**MODELOS LINEARES GENERALIZADOS COM
SUPERDISPERSÃO PARA MORTALIDADE
DE ÁRVORES DE *Eucalyptus grandis***

MARÍA DEL PILAR DÍAZ
Lic. em Matemáticas

Orientador: Prof. Dr. HILTON THADEU ZARATE DO COUTO

Tese apresentada à Escola Superior de
Agricultura "Luiz de Queiroz", da Uni-
versidade de São Paulo, para obtenção
do título de Doutor em Agronomia, Área
de Concentração: Estatística e Experi-
mentação Agronômica.

P I R A C I C A B A
Estado de São Paulo - Brasil
Julho - 1996

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - Campus "Luiz de Queiroz"/USP

Díaz, Maria del Pilar

Modelos lineares generalizados com superdispersão para mortalidade de árvores de *Eucalyptus grandis* / Maria del Pilar Díaz. - - Piracicaba, 1996. 148p. : il.

Tese (doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 1996.

Bibliografia.

1. Eucalipto - Mortalidade - Processo estocástico 2. Método da verosimilhança restrita 3. Modelo linear generalizado I. Título

CDD 519.5

**MODELOS LINEARES GENERALIZADOS COM
SUPERDISPERSÃO PARA MORTALIDADE
DE ÁRVORES DE *Eucalyptus grandis***

MARÍA DEL PILAR DÍAZ

Aprovado em: 03/09/96

Comissão julgadora:

Prof. Dr. Hilton Thadeu Zarate do Couto	ESALQ/USP
Prof ^{ma} . Dr ^a . Clarice Garcia Borges Demétrio	ESALQ/USP
Prof. Dr. Sérgio do Nascimento Kronka	FCAV/UNESP
Prof ^{ma} . Dr ^a . Sheila Zambelo de Pinho	FCA/UNESP
Prof. Dr. José Raúl Matínez	FAMAF/UNC

Prof. Dr. Hilton Thadeu Zarate do Couto
Orientador

AGRADECIMENTOS

Ao Prof. Dr. Hilton Thadeu Zarate do Couto pela orientação, estímulo e confiança.

Ao Departamento de Matemática e Estatística da ESALQ/USP, e à Faculdade de Ciências Agropecuárias da Universidade Nacional de Córdoba pela oportunidade que me deram para a realização deste curso.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio econômico e incentivo permanente.

Ao Prof. Dr. José Eduardo Corrente, do Departamento de Matemática e Estatística da ESALQ/USP pela a gentil colaboração para a implementação de algumas das rotinas utilizadas neste trabalho.

Aos colegas do curso, Roseli, Liciania, João Gil, Doris, Imaculada, Silvio, e Claudia pela amizade e sincero companheirismo.

À Empresa Champion de Papel e Celulose Ltda. por ter cedido os dados experimentais usados neste trabalho.

A todos aqueles que, sem dúvida, colaboraram incondicionalmente nesta etapa da minha vida.

SUMÁRIO

1	INTRODUÇÃO	1
2	PROBLEMA	5
2.1	Material	5
2.2	Método	6
2.2.1	Comentários Gerais	6
2.2.2	Modelo Estatístico	8
2.2.3	Modelo Matemático	11
2.2.4	Inferência e Verificação do Ajuste do Modelo	14
3	EXPLORAÇÃO DO MODELO ESTOCÁSTICO	20
3.1	Modelo Poisson	23
3.2	Teste Quociente Média-Variância	29
3.3	Superdispersão	30
3.4	Teste para Detectar Superdispersão no Modelo Poisson	36
4	MODELO EXTRA-POISSON DE BRESLOW	40
4.1	Introdução	40
4.2	Quase-Verossimilhança	42
4.2.1	Método de Estimação	43
4.3	Equações Generalizadas de Estimação	58
4.4	Enfoque Bayesiano	60
5	MODELO BINOMIAL NEGATIVO	66
5.1	Descrição do Modelo	66
5.2	Estimação por Máxima Verossimilhança	72
5.3	Metodologias Alternativas	92
5.3.1	Iterando a variância	92
5.3.2	Enfoque Intuitivo	101
5.4	Eficiência na Estimação	102
5.4.1	Alguns Comentários Gerais	102
5.4.2	Caso Binomial Negativo	103

6	UM SUPORTE TEÓRICO MAIOR	107
6.1	Introdução	107
6.2	Definições	108
6.2.1	Modelo Exponencial com Dispersão	108
6.2.2	Convolução	110
6.2.3	Modelos com Dispersