(\hat{\mu}_i, y_i)$ o logaritmo da função de verossimilhança calculado utilizando y_i e uma estimativa de μ_i e $l(y_i, y_i)$ é o logaritmo função de verossimilhança calculado utilizando y_i também como uma estimativa de μ_i . Essa medida é denominada *scaled deviance* e

$$D = \sum_{i=1}^n d_i^2,$$

é a *deviance residual*, sendo d_i^2 o *componente de deviance*, dada por

$$d_i^2 = -2 \int_y^{\mu_i} \frac{y - u_i}{V(u_i)} du.$$

Assumindo $V(\mu_i)$ e μ_i para cada distribuição pertencente à família exponencial na forma canônica, obtêm-se diferentes expressões para a *deviance*. Por exemplo, para a distribuição Normal, é fácil ver que assumindo $V(\mu_i) = 1$ e $\mu_i = \eta_i$, tem-se que D é igual à soma de quadrados dos resíduos.

Para o cálculo dos resíduos do modelo ajustado, têm-se as seguintes formas:

1. resíduo componente da *deviance*:

$$r_{d_i} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i^2};$$

2. resíduo de Pearson:

$$r_{P_i} = \frac{y_i - \hat{\mu}_i}{V(\hat{\mu}_i)}.$$

2.1.2.3 Quase-verossimilhança

Em algumas situações, a distribuição probabilística da qual os dados são provenientes não é conhecida. Entretanto, se a relação entre a média e a variância for conhecida, essa informação pode ser utilizada no processo de estimação. Wedderburn (1974) propôs o método da “quase-verossimilhança” (QV), que relaxa a suposição de que a distribuição, necessariamente, pertença à família exponencial.

O logaritmo da função de quase-verossimilhança Q é definido como:

$$\frac{\partial Q(y_i, \mu_i)}{\partial \mu} = \frac{y_i - \mu_i}{V(\mu_i)} \Rightarrow Q(y_i, \mu_i) = \int^{\mu} \frac{y - u}{V(u)} du + f(y_i).$$

Nota-se que a função escore é exatamente a mesma dos MLG, fornecendo estimativas de quase-verossimilhança no lugar de estimativas de máxima verossimilhança. Nesse caso, ϕ é considerado fixo.

A estimação pelo método de QV pode ser usada na modelagem de dados na forma de contagens e proporções. Esse tipo de dados estão sujeitos à superdispersão (HINDE; DEMÉTRIO, 1998), em que a variação observada é muito maior do que aquela prevista pelos modelos baseados nas distribuições Poisson e Binomial. Vieira, Hinde e Demétrio (2000) utilizaram o método QV para modelagem de dados de proporções inflacionadas de zeros, em que a distribuição dos dados era desconhecida e foram consideradas algumas funções de variância em particular, no processo de estimação. Uma restrição ao método de QV é que o parâmetro de dispersão ϕ é assumido constante. Logo, esse método não permite a estimação de efeitos associados ao parâmetro de dispersão, o que demanda o seguinte método de estimação mais geral.

2.1.2.4 Quase-verossimilhança estendida

Para as situações em que se deseja modelar o parâmetro de dispersão ϕ dependente de possíveis covariáveis, ao invés de considerá-lo fixo, Nelder e Pregibon (1987) propuseram o método da *quase-verossimilhança estendida* (QVE). A função QV pode ser estendida para QVE utilizando

$$-2Q^+(y_i, \mu_i) = \sum_{i=1}^n \left\{ \frac{d_i}{\phi_i} + \log [2\pi\phi_i V(y_i)] \right\}. \quad (3)$$

Para as distribuições Normal e Gaussiana Inversa o logaritmo da função de quase-verossimilhança estendida gera funções do logaritmo de verossimilhanças exatas e boas aproximações são obtidas para as demais distribuições. Essa forma de Q^+ é baseada em aproximações de ponto de sela (*saddlepoint approximations*), utilizadas para estudos assintóticos de estimadores de máxima verossimilhança, mas aqui essa função é aplicada apenas para observações individuais sob os modelos lineares generalizados. O método QVE é utilizado nas situações em que a variância apresenta estruturas mais complexas com dispersão variante (LEE; NELDER; PAWITAN, 2006; LEE; NELDER, 1998).

Outra alternativa para a estimação do parâmetro de dispersão seria utilizar a *pseudo-verossimilhança* (PV), descrita como

$$-2P(\phi_i) = \sum_{i=1}^n \left\{ \frac{r_{Pi}^2}{\phi_i} + \log [2\pi\phi_i V(\hat{\mu}_i)] \right\}, \quad (4)$$

em que r_P^2 é o resíduo de Pearson. Nota-se que a expressão da PV é bastante similar à expressão da QVE, em que o componente de *deviance* d_i^2 é substituído por r_{Pi}^2 e $\hat{\mu}_i$ é utilizado no lugar de y_i em $V(\cdot)$. Entretanto, essa função depende de ϕ e não dos parâmetros associados às covariáveis, não podendo ser utilizada para estimar conjuntamente μ . Apesar de esse método produzir estimativas consistentes, o método QVE gera estimadores consistentes e assintoticamente eficientes, quando ϕ não é muito grande. Já o método PV tende a produzir estimativas menos eficientes (LEE; NELDER, 1988). A estimação por PV é utilizada também em modelos com matrizes de covariâncias estruturadas, como em análise de dados longitudinais (MOLENBERGHS; VERBEKE, 2005). Uma análise comparada entre os métodos QVE e PV no contexto do MLG está descrita em LEE et al. (2006).

2.1.3 Modelagem simultânea de média e dispersão

2.1.3.1 Método de Engel

Uma proposta intermediária para a modelagem efetiva da média e da variância foi proposta por Engel (1992). Essa abordagem considera m réplicas para cada uma das i condições experimentais e \bar{y}_i e s_i^2 como variáveis respostas para o seguinte par de modelos lineares generalizados (NELDER; PREGIBON, 1987):

$$\begin{cases} \eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \\ \zeta_i = h(\phi_i) = \mathbf{z}_i^T \boldsymbol{\gamma}, \end{cases}$$

sendo η_i o preditor linear para a média μ_i , \mathbf{x}_i o vetor-linha de \mathbf{X} , $g(\cdot)$ a função de ligação entre o preditor linear e a média, ζ_i o preditor linear que associa os fatores de ruído à dispersão ϕ_i (representada pelos componentes da *deviance* ou pelos componentes da estatística X^2 de Pearson), \mathbf{z}_i o vetor-linha de \mathbf{Z} e $h(\cdot)$ a função de ligação entre o preditor linear e a dispersão. Engel assumiu $g(\cdot) = I(\cdot)$, em que $I(\cdot)$ é a função identidade e $h(\cdot) = \log(\cdot)$.

Duas restrições desaconselham a utilização dessa abordagem. Uma é o fato de que esse modelo utiliza estatísticas sumárias, ao invés de cada observação individual. Isto torna a modelagem um pouco mais complexa e restrita. A verificação da qualidade de ajuste do modelo, através de análise residual não é permitida, mesmo quando m é relativamente grande.

2.1.3.2 MLGS - Modelos lineares generalizados simultâneos

O objetivo do método de modelagem proposto por Nelder e Lee (1991) atende ao objetivo do planejamento robusto proposto por Taguchi, cujo objetivo é encontrar os níveis de fatores experimentais que minimizem a variância experimental e mantenham a média em um valor nominal. Utilizando-se a abordagem MLG, é natural pensar nos parâmetros μ e ϕ como medidas de performance para os fatores de controle e ruído, respectivamente.

Considerando que $E(Y_i) = \mu_i$ e $\text{Var}(Y) = \phi_i V(\mu_i)$, tem-se a seguinte formulação MLG:

$$\begin{cases} \eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \\ \zeta_i = h(\phi_i) = \mathbf{z}_i^T \boldsymbol{\gamma}, \end{cases} \quad (5)$$

sendo $\boldsymbol{\beta}$ o vetor de parâmetros para a média, $\boldsymbol{\gamma}$ o vetor de parâmetros para a dispersão, $g(\cdot)$ e $h(\cdot)$ as funções de ligação para média e dispersão e \mathbf{x}_i e \mathbf{z}_i os vetores dos fatores de controle e ruído, respectivamente. Nesse modelo, ϕ representa a variação *independente* da média; $V(\boldsymbol{\mu})$ representa a variação *dependente* da média.

Algumas vantagens dessa abordagem são:

1. distribuições não-normais podem ser consideradas, pertencendo à família exponencial ou não (utilizando-se estimativas de QV);

2. esse modelo elimina a dependência entre a média e o parâmetro de dispersão, através da escolha certa das funções de ligação para os modelos da média e da dispersão, atendendo assim ao *critério de separação* (BOX, 1988);
3. modelos aditivos mais simples podem ser selecionados, através da inferência correta sobre os parâmetros lineares, atendendo assim ao *critério da parcimônia*;
4. fatores experimentais poderiam ocorrer em \mathbf{Z} , em \mathbf{X} ou, simultaneamente, em ambos;
5. cada observação contribui para o processo de estimação e, conseqüentemente, gera valores individuais de resíduo. Isso permite aplicar todas as técnicas de análise de resíduos e diagnósticos, na verificação da qualidade de ajuste.

Algoritmo de estimação

Considerando que o par de modelos (5) são, individualmente, modelos lineares generalizados, podem-se estimar os vetores de parâmetros β e γ da seguinte forma:

1. considerando o parâmetro ϕ_i fixo, estimar o vetor β pelo algoritmo IRLS e obter os componentes de *deviance* (ou os componentes da estatística de Pearson);
2. considerando β fixo e ϕ como variável resposta, estimar γ pelo algoritmo IRLS, fazendo $\phi = h^{-1}(\tilde{\gamma})$, receber os valores preditos; retorna ao passo 1 até que seja alcançado um critério de convergência sobre os valores das estimativas.

Nelder e Lee (1991) chamam esse algoritmo iterativo de *see-saw*, que, em geral, converge rapidamente. Na estimação da média μ_i com ϕ_i fixo, a função de QVE reduz-se a uma função de QV, com uma função de variância especificada, a menos de uma constante. Logo, as equações de estimação são dadas por

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

O método QVE fornece componentes de deviance escalonados d_i^2/ϕ_i para cada $i = 1, \dots, n$. Dessa forma, tabelas do tipo ANOVA podem ser construídas para modelos encaixados que podem ser comparados usando-se a diferença de *deviances* totais e o teste-F.

No ajuste do modelo de dispersão, considerando μ_i fixo, a função de QVE toma forma de uma função de verossimilhança gama, quando d_i^2 é a variável resposta, tendo $E(d_i^2) = \phi_i$ e $\text{Var}(d_i^2) = 2\phi_i^2$. Um fato importante é que d_i^2 é uma estimativa viciada de ϕ_i , mas afeta, principalmente, o intercepto no modelo de dispersão. Nelder e Lee (1992) demonstram através de um estudo de simulação que o vício não é grande, d_i^2 é melhor do que os resíduos de Pearson e melhores que estimativas de máxima verossimilhança para estimativas de efeitos associados a delineamentos altamente fracionados. Diferente das estimativas QV, as estimativas do método QVE permitem identificar fatores de ruído que influenciam a variabilidade e permitem também comparar diferentes funções de ligação (sendo a função logarítmica, usualmente, utilizada). Nota-se que β e γ são assintoticamente ortogonais, uma vez que

$$E\left(\frac{\partial Q^+}{\partial \beta \partial \gamma}\right) = 0.$$

Logo, com base nos resultados de Cox e Reid (1987) o algoritmo *see-saw* pode ser aplicado sem problemas numéricos sérios, segundo Lee e Nelder (1998).

No caso em que o número de parâmetros é grande, quando comparado com o tamanho da amostra, as estimativas de dispersão são seriamente viciadas. Como solução para esse problema, utilizam-se estimativas de Máxima Verossimilhança Restrita (REML) com o ajuste de Cox e Reid (1987):

$$-2Q_C^+ = \sum_{i=1}^n \left(\frac{d_i^2}{\phi_i} + \log [2\pi\phi_i V(y_i)] \right) + \log \left| \frac{\mathbf{X}^T \mathbf{W}^* \mathbf{X}}{2\pi} \right|,$$

sendo

$$\mathbf{W}^* = \text{diag} \left\{ \frac{1}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \right\}.$$

2.1.3.3 Estratégia de Análise

Em experimentos para os quais cada combinação de níveis de fatores ou tratamentos é repetida, o seguinte procedimento de análise pode ser adotado:

1. fixar um modelo maximal para a média. Isso garante que no ajuste do modelo de dispersão apenas a variação dentro de tratamentos será considerada;
2. ajustar modelos parcimoniosos para a dispersão (considerando o componente de *deviance* ou

o quadrado do componente da estatística de Pearson como variável resposta). Se necessário, utilizar um método de seleção de modelos como *backward elimination*, *stepwise*, dentre outros. Nesse ponto, os erros padrões para os parâmetros da média estão baseados somente na variação dentro de tratamentos, e, portanto, podem estar bastante viciados. Lee e Nelder (1998) sugerem que o modelo final para a média e a dispersão tenham o menor número possíveis de termos comuns. Essa etapa atende ao critério de separação, citado por Shoemaker et al. (1988);

3. encontrado o melhor modelo que ajusta o efeito dos fatores na dispersão, fixá-lo e iniciar a busca do modelo para a média. Esse procedimento atende ao critério de parcimônia, citado por Box (1988).

Em situações para as quais não há repetições, como em experimentos fatoriais fracionados cruzados com outra estrutura fatorial, para acomodar efeitos de ruídos, uma outra abordagem é necessária:

- i. deve-se iniciar o processo de seleção de modelos com um preditor linear maximal para a média, contendo os efeitos principais para os fatores de controle e ruído;
- ii. fazer uma busca por interações de grande significância entre os fatores de controle e os de ruído. Interações de ordem superior mas com efeito negligenciável podem ser desconsideradas do modelo e utilizadas para o modelo de dispersão. Se nenhum fator de ruído se mostrou significativo no modelo para a média, a análise pode ser realizada como se cada fator de controle fosse repetido, com o número de observações correspondentes ao número de combinações referente ao delineamento experimental dos fatores de ruído;
- iii. encontrado um preditor linear inicial para a média, fixá-lo e iniciar a busca de um preditor linear parcimonioso para a dispersão.
- iv. fixado o modelo de dispersão, retomar a busca de um modelo parcimonioso para a média da variável resposta estudada.

Para as duas estratégias de análise, a verificação da qualidade de ajuste do modelo deve ser realizada, com base na análise de resíduos dos modelos de média e dispersão. Algumas implementações computacionais estão disponíveis para os cálculos das estimativas dos parâmetros. Pinto

(2005) disponibilizou código-fonte em FORTRAN para a estimação dos parâmetros do MLGS. Ribatet e Iooss (2006) disponibilizaram uma biblioteca de funções para a linguagem R para o ajuste dos modelos lineares generalizados, de forma similar à função $\text{glm}()$, sendo, adicionalmente, inserido um preditor linear para o parâmetro de dispersão. Lee et al. (2006) incluem um CD-ROM com o software GENSTAT com a função HGANALYSE para a estimação dos parâmetros (PAYNE et Al.,2007).

2.1.4 Modelos com efeitos fixo e aleatório

Quando o pressuposto de independência das observações não pode ser garantido, devido a características intrínsecas do problema tratado (como em estudos multivariados, dados agrupados, medidas repetidas, dados longitudinais e espaço-temporais), modelos apropriados devem ser aplicados, de forma a acomodar a correlação existente entre observações. Neste trabalho, um dos objetivos é modelar simultaneamente a média e a dispersão de forma estruturada, para acomodar dados binários correlacionados e tempos até eventos que ocorrem de forma recorrente. Nas subseções seguintes, será feita uma breve revisão do modelo linear geral misto (MLM), das famílias de modelos para dados longitudinais discretos, sobre o modelo linear generalizado misto (MLGM), alguns algoritmos para otimização de funções e algoritmos para integração numérica.

2.1.4.1 Modelo linear geral misto

Os modelos lineares mistos (MLM) são uma extensão natural dos modelos lineares gerais incorporando-se um vetor de feitos aleatórios. Uma dentre várias aplicações dessa classe de modelos é modelar a estrutura de dependência temporal associada aos dados longitudinais. Essa classe de modelos tem larga aplicação na pesquisa agronômica, como genética animal e de plantas, zootecnia, experimentos com estrutura complexas, etc. Freitas (2007) estudou três abordagens metodológicas para a análise da produtividade de cana-de-açúcar ao longo dos anos. Uma dessas abordagens foi um MLM. Henrique (2004) utilizou um MLM para selecionar genótipos de eucaliptos em melhoramento vegetal. Um estudo amplo dos MLM está descrito em Verbeke e Molenberghs (2000), incluindo a análise de sensibilidade no tratamento de dados incompletos. Em Lee et al. (2006) os MLM são apresentados como um caso particular da grande família dos *Modelos Lineares Generalizados Hierárquicos - MLGH*.

Seja Y_{ij} uma observação do i -ésimo elemento no j -ésimo termo (não necessariamente equiespaçado), sendo $i = 1, \dots, N$ e $j = 1, \dots, n_i$. Portanto, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ é um vetor n_i -dimensional contendo todas as observações disponíveis do i -ésimo elemento. No caso em que a dependência temporal pode ser ignorada, o modelo linear geral pode ser descrito como $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$, sendo que $\boldsymbol{\epsilon}_i$ é um vetor que assume distribuição normal multivariada, com vetor de média $\mathbf{0}$ e matriz de variância-covariância $\boldsymbol{\Sigma}_i = \mathbf{I}_{n_i}\sigma^2$. Logo,

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i). \quad (6)$$

Entretanto, é mais comum que as medidas longitudinais apresentem correlações através do tempo e a matriz $\boldsymbol{\Sigma}$ pode ter diferentes estruturas de covariância. Johnson e Wichern (2007) descrevem a abordagem multivariada para modelar (6); aqui, será apresentada a modelagem de dados longitudinais através do uso dos MLM. Um modelo linear geral misto satisfaz a seguinte estrutura hierárquica (LAIRD; WARE, 1982):

$$\begin{cases} \mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \\ \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}) \\ \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i) \end{cases} \quad (7)$$

sendo \mathbf{X}_i e \mathbf{Z}_i as matrizes de delineamento (contendo fatores e/ou covariáveis) para o vetor p -dimensional $\boldsymbol{\beta}$ que contém os efeitos fixos e o vetor q -dimensional \mathbf{b}_i dos efeitos aleatórios, que possui distribuição normal multivariada com média $\mathbf{0}$ e matriz de covariância \mathbf{D} , $\boldsymbol{\epsilon}_i$ é o vetor de erros aleatórios com distribuição normal multivariada n_i -dimensional com vetor de média $\mathbf{0}$ e matriz de covariância $\boldsymbol{\Sigma}_i$. Assume-se que os elementos de \mathbf{b}_i são independentes dos elementos de $\boldsymbol{\epsilon}_i$.

A estimação e a inferência para (7) na abordagem frequentista, é baseada no modelo marginal. Dado que

$$\begin{aligned} \mathbf{Y}_i|\mathbf{b}_i &\sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \boldsymbol{\Sigma}_i) \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}), \end{aligned} \quad (8)$$

a função densidade de probabilidade marginal de \mathbf{Y}_i é dada por

$$f(y_i) = \int \cdots \int f(y_i|\mathbf{b}_i)f(\mathbf{b}_i)db_{i1} \cdots db_{iq} \quad (9)$$

sendo $f(y_i|\mathbf{b}_i)$ a função densidade de \mathbf{Y}_i dado que \mathbf{b}_i é conhecido e $f(\mathbf{b}_i)$ é a função densidade de \mathbf{b}_i . Pode ser mostrado, facilmente, que a esperança e a variância incondicional de \mathbf{Y}_i são $E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta}$ e $\text{Var}(\mathbf{Y}_i) = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$, e portanto, $\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i)$.

Os estimadores MV de $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})^T$ são obtidos a partir da função de verossimilhança marginal

$$L_{MV}(\boldsymbol{\theta}) = \prod_{i=1}^N \left\{ (2\pi)^{-n_i/2} |\mathbf{V}_i(\boldsymbol{\alpha})|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right] \right\} \quad (10)$$

sendo $\boldsymbol{\alpha}$ um vetor dos parâmetros contidos em $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$. O estimador de máxima verossimilhança para $\boldsymbol{\beta}$, condicionado ao valor conhecido de $\boldsymbol{\alpha}$ é (LAIRD; WARE, 1982 apud VERBEKE; MOLENBERGHS, 2000):

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i^T \mathbf{W}_i \mathbf{y}_i,$$

sendo $\mathbf{W}_i = \mathbf{V}_i^{-1}$.

A estimação do vetor $\boldsymbol{\alpha}$ é, usualmente, feita pelo método de máxima verossimilhança restrita (REML), que garante a obtenção de estimadores não-viciados para os componentes de variância $\boldsymbol{\alpha}$. A função de verossimilhança restrita para $\boldsymbol{\alpha}$ pode ser descrita como:

$$L(\boldsymbol{\alpha}) = C \left| \sum_{i=1}^N \mathbf{X}_i^T \mathbf{W}_i(\boldsymbol{\alpha}) \mathbf{X}_i \right|^{-1/2} L_{MV}(\boldsymbol{\theta}),$$

sendo C uma constante não dependente de $\boldsymbol{\alpha}$. Para maiores informações e detalhamento da inferência sobre os parâmetros de um MLM, ver Verbeke e Molenberghs (2000).

2.1.4.2 Famílias de modelos para dados longitudinais

Nos MLM são utilizados os efeitos aleatórios e as estruturas de covariância D e Σ como estruturas para acomodar fontes de variação e dispersão longitudinal. Entretanto, para dados discretos, existem diversas formas de se tratar esse problema, não havendo uma unidade como no caso Gaussiano. Molenberghs e Verbeke (2005) argumentam que isso é, parcialmente, devido à falta de uma distribuição análoga à distribuição Normal multivariada. Logo, dependendo do tipo de inferência desejada, três grandes famílias de modelos são adotadas:

Modelos Marginais São utilizados quando a variável resposta é modelada marginalmente sobre as demais variáveis respostas. A estrutura de associação existente entre elas são baseadas em medidas de correlação e associação. Uma classe bastante flexível desses modelos foi proposta, denominada Equações de Estimação Generalizadas (GEE), que assume conhecida apenas a forma da estrutura de correlação entre as observações longitudinais (LIANG; ZEGGER, 1986).

Modelos Condicionalmente Especificados São modelos cujas variáveis respostas, medidas repetidamente ao longo do tempo, são modeladas condicionalmente às demais variáveis ou um sub-conjunto delas. Um exemplo são os modelos de transição para dados binários (LARA, 2007).

Modelos para elemento-específico Também conhecidos como *subject-specific models*, são modelos cujas medidas da variável resposta são assumidas independentes, dado um conjunto de parâmetros específicos para cada elemento (*subject*) na amostra.

Os modelos propostos para dados longitudinais neste trabalho são modelos para elemento-específico. Portanto, os demais modelos não serão explorados daqui em diante. Três maneiras de se estimarem os parâmetros elemento-específicos são: (i) tratar todos os parâmetros como fixos e desconhecidos; (ii) usar *inferência condicional*, em que a função de verossimilhança para estimar β dado o vetor de dados y_i é condicionada às estatísticas suficientes para b_i ; (iii) usar os modelos de efeitos aleatórios. A abordagem (iii) é a mais utilizada atualmente para modelagem de dados longitudinais discretos, tanto no enfoque frequentista quanto no Bayesiano. Considerando que os dados foram constituídos de uma amostra da população de elementos, será considerado

também que o vetor \mathbf{b}_i também é amostrado independentemente de uma população com distribuição $Q(\mathbf{b}_i)$. Assim, a estimação do vetor $\boldsymbol{\theta}$ por máxima verossimilhança é feita utilizando-se a função densidade de probabilidade marginal para \mathbf{Y}_i , dada por

$$f_i(\mathbf{y}_i|\boldsymbol{\theta}, Q) = \int \cdots \int f_i(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\theta}) dQ(\mathbf{b}_i), \quad (11)$$

em que o vetor \mathbf{b}_i é eliminado através da operação de integração múltipla.

2.1.4.3 Modelo linear generalizado misto

Uma classe bastante flexível da família de modelos elemento-específico são os modelos lineares generalizados mistos (*Generalized Linear Mixed Models* - MLGM), proposto por Breslow e Clayton (1993). Trata-se de uma extensão direta do MLG proposto por Nelder e Wedderburn (1972), que, dentre outras aplicações, permite acomodar a informação serial contida nas medidas longitudinais. Diversos softwares para análise estatística implementaram rotinas para essa classe de modelos, dentre eles o SAS (SAS Institute, 2008), o ambiente de computação estatística R (R DEVELOPMENT CORE TEAM, 2007) e o GENSTAT (PAYNE et al., 2007).

Formulação do modelo

Seja Y_{ij} a j -ésima medida longitudinal do i -ésimo elemento (*subject*), $j = 1, \dots, n_i$ e $i = 1, \dots, N$; logo, \mathbf{Y}_i é um vetor n_i -dimensional com as medidas longitudinais do i -ésimo elemento; \mathbf{b}_i é o vetor q -dimensional de efeitos aleatórios, independentes e com distribuição $N(\mathbf{0}, \mathbf{D})$, sendo $\mathbf{0}$ um vetor de zeros e \mathbf{D} é uma matriz de variâncias-covariâncias, geralmente, estruturada e pré-especificada; \mathbf{b}_i está associado a $q - 1$ fatores e/ou covariáveis que compõe o vetor \mathbf{z}_{ij} , com dimensão $(q \times 1)$; $\boldsymbol{\beta}$ é o vetor p -dimensional de efeitos fixos, associados a $p - 1$ fatores e/ou covariáveis que compõem o vetor \mathbf{x}_{ij} com dimensão $(p \times 1)$. É assumido que as medidas Y_{ij} são condicionalmente independentes, dado \mathbf{b}_i , com função densidade de probabilidade pertencente à família exponencial de distribuições, ou seja,

$$f_i(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) = \exp \left\{ \phi^{-1} [y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi) \right\}, \quad (12)$$

sendo $\eta(\mu_{ij}) = \eta[\mathbf{E}(Y_{ij}|\mathbf{b}_i)] = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i$, $\eta(\cdot)$ é uma função de ligação pré-especificada e ϕ é o parâmetro de dispersão. Se $\theta_{ij} = \eta(\mu_{ij})$, então η é uma função de ligação canônica. Denota-se também que $f(\mathbf{b}_i|\mathbf{D})$ é a função densidade dos efeitos aleatórios \mathbf{b}_i .

Estimação dos parâmetros - abordagem Bayesiana

Enxergando o modelo (12) de um ponto de vista hierárquico, sendo que a distribuição de Y_{ij} está condicionada ao valor de \mathbf{b}_i (que possui uma função densidade de probabilidade pré-especificada $N(\mathbf{0}, \mathbf{D})$), a abordagem Bayesiana se mostra bastante adequada para lidar com a estrutura naturalmente hierárquica das fontes de variação. Consideram-se como funções densidades *a priori* para \mathbf{D} , ϕ e $\boldsymbol{\beta}$, as funções $f(\mathbf{D})$, $f(\phi)$ e $f(\boldsymbol{\beta})$, respectivamente. Para $\boldsymbol{\beta}$ são, comumente, assumidas a distribuição normal ou a distribuição não informativa de Jeffreys. Entretanto, Fahrmeir e Tutz (1994, apud MOLENBERGHS; VERBEKE, 2005) reportam que densidades *a priori* impróprias como a de Jeffreys, pode levar à obtenção de funções densidades *a posteriori* impróprias. Na prática, são utilizadas funções densidades informativas mas com uma escolha cuidadosa dos hiperparâmetros, para obter uma distribuição *a priori* vaga.

A função densidade de probabilidade *a posteriori* conjunta é dada por:

$$f(\boldsymbol{\beta}, \mathbf{D}, \phi, \mathbf{b}_i | \mathbf{Y}_i) \sim \prod_{i=1}^N \prod_{j=1}^{n_i} f_i(y_{ij} | \boldsymbol{\beta}, \phi, \mathbf{b}_i) \prod_{i=1}^N f(\mathbf{b}_i | \mathbf{D}) f(\mathbf{D}) f(\phi) f(\boldsymbol{\beta}). \quad (13)$$

Com base nas funções densidades *a posteriori* condicionais completas, podem-se obter amostras das funções densidades marginais *a posteriori*, através de algoritmos MCMC (GAMERMAN; LOPES, 2006). Gamerman (1997) propôs um algoritmo MCMC mais eficiente para amostragem das funções densidades marginais *a posteriori*, para os modelos hierárquicos MLGM. Esse algoritmo encontra-se implementado no software OpenBUGS (THOMAS et al, 2006).

Estimação de \mathbf{b}_i via *empirical Bayes*

Se há o interesse em estimar o vetor \mathbf{b}_i (com objetivos preditivos ou como diagnóstico elemento-específico), a abordagem conhecida como *Empirical Bayes* pode ser utilizada. O processo consiste em utilizar as estimativas de \mathbf{D} , ϕ e $\boldsymbol{\beta}$, obtidas por máxima verossimilhança, e aplicá-las

na distribuição marginal *a posteriori* de \mathbf{b}_i , dada por

$$f(\mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\beta}, \mathbf{D}, \phi) = \frac{f(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\beta}, \phi)f(\mathbf{b}_i|\mathbf{D})}{\int f(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\beta}, \phi)f(\mathbf{b}_i|\mathbf{D})d\mathbf{b}_i}.$$

Estimação por máxima verossimilhança

A estimação de $\boldsymbol{\beta}$ e dos parâmetros em \mathbf{D} é feita através da maximização da função de verossimilhança marginal, obtida através da integração da função de verossimilhança sobre os efeitos aleatórios \mathbf{b}_i . A contribuição do i -ésimo elemento da função de verossimilhança é

$$f(\mathbf{y}_i|\boldsymbol{\beta}, \mathbf{D}, \phi) = \int \cdots \int \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi)f(\mathbf{b}_i|\mathbf{D})d\mathbf{b}_i, \quad (14)$$

e a função de verossimilhança marginal para a amostra contendo os N elementos é dada pela expressão

$$\begin{aligned} L(\boldsymbol{\beta}, \mathbf{D}, \phi) &= \prod_{i=1}^N f(\mathbf{y}_i|\boldsymbol{\beta}, \mathbf{D}, \phi) \\ &= \prod_{i=1}^N \int \cdots \int \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi)f(\mathbf{b}_i|\mathbf{D})d\mathbf{b}_i, \end{aligned} \quad (15)$$

sendo que o número de integrais na expressão corresponde ao número de termos em \mathbf{b}_i .

No caso particular em que $Y_{ij} \sim N(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$, as integrais em (15) podem ser tratadas analiticamente. Em geral, porém, essas integrais não possuem solução analítica e métodos numéricos ou estocásticos são necessários para solucioná-las. Uma vez calculado o valor aproximado dessas integrais, a função (15) deve ser otimizada. Novamente, para o caso particular Gaussiano, expressões analíticas são obtidas pelo processo de otimização. Fora esse caso particular, um sistema de equações não-lineares é obtido e métodos numéricos para otimização são utilizados. O cálculo numérico das integrais dentro de cada ciclo iterativo na otimização de (15) torna a obtenção de estimativas um processo computacionalmente intensivo, tomando bastante tempo quando se trata de grandes amostras e/ou quando a estrutura do modelo é complexa, em especial quando a dimensão do espaço paramétrico é grande.

2.1.4.4 Algoritmos para otimização de funções

Otimização de funções reais (ou equivalentemente, solução de equações lineares e não-lineares) é uma atividade crucial para a Inferência Estatística. Frequentemente, tarefas como a obtenção de estimadores de MV, minimização de funções-perda em análise Bayesiana de decisões, dentre outras, são atividades que dependem de otimização de funções. Nesta seção, será apresentado um resumo dos principais algoritmos atualmente implementados nos softwares utilizados nessa pesquisa. Usualmente, o estudo de algoritmos de otimização compete à área de *Análise Numérica* e a grande maioria dos softwares para análise estatística possuem implementações de um subconjunto destes algoritmos. Em todos eles, o problema genérico pode ser resumido em como *otimizar uma função real g aplicada ao seu argumento θ , um vetor p -dimensional*. Naturalmente, quando o problema estatístico é a estimação por máxima verossimilhança, o problema consiste em obter o estimador $\hat{\theta}$ como a solução da equação $l'(\theta) = \mathbf{0}$, sendo $l'(\theta) = \left(\frac{\partial l(\theta)}{\partial \theta_1}, \dots, \frac{\partial l(\theta)}{\partial \theta_p} \right)^T$ e $\mathbf{0}$ um vetor coluna de zeros.

Método de Newton-Raphson

Esse é um método para encontrar raízes de equações não-lineares que é bastante rápido, e largamente utilizado na estimação de parâmetros. Essencialmente, consiste em tomar a primeira derivada da função g (que é contínua e diferenciável, com $g''(x^*) \neq 0$) e aproximá-la para o ponto x^* através da expansão em séries de Taylor (GIVENS; HOETING, 2005):

$$0 = g'(x^*) \approx g'(x^{(t)}) + (x^* - x^{(t)})g''(x^{(t)}). \quad (16)$$

Como essa aproximação constitui a tangente de g no ponto x^* , utiliza-se a raiz da reta tangente como raiz de g' . Solucionando a equação (16) para x^* , tem-se

$$x^* = x^{(t)} - \frac{g'(x^{(t)})}{g''(x^{(t)})}, \quad (17)$$

e para o problema de estimação do parâmetro θ , essa expressão pode ser descrita como

$$\theta^{(t+1)} = \theta^{(t)} - \frac{l'(\theta^{(t)})}{l''(\theta^{(t)})}. \quad (18)$$

A versão multivariada para esse método iterativo, assumindo que $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ é um vetor de parâmetros p -dimensional, é dada por

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \frac{\mathbf{l}'(\boldsymbol{\theta}^{(t)})}{\mathbf{l}''(\boldsymbol{\theta}^{(t)})}, \quad (19)$$

sendo $\mathbf{l}'(\boldsymbol{\theta}^{(t)})$ o vetor escore e $\mathbf{l}''(\boldsymbol{\theta}^{(t)})$ a matriz Hessiana, ambos calculados no ponto $\boldsymbol{\theta}^{(t)}$.

Método *scoring* de Fisher

Consiste em substituir a matriz Hessiana pela matriz de informação esperada de Fisher $I(\boldsymbol{\theta})$ em (19), uma vez que essa pode ser aproximada por $-\mathbf{l}''(\boldsymbol{\theta}^{(t)})$. Assim a equação iterativa que atualiza $\boldsymbol{\theta}^{(t)}$ é descrita como

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \mathbf{l}'(\boldsymbol{\theta}^{(t)})I(\boldsymbol{\theta}^{(t)})^{-1}. \quad (20)$$

Tanto o método de Newton quanto o método *scoring* apresentam as mesmas propriedades assintóticas. Esse último é implementado na maioria das funções e procedimentos para o ajuste de modelos MLG. Quando uma função de ligação canônica é utilizada, ambos os métodos fornecem o mesmo resultado, uma vez que a matriz de informação esperada será igual a matriz Hessiana (PAWITAN, 2001).

Método da secante

Quando o cálculo da matriz Hessiana é difícil de se obter ou quando se deseja minimizar o trabalho de especificação do modelo, o método da secante pode ser bastante útil. Ele difere do método de Newton e *scoring* de Fisher por não necessitar do cálculo das derivadas de segunda ordem $g''(x^{(t)})$. Em seu lugar, são calculadas aproximações por diferença-discreta na forma

$$\frac{g'(x^{(t)}) - g'(x^{(t-1)})}{x^{(t)} - x^{(t-1)}},$$

de onde vem o nome do método por substituir a reta tangente por uma reta secante, para aproximar a função g no ponto $x^{(t)}$. Portanto, a função (17) é adaptada para

$$x^{(t+1)} = x^{(t)} - g'(x^{(t)}) \frac{x^{(t)} - x^{(t-1)}}{g'(x^{(t)}) - g'(x^{(t-1)})}, \quad (21)$$

para $t \geq 1$. Logo, esse método depende de dois valores iniciais $x^{(0)}$ e $x^{(1)}$ que definem a reta secante. A ordem de convergência para esse método é aproximadamente 1,62 (o método da bisseção tem ordem 1 e o método de Newton-Raphson tem ordem 2), sendo portanto, mais lento que o método de Newton-Raphson para convergir. Entretanto, esse método é conhecido por ser eficiente em situações que o método de Newton-Raphson diverge, como no caso de raízes próximas ao limite do espaço paramétrico.

Método Quase-Newton

Uma forma similar ao método da secante para evitar o cálculo da matriz Hessiana é substituí-la por uma matriz de constantes $\mathbf{M}^{(t)}$, de forma que o vetor $\mathbf{x}^{(t)}$ seja conforme a seguinte equação iterativa:

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \mathbf{M}^{-1} \mathbf{g}'(\mathbf{x}^{(t)}). \quad (22)$$

Uma forma eficiente de atualizar os valores de \mathbf{M} é utilizar a informação da direção do vetor gradiente $\mathbf{g}'(\mathbf{M}^{(t)})$ no ciclo iterativo mais recente e aprender sobre a curvatura de \mathbf{g}' , na direção do vetor gradiente próximo a $\mathbf{x}^{(t)}$.

O método *Quase-Newton* utiliza uma condição secante na forma

$$\mathbf{g}'(\mathbf{x}^{(t+1)}) - \mathbf{g}'(\mathbf{x}^{(t)}) = \mathbf{M}^{(t+1)}(\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}) \quad (23)$$

para atualizar $\mathbf{M}^{(t)}$, utilizando a informação da curvatura de \mathbf{g}' e a direção do vetor gradiente, além de utilizar apenas derivadas de primeira ordem. O método de atualização conhecido como método de Broyden-Fletcher-Goldfarb-Shanno (BFGS) pode ser descrito como

$$\mathbf{M}^{(t+1)} = \mathbf{M}^{(t)} - \frac{\mathbf{M}^{(t)} \mathbf{z}^{(t)} (\mathbf{M}^{(t)} \mathbf{z}^{(t)})^T}{(\mathbf{z}^{(t)})^T \mathbf{M}^{(t)} \mathbf{z}^{(t)}} + \frac{\mathbf{y}^{(t)} (\mathbf{y}^{(t)})^T}{(\mathbf{z}^{(t)})^T \mathbf{y}^{(t)}}. \quad (24)$$

Esse método é um dos mais populares na implementação do método Quase-Newton e aceito como

superior a outras alternativas, com base em estudos teóricos e empíricos. Ele possui a característica de gerar matriz definida-positiva hereditária, ou seja, a cada ciclo, uma nova matriz definida-positiva com ascendência para o valor ótimo é garantida. Esse conjunto de características desse algoritmo explica sua popularidade e sua versatilidade. No sistema SAS, esse e outros algoritmos estão implementados no procedimento NLMIXED. Para várias opções de algoritmos, opções para *line search*, *backtracking* e *scaling* estão disponíveis para o controle do analista, o que pode ser necessário para problemas mais complexos.

2.1.4.5 Algoritmos para integração numérica

O cálculo de integrais definidas é sempre um problema que merece atenção em inferência estatística, uma vez que, na maioria dos problemas reais, as integrais são complexas o bastante para permitir a obtenção de resultados analíticos. O problema de integração em (9) nem sempre tem solução analítica e, portanto, métodos numéricos são utilizados dentro dos ciclos iterativos dos métodos de otimização de funções. Os problemas Bayesianos também demandam aproximações numéricas de integrais, uma vez que nem sempre as funções densidades *a posteriori* conjunta possuem formas conhecidas.

Considere uma integral na forma $I = \int_a^b f(x)dx$, definida no intervalo $[a, b]$, que possui n subintervalos $[x_i, x_{i+1}]$ para $i = 1, \dots, n - 1$, com $x_0 = a$ e $x_n = b$. O problema de aproximação de integrais por métodos numéricos, em geral, consiste em obter a seguinte aproximação:

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n-1} A_i f(x_i^*). \quad (25)$$

Quadratura de Newton-Côtes

Seja $h = (b - a)/n$ a largura dos n subintervalos. Três métodos bastante conhecidos do Cálculo Numérico são:

Regra de Riemann Assumindo que os subintervalos são equiespaçados de tamanho h , pode-se escrever que $x_i = a + ih$ e, portanto,

$$\int_a^b f(x)dx \approx h \sum_{i=0}^{n-1} f(a + ih) = \widehat{R}(n). \quad (26)$$

Pode ser visto que, quando $n \rightarrow \infty$, \widehat{R} é uma soma cujo limite é igual ao valor I , conforme a definição da integral de Riemann.

Regra trapezoidal Também assumindo que os n subintervalos de comprimento h , a regra do trapézio aproxima o valor de I através da expressão

$$\int_a^b f(x)dx \approx \frac{h}{2}f(a) + h \sum_{i=1}^{n-1} f(a + ih) + \frac{h}{2}f(b) = \widehat{T}(n). \quad (27)$$

Regra de Simpson Uma forma mais eficiente que a regra anterior, que permite tomar a função do integrando como uma função cúbica. A integral é aproximada através da expressão

$$\int_a^b f(x)dx \approx \frac{h}{3} \sum_{i=1}^{n/2} [f(x_{2i-2}) + 4f(x_{2i-1}) + f(x_{2i})] = \widehat{S}(n/2). \quad (28)$$

Usualmente, essa regra é mais eficiente do que as anteriores, necessitando um número de subintervalos menores para alcançar a mesma precisão.

Aproximação do integrando

Um método que é bastante conhecido na inferência Bayesiana é o método de Laplace, que foi criado para aproximar integrais da forma

$$I = \int e^{Q(\mathbf{b})} d\mathbf{b} \quad (29)$$

sendo $Q(\mathbf{b})$ uma função conhecida, unimodal e limitada, da variável q -dimensional \mathbf{b} . Considera-se $\widehat{\mathbf{b}}$ o valor de \mathbf{b} que maximiza Q . Expandindo Q em séries de Taylor de segunda ordem, obtém-se

$$Q(\mathbf{b}) \approx Q(\widehat{\mathbf{b}}) + \frac{1}{2}(\mathbf{b} - \widehat{\mathbf{b}})^T Q''(\widehat{\mathbf{b}})(\mathbf{b} - \widehat{\mathbf{b}}), \quad (30)$$

considerando $Q''(\widehat{\mathbf{b}})$ a matriz de derivadas de segunda ordem de Q , calculadas no ponto $\widehat{\mathbf{b}}$. Substituindo $Q(\mathbf{b})$ por (30) em (29), obtém-se, então

$$I \approx (2\pi)^{q/2} | -Q''(\widehat{\mathbf{b}}) |^{-1/2} e^{Q(\widehat{\mathbf{b}})}.$$

Na função de verossimilhança em (15) cada integral é proporcional a (29), sendo a função $Q(\mathbf{b})$ dada por

$$Q(\mathbf{b}) = \phi^{-1} \sum_{j=1}^{n_i} [y_{ij}(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}) - \psi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b})] - \frac{1}{2} \mathbf{b}^T \mathbf{D}^{-1} \mathbf{b}$$

e portanto, o método de Laplace pode ser aplicado. O valor de $\hat{\mathbf{b}}$ depende de $\boldsymbol{\beta}$, ϕ e \mathbf{D} , o que significa que o processo é iterativo e a cada ciclo da otimização, o valor desses parâmetros devem ser baseados na sua última atualização. Se $Q(\mathbf{b})$ é igual ao núcleo da função densidade normal, essa função é quadrática e a aproximação é exata. Um grande número n_i de medidas longitudinais aumenta a performance desse método.

Quadratura de Gauss-Hermite

Em algumas situações, os métodos anteriores podem falhar ou não serem eficientes na obtenção de uma aproximação para a integral definida desejada. No contexto dos MLGM, o número de integrais envolvidas pode-se tornar computacionalmente proibitivo. Entretanto, o método da quadratura Gaussiana mostra-se mais eficiente do que a classe de Quadratura de Newton-Côtes, uma vez que os pontos que definem os subintervalos não são necessariamente equiespaçados e os pesos da soma em (25) são escolhidos em função da escolha de classes de polinômios ortogonais (GIVENS; HOETING, 2005). O método é mais efetivo em situações para as quais a integral é da forma

$$I = \int f(x)w(x)dx, \quad (31)$$

sendo $f(x)$ uma função conhecida e $w(x)$ uma função densidade de probabilidade com momentos finitos em torno da origem. Usualmente, $w(x)$ tem função densidade normal padrão multivariada.

Conforme Molenberghs e Verbeke (2005), comparando (31) com (14), nota-se a similaridade dos integrandos, necessitando apenas que seja feita uma padronização em \mathbf{b} . Tomando o efeito aleatório $\boldsymbol{\delta}_i = \mathbf{D}^{-1/2} \mathbf{b}_i$, obtém-se um novo efeito aleatório com distribuição normal q -dimensional, com média $\mathbf{0}$ e matriz de variância-covariância \mathbf{I} . O preditor linear torna-se $\theta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{D}^{-1/2} \boldsymbol{\delta}_i$. A função (14) torna-se

$$f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{D}, \phi) = \int \cdots \int \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{b}_i, \boldsymbol{\beta}, \mathbf{D}, \phi) f(\boldsymbol{\delta}_i) d\boldsymbol{\delta}_i,$$

estando em condição de se utilizar o método de aproximação.

A aproximação pela quadratura de Gauss-Hermite para

$$\int f(z)w(z)dz \approx \sum_{q=1}^Q w_q f(z_q),$$

sendo Q a ordem da aproximação, é baseada na utilização de polinômios ortogonais de Hermite de ordem Q . Os pesos w_q são gerados a partir dos polinômios $p_k(z)$, que possuem a forma geral

$$p_k(z) = (\alpha_k + z\beta_k)p_{k-1}(z) - \gamma_k p_{k-2}(z) \quad (32)$$

com escolhas apropriadas para α_k , β_k e γ_k , que dependem dos valores de k e da função w . Para a quadratura de Gauss-Hermite, $w(z) = \exp(-z^2/2)$, $\alpha_k = 0$, $\beta_k = 1$ e $\gamma_k = k - 1$. Denotam-se as raízes de $P_{Q+1}(z)$ por $a < z_0 < \dots < z_Q < b$. A constante $c_k = 1$ é utilizada na fórmula para gerar os pesos, (GIVENS; HOETING, 2005)

$$A_i = -c_{Q+2}/[c_{Q+1}p_{Q+2}(z_i)p'_{Q+1}(z_i)],$$

podendo, dessa forma, aproximar I como $\sum_{i=0}^Q A_i f(z_i)$.

Conforme Molenberghs e Verbeke (2005), a qualidade da aproximação da integral é maior se $f(z)w(z)$ se comporta de forma similar a $w(z)$, o que nem sempre acontece nos problemas gerais. Nessa situação, um procedimento para reescalonar e deslocar os pontos de quadratura pode aumentar a precisão, sob o preço de um esforço computacional maior. Supõe-se que $f(z)w(z)$ segue, aproximadamente, uma distribuição normal cuja média é a moda \hat{z} de $\ln[f(z)w(z)]$ e a variância é dada por

$$\left[-\frac{\partial^2}{\partial z^2} \ln[f(z)w(z)] \Big|_{z=\hat{z}} \right]^{-1}.$$

Os novos pontos de quadratura (agora adaptativos) são dados por

$$z_i^+ = \hat{z} + \left[-\frac{\partial^2}{\partial z^2} \ln[f(z)w(z)] \Big|_{z=\hat{z}} \right]^{-1/2} z_i$$

com seus respectivos pesos

$$A_i^+ = \left[-\frac{\partial^2}{\partial z^2} \ln[f(z)w(z)] \Big|_{z=\hat{z}} \right]^{-1/2} \frac{w(z_i^+)}{w(z_i)} A_i.$$

Usualmente, o método da quadratura Gaussiana adaptativa necessita de um número bem menor de pontos de quadratura, quando comparado ao método de Gauss-Hermite. Quando apenas um ponto de quadratura é assumido, ou seja, $Q = 1$, mostra-se que o método da quadratura Gaussiana adaptativa reduz-se ao método de Laplace (MOLENBERGHS; VERBEKE, 2005).

2.2 Avaliação da mortalidade de frangos de corte na condição de espera pré-abate utilizando modelos lineares generalizado duplos

2.2.1 Introdução

A avicultura brasileira depara-se, atualmente, com o desafio da redução de perdas, cujo impacto financeiro ultrapassa 100 milhões de Reais anualmente. Porém, existem pontos críticos que devem ser identificados, sendo as operações pré-abate os maiores gargalos de desinformação da cadeia produtiva (KETTLEWELL et al., 2000; SAVENIJE, 2001; BARBOSA FILHO, 2008; VIEIRA, 2008). A morte antes da chegada à linha do abate é o tipo de perda mais crítico dos sistemas de produção avícola, consistindo no indicador mais extremo da falta de bem-estar (NICOL; SCOTT, 1990; BROOM, 1993). Os níveis de mortalidade podem aumentar de zero a valores superiores a 1% em certos trechos da viagem (WARRISS et al., 2005).

Apesar da importância deste tema em estudos de bem-estar e ambiência animal, poucas pesquisas foram realizadas em condições tropicais, evidenciando a necessidade de mais informações nesse setor (BRESSAN; BERAQUET, 2002; VIEIRA et al., 2007; BARBOSA FILHO, 2008). No entanto, a complexidade do problema em questão exige uma estrutura de análise mais robusta, uma vez que o número de variáveis influentes na mortalidade é elevado e o número de aves mortas por caminhão é um evento aleatório raro, com uma grande variabilidade associada aos dados (VIEIRA, 2008).

A mortalidade de aves pode ser tratada na forma de proporções e contagens. Na utilização de modelos de regressão para análise desses tipo de medida, é ainda comum o uso de transformações de dados (BOX; COX, 1964), apesar de várias considerações sobre essa abordagem terem sido discutidas (BOX, 1988). Uma abordagem mais adequada para esse tipo de análise consiste em utilizar modelos de regressão pertencentes à classe de modelos lineares generalizados (ou MLG) (NELDER; WEDDERBURN, 1972) por meio da qual as medidas são modeladas utilizando distribuições probabilísticas previamente assumidas, sem necessidade de transformar as medidas originais para escalas não-informativas.

O número de aves mortas por caminhão é uma proporção de um evento raro (usualmente inferior a 1% e o número de aves transportadas por caminhão costuma ser acima de 800). Apesar de que, proporções podem ser modeladas pelo modelo probabilístico binomial, nessa condição pode-se

utilizar a distribuição Poisson e a abordagem dos MLG pode ser utilizada (NELDER; LEE, 2006). Todavia, uma característica limitante da distribuição Poisson é o fato de a variância ser igual à média. Na situação abordada por Vieira (2008), a variabilidade observada nos dados é muito maior do que a média observada. Esse fenômeno de superdispersão, bastante recorrente em estudos biológicos, necessita ser tratado adequadamente na análise, uma vez que, ignorá-lo acarreta na inflação do erro tipo I e, como consequência, podem-se obter valores de P, artificialmente, significativos em testes de hipóteses (HINDE; DEMÉTRIO, 1998). Uma forma de modelar, simultaneamente, o número de aves mortas por caminhão e a superdispersão que é inerente a esse tipo de medida, é utilizar a classe de modelos lineares generalizados duplos (ou MLGD) (SMYTH, 1989; NELDER; LEE, 1991), que é uma extensão da classe MLG. Apesar desses modelos de regressão serem utilizados na experimentação industrial, são muito pouco conhecidos na pesquisa agrônômica.

Objetivou-se com este trabalho comparar a utilização de diferentes MLG para avaliar a mortalidade de frangos de corte, submetidos à condição de espera em um abatedouro comercial, assim como apresentar e exemplificar a utilidade dos modelos lineares generalizados duplos. Na seção 2.2.2.1 será descrito o estudo de mortalidade de frangos de corte na condição de espera pré-abate, seguido da descrição resumida dos MLG, do modelo para superdispersão com fator de heterogeneidade e dos MLGD, na seção 2.2.2.2. Na seção 2.2.3 será apresentada a comparação dos resultados dos modelos e discutido o impacto de alguns fatores influentes na mortalidade dos frangos.

2.2.2 Materiais e métodos

2.2.2.1 Materiais

A pesquisa foi realizada em um abatedouro comercial de frangos de corte no Estado de São Paulo, apresentando latitude de 22°01'03"S, longitude de 47°53'27"W e altitude média da região de 856m. O clima local é caracterizado como tropical de altitude Cwa, segundo a classificação de Köppen, com temperatura do mês mais frio inferior a 18°C e a do mês mais quente superior a 22°C (EMBRAPA PECUÁRIA SUDESTE, 2007).

O desenvolvimento dessa etapa foi realizado em campo, num galpão de espera, nos meses de outubro e novembro de 2006 e nos meses de janeiro e março de 2007. Foram analisados os caminhões de transporte sob a condição de espera, em cada período estudado (manhã, tarde e noite), totalizando 215 caminhões. Foram considerados, para a análise de dados, 4 dias para o turno da

manhã, 5 para o período da tarde e 5 para o período da noite (Figura 1).

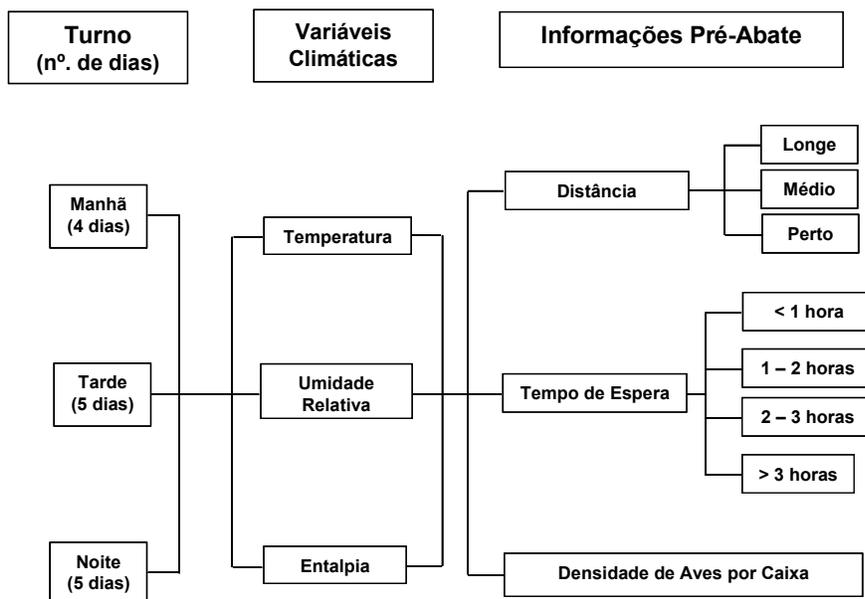


Figura 1 - Fluxograma das variáveis respostas envolvidas nas operações pré-abate

As aves utilizadas no experimento foram da linhagem Cobb, com idades variando entre 6 e 7 semanas. Neste presente estudo, foram evidenciadas as diferentes distâncias entre as granjas integradas e o abatedouro. O tempo de espera consistiu no período de permanência no abatedouro, entre a chegada do caminhão na empresa e a linha de abate. As informações das variáveis foram obtidas por meio das planilhas de dados da empresa e observações no local de estudo, respectivamente.

O ambiente de espera no abatedouro foi caracterizado tipologicamente por um galpão de 23,70m de comprimento por 19,22m de largura, pé direito de 5m e beiral de 0,60m, totalizando uma área livre de 455,5m². Possuía cobertura de telhas galvanizadas, com 6 tesouras metálicas. O ambiente era climatizado com ventiladores, cuja potência era de 0,5 cv, dispostos lateralmente nos pilares (1 ventilador em cada pilar) e nas tesouras da estrutura, em 4 linhas de 7 ventiladores cada uma. Também possuía 8 linhas de nebulização, cada uma com 25 bicos, intercaladas com as linhas de ventilação, cuja distribuição era de 1 bico por metro linear, aproximadamente. As laterais desta sala possuíam telas de polipropileno da marca Sombrite, da cor amarela, para se amenizarem os efeitos da radiação solar direta nos caminhões.

O galpão possuía capacidade para 8 caminhões de transporte de frangos. A chegada e a saída de veículos, assim como o tempo de espera, seguem uma organização que leva em consideração o

deslocamento granja-abatedouro e o período do dia (manhã, tarde e noite). Cada caminhão possuía capacidade para 486 gaiolas, cada uma com uma densidade de 3 a 10 aves por caixa. As caixas de transporte apresentavam as seguintes dimensões: 70cm de comprimento por 60cm de largura.

A variável produtiva utilizada nessa etapa foi o número de aves mortas por caminhão, dentro do intervalo de tempo utilizado no estudo, por meio dos dados de perdas descritos nas planilhas de controle da empresa.

2.2.2.2 Métodos

Os modelos lineares tiveram seu grande *upgrade* com a publicação do relevante artigo de Nelder e Wedderburn (1972), no qual a nova classe de modelos lineares generalizados fazia a unificação de várias técnicas e metodologias estatísticas em uma única moldura. Considerando uma amostra com n observações independente e identicamente distribuídas, \mathbf{X} é uma matriz de delineamento que representa p variáveis explicativas e \mathbf{y} é um vetor de observações amostrado da variável resposta Y , são definidos os três componentes de um MLG (McCULLAGH; NELDER, 1989):

1. Y tem distribuição probabilística que é membro da *família exponencial* de distribuições, com função densidade de probabilidade

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\},$$

com média

$$E(Y_i) = \mu_i = b'(\theta_i)$$

e

$$\text{Var}(Y_i) = a_i(\phi) b''(\theta_i) = a_i(\phi) V(\mu_i) = a_i(\phi) V_i,$$

sendo θ_i o parâmetro canônico, $a_i(\phi) = \phi/w_i$, ϕ o parâmetro de dispersão, w_i um peso *a priori* e V_i a *função de variância* dada por $V_i = d\mu_i/d\theta_i$. As distribuições normal, Poisson, binomial e gama são os membros mais conhecidos dessa família;

2. As covariáveis e os fatores em \mathbf{X} estão relacionados de forma aditiva no *preditor linear*, dado por $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, $i = 1, \dots, n$. Todas as configurações admitidas para os bem conhecidos modelos lineares podem ser aqui utilizados;

3. Uma *função de ligação*, que liga o preditor linear η à média de Y , descrita como $g(\mu_i) = \eta_i$, e portanto, $\mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$.

O método de estimação para o vetor de parâmetros $\boldsymbol{\beta}$ é o método da máxima verossimilhança, utilizando o algoritmo *Iterative Reweighted Least Square (IRLS)*. Esse algoritmo já está implementado em diversos softwares para análise estatística, dentre eles a função `glm()` do ambiente de computação estatística R (R DEVELOPMENT CORE TEAM, 2007), o procedimento `GENMOD` do sistema SAS (SAS INSTITUTE, 2008) e o sistema GENSTAT (PAYNE, 2007). Como medida de discrepância entre o valor observado na amostra e aquele predito pelo modelo estimado, Nelder e Wedderburn (1972) propuseram a medida de *deviance* total $D = \sum_{i=1}^n d_i^2$, sendo que d_i^2 é denominado um *componente de deviance*, calculado para cada observação na amostra por meio da expressão

$$d_i^2 = -2 \int_y^{\mu_i} \frac{y - u_i}{V(u_i)} du.$$

Assumindo $V(\mu_i)$ e μ_i para cada distribuição pertencente à família exponencial na forma canônica, obtêm-se diferentes expressões para a *deviance*. Por exemplo, para a distribuição Normal, é fácil ver que assumindo $V(\mu_i) = 1$ e $\mu_i = \eta_i$, tem-se que D é igual à soma de quadrado dos resíduos. Já para a distribuição Poisson, assumindo $V(\mu_i) = \mu_i$ e $\mu_i = \exp(\eta_i)$, tem-se a estatística G^2 que é utilizada na análise de modelos log-lineares. O componente de *deviance* no caso Poisson é dado por $d_i^2 = w_i [y_i \ln(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)]$. O resíduo componente da *deviance*, utilizado para avaliação da qualidade do ajuste do modelo, é calculado como $r_{d_i} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i^2}$.

Quando a variação observada é muito maior do que a variação esperada, ou seja, $\text{Var}(y) \gg \text{Var}(Y)$, tem-se o fenômeno denominado *superdispersão*, bastante frequente em estudos biológicos. Uma forma de “acomodar” a superdispersão no caso Poisson é adotar uma variância esperada $\text{Var}(Y) = \phi \mu$. Essa modificação no modelo não gera uma distribuição probabilística pertencente à família exponencial e, conseqüentemente, deixa de ser um MLG. Wedderburn (1974) propôs o método da *Quase-Verossimilhança* para estimação dos parâmetros, que permite aproveitar o algoritmo de estimação dos MLG, assumindo apenas a relação entre a média e a variância, sem especificar a distribuição probabilística exata. A obtenção das estimativas dos parâmetros em $\boldsymbol{\beta}$ consiste na

otimização do logaritmo da função de quase-verossimilhança Q definido como:

$$\frac{\partial Q(y_i, \mu_i)}{\partial \mu} = \frac{y_i - \mu_i}{V(\mu_i)} \Rightarrow Q(y_i, \mu_i) = \int^{\mu} \frac{y - u}{V(u)} du + f(y_i).$$

Essas estimativas, nesse caso, coincidem com as estimativas do MLG ordinário. O parâmetro ϕ é estimado igualando-se a estatística de Pearson X^2 ou a *deviance* residual ao número de seus graus de liberdade. Uma vez obtida a estimativa de $\tilde{\phi}$, esta é fixada e o MLG ordinário deve ser reajustado, para que seja possível fazer a inferência sobre os parâmetros. O vetor β será o mesmo mas o erro-padrão estará ajustado em $\sqrt{\tilde{\phi}}$.

Pode-se interpretar que a variância $\text{Var}(Y) = \phi V(\mu)$ possui dois componentes: uma variação dependente da média ($V(\mu)$) e outra variação independente da média (ϕ). Uma extensão desse modelo é adotar um outro preditor linear, ζ , associado à variação não explicada ϕ por meio de uma função de ligação $h(\cdot)$. Smyth (1989) e Nelder e Lee (1991) propuseram o seguinte modelo linear generalizado duplo (MLGD)

$$\begin{cases} \eta_i = g(\mu_i) = \mathbf{x}_i^T \beta \\ \zeta_i = h(\phi_i) = \mathbf{z}_i^T \gamma, \end{cases} \quad (33)$$

sendo que os vetores de covariáveis e/ou fatores \mathbf{z}_i estão associados aos parâmetros em γ , que influenciam na variância da variável resposta. Algumas vantagens dessa abordagem são: (i) distribuições probabilísticas não-normais podem ser consideradas, pertencendo ou não à família exponencial (utilizando-se estimativas de quase-verossimilhança); (ii) preditores lineares mais simples podem ser selecionados, através da inferência correta sobre os parâmetros, atendendo assim ao critério da parcimônia; (iii) fatores experimentais podem ocorrer em \mathbf{Z} , em \mathbf{X} ou, simultaneamente, em ambos e (iv) cada observação contribui para o processo de estimação e, conseqüentemente, gera valores individuais de resíduo. Isso permite aplicar todas as técnicas de análise de resíduos e diagnósticos para a classe MLG, na verificação da qualidade de ajuste.

Para estimação dos parâmetros de (33), um processo iterativo é adotado (SMYTH; VERBYLA, 1996): (i) partindo de valores iniciais para os parâmetros, com γ fixado, estima-se o vetor β através de um MLG ordinário para a variável resposta \mathbf{y} com peso w_i/ϕ_i ; (ii) fixando o vetor β , a estimativa de γ é obtida assumindo ϕ como uma variável resposta com distribuição gama e ajustando um MLG para o preditor linear $\zeta_i = \mathbf{z}_i^T \gamma$, fixando-se o parâmetro de dispersão igual a 2.

Estes dois passos devem ser alternados até que um critério de convergência seja atingido. Neste trabalho foi utilizada a função `dglm()` da biblioteca `MLGD` do ambiente `R` (`R DEVELOPMENT CORE TEAM, 2007`). Esse algoritmo difere da abordagem de Nelder e Lee (1991) e Smyth (1989). Melhorias na aproximação das distribuições dos dados e redução de viés nas estimativas de γ foram propostas (SMYTH; VERBYLA, 1999; LEE; NELDER, 1998). A abordagem de Nelder e Lee já é utilizada na experimentação industrial (PINTO; PONCE DE LEON, 2006), com os algoritmos implementados na linguagem `FORTRAN` para o ajuste dos modelos. Ribatet e Iooss (2006) disponibilizaram a biblioteca de funções `JointModeling` contendo a função `jointfit()` para a linguagem `R`, para o ajuste dos modelos de forma similar à função `glm()`, para o qual, adicionalmente, é inserido um preditor linear para o parâmetro de dispersão. Lee et al. (2006) incluem um CD-ROM com o software `GENSTAT` com a diretiva `HGANALYSE` para a estimação dos parâmetros. Até o momento, esses métodos não foram implementados no sistema `SAS`.

2.2.3 Resultados e discussão

O modelo linear generalizado e suas duas extensões (Quase-Poisson e MLG Duplo) apresentaram diferenças nas estimativas e erros padrões da mortalidade de frangos de corte. O modelo denominado MLG é um modelo linear generalizado ordinário, assumindo distribuição Poisson para o número de aves mortas por caminhão, função de ligação logarítmica e um preditor linear contendo os efeitos principais e interações de segunda ordem. O modelo denominado Quase-Poisson é um MLG adaptado para acomodar o efeito da superdispersão, para o qual a função de variância assumida foi $\text{Var}(Y) = \phi\mu$ e o método da quase-verossimilhança foi utilizado para estimação dos parâmetros (Tabela 1). O Modelo MLGD é um modelo linear generalizado duplo para o qual foi estabelecido um modelo para média (o mesmo assumido para os modelos MLG e Quase-Poisson) e um modelo para a dispersão, contendo efeitos principais.(Tabela 2)

O modelo MLGD completo descrito por Vieira (2008) foi selecionado com base nos testes de Wald (PAWITAN, 2001), validado por uma análise dos resíduos, que indicou uma boa qualidade do ajuste tanto para o modelo da média quanto para o modelo da dispersão.

Para cada modelo foram apresentados os valores das estimativas dos parâmetros e seus respectivos erros padrões e probabilidades de significância (valores de P). No modelo MLG, o parâmetro da dispersão ϕ é fixado em 1; no modelo Quase-Poisson esse parâmetro é estimado com base nos

Tabela 1 - Comparação entre o modelo linear generalizado e Quase-Poisson

| Parâmetros | MLG | | | QUASE-POISSON | | |
|-----------------|-----------------------|-----------------------|------------------------|-----------------------|-----------------------|-----------------------|
| | Estimativa | Erro Padrão | Valor-P | Estimativa | Erro Padrão | Valor-P |
| Intercepto | $-4,228 \cdot 10^1$ | 5,733 | $1,660 \cdot 10^{-13}$ | $-4,228 \cdot 10^1$ | $1,027 \cdot 10^1$ | $3,830 \cdot 10^{-5}$ |
| Tempo Espera | $1,801 \cdot 10^{-2}$ | $5,750 \cdot 10^{-3}$ | 0,002 | $1,801 \cdot 10^{-2}$ | $1,030 \cdot 10^{-2}$ | 0,080 |
| Distância | $2,942 \cdot 10^{-2}$ | $4,084 \cdot 10^{-3}$ | $5,840 \cdot 10^{-13}$ | $2,942 \cdot 10^{-2}$ | $7,313 \cdot 10^{-3}$ | $5,750 \cdot 10^{-5}$ |
| T.Esp:Distância | $1,595 \cdot 10^{-5}$ | $3,671 \cdot 10^{-6}$ | $1,390 \cdot 10^{-5}$ | $1,595 \cdot 10^{-5}$ | $6,575 \cdot 10^{-6}$ | 0,015 |
| ϕ | 1 | - | - | 3,207 | - | - |

Tabela 2 - Modelo linear generalizado duplo (MLGD), para a média e dispersão

| Média | | | |
|-------------------|-----------------------|-----------------------|------------------------|
| Parâmetros | Estimativa | Erro Padrão | Valor-P |
| Intercepto | $-4,971 \cdot 10^1$ | 6,621 | $5,991 \cdot 10^{-14}$ |
| Tempo Espera | $2,365 \cdot 10^{-2}$ | $7,263 \cdot 10^{-3}$ | $1,127 \cdot 10^{-3}$ |
| Distância | $2,807 \cdot 10^{-2}$ | $5,291 \cdot 10^{-3}$ | $1,126 \cdot 10^{-7}$ |
| T.Esp:Distância | $1,621 \cdot 10^{-5}$ | $4,553 \cdot 10^{-6}$ | $3,711 \cdot 10^{-4}$ |
| Dispersão | | | |
| Parâmetros | Estimativa | Erro Padrão | Valor-P |
| Intercepto | -4,825 | 1,580 | $2,262 \cdot 10^{-3}$ |
| Temp. Ext. | 0,094 | 0,029 | $1,427 \cdot 10^{-3}$ |
| poly(T.Esp)1 | -2,912 | 1,679 | $8,297 \cdot 10^{-2}$ |
| poly(T.Esp)2 | -4,393 | 1,466 | $2,739 \cdot 10^{-3}$ |
| poly(Temp. Int.)1 | -4,011 | 2,201 | $6,842 \cdot 10^{-2}$ |
| poly(Temp. Int.)2 | -8,695 | 1,567 | $2,862 \cdot 10^{-8}$ |
| Umid. Rel. Int | 0,038 | 0,011 | $1,326 \cdot 10^{-3}$ |
| Turno tarde | -0,815 | 0,343 | $1,766 \cdot 10^{-2}$ |
| Turno noite | -0,196 | 0,342 | $5,670 \cdot 10^{-1}$ |

dados. No modelo MLGD, a dispersão é uma função de 9 parâmetros, estimada com base nos dados. Um primeiro ponto que chama a atenção são os erros padrões para cada modelo. Dada a forte superdispersão presente nos dados (observando o parâmetro $\phi = 3,207$ no modelo Quase-Poisson), os erros padrões para o MLG ordinário mostraram-se excessivamente pequenos quando comparados com aqueles associados ao modelo Quase-Poisson. Como consequência, as probabilidades de significância sofrem grande alteração. Um exemplo é o caso da interação tempo de espera versus distância. Entretanto, quando comparados com os erros padrões obtidos pelo MLGD, para o qual

a dispersão é modelada e não simplesmente acomodada em um único parâmetro ϕ , nota-se que o modelo Quase-Poisson é mais conservador e o MLGD pode ser visto como um equilíbrio entre o MLG e o modelo Quase-Poisson. Em todos os casos, os valores dos modelos alternativos ao MLG sempre acomodaram uma variabilidade maior, possibilitando que testes de hipóteses e intervalos de confiança possam ser calculados com maior precisão. Em especial, para a interação tempo de espera versus distância percorrida pelos caminhões, essa interação foi considerada significativa em todos os modelos, assumindo o nível de significância de 5%, mas é nítido que o modelo Quase-Poisson foi bastante conservador quando comparado com o MLGD. Pode ser notado também que o modelo Quase-Poisson ajusta apenas o erro-padrão das estimativas e que o MLGD faz ajustes nas estimativas da média em função do modelo de dispersão ajustado.

Uma grande contribuição dos modelos lineares generalizados duplos é a informação gerada pelo modelo de dispersão. A interpretação dessa parte do modelo é similar aos modelos de regressão. Nos parâmetros da dispersão para o MLGD, nota-se que temperatura externa do galpão e umidade relativa interna contribuem linear e positivamente para o aumento da variabilidade do número de aves mortas. Tempo de espera e temperatura interna do galpão climatizado contribuem de forma quadrática na variabilidade. No modelo para a dispersão, essas variáveis foram modeladas através de polinômios ortogonais, para minimizar o mal-condicionamento no processo numérico de estimação. Para ambas as variáveis, o termo quadrático é negativo, indicando que, no caso de temperatura interna, valores menores e maiores tendem a produzir menor variabilidade do que em temperaturas intermediárias. Esse resultado é consistente, uma vez que em temperaturas muito baixas ou muito altas, a mortalidade é maior e menos variável. Turno da tarde reduz a variabilidade no número de aves mortas, assim como o turno da noite, ambos comparados com o turno da manhã. Também, esse resultado é esperado, uma vez que o turno da manhã é um período de transição, variando de baixas temperaturas noturnas para temperaturas maiores no final da manhã.

Por meio das evidências discutidas, verificou-se a contribuição efetiva das variáveis tempo de espera e distância percorrida na mortalidade das aves. Com isso, pode-se utilizar o modelo MLGD ajustado para prever o número de mortes por caminhão em condições diversas (Figura 2). No geral, considerando-se distâncias granjas-abatedouro entre 150 a 280 km, houve diminuição na mortalidade no intervalo de tempo abaixo de 1 hora. No entanto, quando foi observada redução na distância percorrida no mesmo intervalo de tempo, o número de aves mortas dobrou em

relação à primeira situação. Para percursos menores (abaixo de 50 km), a redução na mortalidade correspondeu ao tempo de espera acima de 3 horas de espera.

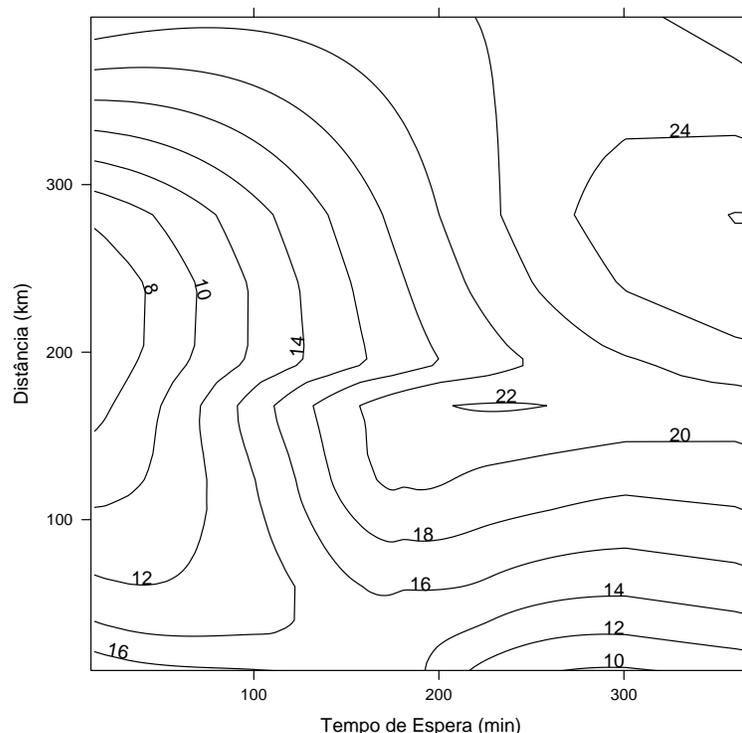


Figura 2 - Número de aves mortas estimado em função da interação entre a distância granja-abatedouro e o tempo de espera

Tais resultados concordam com Bressan e Beraquet (2002), os quais relataram decréscimo na qualidade de carne relacionado com o tempo de espera e com a distância. Os autores descreveram que as aves apresentaram um desgaste acentuado das reservas energéticas nos primeiros 30 minutos de exposição ao estresse, culminando logo após numa adaptação. Todavia, tal efeito adaptativo não acontece nas primeiras horas e, portanto, as aves que percorreram trajetos mais curtos foram encaminhadas ao abate bastante debilitadas quanto aos parâmetros qualitativos. Essa situação justificou a adoção de tempos de espera maiores para distâncias menores, possibilitando às aves o retorno à condição de conforto térmico inicial e diminuindo as proporções de mortalidade no lote.

Vecerek et al. (2006) afirmaram que longas distâncias resultam em aumento do número de aves mortas, na ordem de 0,86% de mortalidade, significativamente maior em relação às distâncias mais curtas (0,15%). Tais resultados concordaram com Warriss et al. (1992), os quais afirmaram que a mortalidade encontrada foi 80% maior em percursos mais longos. No entanto, a dificul-

dade em se manter uma uniformidade quanto à localização das granjas torna inviável a restrição da distância granja-abatedouro para as empresas. Para isso, o tempo de espera é fundamental para se amenizarem os efeitos decorrentes do transporte, desde que o intervalo de tempo seja adequado para cada condição e que a climatização na espera promova o arrefecimento do lote de forma controlada.

2.2.4 Conclusões

A utilização dos modelos lineares generalizados duplos possibilitou um maior entendimento dos fatores influentes na variabilidade do número de aves mortas, consistindo em uma ferramenta robusta de análise em estudos pré-abate de frangos de corte. Por meio dela, foi possível identificar redução da mortalidade sob distâncias longas e tempos de espera curtos, bem como distâncias curtas e tempos de espera longos.

Quanto à modelagem da mortalidade de frangos de corte, os modelos lineares generalizados duplos mostraram-se mais eficientes quando comparados com os modelos lineares generalizados ordinários e sua extensão para acomodar a superdispersão. Essa eficiência caracterizou-se pela obtenção de erros padrões das estimativas dos parâmetros que acomodam a superdispersão, o que gera maior acurácia no cálculo das probabilidades de significância, e, por consequência, maior informação para o pesquisador na tomada de decisões sobre o estudo.

2.3 Modelo linear generalizado duplo: uma perspectiva Bayesiana

2.3.1 Introdução

Experimentos planejados buscam evidências factuais sobre a validade, ou não, de hipóteses científicas previamente estabelecidas. Se esses experimentos são estatisticamente planejados, a análise dos resultados pode ser feita de forma a eliminar viés ou influências externas que comprometam a tomada de decisão sobre os tratamentos estudados. Diversos delineamentos experimentais que são aplicados no dia-a-dia, baseados em métodos conhecidos de longa data (COCHRAN; COX, 1957) assumem que os dados experimentais serão analisados através de modelos que assume variância constante, como um pressuposto inicial. Entretanto, essa pressuposição mostra-se relativamente forte quando se está diante de situações para as quais fatores ambientais ou externos exercem considerável influência sobre as medidas experimentais.

A modelagem explícita da variabilidade na experimentação surge com o trabalho de Taguchi (1985), que enfatiza a necessidade de se tratar, adequadamente, a influência de fatores de controle e ruído em experimentos industriais, como forma de reduzir perdas e otimizar a qualidade de processos industriais. Se fatores ambientais ou mesmo fatores do processo em estudo exercem influência sobre a variância da variável-resposta com distribuição normal, tem-se como consequência que toda a inferência feita sobre o modelo considerado será baseada em uma medida de dispersão inflacionada pelos efeitos não ajustados por esse modelo. Já no caso de medidas na forma de proporções ou contagens, o efeito gerado pelos fatores externos levam a uma subestimação da variância, quando são utilizados modelos probabilísticos como a distribuição binomial e Poisson, levando à indicação de fatores erroneamente significativos (HINDE; DEMÉTRIO, 1998).

Alguns modelos de regressão foram propostos para a média e a dispersão, simultaneamente (AITKIN, 1987; WOLFINGER; TOBIAS, 1998; SMYTH, 1989; NELDER; LEE, 1991). Os dois últimos são baseados na estrutura de modelos lineares generalizados. A inferência sobre os parâmetros desses modelos é baseada na teoria assintótica, e portanto, para grandes amostras, intervalos de confiança e testes de hipóteses podem ser construídos (McCULLAGH; NELDER, 1989). No entanto, em experimentação agrônômica, alguns experimentos apresentam um número grande de parâmetros, em relação ao tamanho da amostra. Sob esse tipo de situação, utilizar as estimativas assintóticas de parâmetros e suas respectivas medidas de incerteza para se fazerem inferências, pode

ser questionável ou até levar a conclusões errôneas.

Neste capítulo, serão apresentados os modelos lineares generalizados duplo (DGLM) e uma proposta para modelagem de proporções com superdispersão, cujo processo de estimação será feito sob o enfoque Bayesiano. Isso permite uma construção conceitualmente simples do modelo e permite incorporar informação *a priori* sobre os parâmetros a serem estimados. As dificuldades para a obtenção de densidades marginais *a posteriori* são transpostas com a utilização de algoritmos eficientes da classe MCMC (GAMERMAN; LOPES, 2006). Uma aplicação será feita para dados de um experimento de cultura de tecidos de maçã, descrito em Ridout e Demétrio (1992).

2.3.2 Modelagem simultânea de média e dispersão

Os modelos propostos por Smyth (1989), Nelder e Lee (1991) estão baseados na estrutura geral dos modelos lineares generalizados – MLG (McCULLAGH; NELDER, 1989). Modelos de regressão que pertencem a essa classe, consideram que a distribuição probabilística da variável resposta pertence à família exponencial (PAWITAN, 2001), com função de probabilidade

$$f(y|\theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}$$

sendo $E(Y) = b'(\theta)$ e $\text{Var}(Y) = \phi b''(\theta)$, $a(\phi) = \phi/w$, ϕ é o parâmetro de dispersão, w é um peso *a priori* e θ é o parâmetro canônico. Assume-se também um preditor linear $\eta = \mathbf{X}\beta$, sendo \mathbf{X} a matriz de planejamento cujas linhas são $\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T$ e β o vetor com $p + 1$ parâmetros desconhecidos. Esse preditor linear é ligado à $E(Y|\theta)$ através de um função de ligação $g(\cdot)$; ou seja, $E(Y_i|\theta) = g^{-1}(\eta_i)$. Neste trabalho, será considerado o caso particular de um MLG com distribuição binomial e com função de ligação logit. Assim, considerando π_i a probabilidade de sucesso para m_i ensaios Bernoulli, tem-se que $\theta_i = \ln [\mu_i/(m_i - \mu_i)]$, $b(\theta_i) = m_i \ln(1 + e^{\theta_i})$ e $\phi = 1$. Portanto, $E(Y_i) = m_i \pi_i = \mu_i$ e $\text{Var}(Y_i) = m_i \pi_i (1 - \pi_i)$ e $g(\mu_i) = \ln [\mu_i/(1 - \mu_i)] = \eta_i$. A estimação dos parâmetros β é feita pelo método da máxima verossimilhança, usando-se o algoritmo de mínimos quadrados ponderados iterativamente (IRLS). Maiores detalhes sobre os MLG's e o método de estimação dos parâmetros estão descritos em Lee, Nelder e Pawitan (2006).

As distribuições binomial e Poisson apresentam o parâmetro de dispersão $\phi = 1$, o que torna a variância dessas distribuições dependente da grandeza da média, unicamente. Entretanto, é bastante

comum que, em situações experimentais, os dados na forma de proporções e contagens apresentem uma variabilidade maior do que a variância teórica dessas distribuições probabilísticas. Esse fenômeno é conhecido como *superdispersão* e são diversas as suas possíveis causas: má-especificação do preditor linear, presença de *outliers*, função de ligação inadequada, proporções com denominador reduzido, dependência longitudinal não ajustada e ocorrência de agrupamentos (*clusters*) que tornam a probabilidade de sucesso π não constante para todas as observações (COLLETT, 1991). Uma fonte particular de superdispersão é quando proporções ou contagens são inflacionadas de zeros (VIEIRA; HINDE; DEMÉTRIO, 2000; RIDOUT; DEMÉTRIO; HINDE, 1998), demandando modelos e testes específicos para essas situações (HALL; BERENHAUT, 2002).

Um desses métodos de estimação foi proposto por Nelder e Pregibon (1987) denominado *quase-verossimilhança estendida*, que consiste em maximizar

$$Q^+ = -\frac{1}{2} \sum_{i=1}^n \left\{ \frac{d(y_i, \mu_i)}{\phi_i} + \ln[2\pi\phi_i V(y_i)] \right\} \quad (34)$$

sendo $V(y_i) = b''(\theta)$ denominada como função de variância e $d(y_i, \mu_i)$ a função *deviance*, que é uma medida de distância entre o valor observado y_i e o valor da média μ_i , calculada como

$$d(y, \mu) = -2 \int_y^\mu \frac{y-t}{V(t)} dt.$$

Dessa forma, o parâmetro de dispersão é indexado nas observações. Isso permite modelar estruturas mais complexas de fatores e/ou covariáveis que influenciam na variabilidade dos dados. Para as distribuições binomial e Poisson, esse método permite modelar a superdispersão em função de um preditor linear, não necessariamente comum ao preditor linear da média.

Algumas propostas foram feitas com o objetivo de modelar simultaneamente a média e a dispersão, dentro da abordagem dos MLG's (SMYTH, 1989; LEE; NELDER, 1991; ENGEL, 1992). Essas propostas possuem igualmente a seguinte estrutura dupla de modelos lineares generalizados. Assumindo $E(Y) = \mu$ e $\text{Var}(Y) = \phi V(\mu)$,

$$\begin{cases} \eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \\ \zeta_i = h(\phi_i) = \mathbf{z}_i^T \boldsymbol{\gamma} \end{cases},$$

sendo β o vetor de parâmetros para a média, γ o vetor de parâmetros para a dispersão, $g(\cdot)$ e $h(\cdot)$ as funções de ligação para a média e a dispersão e \mathbf{x}_i^T e \mathbf{z}_i^T são os vetores-linha das matrizes de planejamento \mathbf{X} e \mathbf{Z} , respectivamente. A matriz \mathbf{X} contém covariáveis e/ou fatores que influenciam na média e a matriz \mathbf{Z} contém as covariáveis e/ou fatores que influenciam no parâmetro de dispersão. Nesse modelo, ϕ representa a variação *independente* da média e $V(\mu)$ a variação *dependente* da média.

Apesar da mesma estrutura conceitual do modelo, as propostas apresentam suas particularidades. Engel (1992) propôs esse modelo, considerando a situação de experimentos industriais, com estruturas de tratamentos fatorial ou fatorial fracionado replicados, sendo que tanto μ quanto ϕ são considerados como medidas sumárias para cada tratamento (média e variância amostral de y para cada tratamento, respectivamente). Essa abordagem é desaconselhada para experimentos não repetidos e para estudos observacionais, por depender de réplicas genuínas. Além disso, não é possível fazer uma análise de diagnósticos para cada observação (LEE; NELDER, 1998). A proposta de Smyth (1989) é mais geral, assim como a de Nelder e Lee (1991), com o modelo construído sobre cada observação. A estimação de parâmetros nessas duas propostas é feita em um esquema iterativo de dois passos: (i) considerando γ fixado, estima-se o vetor β ; (ii) considerando o valor obtido de β , estima-se o valor de γ . Esses dois passos são alternados até que um critério de convergência seja satisfeito. Apesar de as propostas serem diferentes na estimação de cada passo, alguns casos particulares geram resultados similares. Em Smyth (1989) é assumido que Y tem distribuição probabilística pertencente à família de dispersão exponencial (JØRGENSEN, 1987). A estimação do vetor β é feita considerando um MLG ordinário para y considerando pesos *a priori* w_i/ϕ_i . Na estimação do vetor γ é utilizada a aproximação ponto-de-sela (*saddle-point approximation*) como distribuição probabilística de Y e assumindo $d_i = d(y_i, \mu_i)$. Através do cálculo da função geradora de momentos de d_i , pode ser demonstrado que, quando $\phi_i \rightarrow 0$, d_i tem aproximadamente distribuição $\phi\chi_1^2$. Logo, pode-se estimar γ usando-se um modelo linear generalizado para a variável resposta d_i , assumindo que tem distribuição gama e considerando seu parâmetro de dispersão igual a 2 (SMYTH; VERBYLA, 1999). No modelo proposto por Nelder e Lee (1991) todo o processo de estimação é baseado no método QVE. Na fase em que ϕ é fixado, o método de estimação coincide com o método de Smyth (1989), tornando-se uma estimação pelo método QV. Quando β é fixado, a função QVE toma forma de uma função de verossimilhança gama, quando d_i é a variável resposta.

Smyth e Verbyla (1999) apresentam uma versão do método de máxima verossimilhança residual (MVR) para a estimação de γ , já que as estimativas MV sobre d_i são viciadas, principalmente se o número de parâmetros é grande relativamente ao número de observações. Lee e Nelder (1998) também apresentaram uma alternativa do método de estimação baseada no método MVR com o ajuste proposto por Cox e Reid (1987).

2.3.3 Modelo linear generalizado duplo Bayesiano

2.3.3.1 Modelo para medidas com distribuição normal

Os modelos frequentistas de Smyth (1989), Nelder e Lee (1991) são claramente dependentes de suposições assintóticas e aproximadas, uma vez que todo o procedimento inferencial está baseado nessas suposições (McCULLAGH; NELDER, 1989). Em situações experimentais agrônômicas, o número de unidades experimentais é, às vezes, limitado devido a restrições de espaço físico, recursos ou por questões de ética quando se trata de experimentação animal. É comum, também, que o número de parâmetros seja relativamente grande quando comparado com o número de observações. Nos modelos frequentistas isso gera estimativas fortemente viciadas (SMYTH; VERBYLA, 1999). Considerando esse fato, propõe-se aqui um modelo linear generalizado duplo (MLGD) de um ponto de vista Bayesiano.

Uma proposta de um DGLM Bayesiano foi proposto (CEPEDA; GAMERMAN, 2000) com a seguinte estrutura:

$$\begin{aligned}
 y &= \mu + e \\
 e &\sim N(0, \sigma_i^2) \\
 \mu &= \mathbf{X}\beta \\
 g(\sigma_i^2) &= \mathbf{Z}\gamma
 \end{aligned} \tag{35}$$

sendo μ o preditor linear da média, e um componente aleatório com distribuição normal com média zero e variância σ_i^2 . Essa variância é ligada ao preditor linear $\mathbf{Z}\gamma$ de forma não-linear, por meio da função de ligação $g(\cdot)$. \mathbf{X} e \mathbf{Z} são as respectivas matrizes de planejamento da média e da dispersão. Foi assumida como função densidade de probabilidade conjunta *a priori* para β e γ

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{b}_0 \\ \mathbf{g}_0 \end{pmatrix}, \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{G} \end{pmatrix} \right]$$

cujos hiperparâmetros \mathbf{b}_0 , \mathbf{g}_0 , \mathbf{B} , \mathbf{C} e \mathbf{G} são fixos e conhecidos. A função de densidade de probabilidade conjunta *a posteriori* é dada por

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{X}, \mathbf{Z}) &= \frac{p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\beta}, \boldsymbol{\gamma}) p(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\int_{\boldsymbol{\beta}} \int_{\boldsymbol{\gamma}} p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\beta}, \boldsymbol{\gamma}) p(\boldsymbol{\beta}, \boldsymbol{\gamma}) \partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}} \\ &\propto p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\beta}, \boldsymbol{\gamma}) p(\boldsymbol{\beta}, \boldsymbol{\gamma}). \end{aligned} \quad (36)$$

Como essa função assume uma forma analiticamente intratável para as operações de integração, foi utilizado o algoritmo Metropolis-Hastings e um esquema “Blockwise” para a obtenção de amostras das funções densidade marginais *a posteriori* dos parâmetros (GAMERMAN, 1997).

2.3.3.2 Modelo para proporções com superdispersão

Para um DGLM binomial Bayesiano, não se pode tomar diretamente a idéia do DGLM Bayesiano com distribuição normal, uma vez que a distribuição binomial não possui um parâmetro exclusivo para expressar a variabilidade. Hinde e Demétrio (1998) descrevem um modelo logístico-normal com a seguinte estrutura:

$$\begin{aligned} y_i &\sim \text{Bin}(m_i, p_i) \\ \text{logit}(p_i) = \eta_i &= \mathbf{x}_i^T \boldsymbol{\beta} + \sigma \mathbf{z}_i, \\ \mathbf{z}_i &\sim \text{N}(0, 1) \end{aligned} \quad (37)$$

com o objetivo de acomodar o efeito de superdispersão em dados de proporções, usando um parâmetro de efeito aleatório σ . Uma versão Bayesiana de (37) também foi proposta, descrita como

$$\begin{aligned} y_i &\sim \text{Bin}(m_i, p_i) \\ \text{logit}(p_i) = \eta_i &= \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{b}_i, \\ \mathbf{b}_i &\sim \text{N}(0, \sigma^2) \end{aligned} \quad (38)$$

sendo que $\boldsymbol{\beta}$ e $\tau = \sigma^{-2}$ assumem distribuições *a priori* “não-informativas”. O objetivo deste trabalho é apresentar uma generalização de (38), de forma a acomodar os efeitos de covariáveis e/ou fatores na dispersão das proporções. Seja $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ um vetor com n observações da va-

riável aleatória Y , que é igual ao número de sucessos dentre m ensaios Bernoulli, cuja probabilidade de sucesso é p . Então, assume-se o seguinte modelo linear generalizado duplo Bayesiano:

$$\begin{aligned}
 Y_i &\sim \text{Bin}(m_i, p_i) \\
 \text{logit}(p_i) = \eta_i &= \mathbf{x}_i^T \boldsymbol{\beta} + \delta_i \\
 \delta_i &\sim \text{N}(a, \tau_i) \\
 \tau_i &= \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \quad i = 1, \dots, n \\
 \beta_j &\sim \text{N}(b, c) \quad j = 0, \dots, r \\
 \gamma_k &\sim \text{N}(d, e) \quad k = 0, \dots, s
 \end{aligned} \tag{39}$$

sendo \mathbf{x}_i e \mathbf{z}_i os respectivos vetores-linha das matrizes de planejamento \mathbf{X} e \mathbf{Z} ; δ_i um efeito aleatório com distribuição normal e com média a e variância τ_i^{-1} , ou seja, com precisão τ_i ; $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ os vetores de parâmetros desconhecidos e que assumem distribuições *a priori* normais com hiperparâmetros conhecidos b , c , d e e . Assume-se, também, que β_j e γ_k são independentes, ou seja, $p(\beta_j, \gamma_k) = p(\beta_j)p(\gamma_k)$, considerando a dificuldade prática em estabelecer “a priori” uma estrutura de dependência entre estes parâmetros. Nesse modelo, a função de ligação para a média de y_i é $\text{logit}(p_i) = \ln[p_i/(1 - p_i)] = \ln[\mu_i/(m_i - \mu_i)]$. A função de ligação da variância do efeito aleatório δ_i assumida aqui é $\ln \sigma_i^2$.

Considerando-se a abordagem Bayesiana, a função densidade de probabilidade conjunta *a posteriori* para o modelo (39) pode ser descrita como

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta} | \mathbf{X}, \mathbf{Z}, \mathbf{y}) \propto L(\boldsymbol{\beta}, \boldsymbol{\delta} | \mathbf{y}, \mathbf{X}) p(\boldsymbol{\delta} | \boldsymbol{\gamma}, \mathbf{Z}) p(\boldsymbol{\beta}) p(\boldsymbol{\gamma}), \tag{40}$$

ou seja, baseada na independência condicional de alguns parâmetros. Dessa forma o modelo pode ser descrito por um modelo gráfico acíclico direcionado (GAD) conforme descrito em Best e Green (2007). A vantagem em se apresentar um modelo na forma de um GAD é que a essência da estrutura do modelo é preservada, clareando a informação da forma funcional do modelo e “escondendo” em primeira instância as suposições distribucionais e relações determinísticas entre variáveis e parâmetros. Além disso, a estrutura visual do modelo gráfico sugere uma independência condicional para a implementação matemática e computacional. O ambiente de computação Bayesiana OpenBUGS

(THOMAS et al., 2006) está construído para executar os cálculos das distribuições *a posteriori* dos parâmetros para um GAD descrito graficamente ou na linguagem BUGS (SPIEGELHALTER et al., 1996). Resumidamente, um modelo gráfico é uma técnica bastante útil para especificar e manipular modelos probabilísticos multivariados. Ele consiste de *nós* (representados por elipses e retângulos), *setas* e *camadas*. Os nós na forma de elipses representam variáveis aleatórias ou relações determinísticas entre outros nós. Os retângulos representam valores constantes. As setas direcionam o fluxo de informação entre os nós. As camadas podem representar as observações para cada unidade observacional, parâmetros de diferentes agrupamentos, etc. Diversas camadas podem ser estabelecidas, sobrepostas ou independentes.

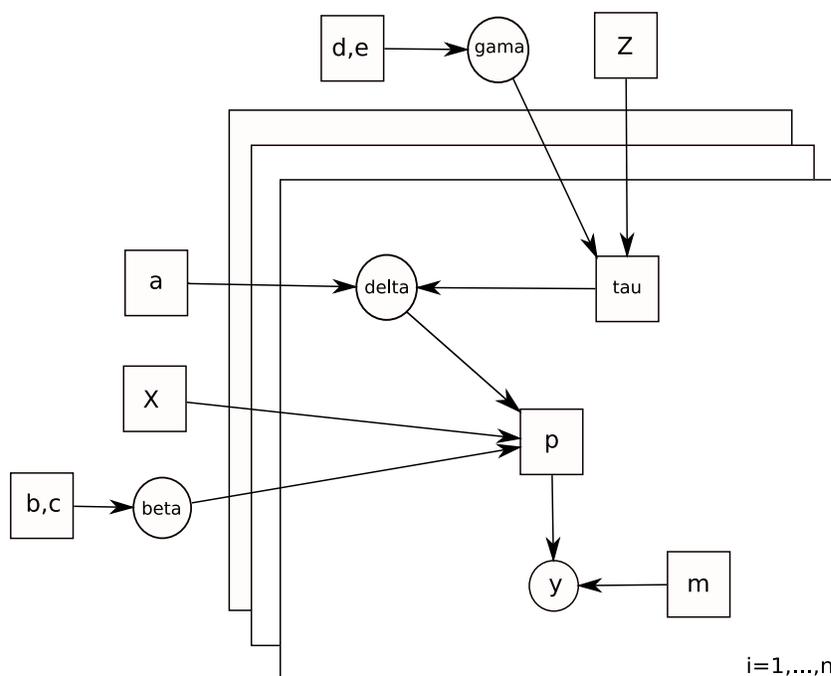


Figura 3 - Modelo gráfico acíclico direcionado para o modelo linear generalizado duplo binomial Bayesiano

Estimação das funções densidades marginais *a posteriori*

Serão descritos a seguir, a construção probabilística do modelo (39) e o algoritmo de estimação das funções densidade de probabilidade *a posteriori* dos parâmetros. A função densidade *a posteriori* para esse modelo pode ser construída a partir do GAD descrito na Figura 3 como (40).

As funções densidades *a priori* de β_j e γ_k são

$$\begin{aligned} p(\beta_j) &= \sqrt{\frac{c^{-1}}{2\pi}} \exp\left[-\frac{c^{-1}}{2}(\beta_j - b)^2\right] \\ &\propto \exp\left[-\frac{c^{-1}}{2}(\beta_j - b)^2\right], \quad j = 0, \dots, r \end{aligned} \quad (41)$$

e

$$\begin{aligned} p(\gamma_k) &= \sqrt{\frac{e^{-1}}{2\pi}} \exp\left[-\frac{e^{-1}}{2}(\gamma_k - d)^2\right] \\ &\propto \exp\left[-\frac{e^{-1}}{2}(\gamma_k - d)^2\right], \quad k = 0, \dots, s. \end{aligned} \quad (42)$$

Considerando que $y_i \sim \text{Bin}(m_i, p_i)$ e que $\ln[p_i/(1 - p_i)] = \mathbf{x}_i^T \boldsymbol{\beta} + \delta_i$, tem-se como função de verossimilhança

$$\begin{aligned} p(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \delta_i) &= \binom{m_i}{y_i} \left[\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)} \right]^{y_i} \left[1 - \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)} \right]^{m_i - y_i} \\ &= \binom{m_i}{y_i} \left[\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)} \right]^{y_i} \left[1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i) \right]^{y_i - m_i} \\ &= \binom{m_i}{y_i} \frac{[\exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{y_i}}{[1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{m_i}}; \end{aligned}$$

e, portanto,

$$p(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \delta_i) \propto [\exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{y_i} [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{-m_i}. \quad (43)$$

A função densidade condicional de δ dado o vetor de parâmetros $\boldsymbol{\gamma}$ é

$$\begin{aligned} p(\delta_i | \boldsymbol{\gamma}) &= \sqrt{\frac{\tau_i}{2\pi}} \exp\left\{-\frac{\tau_i}{2}(\delta_i - a)^2\right\} \\ &= \sqrt{\frac{1}{2\pi \exp(\mathbf{z}_i^T \boldsymbol{\gamma})}} \exp\left\{-\frac{1}{2}[\exp(\mathbf{z}_i^T \boldsymbol{\gamma})]^{-1}(\delta_i - a)^2\right\} \end{aligned}$$

$$\begin{aligned}
p(\delta_i|\boldsymbol{\gamma}) &= (2\pi)^{-\frac{1}{2}}[\exp(\mathbf{z}_i^T \boldsymbol{\gamma})]^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \exp(-\mathbf{z}_i^T \boldsymbol{\gamma})(\delta_i - a)^2 \right\} \\
&= (2\pi)^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma} \right) \exp \left\{ -\frac{1}{2} \exp(-\mathbf{z}_i^T \boldsymbol{\gamma})(\delta_i - a)^2 \right\} \\
&= (2\pi)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma} - \frac{1}{2} \exp(-\mathbf{z}_i^T \boldsymbol{\gamma})(\delta_i - a)^2 \right\}
\end{aligned}$$

e, portanto,

$$p(\delta_i|\boldsymbol{\gamma}) \propto \exp \left\{ -\frac{1}{2} [\mathbf{z}_i^T \boldsymbol{\gamma} + \exp(-\mathbf{z}_i^T \boldsymbol{\gamma})(\delta_i - a)^2] \right\}. \quad (44)$$

Aplicando (41), (42), (43) e (44) em (40), considerando o vetor de observações \mathbf{y} , tem-se como função densidade de probabilidade conjunta *a posteriori*

$$\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta} | \mathbf{X}, \mathbf{Z}, \mathbf{y}) &\propto \prod_{i=1}^n p(y_i | \boldsymbol{\beta}, \mathbf{x}_i, \delta_i) p(\delta_i | \boldsymbol{\gamma}) \prod_{j=0}^r p(\beta_j) \prod_{k=0}^s p(\gamma_k) \\
&\propto \prod_{i=1}^n \frac{[\exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{y_i}}{[1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{m_i}} \exp \left\{ -\frac{1}{2} [\mathbf{z}_i^T \boldsymbol{\gamma} + \exp(-\mathbf{z}_i^T \boldsymbol{\gamma})(\delta_i - a)^2] \right\} \times \\
&\quad \times \prod_{j=0}^r \exp \left[-\frac{c^{-1}}{2} (\beta_j - b)^2 \right] \prod_{k=0}^s \exp \left[-\frac{e^{-1}}{2} (\gamma_k - d)^2 \right] \\
&\propto \prod_{i=1}^n \frac{\exp[y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]}{[1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{m_i}} \exp \left\{ -\frac{1}{2} [\mathbf{z}_i^T \boldsymbol{\gamma} + \exp(-\mathbf{z}_i^T \boldsymbol{\gamma})(\delta_i - a)^2] \right\} \times \\
&\quad \times \prod_{j=0}^r \exp \left[-\frac{c^{-1}}{2} (\beta_j - b)^2 \right] \prod_{k=0}^s \exp \left[-\frac{e^{-1}}{2} (\gamma_k - d)^2 \right] \\
&\propto \frac{\prod_{i=1}^n \exp \left\{ y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i) - \frac{1}{2} [\mathbf{z}_i^T \boldsymbol{\gamma} + \exp(-\mathbf{z}_i^T \boldsymbol{\gamma})(\delta_i - a)^2] \right\}}{\prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{m_i}} \times \\
&\quad \times \exp \left\{ -\frac{1}{2} \left[\frac{1}{c} \sum_{j=0}^r (\beta_j - b)^2 + \frac{1}{e} \sum_{k=0}^s (\gamma_k - d)^2 \right] \right\} \\
&\propto \exp \left\{ \sum_{i=1}^n \mathbf{x}_i^T \boldsymbol{\beta} y_i + \sum_{i=1}^n y_i \delta_i - \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\gamma} - \frac{1}{2} \sum_{i=1}^n \exp(-\mathbf{z}_i^T \boldsymbol{\gamma})(\delta_i - a)^2 \right\} \times \\
&\quad \times \prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{-m_i} \exp \left\{ -\frac{1}{2c} \sum_{j=0}^r (\beta_j - b)^2 - \frac{1}{2e} \sum_{k=0}^s (\gamma_k - d)^2 \right\}
\end{aligned}$$

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta} | \mathbf{X}, \mathbf{Z}, \mathbf{y}) \propto \exp \left\{ \sum_{i=1}^n \mathbf{x}_i^T \boldsymbol{\beta} y_i + \sum_{i=1}^n y_i \delta_i - \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\gamma} - \frac{1}{2} \sum_{i=1}^n \exp(-\mathbf{z}_i^T \boldsymbol{\gamma}) (\delta_i - a)^2 - \frac{1}{2c} \sum_{j=0}^r (\beta_j - b)^2 - \frac{1}{2e} \sum_{k=0}^s (\gamma_k - d)^2 \right\} \prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{-m_i}$$

e, finalmente,

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta} | \mathbf{X}, \mathbf{Z}, \mathbf{y}) \propto \exp \left\{ \mathbf{y}^T \mathbf{X} \boldsymbol{\beta} + \mathbf{y}^T \boldsymbol{\delta} - \frac{1}{2} \mathbf{1}^T \mathbf{Z} \boldsymbol{\gamma} - \sum_{i=1}^n \exp(-\mathbf{z}_i^T \boldsymbol{\gamma}) (\delta_i - a)^2 - \frac{1}{2c} \sum_{j=0}^r (\beta_j - b)^2 - \frac{1}{2e} \sum_{k=0}^s (\gamma_k - d)^2 \right\} \times \prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{-m_i}. \quad (45)$$

Essa expressão não permite a obtenção de funções densidades marginais para $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ e $\boldsymbol{\delta}$ diretamente via integração. Também é inviável a utilização de métodos de integração numérica, dada a multidimensionalidade dessa função. A solução pode ser obtida através de simulação estocástica de amostras das funções densidades marginais *a posteriori*. Os algoritmos de simulação utilizados são denominados Monte Carlo via Cadeias de Markov (MCMC), sendo os mais conhecidos o amostrador de Gibbs (GELFAND; SMITH, 1991) e o algoritmo Metropolis-Hastings (GELMAN et al., 2000). Esse último consiste em definir uma distribuição de transição $J(\cdot, \cdot)$ no espaço paramétrico de um vetor de parâmetros $\boldsymbol{\theta}$ (nesse problema tem-se $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \boldsymbol{\delta}^T)$). Os valores de $\boldsymbol{\theta}$ são gerados sequencialmente ($\boldsymbol{\theta}^t$, $t = 1, 2, 3, \dots$). Um novo valor $\boldsymbol{\theta}^*$ é amostrado e será aceito como um valor $\boldsymbol{\theta}^t$ com probabilidade de aceitação α : se aceito, $\boldsymbol{\theta}^t = \boldsymbol{\theta}^*$; caso contrário, $\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1}$. Essa sequência tem a propriedade Markoviana e, portanto, quando uma longa sequência é gerada e confirmada que se trata de uma cadeia irreduzível, aperiódica e recorrente positivamente (PAULINO et al., 2003), a distribuição de equilíbrio resultante dos valores de $\boldsymbol{\theta}^t$ é uma aproximação para a distribuição estacionária *a posteriori* de $p(\boldsymbol{\theta} | \mathbf{X}, \mathbf{Z}, \mathbf{y})$. A construção do núcleo das cadeias de Markov

depende das funções densidades marginais condicionais completas *a posteriori*, dadas por

$$p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\delta}, \mathbf{X}, \mathbf{Z}, \mathbf{y}) \propto \exp \left\{ \mathbf{y}^T \mathbf{X} \boldsymbol{\beta} - \frac{1}{2c} \sum_{j=0}^r (\beta_j - b)^2 \right\} \prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{-m_i} \quad (46)$$

$$p(\boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{X}, \mathbf{Z}, \mathbf{y}) \propto \frac{\exp \left\{ -\frac{1}{2} \sum_{i=1}^n \exp(-\mathbf{z}_i^T \boldsymbol{\gamma}) (\delta_i - a)^2 - \frac{1}{2e} \sum_{k=0}^s (\gamma_k - d)^2 \right\}}{\sqrt{\exp(\mathbf{1}^T \mathbf{Z} \boldsymbol{\gamma})}} \quad (47)$$

$$p(\boldsymbol{\delta}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{X}, \mathbf{Z}, \mathbf{y}) \propto \frac{\exp \left\{ \mathbf{y}^T \boldsymbol{\delta} - \frac{1}{2} \sum_{i=1}^n \exp(-\mathbf{z}_i^T \boldsymbol{\gamma}) (\delta_i - a)^2 \right\}}{\prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \delta_i)]^{m_i}}. \quad (48)$$

Como essas funções densidades não apresentam formas conhecidas e fáceis de simular, não é possível utilizar o amostrador de Gibbs, diretamente. Torna-se necessária a utilização de algoritmos MCMC mais gerais, como o Metropolis-Hastings ou Slice Sampling (NEAL, 2003) dentro do amostrador de Gibbs. Nesse trabalho foi utilizado o algoritmo de Metropolis-Hastings, descrito a seguir. Considera-se θ o parâmetro de interesse e $p(\theta|y)$ a distribuição da qual se deseja amostrar (não necessariamente uma função densidade normalizada, que integre em 1). O algoritmo Metropolis-Hastings gerará uma seqüência $\theta^1, \theta^2, \dots$, a partir de um ponto inicial θ^0 , da seguinte forma:

1. amostrar um ponto-candidato θ^* da distribuição-proposta no tempo t , $J(\theta^*|\theta^{t-1})$;
2. calcular a probabilidade de aceitação do ponto-candidato θ^* :

$$\alpha = \frac{p(\theta^*|y)/J_t(\theta^*|\theta^{t-1})}{p(\theta^{t-1}|y)/J_t(\theta^{t-1}|\theta^*)};$$

3. Gerar uma observação $u \sim U(0, 1)$. Faça

$$\theta^t = \begin{cases} \theta^* & \text{se } \alpha < u; \\ \theta^{t-1} & \text{caso contrário.} \end{cases}$$

Quando a seqüência θ^t , $t = 1, 2, \dots$ alcança a convergência, a amostra de θ^t pode ser considerada como uma amostra de $p(\theta|y)$. Teoricamente, pode ser demonstrada a convergência da seqüência para a distribuição estacionária $p(\theta|y)$ mas na prática de análise de dados, gráficos devem ser construídos e medidas de diagnósticos devem ser calculadas para atestar a convergência (GELMAN et al., 2000).

Diagnóstico das cadeias e seleção de modelos

A abordagem Bayesiana, via métodos de simulação estocástica, é bastante flexível para a modelagem estatística de problemas bastante complexos, que são às vezes difíceis de lidar do ponto de vista freqüentista. Entretanto, há um preço a pagar: as cadeias de Markov simuladas para cada parâmetro devem ser diagnosticadas, com o objetivo de garantir que amostras geradas a partir das cadeias contêm as características principais da distribuição de equilíbrio, como forma, primeiros momentos amostrais, etc. Diversos métodos formais (baseados em testes) e informais (baseados em análise gráfica) são indicadores da qualidade das amostras simuladas. Neste trabalho, foram utilizados métodos gráficos na fase de seleção de modelos e métodos gráficos e testes formais na análise do modelo final. Os testes utilizados foram os de Geweke (1992), Raftery e Lewis (1992) e Heidelberger e Welch (1983), todos implementados na biblioteca `coda` (PLUMMER et al., 2007).

Para a seleção de modelos, o critério DIC (Deviance Information Criteria) foi utilizado (SPIEGELHALTER et al., 2002). Este índice é calculado como

$$\text{DIC} = p_D + E_\theta[D(\theta)]$$

sendo $p_D = E_\theta[D(\theta)] - D(E_\theta[p(\theta|y)])$, representando o número efetivo de parâmetros, $E_\theta[D(\theta)]$ é a média de D calculado sobre todos os valores de θ da amostra gerada pelos métodos MCMC e $D(E_\theta[p(\theta|y)])$ a medida de *deviance* calculada sobre a média dos valores amostrados de θ . Essa medida, assim como os critérios AIC e BIC, quantifica a qualidade do ajuste do modelo e, ao mesmo tempo, penaliza a complexidade do mesmo.

2.3.3.3 Aplicação: experimento de cultura de tecidos

O modelo (39) foi usado para a análise dos dados obtidos de um experimento com cultura de tecido vegetal, especificamente tecido de maçã, descrito em Ridout e Demétrio (1992). A estrutura de tratamentos era fatorial 2×5 e a estrutura de parcelas foi completamente aleatorizada. A unidade experimental eram placas de Petri divididas em uma matriz 5×5 , gerando 25 compartimentos individuais. Em cada compartimento foi colocado um meio de cultura com volume padronizado, no qual foi colocado um pequeno pedaço de tecido vegetal, denominado *explante*. As placas de Petri foram mantidas em uma incubadora por algumas semanas. Nesse período, novos brotos podem

surgir dos explantes, o que efetiva o processo de regeneração. Um dos objetivos dos pesquisadores foi identificar se algum dos 5 tipos de explantes e/ou um dos 2 tipos de meio de cultura influenciam na proporção de regeneração dos explantes. Os dados estão reproduzidos na Tabela 3. Na Figura 4 é apresentado um gráfico das médias do número de explantes regenerados em função dos meios de cultura e dos tipos de explantes. Nota-se a forte variabilidade das médias de explantes regenerados no meio de cultura X, quando comparadas com as médias no meio de cultura Y. O explante E parece ter um comportamento diferenciado dos demais quando é observado no meio de cultura X, sugerindo uma possível interação. Na Figura 5, é apresentado o mesmo tipo de gráfico, mas ao invés de média como medida sumária, foi utilizado o desvio-padrão. Para os explantes com repetições, o gráfico sugere que o explante D deve exercer uma forte influência na dispersão.

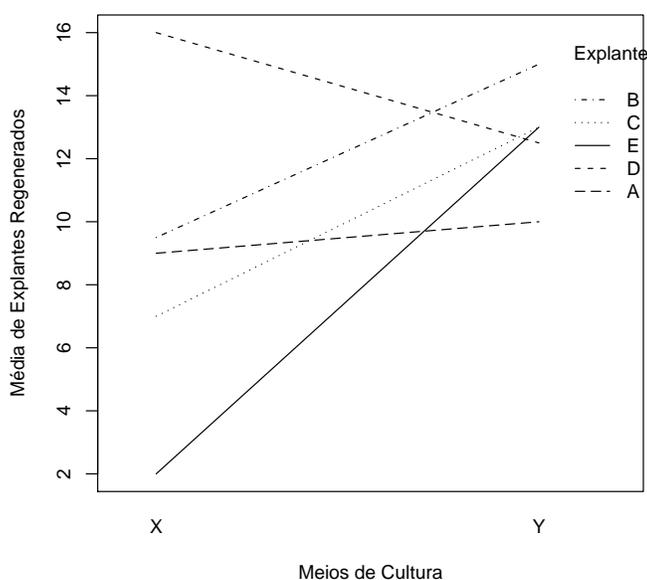


Figura 4 - Média do número de explantes regenerados em função do meio de cultura e do tipo de Explante

Para as análises, foi utilizado o ambiente de computação estatística R (R DEVELOPMENT CORE TEAM, 2007) em conjunto com as bibliotecas `BRugs` (THOMAS et al., 2006) e `coda` (PLUMMER et. al., 2007). O modelo estatístico (39) foi adotado e, para o processo de estimação das funções densidades marginais *a posteriori*, foram utilizadas 3 cadeias de Markov, com pontos iniciais dispersos no espaço paramétrico. Foram realizadas 110000 simulações, sendo as 40000 primeiras apenas para aquecimento das cadeias e das 70 mil restantes foram tomadas uma observação

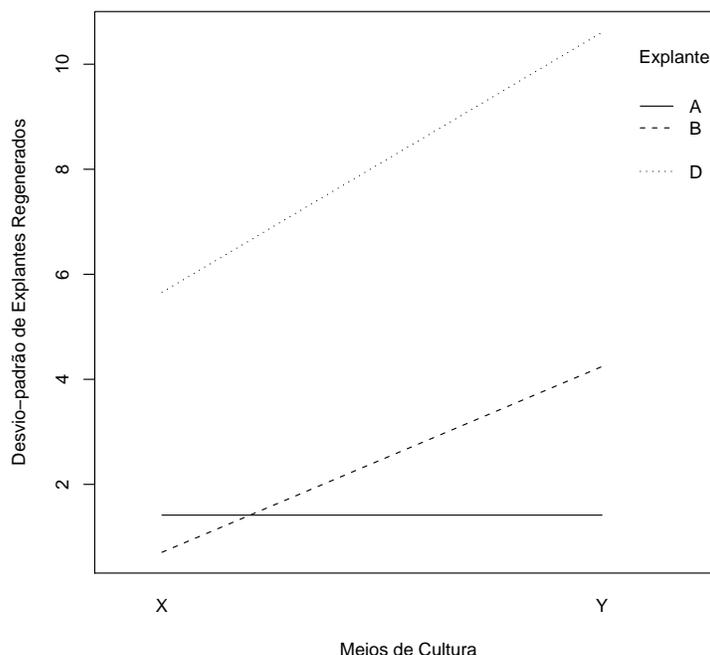


Figura 5 - Desvio padrão do número de explantes regenerados em função do meio de cultura e do tipo de Explante

a cada 50, para minimizar o efeito de autocorrelação das cadeias.

Modelos sucessivos foram ajustados, do modelo maximal até o modelo mais parcimonioso. A seleção de modelos se baseou no DIC. O modelo mais complexo ajustado foi assumindo

$$\eta = \text{cultura} + \text{explante} + \text{cultura} \times \text{explante}$$

$$\zeta = \text{cultura} + \text{explante} + \text{cultura} \times \text{explante}.$$

Na Tabela 4, pode ser notado que o modelo que minimizou tanto a medida de qualidade de ajuste quanto a complexidade do modelo (conseqüentemente, minimizando o índice DIC) foi o modelo 5, que contém como preditor linear para a dispersão a variável indicadora do explante D e no preditor linear da média os efeitos principais de cultura, explante e o termo de interação entre o meio de cultura Y e o explante E. O modelo 6 contém apenas um termo constante no preditor para a dispersão, o modelo 7 contém apenas o efeito aleatório δ com distribuição $N(0, 1)$ e o modelo 8 é um GLM binomial ordinário, sem o efeito aleatório, respectivamente. É bem destacado o forte efeito da superdispersão, dada a má qualidade do ajuste para esse modelo. Os modelos 9, 10 e

Tabela 3 - Número de explantes (y_i) de maçã que regeneraram, dentre 25, considerando 16 placas de Petri

| Placa | Tipo Explante | Meio de Cultura | Número Regenerados |
|-------|---------------|-----------------|--------------------|
| 1 | A | X | 8 |
| 2 | A | X | 10 |
| 3 | A | Y | 9 |
| 4 | A | Y | 11 |
| 5 | B | X | 9 |
| 6 | B | X | 10 |
| 7 | B | Y | 18 |
| 8 | B | Y | 12 |
| 9 | C | X | 7 |
| 10 | C | Y | 13 |
| 11 | D | X | 20 |
| 12 | D | X | 12 |
| 13 | D | Y | 20 |
| 14 | D | Y | 5 |
| 15 | E | X | 2 |
| 16 | E | Y | 13 |

FONTE: Ridout e Demétrio (1992).

11 são modelos encaixados para o preditor linear da média, mostrando pelos seus índices, uma performance inferior ao modelo 5.

Tomando a mediana como uma estimativa pontual dos parâmetros do modelo, tem-se que o explante D aumenta a variância do efeito aleatório δ para 2,8, enquanto os outros tipos de explantes reduzem a variância deste efeito aleatório para praticamente zero. Apenas o explante D é responsável pela superdispersão na proporção de explantes regenerados. No efeito dos fatores na média de explantes regenerados, a influência maior é da interação entre o meio de cultura Y versus o explante E. Quando essa combinação ocorre, ela aumenta em aproximadamente 1,13 vezes a chance de regeneração do explante, em relação aos demais explantes e o meio de cultura X. Esses efeitos na média e na dispersão do efeito aleatório podem ser visualizados nas Figuras 6 e 7.

2.3.4 Conclusões

O objetivo deste trabalho foi apresentar uma proposta Bayesiana do modelo linear generalizado duplo para aplicação a dados de proporções com superdispersão, como alternativa aos modelos

Tabela 4 - Medidas de qualidade de ajuste e complexidade dos modelos MLGD Binomial Bayesiano ajustados

| Modelos Ajustados | DIC | pD | \bar{D} |
|--|--------|-------|-----------|
| Modelo 1: $\left\{ \begin{array}{l} \eta = \text{cultura} + \text{explante} + \text{cultura} \times \text{explante} \\ \zeta = \text{cultura} + \text{explante} + \text{cultura} \times \text{explante} \end{array} \right.$ | 86,31 | 13,15 | 73,16 |
| Modelo 2: $\left\{ \begin{array}{l} \eta = \text{cultura} + \text{explante} + \text{cultura} \times \text{explante} \\ \zeta = \text{cultura} + \text{explante} \end{array} \right.$ | 86,10 | 13,44 | 72,66 |
| Modelo 3: $\left\{ \begin{array}{l} \eta = \text{cultura} + \text{explante} + \text{cultura} \times \text{explante} \\ \zeta = \text{explante} \end{array} \right.$ | 84,69 | 12,79 | 71,9 |
| Modelo 4: $\left\{ \begin{array}{l} \eta = \text{cultura} + \text{explante} + \text{cultura} \times \text{explante} \\ \zeta = \gamma_0 + \text{explante}D \end{array} \right.$ | 84,49 | 18,81 | 71,67 |
| Modelo 5: $\left\{ \begin{array}{l} \eta = \text{cultura} + \text{explante} + \text{cultura}Y \times \text{explante}E \\ \zeta = \gamma_0 + \text{explante}D \end{array} \right.$ | 82,55 | 10,96 | 71,55 |
| Modelo 6: $\left\{ \begin{array}{l} \eta = \text{cultura} + \text{explante} + \text{cultura}Y \times \text{explante}E \\ \zeta = \gamma_0 \end{array} \right.$ | 87,16 | 14,52 | 72,64 |
| Modelo 7: $\left\{ \begin{array}{l} \eta = \text{cultura} + \text{explante} + \text{cultura}Y \times \text{explante}E \\ \zeta = 0 \Rightarrow \delta \sim N(0, 1) \end{array} \right.$ | 86,16 | 14,77 | 71,39 |
| Modelo 8: $\left\{ \begin{array}{l} \eta = \text{cultura} + \text{explante} + \text{cultura}Y \times \text{explante}E \end{array} \right.$ | 107,10 | 7,03 | 100,10 |
| Modelo 9: $\left\{ \begin{array}{l} \eta = \text{cultura} + \text{explante} \\ \zeta = \gamma_0 + \text{explante}D \end{array} \right.$ | 85,77 | 11,19 | 74,58 |
| Modelo 10: $\left\{ \begin{array}{l} \eta = \text{explante} \\ \zeta = \gamma_0 + \text{explante}D \end{array} \right.$ | 88,66 | 14,37 | 74,29 |
| Modelo 11: $\left\{ \begin{array}{l} \eta = \text{cultura} \\ \zeta = \gamma_0 + \text{explante}D \end{array} \right.$ | 85,18 | 8,82 | 76,30 |

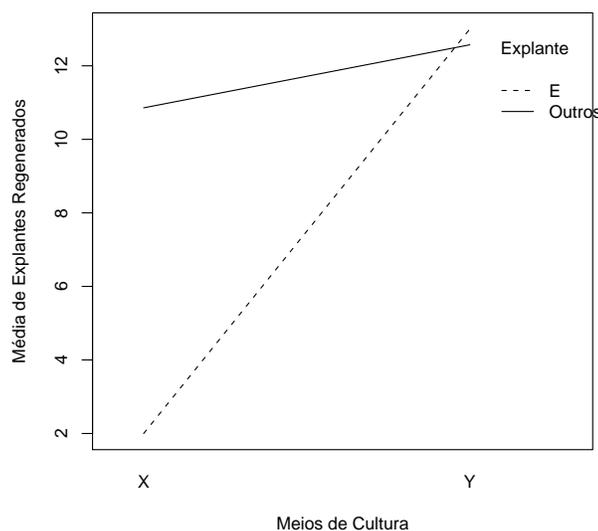


Figura 6 - Média do número de explantes regenerados em função do meio de cultura e do tipo de Explante para o modelo 5

freqüentistas propostos por Smyth (1989), Nelder e Lee (1991). A construção deste modelo foi baseada na inclusão de um efeito aleatório no preditor linear do modelo linear generalizado, sendo

Tabela 5 - Estatísticas sumárias das amostras das densidades marginais dos parâmetros do modelo 5, calculadas com base em 4.200 observações

| Parâmetros | Média | Desvio- Padrão | Erro | | Mediana | Perc 97,5% |
|---------------|-------|-------------------|----------|-----------|---------|------------|
| | | | Padrão | Perc 2,5% | | |
| β_0 | -0,81 | 0,28 | 0,004383 | -1,36 | -0,80 | -0,26 |
| β_Y | 0,59 | 0,29 | 0,004598 | 0,02 | 0,59 | 1,16 |
| β_B | 0,48 | 0,33 | 0,004815 | -0,19 | 0,48 | 1,13 |
| β_C | 0,09 | 0,41 | 0,006554 | -0,70 | 0,11 | 0,85 |
| β_D | 0,82 | 1,09 | 0,021200 | -1,41 | 0,83 | 2,96 |
| β_E | -1,92 | 0,90 | 0,012980 | -3,90 | -1,85 | -0,35 |
| $\beta_{Y,E}$ | 2,21 | 1,03 | 0,014570 | 0,38 | 2,18 | 4,39 |
| γ_0 | -6,31 | 3,90 | 0,178600 | -15,80 | -5,55 | -1,09 |
| γ_D | 7,22 | 3,99 | 0,179400 | 1,47 | 6,58 | 16,75 |

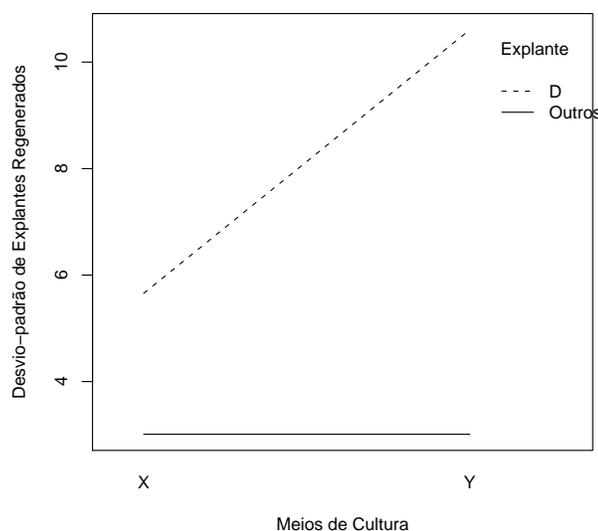


Figura 7 - Desvio-padrão do número de explantes regenerados em função do meio de cultura e do tipo de Explante para o modelo 5

que a variância do efeito aleatório é dependente de outro preditor linear. Este modelo foi aplicado com sucesso a uma situação bastante comum na experimentação agrônômica, e extensiva a outras situações semelhantes. Experimentos com o objetivo de identificar fatores que influenciam na variabilidade de um processo industrial (como os Experimentos de Taguchi) podem ser analisados utilizando essa abordagem.

A abordagem Bayesiana baseada em simulação estocástica é bastante flexível para o ajuste dessa classe de modelos, sem ter que se limitar às distribuições pertencentes à família exponencial,

a grandes amostras, funções de ligação específicas e, conseqüentemente, aproximações assintóticas. A construção de modelos gráficos direcionados acíclicos facilita a apresentação do conceito do modelo e também a sua especificação probabilística, baseada no condicionamento dos valores aleatórios imediatamente anteriores. A utilização dos algoritmos MCMC contornam de forma eficiente o problema do cálculo de integrais multidimensionais, permitindo a obtenção de amostras das densidades marginais *a posteriori* associadas aos parâmetros do modelo. Para essas amostras foram calculadas estatísticas sumárias que permitem fazer inferências. Entretanto, essa abordagem teve seu preço: os métodos são computacionalmente intensivos e uma análise dos resultados das cadeias de Markov deve ser feita, para garantir consistência nos procedimentos de inferências. O ambiente de computação estatística R, juntamente com as bibliotecas BRugs e coda se mostraram bastante eficientes e flexíveis para o processo de estimação, de simulação e para a análise dos dados.

Outras possibilidades poderão ser desenvolvidas, como extensão desses modelos. A incorporação da dependência temporal em medidas repetidas, os modelos para dados de contagens e de sobrevivência são algumas situações que podem ser estudadas. Os modelos estudados utilizaram densidades *a priori* normais, que se mostraram insensíveis a variações nos hiperparâmetros para o preditor da média mas sensíveis a prioris informativas muito vagas para o preditor da variância do efeito aleatório, o que exige que alguma informação sobre a amplitude desse parâmetros seja fornecida pelo pesquisador. Aspectos sobre identificabilidade, reparametrização, e teste com outras densidades *a priori* merecem mais estudos.

2.4 Modelos para dados binários longitudinais com superdispersão

2.4.1 Introdução

Resultados experimentais e observacionais medidos na forma binária (seja presença ou ausência de alguma característica em uma unidade observacional) são bastante recorrentes em diversas áreas da Ciência e os modelos lineares generalizados (MLG) trouxeram unidade e flexibilidade para a análise deste tipo de dados. A transformação de dados (BOX; COX, 1964) foi, por muito tempo, utilizada como a única solução para lidar com os pressupostos necessários para a utilização dos modelos lineares clássicos, mas raramente atendia, simultaneamente, aos requisitos de normalidade e homogeneidade da variância. A classe dos modelos lineares generalizados (NELDER; WEDERBURN, 1972) engloba um grande número de modelos conhecidos até então, como métodos estatísticos independentes, dentre eles o modelo probit e o modelo de regressão logística (DEMÉTRIO, 2002). A generalidade dessa classe de modelos, sua simplicidade conceitual, sofisticação e eficiência computacional tornaram-na referência metodológica para a análise de dados binários no final do século 20.

Para medidas binárias, é natural assumir a distribuição Bernoulli como referência probabilística na construção de modelos. Entretanto, a relação limitada entre a esperança e a variância dessa distribuição e a ausência de um parâmetro para quantificar a variabilidade extra-Bernoulli, torna necessária a utilização de modelos específicos para acomodar o efeito de superdispersão. Um desses modelos acomoda a correlação entre medidas binárias de um mesmo elemento (HINDE; DEMÉTRIO, 1998). Se as diversas medidas feitas sobre um elemento forem longitudinais, a classe de modelos lineares generalizados mistos (MLGM) (BRESLOW; CLAYTON, 1993) pode ser utilizada, na qual a estrutura de dependência temporal é acomodada adicionando-se efeitos aleatórios ao preditor linear. Outras formas de modelar a estrutura de variação longitudinal podem ser aplicadas, dependendo do interesse na inferência (MOLENBERGHS; VERBEKE, 2005).

Entretanto, a superdispersão e a estrutura longitudinal são, usualmente, modeladas separadamente. Situações clínicas e agronômicas para as quais diversos elementos sob agrupamentos não-especificados são medidos longitudinalmente, não são contemplados pelos modelos separados.

Objetivou-se com esse trabalho propor dois modelos combinados, que estendem os MLGM para acomodar essas duas fontes de variação, através de dois efeitos aleatórios. Na seção 2.4.2.1

será apresentado um estudo clínico humano, em que medidas longitudinais binárias são registradas para cada paciente. Na Seção 2.4.2.2 serão revisados os MLG, MLGM e modelos para superdispersão e na Seção 2.4.2.2.4 serão apresentadas as propostas dos modelos combinados. Os métodos de estimação dos parâmetros para os modelos estudados são apresentados na Seção 2.4.2.3 e a aplicação aos dados do estudo clínico é apresentada na seção 2.4.2.4, seguida da discussão dos resultados.

2.4.2 Desenvolvimento

2.4.2.1 Um estudo clínico em dermatologia

Um estudo aleatorizado multi-centro, duplo cego e com grupos paralelos foi realizado com o objetivo de comparar dois medicamentos via oral (denominados aqui *A* e *B*) para tratamento da *toenail dermatophite onychomycosis* (TDO). Essa dermatite é uma infecção de unha bastante comum e de difícil tratamento. Os compostos antifúngicos tradicionais para tratamento de TDO necessitam que o paciente administre a medicação até que a(s) unha(s) se desenvolva(m) saudavelmente, por completo. Os novos medicamentos têm reduzido esse período de administração do medicamento para três meses. Neste estudo, foram comparadas duas drogas quanto à sua eficácia e segurança durante o tratamento, por 12 semanas de terapia contínua. Este estudo foi descrito em De Backer et al. (1996) e seu conjunto de dados explorado em Molenberghs e Verbeke (2005).

Foram estudados 378 pacientes associados a 36 centros de atendimento de saúde. Os pacientes foram acompanhados por 12 semanas durante o tratamento e, posteriormente, acompanhados até um total de 48 semanas (12 meses). As medidas foram tomadas no início do tratamento, a cada mês durante o tratamento e a cada 3 meses após o tratamento, o que resultou de um máximo de 7 medidas para cada paciente. Antes de iniciar o tratamento, o médico responsável selecionou uma unha para ser acompanhada durante o período. O estudo quantitativo se limitou aos pacientes cujas unhas selecionadas para acompanhamento correspondem a uma dos primeiros pododáctilos (“dedão do pé”). Dessa forma, o número de pacientes para os tratamentos A e B reduziu para 294. Os tratamentos foram alocados aos pacientes por processo aleatório. Foram medidos 7 pontos no tempo para 76% dos pacientes. Uma das medidas de interesse foi o grau de severidade codificados como 0 (não-severo) e 1 (severo). As questões de interesse são (i) como o percentual de severidade decresce com o tempo e (ii) se os dois tratamentos diferem ou não. A Figura 8 apresenta a evolução do percentual de severidade para os dois tratamentos, ao longo do tempo. Nota-se a clara influência

do tempo no percentual de severidade para ambos os tratamentos mas não é claro se os tratamentos diferem significativamente.

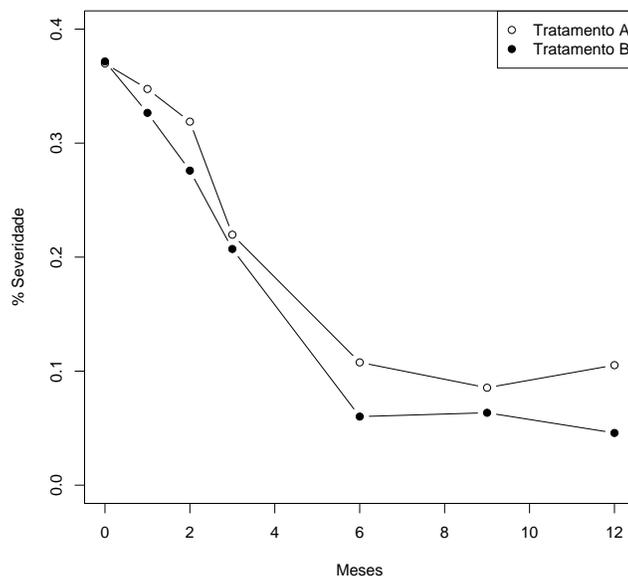


Figura 8 - Proporção de severidade de TDO para os tratamentos estudados

2.4.2.2 Modelos lineares generalizados e extensões

Nesta seção, serão revisados alguns dos ingredientes importantes para a construção do modelo geral que acomoda tanto a variação longitudinal quanto a superdispersão.

2.4.2.2.1 Modelos lineares generalizados

Os MLG (NELDER; WEDDERBURN, 1972) são uma importante extensão de vários modelos e métodos conhecidos como os modelos lineares, modelos log-lineares, alguns modelos paramétricos de sobrevivência, regressão Poisson e regressão logística, dentre outros. Considerando uma amostra com n variáveis independente e identicamente distribuídas, \mathbf{X} uma matriz com $p + 1$ colunas que representam p variáveis explicativas e y um vetor de observações amostrado de Y , são definidos os 3 componentes de um MLG:

1. Y tem distribuição probabilística como membro da **Família Exponencial** de distribuições, com uma função de probabilidade ou função densidade de probabilidade (para variáveis ale-

atórias discretas e contínuas, respectivamente)

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{1}{a(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\},$$

com média

$$E(Y_i) = \mu_i = b'(\theta_i)$$

e

$$\text{Var}(Y_i) = a_i(\phi) b''(\theta_i) = a_i(\phi) V(\mu_i) = a_i(\phi) V_i,$$

sendo θ_i o parâmetro canônico, $a_i(\phi) = \phi/w_i$, ϕ o parâmetro de dispersão, w_i um peso “a priori” e V_i a *função de variância* dada por $V_i = d\mu_i/d\theta_i$.

Algumas distribuições-membros da Família Exponencial são a Normal, Poisson, Binomial, Gama, Gaussiana Inversa, Binomial Negativa (com o parâmetro k fixado), algumas distribuições truncadas, etc.

2. As covariáveis e os fatores em \mathbf{X} estão relacionados de forma aditiva no **preditor linear**

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta};$$

Todas as configurações admitidas para os bem conhecidos Modelos Lineares podem ser aqui utilizados.

3. Uma **Função de Ligação** monotônica e diferenciável $g(\cdot)$, que liga o preditor linear η à média de Y , e portanto,

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \Rightarrow \mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}). \quad (49)$$

Um caso particular dos MLG é quando se assume que $Y \sim \text{Bernoulli}(\pi)$ e função de ligação $g(\mu) = \log[\pi/(1 - \pi)]$, conhecida como função logit. Na forma da família exponencial tem-se que $\theta = \log[\pi/(1 - \pi)]$, $\phi = 1$ e $V(\mu) = \pi(1 - \pi)$. Definindo-se um preditor linear η em particular, tem-se, então, o conhecido modelo de regressão logística (COLLETT, 1991). Outra alternativa é considerar a função de ligação probit, definida por $g(\mu) = \Phi^{-1}(\mu)$ sendo Φ a função de distribuição acumulada da distribuição normal padronizada.

2.4.2.2.2 Modelos para superdispersão

Quando um MLG é ajustado a dados na forma de proporções (supondo distribuição binomial para Y) e assumido que o preditor linear resultante está adequado, é esperado que a *deviance*² média seja próxima de 1, uma vez que, sob determinadas condições de regularidade, essa tem aproximadamente distribuição $\chi^2_{(n-p)}$, sendo n o número de observações e p o número de parâmetros estimados. Entretanto, algumas situações práticas em que esses modelos são utilizados, obtém-se uma *deviance* média muito maior do que 1, caracterizando assim a variação *extra-binomial* ou *superdispersão*. Observando-se a função de variância para a distribuição binomial com denominador m , dada por $V(\mu) = m\pi(1 - \pi)$, nota-se, que, uma vez estabelecida a esperança $m\pi$, tem-se automaticamente definida a variância esperada para essa distribuição. Em muitas situações práticas a variância observada pode ser bem maior.

São diversas as possíveis causas da superdispersão: má-especificação do preditor linear, presença de *outliers*, função de ligação inadequada, proporções com denominador reduzido, dependência longitudinal não ajustada e ocorrência de agrupamentos (*clusters*) que tornam a probabilidade de sucesso π não constante para todas as observações (COLLETT, 1991). Uma fonte particular de superdispersão é quando proporções ou contagens são inflacionadas de zeros (VIEIRA; HINDE; DEMÉTRIO, 2000), demandando modelos e testes específicos para essas situações (HALL; BERENHAUT, 2002).

Foram propostos alguns métodos para modelar a superdispersão (HINDE; DEMÉTRIO, 1998a) e um desses métodos consiste em considerar uma abordagem em dois estágios. Seja $Y_i|P_i \sim \text{Bin}(m_i, P_i)$ uma variável aleatória com distribuição binomial, com parâmetro m_i e probabilidade de sucesso P_i , sendo que $P_i \sim \text{Beta}(\alpha_i, \beta_i)$, $i = 1, \dots, N$. Portanto, a função de probabilidade condicional de $Y_i|P_i$ é

$$f_{Y_i|P_i}(y_i|p_i) = \binom{m_i}{y_i} p_i^{y_i} (1 - p_i)^{m_i - y_i}, \quad y_i = 0, 1, \dots, m_i,$$

sendo a função densidade para P_i dada por

$$f_{P_i}(p_i) = \frac{p_i^{\alpha_i - 1} (1 - p_i)^{\beta_i - 1}}{B(\alpha_i, \beta_i)}, \quad 0 \leq p_i \leq 1$$

²Ver a definição na seção 2.1.2.2

com $B(\alpha_i, \beta_i) = \int_0^1 x^{\alpha_i-1}(1-x)^{\beta_i-1}dx$ a função Beta. A função de probabilidade incondicional para Y_i , é dada por

$$f_{Y_i}(y_i) = \binom{m_i}{y_i} \frac{B(\alpha_i + y_i, m_i + \beta_i - y_i)}{B(\alpha_i, \beta_i)}, \quad (50)$$

conhecida como distribuição *Beta-Binomial* e para qual obtém-se $E(Y_i) = m_i\alpha/(\alpha + \beta) = m_i\pi_i$ e $\text{Var}(Y_i) = m_i\pi_i(1 - \pi_i)[1 + (m_i - 1)(\alpha_i + \beta_i + 1)^{-1}]$. Pode ser mostrado que a correlação entre duas medidas Y_{ij} e Y_{ik} , $j \neq k$, do mesmo elemento i é igual a $\rho = \text{Corr}(Y_{ij}, Y_{ik}) = (\alpha + \beta + 1)^{-1}$ (HINDE; DEMÉTRIO, 1998; MOLENBERGHS; VERBEKE, 2005). Nota-se a semelhança dessa função de variância com aquela da distribuição binomial, sendo que o termo $1 + (m_i - 1)(\alpha_i + \beta_i + 1)^{-1}$ é o termo que inflaciona a variância, dependente dos parâmetros α_i e β_i . Essa distribuição pode ser utilizada para modelar dados na forma de proporções com superdispersão. Porém, para dados binários, em que $m_i = 1$, a variância de (50) reduz-se à variância da distribuição Bernoulli e portanto, não é possível utilizar a função de probabilidade (50) para modelar medidas binárias com superdispersão. Dentre os métodos mais conhecidos para tratar a superdispersão, o único que pode ser utilizado em dados binários com repetições é baseado em modelos contendo efeitos fixos e aleatórios (COLLETT, 1991), descrito a seguir.

2.4.2.2.3 Modelos com efeitos fixos e aleatórios

Quando o pressuposto de independência das observações não pode ser garantido devido às características intrínsecas do problema tratado (como em estudos multivariados, dados agrupados, experimentos *cross-over*, dados longitudinais e espaço-temporais), modelos apropriados devem ser aplicados, de forma a acomodar a correlação existente entre observações. Dentre as diversas formas de analisar dados longitudinais, a utilização de modelos contendo efeitos fixos e aleatórios mostra-se uma forma eficiente para ajustar as diversas fontes de variação.

Ao contrário dos modelos lineares mistos, que encontra na distribuição normal multivariada uma forma flexível e abrangente para tratar os efeitos aleatórios associados a medidas contínuas Gaussianas (VERBEKE; MOLENBERGHS, 2000), as abordagens utilizadas para dados longitudinais discretos dependem do objetivo da inferência e, conseqüentemente, das famílias de modelos associadas³.

³Essas famílias estão descritas na seção 2.1.4.2.

Uma classe bastante flexível da família de modelos elemento-específico são os modelos lineares generalizados mistos (MLGM), proposto por Breslow e Clayton (1993). Trata-se de uma extensão direta dos MLG, que, dentre outras aplicações, permite acomodar a informação serial contida nas medidas longitudinais.

Seja Y_{ij} a j -ésima medida longitudinal do i -ésimo elemento (*subject*), $j = 1, \dots, n_i$ e $i = 1, \dots, N$; logo, \mathbf{Y}_i é um vetor n_i -dimensional com as medidas longitudinais do i -ésimo elemento. Seja \mathbf{b}_i o vetor q -dimensional de efeitos aleatórios, independentes e com distribuição $N(\mathbf{0}, \mathbf{D})$, em que $\mathbf{0}$ é um vetor de zeros e \mathbf{D} uma matriz de variância-covariância, geralmente estruturada e pré-especificada; \mathbf{b}_i está associado a $q - 1$ fatores e/ou covariáveis que compõe o vetor \mathbf{z}_{ij} , com dimensão $(q \times 1)$; $\boldsymbol{\beta}$ é o vetor p -dimensional de efeitos fixos, associados a $p-1$ fatores e/ou covariáveis que compõem o vetor \mathbf{x}_{ij} com dimensão $(p \times 1)$. É assumido que as medidas Y_{ij} são condicionalmente independentes, dado \mathbf{b}_i , com função (densidade) de probabilidade pertencente à família exponencial de distribuições, ou seja,

$$f_i(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) = \exp \left\{ \phi^{-1} [y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi) \right\}, \quad (51)$$

sendo $\eta(\mu_{ij}) = \eta[\mathbf{E}(Y_{ij}|\mathbf{b}_i)] = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i$, $\eta(\cdot)$ é uma função de ligação pré-especificada e ϕ é o parâmetro de dispersão. Se $\theta_{ij} = \eta(\mu_{ij})$, então η é uma função de ligação canônica. Denota-se também que $f(\mathbf{b}_i|\mathbf{D})$ é a função densidade dos efeitos aleatórios \mathbf{b}_i .

Assumindo um modelo probabilístico Bernoulli para as medidas longitudinais e utilizando a função de ligação logit, descrita na Seção 2.4.2.2, tem-se um modelo linear generalizado misto logit-normal-Bernoulli, definido por

$$\begin{aligned} Y_{ij}|\mathbf{b}_i &\sim \text{Bernoulli}(\pi_{ij}) \\ \pi_{ij} &= \frac{\exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)}{1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)} \\ \mathbf{b}_i &\sim N_q(\mathbf{0}, \mathbf{D}). \end{aligned} \quad (52)$$

Quando é adequado assumir a função de ligação probit, tem-se então um MLGM probit-

normal-Bernoulli, dado por

$$\begin{aligned}
 Y_{ij} | \mathbf{b}_i &\sim \text{Bernoulli}(\pi_{ij}) \\
 \pi_{ij} &= \Phi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i) \\
 \mathbf{b}_i &\sim N_q(\mathbf{0}, \mathbf{D}),
 \end{aligned} \tag{53}$$

sendo $\Phi(x) = \int_{-\infty}^x (2\pi)^{-1/2} \exp[-(1/2)u^2] du$.

2.4.2.2.4 Modelos combinando superdispersão e estrutura longitudinal

Quando dados binários estão sujeitos à superdispersão e, ao mesmo tempo, são medidos longitudinalmente, uma extensão particular dos MLGM se faz necessária. Aqui serão combinadas as idéias da Seção 2.4.2.2 para produzir um modelo que permita acomodar essas duas fontes de variação simultaneamente, de forma similar àquela descrita em Molenberghs, Verbeke e Demétrio (2007).

Seja y_{ij} uma medida binária (codificado como 0 igual à não-ocorrência do evento e 1 igual à ocorrência do evento) do i -ésimo elemento (parcelas, pessoas, animais, equipamento, etc) no j -ésimo tempo; θ_{ij} é o parâmetro de superdispersão para a (ij) -ésima observação. \mathbf{b}_i é o vetor q -dimensional de efeitos aleatórios, associados ao vetor \mathbf{z}_{ij} com q fatores e/ou covariáveis; o vetor de efeitos fixos $\boldsymbol{\beta}$ com dimensão p está associado ao vetor \mathbf{x}_{ij} contendo p linhas referentes a fatores e/ou covariáveis. A probabilidade de ocorrer sucesso para a (ij) -ésima medida é $P(Y_{ij} = 1) = \pi_{ij}$.

Modelo logit-normal-Bernoulli-beta

Assumindo o modelo probabilístico Bernoulli para as medidas binárias, a distribuição beta para o efeito aleatório que acomodará a superdispersão e a distribuição normal para o efeito aleatório que modelará a correlação entre as medidas longitudinais e considerando ainda a função de ligação

⁴Chama-se a atenção para o símbolo π , que nessa integral significa o número $\pi = 3,1415\dots$; a probabilidade de sucesso Bernoulli π_{ij} estará sempre acompanhada de um ou dois sub-índices.

logit, propõe-se o modelo logit-normal-Bernoulli-beta que se segue:

$$\begin{aligned}
 Y_{ij} | \mathbf{b}_i &\sim \text{Bernoulli}(\pi_{ij}) \\
 \pi_{ij} &= \theta_{ij} \frac{\exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)}{1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)} \\
 \theta_{ij} &\sim \text{Beta}(\alpha, \beta) \\
 \mathbf{b}_i &\sim N_q(\mathbf{0}, \mathbf{D}),
 \end{aligned} \tag{54}$$

sendo que α e β são os parâmetros da distribuição beta e \mathbf{D} é a matriz de variância-covariância referente a \mathbf{b}_i .

Modelo probit-normal-Bernoulli-beta

O modelo probit-normal-Bernoulli-beta é, essencialmente, igual ao modelo (54), apenas diferindo pela função de ligação adotada. É descrito como

$$\begin{aligned}
 Y_{ij} | \mathbf{b}_i &\sim \text{Bernoulli}(\pi_{ij}) \\
 \pi_{ij} &= \theta_{ij} \Phi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i) \\
 \theta_{ij} &\sim \text{Beta}(\alpha, \beta) \\
 \mathbf{b}_i &\sim N_q(\mathbf{0}, \mathbf{D}).
 \end{aligned} \tag{55}$$

2.4.2.3 Estimação

2.4.2.3.1 Modelos lineares generalizados

O método de estimação para o vetor parâmetros $\boldsymbol{\beta}$, proposto por Nelder e Wedderburn (apud DEMÉTRIO, 2002), é o método da máxima verossimilhança. Utilizando o logaritmo da função de verossimilhança da família exponencial,

$$l(\boldsymbol{\theta} | \mathbf{y}) = \log \left[\prod_{i=1}^n f(y_i; \theta_i, \phi) \right] = \sum_{i=1}^n \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\},$$

o j -ésimo componente do vetor escore pode ser descrito como

$$U_j = \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^n \frac{1}{a_i(\phi)} (y_i - \mu_i) \frac{1}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} \quad j = 1, \dots, p. \quad (56)$$

Nota-se que somente μ_i and $V(\mu_i)$ estão sendo utilizadas no processo de estimação, não sendo necessária a forma funcional completa da família exponencial. Para obter estimativas de máxima verossimilhança para $\boldsymbol{\beta}$, o algoritmo de mínimos quadrados iterativamente reponderados (MQIR)⁵, pode ser utilizado para maximizar a função de verossimilhança.

Como medida de discrepância entre o valor observado na amostra e aquele predito pelo modelo estimado, foi proposta a seguinte medida de desvio:

$$\frac{D}{\phi} = -2 \sum_{i=1}^n \{l(\hat{\mu}_i, y_i) - l(y_i, y_i)\},$$

sendo $l(\hat{\mu}_i, y_i)$ o logaritmo da função de verossimilhança calculado utilizando y_i e uma estimativa de μ_i e $l(y_i, y_i)$ o logaritmo da função de verossimilhança calculada utilizando y_i também como uma estimativa de μ_i . Essa medida é denominada *scaled deviance* e $D = \sum_{i=1}^n d_i^2$ é a *deviance total*, sendo d_i^2 o *componente de deviance*, dado por $d_i^2 = -2 \int_y^{\mu_i} \frac{y-u_i}{V(u_i)} du$. Essa medida pode ser utilizada na seleção de modelos mas não como medida de qualidade de ajuste, no caso de dados binários. Rotinas e funções para ajustar os MLG estão implementados e diversos softwares para análise estatística, como o R, SAS e GENSTAT.

2.4.2.3.2 Modelos lineares generalizados mistos

A estimação de $\boldsymbol{\beta}$ e dos parâmetros em \mathbf{D} é feita pela maximização da função de verossimilhança marginal, obtida pela integração da função de verossimilhança sobre os efeitos aleatórios \mathbf{b}_i . A contribuição do i -ésimo elemento para a função de verossimilhança é

$$f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{D}, \phi) = \int \cdots \int \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i | \mathbf{D}) d\mathbf{b}_i, \quad (57)$$

⁵descrito na p. 21.

e a função de verossimilhança marginal para a amostra contendo os N elementos é dada pela expressão

$$\begin{aligned} L(\boldsymbol{\beta}, \mathbf{D}, \phi) &= \prod_{i=1}^N f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{D}, \phi) \\ &= \prod_{i=1}^N \int \cdots \int \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i | \mathbf{D}) d\mathbf{b}_i, \end{aligned} \quad (58)$$

sendo que o número de integrais na expressão corresponde ao número de termos em \mathbf{b}_i .

No caso particular em que $Y_{ij} \sim N(\mathbf{X}_i \boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$, as integrais em (58) podem ser tratadas analiticamente. Em geral, porém, essas integrais não possuem solução analítica e métodos numéricos e estocásticos são necessários para solucioná-las. Uma vez calculado o valor aproximado dessas integrais, a função (58) deve ser otimizada. Novamente, para o caso particular Gaussiano, expressões analíticas são obtidas pelo processo de otimização. Fora esse caso particular, um sistema de equações não-lineares é obtido e métodos numéricos para otimização são utilizados. O cálculo numérico das integrais dentro de cada ciclo iterativo na otimização de (58) torna a obtenção de estimativas um processo computacionalmente intensivo, tomando bastante tempo quando se trata de grandes amostras e/ou quando a estrutura do modelo é complexa, em especial quando a dimensão do espaço paramétrico é grande.

2.4.2.3.3 Modelos combinados

Modelo logit-normal-Bernoulli-beta

Uma forma de acomodar a superdispersão e a correlação presente nos dados longitudinais foi modificar o modelo (52), acrescentando um efeito aleatório multiplicativo θ_{ij} na expressão de π_{ij} . O método de estimação adotado foi o da máxima verossimilhança. A função de probabilidade condicional para $Y_{ij} | \theta_{ij}, \mathbf{b}_i$ é dada por

$$f(y_{ij} | \theta_{ij}, \mathbf{b}_i) = \left[\theta_{ij} \frac{\exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)}{1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)} \right]^{y_{ij}} \left[1 - \theta_{ij} \frac{\exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)}{1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)} \right]^{1-y_{ij}}, \quad y_{ij} \in \mathbb{R} \quad (59)$$

sendo o parâmetro β fixo e desconhecido. A função densidade de probabilidade para o efeito aleatório \mathbf{b}_i é descrita como

$$f(\mathbf{b}_i) = \frac{1}{\sqrt{(2\pi)^{n_i}}} \frac{1}{\sqrt{|\mathbf{D}|}} \exp\left(-\frac{1}{2}\mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i\right), \quad \mathbf{b}_i \in \mathbb{R}^q \quad (60)$$

sendo \mathbf{D} uma matriz contendo parâmetros fixos e desconhecidos. A função densidade de probabilidade para o efeito aleatório θ_{ij} é

$$f(\theta_{ij}) = \frac{\theta_{ij}^{\alpha-1} (1 - \theta_{ij})^{\beta-1}}{B(\alpha, \beta)}, \quad \theta_{ij} \in [0, 1]. \quad (61)$$

sendo α e β também parâmetros fixos e desconhecidos. A contribuição da (ij) -ésima observação para a função de verossimilhança, assumindo a independência entre θ_{ij} e \mathbf{b}_i , é dada por

$$f(y_{ij}, \theta_{ij}, \mathbf{b}_i) = f(y_{ij} | \theta_{ij}, \mathbf{b}_i) f(\mathbf{b}_i) f(\theta_{ij}) \quad (62)$$

e a contribuição do i -ésimo elemento para a função de verossimilhança, assumindo independência condicional⁶, baseado em (62), pode ser descrita por meio da seguinte função de densidade conjunta:

$$f(\mathbf{y}_i, \boldsymbol{\theta}_i, \mathbf{b}_i) = \prod_{j=1}^{n_i} f(y_{ij} | \theta_{ij}, \mathbf{b}_i) f(\mathbf{b}_i) f(\theta_{ij}). \quad (63)$$

Essa função densidade está condicionada aos dois efeitos aleatórios. Para a estimação por máxima verossimilhança de β e \mathbf{D} , a função (63) deverá ser integrada considerando os domínios de $\boldsymbol{\theta}_i$ e \mathbf{b}_i .

Integrando essa função em \mathbf{b}_i , obtém-se

$$f(\mathbf{y}_i, \boldsymbol{\theta}_i) = \int \prod_{j=1}^{n_i} f(y_{ij} | \theta_{ij}, \mathbf{b}_i) f(\mathbf{b}_i) f(\theta_{ij}) d\mathbf{b}_i, \quad (64)$$

⁶Uma vez que a dependência longitudinal e a superdispersão foram levadas em conta na função de verossimilhança, as medidas são consideradas independentes.

e, portanto, utilizando as funções (59), (60) e (61), a função de verossimilhança para β e \mathbf{D} condicional ao valor de θ , para N elementos é descrita como

$$\begin{aligned}
L(\beta, \mathbf{D}, \alpha, \beta | \theta) &= \prod_{i=1}^N \int \prod_{j=1}^{n_i} \left[\theta_{ij} \frac{\exp(\mathbf{x}_{ij}^T \beta + \mathbf{z}_{ij}^T \mathbf{b}_i)}{1 + \exp(\mathbf{x}_{ij}^T \beta + \mathbf{z}_{ij}^T \mathbf{b}_i)} \right]^{y_{ij}} \\
&\quad \times \left[1 - \theta_{ij} \frac{\exp(\mathbf{x}_{ij}^T \beta + \mathbf{z}_{ij}^T \mathbf{b}_i)}{1 + \exp(\mathbf{x}_{ij}^T \beta + \mathbf{z}_{ij}^T \mathbf{b}_i)} \right]^{1-y_{ij}} \\
&\quad \times \frac{1}{\sqrt{(2\pi)^{n_i}}} \frac{1}{\sqrt{|\mathbf{D}|}} \exp\left(-\frac{1}{2} \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i\right) \\
&\quad \times \frac{\theta_{ij}^{\alpha-1} (1 - \theta_{ij})^{\beta-1}}{B(\alpha, \beta)} d\mathbf{b}_i. \tag{65}
\end{aligned}$$

Essa função de verossimilhança continua condicionada ao efeito aleatório θ_{ij} . Uma alternativa é buscar a solução analítica da integral de $f(y_{ij} | \theta_{ij}, \mathbf{b}_i)$ sobre o domínio de θ_{ij} . Para o caso Bernoulli, pode-se escrever que $f(y_{ij} | \theta_{ij}, \mathbf{b}_i) = f(y_{ij} = 0 | \theta_{ij}, \mathbf{b}_i)^{y_{ij}} f(y_{ij} = 1 | \theta_{ij}, \mathbf{b}_i)^{1-y_{ij}}$. Denotando $k_{ij} = [\exp(\mathbf{x}_{ij}^T \beta + \mathbf{z}_{ij}^T \mathbf{b}_i)] [1 + \exp(\mathbf{x}_{ij}^T \beta + \mathbf{z}_{ij}^T \mathbf{b}_i)]^{-1}$, tem-se:

$$\begin{aligned}
f(y_{ij} = 1 | \mathbf{b}_i) &= \int_0^1 f(y_{ij} = 1 | \theta_{ij}, \mathbf{b}_i) f(\theta_{ij}) d\theta_{ij} \\
&= \int_0^1 \theta_{ij} k_{ij} \theta_{ij}^{\alpha-1} (1 - \theta_{ij})^{\beta-1} [B(\alpha, \beta)]^{-1} d\theta_{ij} \\
&= \frac{k_{ij}}{B(\alpha, \beta)} \int_0^1 \theta_{ij}^{(\alpha-1)+1} (1 - \theta_{ij})^{\beta-1} d\theta_{ij} \\
&= k_{ij} \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} \\
&= k_{ij} \frac{\Gamma(\alpha + 1) \Gamma(\beta)}{\Gamma(\alpha + \beta + 1)} \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \\
&= k_{ij} \alpha \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma[(\alpha + \beta) + 1]} \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \\
&= k_{ij} \alpha \frac{\Gamma(\alpha + \beta)}{(\alpha + \beta) \Gamma(\alpha + \beta)} \\
&= \alpha k_{ij} (\alpha + \beta)^{-1}; \tag{66}
\end{aligned}$$

Da mesma forma, calcula-se

$$\begin{aligned}
f(y_{ij} = 0 | \mathbf{b}_i) &= \int_0^1 f(y_{ij} = 0 | \theta_{ij}, \mathbf{b}_i) f(\theta_{ij}) d\theta_{ij} \\
&= \int_0^1 (1 - \theta_{ij} k_{ij}) \theta_{ij}^{\alpha-1} (1 - \theta_{ij})^{\beta-1} [B(\alpha, \beta)]^{-1} d\theta_{ij} \\
&= \int_0^1 \theta_{ij}^{\alpha-1} \frac{(1 - \theta_{ij})^{\beta-1}}{B(\alpha, \beta)} d\theta_{ij} - \int_0^1 k_{ij} \theta_{ij}^{(\alpha-1)+1} \frac{(1 - \theta_{ij})^{\beta-1}}{B(\alpha, \beta)} d\theta_{ij} \\
&= 1 - k_{ij} \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} = 1 - k_{ij} \frac{\alpha}{\alpha + \beta} \\
&= \frac{(1 - k_{ij})\alpha + \beta}{\alpha + \beta}. \tag{67}
\end{aligned}$$

Logo, a função de verossimilhança marginal, apenas condicionada a \mathbf{b}_i , é

$$L(\boldsymbol{\beta}, \mathbf{D}, \alpha, \beta) = \prod_{i=1}^N \int \prod_{j=1}^{n_i} \left(\frac{\alpha k_{ij}}{\alpha + \beta} \right)^{y_{ij}} \cdot \left[\frac{(1 - k_{ij})\alpha + \beta}{\alpha + \beta} \right]^{1-y_{ij}} f(\mathbf{b}_i | \mathbf{D}) d\mathbf{b}_i \tag{68}$$

$$= \prod_{i=1}^N \int \prod_{j=1}^{n_i} \frac{1}{\alpha + \beta} (\alpha k_{ij})^{y_{ij}} [(1 - k_{ij})\alpha + \beta]^{1-y_{ij}} f(\mathbf{b}_i | \mathbf{D}) d\mathbf{b}_i; \tag{69}$$

Essa função claramente não permite a obtenção de estimadores analíticos, o que demanda algoritmos iterativos como Newton-Raphson, Gauss-Newton, dentre outros⁷. Dentro de cada ciclo iterativo é necessário que a integral sobre \mathbf{b}_i seja solucionada. Como a integral em (69) é da forma $\int f(z)\phi(z)dz$ o algoritmo de quadratura de Gauss-Hermite adaptativa⁸ pode ser aplicado para obter uma solução numérica da integral. Esse algoritmo, junto com alguns algoritmos de otimização já estão implementados no flexível e conveniente procedimento NLMIXED do sistema SAS, utilizado neste trabalho. Após os critérios de convergência terem sido atingidos, estimativas de máxima verossimilhança podem ser utilizadas para inferência estatística (PAWITAN, 2001). Um comentário importante é sobre a identificabilidade dos parâmetros. Os parâmetros α e β não são identificáveis simultaneamente, quando o preditor linear contém um intercepto. Uma solução é estabelecer uma restrição do tipo $\beta/\alpha = c$, que soluciona o problema de identificabilidade.

⁷Alguns algoritmos iterativos para otimização de funções estão descritos na seção 2.1.4.4.

⁸Esse algoritmo está revisado na seção 2.1.4.5.

Modelo probit-normal-Bernoulli-beta

O processo de estimação para o modelo (55) modifica a função de probabilidade condicional para $Y_{ij}|\theta_{ij}, \mathbf{b}_i$, sendo descrita como

$$f(y_{ij}|\theta_{ij}, \mathbf{b}_i) = (\theta_{ij}k_{ij})^{y_{ij}} (1 - k_{ij})^{1-y_{ij}}, \quad (70)$$

sendo $k_{ij} = \Phi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)$ e $\Phi(x) = P(X \leq x) = \int_{-\infty}^x (2\pi)^{-1/2} \exp[-t^2/2] dt$. Utilizando os resultados (66) e (67), tem-se a função densidade de probabilidade de Y_{ij} condicional a \mathbf{b}_i , dada por

$$f(y_{ij}|\mathbf{b}_i) = \frac{1}{\alpha + \beta} (\alpha k_{ij})^{y_{ij}} [(1 - k_{ij})\alpha + \beta]^{1-y_{ij}}$$

e dessa forma, a função de verossimilhança marginal, condicional a \mathbf{b}_i , para N elementos é dada por

$$\begin{aligned} L(\boldsymbol{\beta}, \mathbf{D}, \alpha, \beta) &= \prod_{i=1}^N \int \prod_{j=1}^{n_i} \frac{1}{\alpha + \beta} (\alpha k_{ij})^{y_{ij}} [(1 - k_{ij})\alpha + \beta]^{1-y_{ij}} f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i \\ &= \prod_{i=1}^N \int \prod_{j=1}^{n_i} \frac{[\Phi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)\alpha]^{y_{ij}}}{\alpha + \beta} \{ [1 - \Phi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)]\alpha + \beta \}^{1-y_{ij}} \\ &\quad \times (2\pi)^{-q/2} |\mathbf{D}|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{b}_i^T \mathbf{D} \mathbf{b}_i \right\} d\mathbf{b}_i. \end{aligned} \quad (71)$$

Para essa função de verossimilhança, pode-se adotar o mesmo método utilizado para a otimização de (69), utilizando um algoritmo para integração numérica em conjunto com um algoritmo de otimização iterativo, adotando alguma restrição sobre os parâmetros α e β .

Entretanto, para esse modelo, também é possível obter uma solução analítica para a integral em (71), o que permite obter uma expressão para função densidade de probabilidade marginal de \mathbf{Y}_i .

Reescrevendo a função densidade de probabilidade conjunta de \mathbf{Y}_i , \mathbf{b}_i , e $\boldsymbol{\theta}_i$,

$$f(\mathbf{y}_i, \mathbf{b}_i, \boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \theta_{ij}) f(\mathbf{b}_i) \prod_{j=1}^{n_i} f(\theta_{ij}),$$

a função densidade marginal para $\mathbf{Y}_i = \mathbf{1}$ pode ser escrita como

$$\begin{aligned}
f(\mathbf{y}_i = \mathbf{1}) &= \int \cdots \int \prod_{j=1}^{n_i} f(y_{ij} = 1 | \mathbf{b}_i, \theta_{ij}) f(\mathbf{b}_i) \prod_{j=1}^{n_i} f(\theta_{ij}) d\mathbf{b}_i d\theta_{ij} \\
&= \int \cdots \int \prod_{j=1}^{n_i} \theta_{ij} \Phi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i) f(\mathbf{b}_i) \prod_{j=1}^{n_i} f(\theta_{ij}) d\mathbf{b}_i d\theta_{ij} \\
&= \int \prod_{j=1}^{n_i} \Phi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \int \prod_{j=1}^{n_i} \theta_{ij} f(\theta_{ij}) d\theta_{ij} \\
&= \prod_{j=1}^{n_i} \int \Phi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \prod_{j=1}^{n_i} \int \theta_{ij} f(\theta_{ij}) d\theta_{ij}. \tag{72}
\end{aligned}$$

A segunda integral em (72) é, na realidade, um produtório de esperanças da distribuição beta. Logo,

$$f(\mathbf{y}_i = \mathbf{1}) = \left(\frac{\alpha}{\alpha + \beta} \right)^{n_i} \prod_{j=1}^{n_i} \int \Phi(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i. \tag{73}$$

O fator-produto em (73) é a função de verossimilhança de um MLGM probit-normal-Bernoulli, descrito em (53). A função densidade de probabilidade marginal de $\mathbf{Y}_i = \mathbf{1}$ para esse modelo pode ser descrita como

$$\begin{aligned}
f(\mathbf{y}_i = \mathbf{1}) &= f(\mathbf{y}_i = \mathbf{1} | \mathbf{b}_i) f(\mathbf{b}_i) = \prod_{j=1}^{n_i} f(y_{ij} = 1 | \mathbf{b}_i) f(\mathbf{b}_i) \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}_i} \frac{1}{(2\pi)^{q/2} |\mathbf{D}|^{1/2}} \frac{1}{\exp \left\{ -\frac{1}{2} \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right\}} \\
&\quad \times \frac{1}{(2\pi)^{n_i/2}} \exp \left[-\frac{1}{2} \sum_{j=1}^{n_i} t_j^2 \right] dt d\mathbf{b}_i \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}_i} \frac{(2\pi)^{-\frac{n_i+q}{2}}}{|\mathbf{D}|^{1/2}} \exp \left\{ -\frac{1}{2} [\mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i + \mathbf{t}^T \mathbf{t}] \right\} dt d\mathbf{b}_i. \tag{74}
\end{aligned}$$

Fazendo uma mudança para o vetor $\mathbf{s} = \mathbf{t} - \mathbf{Z}_i \mathbf{b}_i$, tem-se que $\mathbf{t} = \mathbf{s} + \mathbf{Z}_i \mathbf{b}_i$ e o diferencial $d\mathbf{t} = d\mathbf{s}$.

Assim, a densidade (74) passa a ser descrita como

$$\begin{aligned}
f(\mathbf{y}_i = \mathbf{1}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{\mathbf{X}\boldsymbol{\beta}} \frac{(2\pi)^{-\frac{n_i+q}{2}}}{|\mathbf{D}|^{1/2}} \\
&\quad \times \exp \left\{ -\frac{1}{2} [\mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i + (\mathbf{s} + \mathbf{Z}_i \mathbf{b}_i)^T (\mathbf{s} + \mathbf{Z}_i \mathbf{b}_i)] \right\} ds d\mathbf{b}_i.
\end{aligned}$$

Dessa expressão tem-se que:

$$\begin{aligned}
\mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i + (\mathbf{s} + \mathbf{Z}_i \mathbf{b}_i)^T (\mathbf{s} + \mathbf{Z}_i \mathbf{b}_i) &= \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i + \mathbf{s}^T \mathbf{s} + \mathbf{s}^T \mathbf{Z}_i \mathbf{b}_i + \mathbf{b}_i^T \mathbf{Z}_i^T \mathbf{s} + \mathbf{b}_i^T \mathbf{Z}_i^T \mathbf{Z}_i \mathbf{b}_i \\
&= \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i + 2\mathbf{s}^T \mathbf{Z}_i \mathbf{b}_i + \mathbf{b}_i^T \mathbf{Z}_i^T \mathbf{Z}_i \mathbf{b}_i + \mathbf{s}^T \mathbf{s} \\
&= \mathbf{b}_i^T (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i) \mathbf{b}_i + 2\mathbf{s}^T \mathbf{Z}_i \mathbf{b}_i + \mathbf{s}^T \mathbf{s} \\
&= \mathbf{b}_i^T \mathbf{K}_i \mathbf{b}_i + 2\mathbf{s}^T \mathbf{Z}_i \mathbf{b}_i + \mathbf{s}^T \mathbf{s},
\end{aligned} \tag{75}$$

sendo $\mathbf{K}_i = \mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i$. Nota-se que a expressão (75) é similar ao quadrado incompleto

$$(\mathbf{b}_i - \mathbf{h})^T \mathbf{K}_i (\mathbf{b}_i - \mathbf{h}) + 1 = \mathbf{b}_i^T \mathbf{K}_i \mathbf{b}_i - 2\mathbf{b}_i^T \mathbf{K}_i \mathbf{h} + \mathbf{h}^T \mathbf{K}_i \mathbf{h} + 1. \tag{76}$$

comparando (75) com (76), tem-se a igualdade $-2\mathbf{b}_i^T \mathbf{K}_i \mathbf{h} = 2\mathbf{s}^T \mathbf{Z}_i \mathbf{b}_i \Rightarrow \mathbf{h} = -\mathbf{K}_i^{-1} \mathbf{Z}_i^T \mathbf{s}$, e, $\mathbf{h}^T \mathbf{K}_i \mathbf{h} + 1 = \mathbf{s}^T \mathbf{s} \Rightarrow 1 = \mathbf{s}^T \mathbf{s} - \mathbf{h}^T \mathbf{K}_i \mathbf{h}$, ou seja,

$$\mathbf{K}_i = \mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i \tag{77}$$

$$\mathbf{h} = -(\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{s} \tag{78}$$

$$\begin{aligned}
\mathbf{l} &= \mathbf{s}^T \mathbf{s} - \mathbf{s}^T \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i) (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{s} \\
&= \mathbf{s}^T \mathbf{s} - \mathbf{s}^T \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{s} \\
&= \mathbf{s}^T [\mathbf{I} - \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T] \mathbf{s}.
\end{aligned} \tag{79}$$

Substituindo-se (77), (78) e (79) em (74), a função densidade pode ser escrita como

$$\begin{aligned}
f(\mathbf{y}_i = \mathbf{1}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{\mathbf{X}\boldsymbol{\beta}} \frac{(2\pi)^{-\frac{n_i+q}{2}}}{|\mathbf{D}|^{1/2}} \exp \left\{ -\frac{1}{2} [(\mathbf{b}_i - \mathbf{h})^T \mathbf{K}_i (\mathbf{b}_i - \mathbf{h}) + 1] \right\} d\mathbf{b}_i ds \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{\mathbf{X}\boldsymbol{\beta}} \frac{(2\pi)^{-\frac{n_i+q}{2}}}{|\mathbf{D}|^{1/2}} \exp \left\{ -\frac{1}{2} [(\mathbf{b}_i - \mathbf{h})^T (\mathbf{K}_i^{-1})^{-1} (\mathbf{b}_i - \mathbf{h})] \right\} \\
&\quad \times \exp \left\{ -\frac{1}{2} \{ \mathbf{s}^T [\mathbf{I} - \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T] \mathbf{s} \} \right\} d\mathbf{b}_i ds \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{\mathbf{X}\boldsymbol{\beta}} \frac{(2\pi)^{-\frac{n_i+q}{2}}}{|\mathbf{D}|^{1/2}} \exp \left\{ -\frac{1}{2} [(\mathbf{b}_i - \mathbf{h})^T (\mathbf{K}_i^{-1})^{-1} (\mathbf{b}_i - \mathbf{h})] \right\} \\
&\quad \times \exp \left\{ -\frac{1}{2} \{ \mathbf{s}^T \mathbf{L} \mathbf{s} \} \right\} d\mathbf{b}_i ds,
\end{aligned}$$

sendo $\mathbf{L} = \mathbf{I} - \mathbf{Z}_i(\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T$. Multiplicando e dividindo essa última expressão por constantes adequadas, obtém-se

$$\begin{aligned}
f(\mathbf{y}_i = \mathbf{1}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{\mathbf{X}\beta} \frac{(2\pi)^{-\frac{n_i+q}{2}} |\mathbf{K}_i^{-1}|^{-1/2}}{|\mathbf{D}|^{1/2} |\mathbf{K}_i^{-1}|^{-1/2}} \cdot \frac{|\mathbf{L}_i^{-1}|^{-1/2}}{|\mathbf{L}_i^{-1}|^{-1/2}} \\
&\quad \times \exp \left\{ -\frac{1}{2} [(\mathbf{b}_i - \mathbf{h})^T (\mathbf{K}_i^{-1})^{-1} (\mathbf{b}_i - \mathbf{h})] \right\} \exp \left\{ -\frac{1}{2} \{ \mathbf{s}^T [\mathbf{L}^{-1}]^{-1} \mathbf{s} \} \right\} d\mathbf{b}_i ds, \\
&= |\mathbf{D}|^{-1/2} |\mathbf{K}_i|^{-1/2} |\mathbf{L}_i|^{-1/2} \\
&\quad \times \int_{-\infty}^{+\infty} (2\pi)^{-\frac{q}{2}} |\mathbf{K}_i^{-1}|^{-1/2} \exp \left\{ -\frac{1}{2} [(\mathbf{b}_i - \mathbf{h})^T (\mathbf{K}_i^{-1})^{-1} (\mathbf{b}_i - \mathbf{h})] \right\} d\mathbf{b}_i \\
&\quad \times \int_{-\infty}^{\mathbf{X}\beta} (2\pi)^{-\frac{n_i}{2}} |\mathbf{L}_i^{-1}|^{-1/2} \exp \left\{ -\frac{1}{2} \{ \mathbf{s}^T [\mathbf{L}^{-1}]^{-1} \mathbf{s} \} \right\} ds. \tag{80}
\end{aligned}$$

A primeira integral na expressão (80) é igual a 1, uma vez que o integrando é uma densidade normal multivariada com média \mathbf{h} e matriz de variância-covariância \mathbf{K}^{-1} e os limites de integração cobrem todo o domínio de \mathbf{b}_i . A segunda integral em (80) é a função de distribuição acumulada da normal multivariada com vetor de media $\mathbf{0}$ e matriz de variâncias-covariâncias \mathbf{L}^{-1} . Assim, a função de probabilidade marginal para $\mathbf{Y}_i = \mathbf{1}$ do modelo (53) pode ser escrita como

$$\begin{aligned}
f(\mathbf{y}_i = \mathbf{1}) &= |\mathbf{D}|^{-1/2} |\mathbf{K}_i|^{-1/2} |\mathbf{L}_i|^{-1/2} \Phi(\mathbf{X}\beta; \mathbf{L}^{-1}) \\
&= \frac{\Phi(\mathbf{X}\beta; \mathbf{L}^{-1})}{|\mathbf{D} \cdot \mathbf{K} \cdot \mathbf{L}|^{1/2}}. \tag{81}
\end{aligned}$$

Expandindo o denominador de (81), tem-se

$$\begin{aligned}
|\mathbf{D} \cdot \mathbf{K} \cdot \mathbf{L}|^{1/2} &= |\mathbf{D}(\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)[\mathbf{I} - \mathbf{Z}_i(\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T]|^{1/2} \\
&= |\mathbf{I} - \mathbf{Z}_i(\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T + \mathbf{DZ}_i^T \mathbf{Z}_i \\
&\quad - \mathbf{DZ}_i^T \mathbf{Z}_i \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T|^{1/2}, \tag{82}
\end{aligned}$$

tomando a expressão (82) no caso univariado, ou seja, assumindo um único efeito aleatório $b_i \sim N(0, \sigma^2)$, $Z_i = 1$ e $D = \sigma^2$, obtém-se o resultado

$$|1 - 1 \cdot (\sigma^{-2} + 1)^{-1} \cdot 1 + \sigma^2 \cdot 1 - \sigma^2 \cdot 1 \cdot (\sigma^{-2} + 1)^{-1} \cdot 1|^{1/2} = 1.$$

Logo, é esperado que o resultado multivariado seja igual ao caso univariado. O seguinte corolário será utilizado (RENCHEER, 2000, pág. 34):

COROLÁRIO 1: Seja $A = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$ uma matriz subdividida, sendo \mathbf{A}_{11} e \mathbf{A}_{22} submatrizes quadradas. O determinante de \mathbf{A} pode ser calculado como:

$$\begin{aligned} |\mathbf{A}| &= |\mathbf{A}_{11}| \cdot |\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}| \\ &= |\mathbf{A}_{22}| \cdot |\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|. \end{aligned}$$

Temos da expressão (82), que

$$|\mathbf{D} \cdot \mathbf{K} \cdot \mathbf{L}|^{1/2} = |\mathbf{D}|^{1/2} \cdot |\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i|^{1/2} \cdot |\mathbf{I} - \mathbf{Z}_i(\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T|^{1/2}.$$

Comparando os dois últimos determinantes do lado direito da igualdade com o corolário 1, nota-se o determinante de uma matriz $\mathbf{A} = \begin{bmatrix} \mathbf{K} & \mathbf{Z}_i^T \\ \mathbf{Z}_i & \mathbf{I} \end{bmatrix}$. Logo, tem-se que

$$\begin{aligned} |\mathbf{A}|^{1/2} &= |\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i|^{1/2} \cdot |\mathbf{I} - \mathbf{Z}_i(\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T|^{1/2} \\ &= |\mathbf{I}|^{1/2} \cdot |\mathbf{K} - \mathbf{Z}_i^T \mathbf{I}^{-1} \mathbf{Z}_i|^{1/2} \\ &= |\mathbf{K} - \mathbf{Z}_i^T \mathbf{Z}_i|^{1/2} \\ &= |\mathbf{D}^{-1}|^{1/2}. \end{aligned}$$

Portanto,

$$\begin{aligned} |\mathbf{D} \cdot \mathbf{K} \cdot \mathbf{L}|^{1/2} &= |\mathbf{D}|^{1/2} |\mathbf{K} \cdot \mathbf{L}|^{1/2} \\ &= |\mathbf{D}|^{1/2} |\mathbf{D}^{-1}|^{1/2} \\ &= |\mathbf{D} \cdot \mathbf{D}^{-1}|^{1/2} \\ &= 1. \end{aligned}$$

Com esse resultado, a expressão da função de probabilidade marginal de $\mathbf{Y}_i = 1$, descrita em (81)

passa a ser

$$f(\mathbf{y}_i = \mathbf{1}) = \Phi(\mathbf{X}_i\boldsymbol{\beta}; \mathbf{L}_i^{-1}), \quad \mathbf{L}_i = \mathbf{I} - \mathbf{Z}_i(\mathbf{D}^{-1} + \mathbf{Z}_i^T\mathbf{Z}_i)^{-1}\mathbf{Z}_i^T. \quad (83)$$

Essa expressão é a função de probabilidade do modelo MLGM probit-Bernoulli-normal. Utilizando esse resultado em (73), tem-se como a função de probabilidade para um modelo probit-normal-Bernoulli-beta a função

$$f(\mathbf{y}_i = \mathbf{1}) = \left(\frac{\alpha}{\alpha + \beta}\right)^{n_i} \Phi(\mathbf{X}_i\boldsymbol{\beta}; \mathbf{L}_i^{-1}), \quad \mathbf{L}_i = \mathbf{I} - \mathbf{Z}_i(\mathbf{D}^{-1} + \mathbf{Z}_i^T\mathbf{Z}_i)^{-1}\mathbf{Z}_i^T. \quad (84)$$

Essa expressão pode ser utilizada diretamente no processo de estimação pelo método de máxima verossimilhança. Uma aproximação da função distribuição Gaussiana acumulada para a distribuição logística pode ser utilizada para aproximar as estimativas do modelo probit para o modelo logit, por meio da multiplicação do preditor linear com a constante $c = 16\sqrt{3}/15\pi$, o que permite o cálculo de medidas como a razão das chances e seus intervalos de confiança (ZEGGER; LIANG; ALBERT, 1988).

Esse processo de estimação adotado é apenas uma das opções disponíveis. A inferência estatística do ponto de vista Bayesiano é uma opção bastante flexível, que pode ser também aplicada e o resultado em (84) pode ser utilizado, simplificando a função densidade de probabilidade conjunta *a posteriori*.

2.4.2.4 Aplicação e resultados

Os modelos (49), (50), (52) e (54) foram aplicado aos dados de TDO, descritos na Seção 2.4.2.1, utilizando o procedimento NLMIXED do sistema SAS. Como o objetivo da pesquisa era verificar se os dois tratamentos estudados diferem ou não, foi utilizado os seguinte preditor linear:

$$\text{Logit}(\mu_{ij}) = \begin{cases} (\beta_0 + b_i) + \beta_1 t_{ij}, & \text{Tratamento A} \\ (\beta_2 + b_i) + \beta_3 t_{ij}, & \text{Tratamento B} \end{cases} \quad (85)$$

Os resultados estão apresentados na Tabela 6 e o código SAS é apresentado nos anexos. As estatísticas do logaritmo da função de verossimilhança (multiplicada por -2) e o valor do critério de informação de Akaike (PAWITAN, 2001) são apresentados para auxiliar na seleção de modelos. Os parâmetros do modelo combinado foram estimados utilizando a função (71), que foi programada

Tabela 6 - Resultados dos ajustes dos modelos MLG, beta-binomial, MLGM e o modelo combinado, assumindo função de ligação logit

| Efeito | Parâmetro | Bernoulli | Beta-binomial |
|------------------------------|----------------|-------------------|-------------------|
| | | Estimativa (e.p.) | Estimativa (e.p.) |
| Intercepto Tratamento A | β_0 | -0,5571 (0,1090) | 17,9714 (1482,6) |
| Coef. Angular Tratamento A | β_1 | -0,1769 (0,0246) | 5,2454 (12970,0) |
| Intercepto Tratamento B | β_2 | -0,5335 (0,1122) | 18,6744 (2077,13) |
| Coef. Angular Tratamento B | β_3 | -0,2549 (0,0309) | 4,7775 (12912,0) |
| Desvio Padrão Ef. Aleatórios | σ | — | — |
| Razão | α/β | — | 3,6739 (0,2051) |
| -2 log-likelihood | | 1811,7 | 1980,0 |
| Critério AIC | | 1815,7 | 1990,0 |
| Efeito | Parâmetro | Bernoulli-normal | Combinado |
| | | Estimativa (e.p.) | Estimativa (e.p.) |
| Intercepto Tratamento A | β_0 | -1,6299 (0,4354) | -1,6042 (4,0263) |
| Coef. Angular Tratamento A | β_1 | -0,4042 (0,0460) | -6,4783 (1,4386) |
| Intercepto Tratamento B | β_2 | -1,7486 (0,4478) | -16,2079 (3,5830) |
| Coef. Angular Tratamento B | β_3 | -0,5634 (0,0602) | -8,0745 (1,5997) |
| Desvio Padrão Ef. Aleatório | σ | 4,0150 (0,3812) | 60,8835 (14,2237) |
| Razão | α/β | — | 0,2805 (0,0350) |
| -2 log-likelihood | | 1247,5 | 1239,5 |
| Critério AIC | | 1257,5 | 1251,5 |

no procedimento NLMIXED do sistema SAS.

Nota-se que o modelo combinado apresentou o menor valor do critério de Akaike e o modelo Beta-Binomial apresentou o maior valor. Entretanto, este último apresentou valores de grande magnitude para os erros padrões, reflexo da dificuldade de se alcançar uma estabilidade numérica na estimação dos parâmetros desse modelo. Os modelos com efeitos aleatórios apresentaram resultados melhores que o modelo MLG Bernoulli, com uma ligeira vantagem para o modelo combinado. Porém, notam-se, também, grandes valores dos erros padrões e um valor muito grande para o desvio padrão do efeito aleatório b_i . Para os modelos Beta-Binomial, MLGM Bernoulli-normal e o modelo combinado, foram encontradas dificuldades no processo numérico de estimação dos parâmetros, uma vez que a convergência era fortemente dependente dos valores iniciais do algoritmo. Para o modelo beta-binomial e o modelo combinado foi adotada a restrição $\alpha/\beta = c$ para solucionar o problema de identificabilidade do modelo. Outra alternativa foi utilizar a restrição $\alpha + \beta = c$ (HINDE; DEMÉTRIO, 1998), que também foi testada mas não foi obtida uma convergência dos algoritmos, especificamente para o conjunto de dados utilizados.

A análise dos dados de TDO também foi feita, utilizando-se os modelos baseados na função de ligação probit, com os resultados apresentados na Tabela 7. Como no caso logit, foram ajusta-

Tabela 7 - Resultados dos ajustes dos modelos MLG, beta-binomial, MLGM e o modelo combinado, assumindo função de ligação probit

| Efeito | Parâmetro | Bernoulli | Beta-binomial |
|-------------------------------|-----------|--------------------|-------------------|
| | | Estimativa (e.p.) | Estimativa (e.p.) |
| Intercepto Tratamento A | β_0 | -0,3678 (0,06493) | — |
| Coef. Angular Tratamento A | β_1 | -0,09647 (0,01276) | -0,1125 (0,01136) |
| Intercepto Tratamento B | β_2 | -0,3661 (0,06592) | — |
| Coef. Angular Tratamento B | β_3 | -0,1335 (0,01502) | -0,1537 (0,01320) |
| Desvio Padrão Ef. Aleatórios | σ | — | — |
| Parâmetro Superdispersão | α | — | 7,0493 (0,3507) |
| -2 log-verossimilhança | | 1815,1 | -6965 |
| Critério AIC | | 1823,1 | -6959 |

| Efeito | Parâmetro | Bernoulli-normal | Combinado |
|-------------------------------|-----------|-------------------|-------------------|
| | | Estimativa (e.p.) | Estimativa (e.p.) |
| Intercepto Tratamento A | β_0 | -0,9175 (0,2296) | -1,0714 (0,6774) |
| Coef. Angular Tratamento A | β_1 | -0,1961 (0,02117) | -0,8335 (0,1631) |
| Intercepto Tratamento B | β_2 | -0,9970 (0,2361) | -1,8020 (0,4939) |
| Coef. Angular Tratamento B | β_3 | -0,2723 (0,02626) | -1,2039 (0,2099) |
| Desvio Padrão Ef. Aleatório | σ | 2,1173 (0,1975) | 8,0835 (1,1202) |
| Parâmetro Superdispersão | α | — | 7,8882 (0,2189) |
| -2 log-verossimilhança | | 1271,7 | -7523 |
| Critério AIC | | 1281,7 | -7511 |

dos os modelos MLG Bernoulli, beta-binomial, MLGM Bernoulli-normal e o modelo combinado Bernoulli-beta-normal. Foi assumida para os modelos beta-binomial e o combinado a restrição $\alpha + \beta = c$ para eliminar o problema da identificabilidade, uma vez que a restrição $\beta/\alpha = c$ não levou a uma convergência dos algoritmos. Foram obtidos resultados diferentes do modelo que considerou a função de ligação logit: com base na estatística de $-2 \times \log$ -verossimilhança e no critério de Akaike, os modelos beta-binomial e o modelo combinado foram bastante superiores aos modelos MLG e MLGM. O modelo beta-binomial foi considerado apenas com os coeficientes angulares dos modelos, uma vez que a presença dos interceptos e do parâmetro de superdispersão α , simultaneamente no modelo, gerava forte correlação entre os parâmetros estimados e, conseqüentemente, instabilidade numérica no processo iterativo de estimação. Nota-se também que os modelos alternativos ao modelo MLG Bernoulli foram todos mais informativos, indicando a necessidade de se modelar a variação extra-Bernoulli. Porém, a utilização do modelo combinado mostra-se mais geral e abrangente, acomodando tanto o efeito de superdispersão quanto a dependência temporal dos dados binários. Tanto o parâmetro de superdispersão α quanto o desvio padrão do efeito aleatório mostraram-se significativos e com valores mais razoáveis, quando comparado ao modelo com função de ligação logit.

Tabela 8 - Resultados do ajuste do modelo combinado probit-normal-Bernoulli-beta

| Parâmetro | Estimativa | Erro Padrão | G.L. | Valor-P |
|-----------|------------|-------------|------|---------|
| β_0 | -1.8220 | 1.1518 | 293 | 0.1147 |
| β_1 | -1.4173 | 0.2773 | 293 | <.0001 |
| β_2 | -3.0641 | 0.8398 | 293 | 0.0003 |
| β_3 | -2.0471 | 0.3569 | 293 | <.0001 |
| α | 7.8882 | 0.2189 | 293 | <.0001 |
| σ | 13.7456 | 1.9049 | 293 | <.0001 |

A aproximação do modelo probit para o modelo logit, em que a constante $c = 16\sqrt{3}/15\pi$ é multiplicada pelo preditor linear, permite que inferências sobre a razão das chances sejam estabelecidas. Os resultados do modelo logit aproximado são apresentados na Tabela 8. A significância dos coeficientes angulares já indicam que ambos os tratamentos levam à redução da severidade ao longo dos 12 meses. A diferença entre os coeficientes angulares do tratamento A e B foi testada e foi concluído que os tratamentos diferem na intensidade de redução da severidade de ocorrência do TDO ($P=0,0255$) e, portanto, o tratamento B é mais rápido na redução da severidade. Calculando a razão das chances, tem-se que, para o tratamento A, a chance de redução da severidade de TDO é de $1/\exp(-1,4173) = 4.13$ vezes, para cada mês após o início do tratamento. Para o tratamento B a chance de redução da severidade é de 7,8 vezes, para cada mês que passa, após o início do tratamento.

Os modelos da família elemento-específico (como a classe MLGM) diferem dos modelos marginais (como os modelos obtidos por equações de estimação generalizadas); esses últimos estimam a tendência populacional esperada. Os modelos elemento-específicos estimam uma curva temporal para cada elemento da população estudada, e portanto, ambas as abordagens podem diferir, principalmente, se a variância de \mathbf{b}_i é grande. Entretanto, pode-se obter uma aproximação do modelo marginal através do modelo elemento-específico. No caso do modelo probit-normal-Bernoulli-beta aproximado ao modelo (54), deve-se gerar uma amostra de tamanho M de θ_{ij} e de \mathbf{b}_i , aplicar o modelo estimado a essas amostras e calcular a média aritmética dos M valores gerados. Na Figura 9 são apresentadas duas curvas referentes à tendência populacional aproximada de cada tratamento, baseado em 10.000 amostras dos dois efeitos aleatórios.

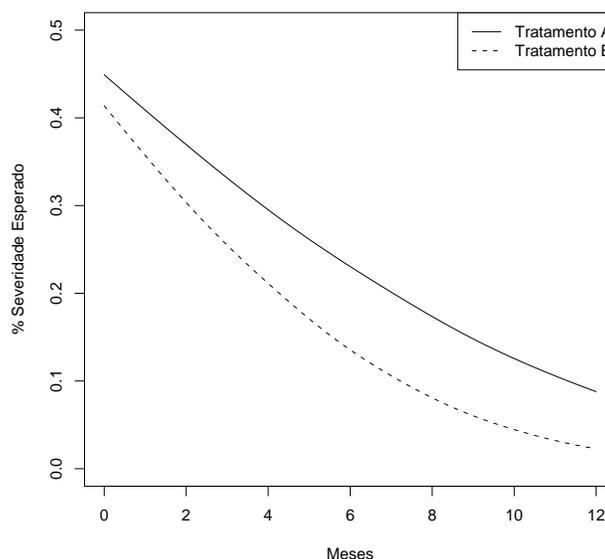


Figura 9 - Tendência populacional esperada para o percentual de severidade dos dois tratamentos, com base nos parâmetros da Tabela 8

2.4.3 Considerações finais

Foram propostos nesse trabalho, dois modelos para dados binários medidos longitudinalmente e com superdispersão. Ambos os modelos são generalizações do modelo linear generalizado com função de ligação logit e probit e distribuição Bernoulli. O modelo também se reduz ao modelo beta-binomial e a um modelo linear generalizado misto.

O processo de estimação adotado para os modelos é baseado no método da máxima verossimilhança. Para o modelo logit foi apresentada a função de verossimilhança que deve ser otimizada utilizando os algoritmos quase-Newton ou Newton-Raphson, que em cada ciclo iterativo, soluciona integrais numéricas através do algoritmo de quadratura Gaussiana adaptativa. Para o modelo probit foi possível obter a expressão analítica para a densidade marginal incondicional para Y , que pode ser otimizada usando o método da máxima verossimilhança sem a necessidade de solucionar integrais numéricas.

Os modelos foram aplicados ao problema da análise de severidade do TDO, em que tanto no caso logit quanto no caso probit se mostraram mais explicativos que o modelo MLG Bernoulli, e também quando comparado aos modelos beta-binomial e MLGM Bernoulli-normal. Entretanto, o modelo probit se mostrou mais robusto do que o modelo logit, tanto no processo numérico quanto

nos resultados obtidos na forma dos parâmetros estimados.

A aplicação clínica aqui apresentada foi a motivação para o estudo desses modelos mas diversas situações são também contempladas por esses modelos. Uma aplicação agronômica que caracteriza essa situação é o estudo comportamental de suínos em uma granja, em que o número de animais é grande e muitos deles pertencem à mesma leitegada, além de cada animal ter seu comportamento classificado como ‘estressado’ ou não ‘estressado’, com base na pressão sonora emitida pelos animais em diferentes horários do dia. Essa e outras aplicações agronômicas serão desenvolvidas em um futuro próximo.

Os modelos combinados foram desenvolvidos fazendo algumas escolhas particulares. O método de estimação, por exemplo, pode ser aplicado do ponto de vista Bayesiano, que permite bastante flexibilidade e robustez quando se trata de pequenas amostras. A distribuição normal multivariada, utilizada como distribuição para os efeitos aleatórios em b_i foi escolhida em um primeiro momento, tanto pela disponibilidade computacional quanto por ser razoável no tipo de problema estudado. Entretanto, outras distribuições poderiam ser aplicadas e avaliadas quanto à má-especificação da distribuição para os efeitos aleatórios (LITIÈRE; MOLENBERGHS; ALONSO, 2006). A base de dados TDO é adequada para que os resultados assintóticos aqui obtidos sejam válidos. Para problemas experimentais com pequenas amostras, a performance dos estimadores de máxima verossimilhança desses modelos deve ser avaliada em estudos de simulação.

3 COMENTÁRIOS FINAIS

Objetivou-se nesse trabalho de tese, em um sentido mais amplo, trazer para a pesquisa agrônômica a potencialidade que a modelagem simultânea de média e dispersão pode propiciar, aumentando a precisão das inferências sobre o estudo e ampliando o entendimento sobre a variabilidade inerente aos fenômenos e estudos biológicos. A classe dos modelos lineares generalizados duplos já está razoavelmente amadurecida quanto à especificação do modelo, contando com a implementação em softwares bem conhecidos e confiáveis para a aplicação no dia-a-dia do pesquisador. O que é necessário nesse momento é a compreensão da importância de se conhecer melhor a forma com que fatores e variáveis influenciam na variabilidade. As idéias de Taguchi são perfeitamente aplicáveis à pesquisa agrônômica e biológica contemporânea, em que produtos vegetais e animais devem apresentar homogeneidade e alta qualidade.

Em problemas de grande dimensão, como o caso de mortalidade de frangos de corte na condição pré-abate, muitas fontes de variação estão presentes e influenciando nas respostas de interesse. O MLGD foi eficiente tanto para modelar os fatores que influenciam na mortalidade quanto para modelar explicitamente a superdispersão gerada por fatores bioclimáticos, que nem sempre são possíveis de serem controlados.

A situação de experimentos com pequenas amostras, em que se tem um número relativamente grande de parâmetros a serem estimados, foi abordada do ponto de vista Bayesiano, que se mostrou robusto e eficiente na análise de um problema com tão poucas observações. A abordagem Bayesiana atualmente conta com poderosas funções e rotinas computacionalmente intensivas que habilitam o estudo de modelos em situações práticas que anteriormente seriam proibitivos pela abordagem frequentista.

O problema clínico de dermatologia abordado na Seção 2.4 motivou e exemplificou as propostas para modelagem de dados binários longitudinais com superdispersão. Em particular, para o caso probit, foi obtida a expressão da função densidade de probabilidade marginal incondicional para a variável resposta, o que simplifica consideravelmente o esforço computacional para a estimação dos parâmetros.

As classes de modelos MLGS, MLGD e os modelos combinados para dados longitudinais com superdispersão são casos particulares de uma classe muito maior e que não foi estudada aqui, denominada *modelos lineares generalizados hierárquicos duplos (MLGHD)*. Essa classe traz a ge-

neralidade dos dois mundos (os MLG e os modelos de efeitos aleatórios) para solucionar uma gama enorme de problemas. Entretanto, essa generalidade tem seu preço: o entendimento do algoritmos de estimação não é trivial e as propriedades gerais dessa grande classe está sob estudo.

Existem muitas perspectivas para a continuação desta pesquisa. Uma delas é estudar extensões do modelo DGLM baseado nos MLG aditivos, em que a forma funcional do preditor linear é não paramétrica e apenas aditiva para cada variável independente. Isso garantirá uma generalidade, quando o interesse é na predição ou na exploração da relação das covariáveis na média e na dispersão. No problema da mortalidade de aves pré-abate pode-se tentar incorporar a informação da granja das aves, por meio dos MLGM, uma vez que essa informação carrega muitas fontes de variação e a incorporação desse fator através de um efeito aleatório parece ser bastante razoável. A exemplificação dos modelos longitudinais para dados binários com superdispersão foi feita sobre um problema clínico humano, mas sendo de grande interesse, a aplicação desta classe de modelos a problemas agronômicos.

REFERÊNCIAS

AASLYNG, M.D.; GADE, P.B. Low stress pre-slaughter handling: effect of lairage time on the meat quality of pork. **Meat Science**, London, v. 57, p. 87-92, 2001.

ASSOCIAÇÃO BRASILEIRA DOS PRODUTORES E EXPORTADORES DE FRANGOS. **Relatório anual 2006**. Disponível em: <<http://www.abef.com.br>>. Acesso em: 15 dez. 2007.

AITKIN, M. Modelling variance heterogeneity in normal regression using GLIM. **Applied Statistics**, London, v. 36, n. 3, p.332-339, 1987.

ANDERSSON, B.E.; JÓNASSON, H. Regulação da temperatura e fisiologia ambiental. In: SWENSON, M.J.; REECE, W. O. (Ed.). **Dukes - fisiologia dos animais domésticos**. 11 ed. Rio de Janeiro: Guanabara Koogan, 1996. cap. 47, p. 805 - 813.

BARBOSA FILHO, J.A.D. **Caracterização quantiquantitativa das condições bioclimáticas e produtivas nas operações pré-abate de frangos de corte**. 2008. 174p. Tese (Doutorado em Física do Ambiente Agrícola) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2008.

BARBOSA FILHO, J.A.D.; VIEIRA, F.M.C.; GARCIA, D.B.; SILVA, M.A.N. da; SILVA, I.J.O. da. **Mudanças e uso das tabelas de entalpia**. Piracicaba, 2007. Disponível em: <<http://www.nupea.esalq.usp.br>>. Acesso em: 09 ago. 2007.

BEST, N.; GREEN, P. Structure and uncertainty: graphical models for understanding complex data. **Significance**, London, p. 177-181, Dec. 2005.

BOX, G. Signal-to-noise ratios, performance criteria and transformations. **Technometrics**, Alexandria, v. 30, n. 1, p.1-40, feb. 1988.

BOX, G.E.P., COX, D.R. An analysis of transformations. **Journal of the Royal Statistical Society series B**, London, v. 26, p.211-52, 1964.

BRESLOW, N.E.; CLAYTON, D.G. Approximate inference in generalized linear mixed models. **Journal of the American Statistical Association**, Boston, v.88, p.9-25, 1993.

BRESSAN, M.C.; BERAQUET, N.J. Efeito de fatores pré-abate sobre a qualidade da carne de Peito de frango. **Ciência Agrotécnica**, Lavras, v. 26, n. 5, p. 1049-1059, set. / out. 2002.

BROOM, D.M. Welfare assessment and welfare problem areas. In: GRANDIN, T. (Ed.). **Livestock handling and transport**. Wallingford: CAB International, 1993, 320 p.: ill.

CEPEDA, E.; GAMERMAN, D. Bayesian modeling of variance heterogeneity in normal regression models. **Brazilian Journal of Probability and Statistics**, São Paulo, v. 14, n. 2, p. 207-221, Dec. 2000.

COCHRAN, W.G.; COX, G.M. **Experimental designs**. 2.ed. New York: John Wiley & Sons, 1957.

COLLETT, D. **Modelling binary data**. London: Chapman & Hall, 1991.

COX, D.R., REID, N. Parameter orthogonality and approximate conditional inference. **Journal of the Royal Statistical Society, Series B**, London, v. 49, p.1-39, 1987.

DE BACKER, M.; DE KEYSER, P.; DE VROEY, c.; LESAFFRE, E. A 12-week treatment for dermatophyte toe onychomycosis: terbinafine 250mg/day vs itraconazole 200mg/day – a double-blind comparative trial. **British Journal of Dermatology**, London, v. 134, p.16-17, 1996.

DELLAPORTAS, P; SMITH, A.F.M. Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. **Applied Statistics**, London, v. 42, n. 3, p. 443-459, 1993.

DEMÉTRIO, C.G.B. **Modelos lineares generalizados em experimentação agrônômica**. Piracicaba: Departamento de Ciências Exatas, ESALQ/USP, 2002. 117 p.

EMBRAPA PECUÁRIA SUDESTE. **Clima e hidrologia**. São Carlos: Embrapa Pecuária Sudeste, 2007. Disponível em: <<http://www.cppse.embrapa.br/servicos/dados-meteorologicos>>. Acesso em: 21 mar. 2007.

ENGEL, J. Modelling variation in industrial experiments. **Applied Statistics**, London, v. 41, n. 3, p.579-593, 1992.

ERNST, M.D. Permutation methods: a basis for exact inference. **Statistical Science**, Baltimore, v. 19, n. 4, p. 676-685, 2004.

FERRARI, S.L.P.; SILVA, A.F. Analytical and resampling-based bias corrections for one-parameter models, **Brazilian Journal of Probability and Statistics**, São Paulo, v. 13, n. 1, p. 13-27, june 1999.

FERREIRA, D.F.; DEMETRIO, C.G.B.; MANLY, B.F.J.; MACHADO, A.A., VENCOVSKY, R. Statistical models in agriculture: biometrical methods for evaluating the phenotypic stability in plant breeding. **Cerne**, Lavras, v.12, n.4, p.373-88, oct/dec. 2006.

FISHER, R.A. **The design of experiments**. Edingburgh: Oliver and Boyd, 1935.

FISHER, R.A. On the mathematical foundations of theoretical statistics. **Philosophical Transactions of the Royal Society of London, Series A**, London, v. 222, p. 309-368, 1922.

FREITAS, E.G. **Análise de dados longitudinais em experimentos com cana-de-açúcar**. 2007. 74 p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidadee São Paulo, Piracicaba. 2007.

GAMERMAN, D. Sampling from the posterior distribution in generalized linear mixed models. **Statistics and Computing**, London, v. 7, p. 57-68, 1997.

GAMERMAN, D.; LOPES, H. **Markov chain monte carlo: stochastic simulation for bayesian inference**. London: Chapman & Hall, 2006.

GARCIA, C.H. **Aplicação de equações de modelos mistos em testes clonais de *Eucalyptus SPP.*** 2004. 74 p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidadee São Paulo, Piracicaba. 2004.

GARCIA, D.B.; HILDEBRAND, A.; BARBOSA FILHO, J.A.D.; VIEIRA, F.M.C.; VIEIRA, A.M.C.; SILVA, I.J.O. Avaliação da temperatura retal durante a operação de pega pré-abate de frangos de corte. In: SIMPÓSIO INTERNACIONAL DE INICIAÇÃO CIENTÍFICA DA USP, 15., 2007, Pirassununga. **Anais...**São Paulo: Universidade de São Paulo, 2007. 1 CD-ROM.

HARVEY, A.C. Estimating regression models with multiplicative heteroscedasticity. **Econometrica**, v. 44, n. 3, p.461-465, may 1976.

GIVENS, G.H.; HOETING, J.A. **Computational statistics**. Hoboken: John Wiley and sons. 2005

HALL, D.B.; BERENHAUT, K.S. Score tests for heterogeneity and overdispersion in zero-inflated Poisson and binomial regression models. **The Canadian Journal of Statistics**, Ottawa, v. 30, n. 3, p. 1-16, 2002.

HEIDELBERGER; WELCH. Simulation run length control in presence of a initial transient. **Operations Research**, Linthicum, v. 31, n. 6, 1109-1144, Nov. 1983.

HINDE, J.; DEMÉTRIO, C.G.B. Overdispersion: models and estimation. **Computational Statistics & Data Analysis**, Amsterdam, v. 27, n. 2, p. 151-170, 1998.

HINDE, J.; DEMÉTRIO, C.G.B. **Overdispersion**: models and estimation. In: Simpósio Nacional de Probabilidade e Estatística, 13, 1998, Caxambú: Associação Brasileira de Estatística, 73 p.

HO, L.L.; da SILVA, A.F. Bias correction for mean time to failure and p -quantiles in a Weibull distribution by bootstrap procedures. **Communications in Statistics - Simulation and Computation**, Hamilton, v. 34, p. 617-629, 2005.

JOHNSON, R.A.; WICHERN, W.D. **Applied multivariate statistical analysis**. 6^a ed. Upper Saddle River: Prentice-Hall, 2007. 773 p.

JØRGENSEN, B. Exponential dispersion models. **Journal of the Royal Statistical Society, Series B (Methodological)**, London, v. 49, n. 2, p. 127-162, 1987.

KETTLEWELL, P.J.; HOXEY, R.P.; MITCHELL, M.A. Heat produced by broiler chickens in a commercial transport vehicle. **Journal of Agricultural Engineering Research**, St. Joseph, v. 75, n. 3, p. 315-326, Mar. 2000.

LAIRD, N.M.; WARE, J.H. Random-effects models for longitudinal data. **Biometrics**, Washington, v. 38, n. 4, p. 963-974, Dec. 1982.

LARA, I.A.R. **Modelos de transição para dados binários**. 2007. 128 p. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade São Paulo, Piracicaba, 2007.

LEE, Y.; NELDER, J.A. Generalized linear models for the analysis of quality-improvement experiments. **The Canadian Journal of Statistics**, Ottawa, v. 26, n. 1, p. 95-105, 1998.

LEE, Y.; NELDER, J.A.; PAWITAN, Y. **Generalized linear models with random effects: Unified analysis via H-likelihood**. London: Chapman & Hall, 2006. 396 p.

LEÓN, R.V.; SHOEMAKER, A.C.; KACKER, R.N. Performance measures independent of adjustment: an explanation and extension of Taguchi's signal-to-noise ratios. **Technometrics**, Alexandria, v. 29, n. 3, p.253-285, Aug. 1987.

LIANG, K.-Y.; ZEGER, S.L. Longitudinal data analysis using generalized linear models. **Biometrika**, London, v. 73, p. 13-22, 1986.

LITIÈRE, S.; MOLENBERGHS, G.; ALONSO, A. **The impact of a misspecified random-effects distribution on maximum likelihood estimation in generalized linear mixed models**. Diepenbeek: Universiteit Hasselt, 2006. 35 diapositivos: color.

LOGOTHETIS, N. Box-Cox transformations and the Taguchi method. **Applied Statistics**, London, v. 39, n. 1, p.31-48, 1990.

MANLY, B.J.F. **Randomization, bootstrap and monte carlo methods in biology**. 2. ed. New York: Chapman & Hall, 2004.

McCULLAGH, P.; NELDER, J.A. **Generalized linear models**. 2nd.ed. London: Chapman & Hall, 1989.

MOLENBERGHS, G.; VERBEKE, G. **Models for discrete longitudinal data**. New York: Springer, 2005.

MOLENBERGHS, G.; VERBEKE, G.; DEMÉTRIO, C.G.B. An extended random-effects approach to modeling repeated, overdispersion count data. **Lifetime Data Analysis**, Hingham, v. 13, n. 4, Dec. 2007.

MONTGOMERY, D.C. **Introdução ao controle estatístico da qualidade**. 4.ed. Rio de Janeiro: LTC Editora, 2004.

NAIR, V.N.; PREGIBON, D. Analysing dispersion effects from replicated factorial experiments. **Technometrics**, Alexandria, v. 30, n. 3, p.247-257, Aug. 1988.

NEAL, R.M. Slice sampling. **The Annals of Statistics**, Hayward, v. 31, n. 3, p. 705-767, 2003.

NELDER, J.A., LEE, Y. Generalized linear models for the analysis of Taguchi-type experiments. **Applied Stochastic Models and Data Analysis**, Chichester, v. 7, p. 107-120, 1991.

NELDER, J.P., LEE, Y. Likelihood, quasi-likelihood and pseudo-likelihood: Some comparisons. **Journal of the Royal Statistical Society Series B**, London, v. 54, p. 273-284. 1992.

NELDER, J.P., PREGIBON, D. An extended quasi-likelihood function. **Biometrika**, London, v. 74, n. 2, p.221-232, 1987.

NELDER, J.A., WEDDERBURN, W.M. Generalised linear models. **Journal of the Royal Statistical Society Series A**, London, v. 135, n. 3, p. 370-384, 1972.

NICOL, C.J.; SCOTT, G.B. Pre-slaughter handling and transport of broiler chickens. **Applied Animal Behaviour Science**, London, v. 28, p. 57-73, 1990.

PAYNE, R.W.; MURRAY, D.A.; HARDING, S.A.; BAIRD, D.B.; SOUTAR, D.M. **GenStat for Windows: Introduction**. 10.ed. Hemstead: VSN International, 2007.

PAWITAN, Y. **In all likelihood: statistical modelling and inference using likelihood**, Oxford: Clarendon Press, 2001.

PINTO, E.R. **Planejamento de experimentos ótimos para modelos lineares generalizados com parâmetro de dispersão variando**. 2005. 147p. Tese (Doutorado em Engenharia de Produção) – Coordenação dos Programas de Pós-Graduação em Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

PINTO, E. R.; PONCE DE LEON, A. C. M. Modelagem conjunta da média e dispersão de Nelder e Lee como alternativa aos métodos de Taguchi. **Pesquisa Operacional**, Rio de Janeiro, v. 26, n. 2, p. 203-224, 2006.

PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. **coda: output analysis and diagnostics for MCMC**. R package version 0.12-1, 2007.

PREGIBON, D. Review of generalized linear models by P. McCullagh and J. Nelder, **The Annals of Statistics**, Hayward, v. 12, n. 4, p. 1589-1596, Dec. 1984.

RAFTERY, A.E.; LEWIS, S.M. One long run with diagnostics: implementations strategies for Monte Carlo Markov Chains. **Statistical Science**, Hayward, v. 7, p. 493-497, 1992.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2006. Disponível em: <<http://www.R-project.org>>. Acesso em: 01 mar. 2007.

RENCHER, A.C. **Linear models in statistics**. New York: Wiley-Interscience, 2000.

RIBATET, M., IOOSS, B. **JointModeling: joint modelling of mean and dispersion**. R package version 1.0-0, 2006. URL <http://www.R-project.org>.

RIDOUT, M.; DEMÉTRIO, C.G.B. Generalized linear models for positive count data. **Revista de Matemática e Estatística**, Jaboticabal, v. 10, p. 139-148, 1992.

RIDOUT, M.; DEMETRIO, C.G.B.; HINDE, J. Models for count data with many zeros. In: 19th INTERNATIONAL BIOMETRIC CONFERENCE, 19., Cape Town. **Proceedings...**, Cape Town: International Biometric Society, 1998. 1 CD-ROM.

RODRIGUES, V.C. **Distribuição espacial e bem-estar de aves poedeiras em condições de estresse e conforto térmico utilizando visão computacional e inteligência artificial**. 2006. 101 p. Dissertação (Mestrado em Física do Ambiente Agrícola) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2006.

SAS INSTITUTE. **SAS Software Release 9.1.3**. Cary: SAS Institute, 2008.

SAVENIJE, B. **Metabolic parameters as indicators of broiler chicken welfare and meat quality**. Groningen: University of Groningen, 2001. Disponível em: <<http://irs.ub.rug.nl/ppn/235186198>>. Acesso em: 23 set. 2005.

SHOEMAKER, A.C.; TSUI, K.; LEÓN, R. V. Signal-to-noise ratios, performance criteria, and transformation: discussion. **Technometrics**, Alexandria, v. 30, n. 1, p. 19-21, Feb. 1988.

SMYTH, K. Generalized linear models with varying dispersion. **Journal of the Royal Statistical Society Series B**, London, v. 51, n. 1, p. 47-60, 1989.

SMYTH, G.K., VERBYLA, A.P. A conditional likelihood approach to REML in GLM. **Journal of the Royal Statistical Society, Series B**, London, v. 58, n.3, p. 565-572, 1996.

SMYTH, G.K., VERBYLA, A.P. Adjusted likelihood methods for modelling dispersion in generalized linear models. **Environmetrics**, v. 10, p. 696-709, 1999.

SPIEGELHALTER, D.; THOMAS, A.; BEST, N.; GILKS, W. **BUGS 0.5**: Bayesian inference using Gibbs sampling. Cambridge: Medical Research Council Biostatistics Unit, 1996. (Manual – version ii)

SPIEGELHALTER, D.; BEST, N.G.; CARLIN, B.P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society, Series B (Methodological)**, London, v. 64, n. 4, p. 583-639, 2002.

TAGUCHI, G. Quality engineering in Japan. **Communication in Statistics: Theory and Methods**, Hamilton, v. 14, p. 2785-2801, 1985.

THOMAS, A.; O'HARA, B.; LIGGES, U.; STURTZ, S. Making BUGS Open, **R News**, London, v. 6, n. 1, p. 12-17, 2006.

VECEREK, V.; GRBALOVA, S.; VOŠLAROVA, E.; JANACKOVA, B.; MALENA, M. Effects of travel distance and the season of the year on death rates of broilers transported to poultry processing plants. **Poultry Science**, Stanford, v. 85, p. 1881-1884, 2006.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. New York: Springer Science + Business Media, 2000.

VIEIRA, A. M. C., HINDE, J.P., DEMÉTRIO, C.G.B. Zero-inflated proportion data models applied to a biological control assay. **Journal of Applied Statistics**, Abingdon, v. 27, n.3, p. 373-389, 2000.

VIEIRA, F.M.C. **Avaliação das perdas e dos fatores bioclimáticos atuantes na condição de espera pré-abate de frangos de corte**. 2008. 176 p. Dissertação (Mestrado em Física do Ambiente Agrícola) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2008.

VIEIRA, F.M.C.; BARBOSA FILHO, J.A.D.; VIEIRA, A.M.C.; GARCIA, D.B.; HILDEBRAND, A.; SILVA, I.J.O. Influência do tempo de espera pré-abate na temperatura retal de frangos de corte. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 44., 2007, Jaboticabal. **Anais...Viçosa**: Sociedade Brasileira de Zootecnia, 2007. 1 CD-ROM.

WARRISS, P.D.; KNOWLES, T.G.; BROWN, S.N.; EDWARDS, J.E.; KETTLEWELL, P.J.; MITCHELL, M.A.; BAXTER, C.A. Effects of lairage time on body temperature and glycogen reserves of broiler chickens held in transport modules. **Veterinary Record**, London, v. 145, p. 218-212, 1999.

WEDDERBURN, R.W.M. Quasi-likelihood functions, generalized linear models, and Gauss-Newton method. **Biometrika**, London, v. 61, n. 3, p. 439-447, Dec. 1974.

WOLFINGER, R.D.; TOBIAS, R.D. Joint estimation of location, dispersion and random effects in robust design. **Technometrics**, Alexandria, v. 40, n. 1, p. 62-70, Feb. 1998.

ZEGER, S.L.; LIANG, K.; ALBERT, P.S. Models for longitudinal data: a generalized estimating equation approach. **Biometrics**, Washington, v. 44, n. 4, p. 1049-1060, Dec. 1988.

ANEXOS

Código R para o modelo DGLM aplicado ao problema de espera pré abate

```

# AnaliseEtapa2.r
# OBJETIVO: Este programa gera a análise dos dados de Espera Pré-abate
#           através de um modelo linear generalizado duplo, utilizando
#           a biblioteca dglm.
# DATA: 01/2008 - Depto de Ciências Exatas - ESALQ/USP
#####
# Comandos gerais

GERAL.COMP <- read.table("ETAPA2GERAL.txt", head=T)
head(GERAL.COMP) # Visualizando o cabeçalho
names(GERAL.COMP) # O nome das colunas
GERAL<-GERAL.COMP[c(-121,-126,-194),] #Excluindo outliers

# Conversao de variáveis para o formato de data
require (chron)
GERAL$TempoEspera <- chron(times=(GERAL$TempoEspera))
GERAL$Tempototal <- chron(times=(GERAL$Tempototal))
GERAL$TempoEspera <- minutes(GERAL$TempoEspera)+hours(GERAL$TempoEspera)*60
GERAL$Tempototal <- minutes(GERAL$Tempototal)+hours(GERAL$Tempototal)*60
class(GERAL$Tempototal)
class(GERAL$TempoEspera)

#Algumas modificações no tipo de dados...
GERAL$Turno <- factor(GERAL$Turno, levels=c("manha", "tarde", "noite"))
GERAL$Data <- chron(as.character(GERAL$Data),format=c(dates="d/m/y"))
attach(GERAL)
TRDif <- TRFinal - TRInicial
GERAL <- cbind(GERAL, TRDif)
head(GERAL)

library(dglm)

# MODELO 4: modelo novo, considerando temperatura no
# modelo da média e outros fatores no modelo de dispersão

modelo4 <- dglm(NumMort~(TempoEspera+NumAves+TInt+URInt+Distancia
+Turno+TExt+URExt+TRDif)^2
-Turno:TRDif
-Distancia:TRDif
-URInt:URExt
-URInt:Distancia
-TInt:URExt
-TInt:TExt
-TInt:Distancia
-NumAves:TRDif
-NumAves:Distancia
-NumAves:URInt
-TempoEspera:TRDif
-TempoEspera:URInt
-TempoEspera:TInt
-TempoEspera:NumAves
-NumAves:Turno,
~TExt+poly(TempoEspera,2)+poly(TInt,2)
+URInt+Turno,family=poisson)

summary(modelo4)

```

```
# ANOVA para os modelos candidatos
anova(modelo4)

##### ANÁLISE DOS RESÍDUOS - MODELOS CANDIDATOS #####

# MODELO 4
pdf("residuodisp1.pdf")
par(mfrow=c(2,2))
plot(modelo4$dispersion.fit)
dev.off()
pdf("residuol.pdf")
par(mfrow=c(2,2))
plot(modelo4)
dev.off()
```

Código R para o modelo DGLM Bayesiano aplicado ao problema cultura de tecidos

```

# ExemploDGLM1bayesiano.r
# DESCRICAO: Analise de dados do experimento com cultura de tecido. Dados de
#             proporções com Superdispersão, Exemplo A2 do livro OVERDISPERSION:
#             MODELS AND ESTIMATION, de Hinde & Demétrio, pag.66.
#             Analise via DGLM Bayesiano.
# ESTRUTURA DE DADOS:
#
# RESPOSTA: Número de explantes de maçã regenerados em 25.
#
# DELINEAMENTO EXPERIMENTAL: Fatorial 2x5
# Fatores:
#             TipoExplante: 5 níveis
#             MeioCultura: 2 níveis
#
# DATA: 20/10/2007
# AUTOR: Afrânio Vieira / Estatística / IE / UnB
#####
windows(record=T)
rm(list=ls(all=TRUE))
library(BRugs)
library(sound)

setwd("E:/Doutorado/PesquisaTese/AnaliseDados/ExemploDGLM1")

# LEITURA DOS DADOS
y <- c(8,10,9,11,9,10,18,12,7,13,20,12,20,5,2,13)
m <- rep(25,16)
Cultura <- factor(c(1,1,2,2,1,1,2,2,1,2,1,1,2,2,1,2), labels=c("X", "Y"))
Explante <- factor(c(1,1,1,1,2,2,2,2,3,3,4,4,4,4,5,5), labels=c("A", "B", "C", "D", "E"))
MatrizPlanejamento <- model.matrix( ~ Cultura*Explante,
                                     contrasts=list(Cultura="contr.treatment",
                                                  Explante="contr.treatment"))

Ncol <- ncol(MatrizPlanejamento)
#invisible(dput(list(N=length(y), Ncol=Ncol, Y=y, M=m, X=t(MatrizPlanejamento)),
#                  file="DadosCulturaTecidos.txt"))
#
# ANÁLISE VIA BIBLIOTECA BRugs...

##### Modelo com priori normais;
##### Preditor linear para média: Cultura+Explante+Cultura:Explante
##### Preditor linear para dispersão: intercepto + Explante D
# DEFININDO O MODELO GRAFICO DGLM BINOMIAL...
ModBinomial <- function(){
  for(i in 1:N){
    Y[i] ~ dbin(p[i],M[i])
    logit(p[i]) <- inprod(alfa[],X[i,]) + delta[i]
    delta[i] ~ dnorm(0,Tau[i])
    Tau[i] <- 1/exp(zeta[i])
    zeta[i] <- gama[1]*X[i,1] + gama[5]*X[i,5]
  }
  tauP <- 1/pow(10,2)

# Densidades a Priori de alpha e gamma...
  for(j in 1: Ncol){
    alfa[j]~ dnorm(0,tauP)
    gama[j]~ dnorm(0,tauP)
  }
}
writeModel(ModBinomial, "ModeloGraficoDGLM1.txt")

#Valores Iniciais...
alfa <- vector("numeric",Ncol)

```

```

gama <- vector("numeric",Ncol)
ValIniciais <- list(list(alfa=alfa,gama=gama,delta=rep(0,length(y))))
NomeFileInits <- c("DGLM1_init1.txt")
bugsInits(inits=ValIniciais, numChains=1, fileName=NomeFileInits)

##### PASSO-A-PASSO...
##Leitura do modelo, dados, número de cadeias e valores iniciais...
modelCheck("ModeloGraficoDGLM1.txt") # verifica sintaxe modelo
modelData("DadosCulturaTecidos.txt") # lê arquivo de dados
modelCompile(numChains=1) # compila modelo com três cadeia de Markov
#
## valores iniciais da cadeia
modelInits(c("DGLM1_init1.txt"))
#
modelUpdate(40000); #play(Sine(840,1,channels=1)) # burn in
dicSet() # Ativa o monitoramento do DIC
samplesSet(c("alfa[]","gama[]")) # parâmetros monitorados
modelUpdate(70000); play(Sine(840,3,channels=1)) # iterações da cadeia
## Alguns gráficos descritivos...
#samplesHistory("*") # desenha a cadeia de Markov,
#samplesBgr("*") # constrói o gráfico da estatística brg
#samplesAutoC("*",1,thin=20) # plot autocorrelations for the 1st chain
#samplesDensity("*",thin=20) # desenha densidades estimadas
#
## ESTATÍSTICAS DESCRITIVAS...
samplesStats("*", thin=50)
dicStats() # Calcula o DIC

```

Código SAS para os modelos GLM Logit-Bernoulli, Beta-Binomial, GLMM Logit-Bernoulli-Normal e o modelo combinado Logit-Bernoulli-Beta-Normal

```

/* Overdispersed GLMM for Binary Data
   Logistic-Beta-Bernoulli GLMM
   Dataset: Toenail - Molemberghs & Verbeke (2005)
   Date: 04/2008 - Universiteit Hasselt, Diepenbeek, België
*/

/* Fitting a simple Logistic-Bernoulli GLM */
proc genmod data=toenail descending;
  title 'Logistic-Bernoulli GLM';
  class treatn;
  model y = treatn time(treatn)/ dist=binomial noint;
run;

/* Fitting an overdispersed GLM */
proc genmod data=toenail descending;
  class treatn;
  title 'Logistic-Overdispersed Bernoulli GLM';
  title2 'Constant overdispersion parameter';
  model y = treatn time(treatn) /
    dist=binomial noint scale=1.2292;
    *dist=binomial noint scale=Pearson aggregate=(treatn time);
run;

/* Beta-Binomial model with the restriction beta/alpha = const */
proc nlmixed data=toenail qpoints=50;
  title 'Beta-Binomial GLM to model the Overdispersion';
  title3 'Restriction beta/alpha = const';
  parms Beta_0 =1 Beta_1=1 Beta_2=1 Beta_3=1 const=1;
  eta = Beta_0*(treatn=0) + Beta_1*time*(treatn=0)
    + Beta_2*(treatn=1) + Beta_3*time*(treatn=1)
    ;
  expeta = exp(eta);
  ll = -log(1+const) + y*eta - y*log(1+expeta)
    + (1-y)*log((1-expeta)/(1+expeta)) + const);
  model y ~ general(ll);
run;

/* Fitting an simple GLMM */
proc nlmixed data=toenail qpoints=50;
  title 'Logistic-Bernoulli GLMM';
  title2 'Random Effect b1 ~ Normal(0, s^2)';
  parms Beta_0=-0.7 Beta_1=-0.22 Beta_2=-0.7 Beta_3=-0.31 sigma=10;
  eta = Beta_0*(treatn=0) + Beta_1*time*(treatn=0)
    + Beta_2*(treatn=1) + Beta_3*time*(treatn=1)

```

```

        + b1;
expeta = exp(eta);
p = expeta / (1 + expeta);
model y ~ binary(p);
random b1 ~ normal(0,sigma**2) subject = idnew;
run;

/* Combined model with the restriction beta/alpha = const */
proc nlmixed data=toenail qpoints=50;
  title 'Overdispersed-Logistic-Bernoulli GLMM';
  title2 'Random Effect b1 ~ Normal(0,sigma**2)';
  title3 'Restriction beta/alpha = const';
  parms  Beta_0 =-1.54 Beta_1=-6.49 Beta_2=-16.27
         Beta_3=-8.11 const=0.3 sigma=61;
  eta = Beta_0*(treatn=0) + Beta_1*time*(treatn=0)
        + Beta_2*(treatn=1) + Beta_3*time*(treatn=1)
        + b1;
  expeta = exp(eta);
  ll = -log(1+const) + y*eta - y*log(1+expeta)
        + (1-y)*log((1-expeta)/(1+expeta)) + const);
  model y ~ general(ll);
  random b1 ~ normal(0,sigma**2) subject = idnew;
run;

/*****
/*                               END OF PROGRAM                               */
*****/

```

Código SAS para os modelos GLM Probit-Bernoulli, Beta-Binomial, GLMM probit-Bernoulli-Normal e o modelo combinado Probit-Bernoulli-Beta-Normal

```

/* Overdispersed GLMM for Binary Data
   Probit-Beta-Bernoulli GLMM
   Dataset: Toenail - Molemberghs & Verbeke (2005)
   Date: 07/2008 - Universiteit Hasselt, Diepenbeek, België
*/

options nocenter papersize=A4 pagesize=max;

data toenail;
  set d.toenail2;
  timeclass=time;
run;

/* Fitting a simple Probit-Bernoulli GLM */
proc genmod data=toenail descending;
  title 'Probit-Bernoulli GLM';
  class treatn;
  model y = treatn time(treatn) / dist=binomial link=probit point;
run;

/* Fitting an overdispersed GLM */
proc genmod data=toenail descending;
  class treatn;
  title 'Probit-Overdispersed Bernoulli GLM';
  title2 'Constant overdispersion parameter';
  model y = treatn time(treatn) /
          dist=binomial link=probit point scale=1.3518;
          /*dist=binomial link=probit point
          scale=Pearson aggregate=(treatn time);*/
run;

/* Probit-Bernoulli GLM via NLMIXED */
proc nlmixed data=toenail tech=newrap cov corr;
  title 'Probit-Bernoulli GLM';
  parms Beta_0=-0.7 Beta_1=-0.22 Beta_2=-0.7 Beta_3=-0.31;
  if (treatn=0) then eta=Beta_0 + Beta_1*time;
  if (treatn=1) then eta=Beta_2 + Beta_3*time;
  p = probnorm(eta);
  model y ~ binary(p);
run;

/* Fitting an simple GLMM */
proc nlmixed data=toenail qpoints=10;
  title 'Probit-Bernoulli GLMM';
  title2 'Random Effect b1 ~ Normal(0,s^2)';
  parms Beta_0=-0.7 Beta_1=-0.22 Beta_2=-0.7
        Beta_3=-0.31 sigma=10;
  if (treatn=0) then eta=Beta_0 + Beta_1*time + b1;

```

```

    if (treatn=1) then eta=Beta_2 + Beta_3*time + b1;
    p = probnorm(eta);
    model y ~ binary(p);
    random b1 ~ normal(0,sigma**2) subject = idnew;
    run;

/* Probit-Bernoulli-Beta model
   Constraint: alpha + beta=10 */
proc nlmixed data=toenail cov corr technique=quanew
            maxiter=1000 maxfunc=1000;
    title 'Probit-Bernoulli-Beta Model';
    parms   Beta_0=8 Beta_1=-0.9
            Beta_2=2  Beta_3=-0.5
            alpha=10;
    bounds Beta_3 < 0, Beta_1 < 0;
    const=10;
    if (treatn=0) then eta=Beta_0 + Beta_1*time;
    if (treatn=1) then eta=Beta_2 + Beta_3*time;
    probit_eta = probnorm(eta);
    ll = y*log(probit_eta) + y*log(alpha)
        + (1-y)*log(const-probit_eta*alpha);
    model y ~ general(ll);
    run;

/* Probit-Bernoulli-Beta model
   Keeping alpha + beta=10 and estimating alpha, Beta_1 and Beta_3 */
proc nlmixed data=toenail cov corr technique=newwrap
            maxiter=1000 maxfunc=1000;
    title 'Probit-Bernoulli-Beta Model';
    parms Beta_1=-0.9 Beta_3=-0.5 alpha=3.2;
    bounds Beta_3 < 0, Beta_1 < 0, alpha < 50;
    const=10;
    if (treatn=0) then eta=Beta_1*time;
    if (treatn=1) then eta=Beta_3*time;
    probit_eta = probnorm(eta);
    ll = y*log(probit_eta) + y*log(alpha)
        + (1-y)*log(const-probit_eta*alpha);
    model y ~ general(ll);
    run;

/* Combined model with the restriction alpha + beta = const */
proc nlmixed data=toenail qpoints=10 cov corr
            technique=quanew maxiter=1000 maxfunc=1000;
    title 'Probit-Bernoulli-Beta-Normal Model';
    title2 'Random Effect b1 ~ Normal(0,sigma**2)';
    title3 'Restriction alpha + beta = const';
    parms   Beta_0=0 Beta_1=-0.9
            Beta_2=0 Beta_3=-0.8
            alpha=3.6 sigma=10;
    const=10;
    if (treatn=0) then eta=Beta_0 + Beta_1*time + b1;
    if (treatn=1) then eta=Beta_2 + Beta_3*time + b1;
    probit_eta = probnorm(eta);
    ll = y*log(probit_eta) + y*log(alpha)

```

```

        + (1-y)*log(const-probit_eta*alpha);
model y ~ general(ll);
random b1 ~ normal(0,sigma**2) subject = idnew;
run;

/* Combined model with the restriction alpha + beta = const
   The aproximation to the logistic model was applied, through
   the multiplication of eta by a constant */
proc nlmixed data=toenail qpoints=10 cov corr
        technique=quanew maxiter=1000 maxfunc=1000;
title 'Probit-Bernoulli-Beta-Normal Model with Logit Approximation';
title2 'Random Effect b1 ~ Normal(0,sigma**2)';
title3 'Retriction alpha + beta = const';
parms  Beta_0=-1 Beta_1=-1
        Beta_2=-1 Beta_3=-1
        alpha=7.5 sigma=8;
const=10;
if (treatn=0) then eta=Beta_0 + Beta_1*time + b1;
if (treatn=1) then eta=Beta_2 + Beta_3*time + b1;
ApproxConst = (16/15)*(sqrt(3)/3.141593);
probit_eta = probnorm(ApproxConst*eta);
ll = y*log(probit_eta) + y*log(alpha)
      + (1-y)*log(const-probit_eta*alpha);
model y ~ general(ll);
random b1 ~ normal(0,sigma**2) subject = idnew;
contrast 'Beta1 - Beta3' Beta_1 - Beta_3;
contrast 'Beta1/Beta3' Beta_1/Beta_3;
contrast 'beta' const-alpha;
run;

/*****
/*                               END OF PROGRAM                               */
/*****/

```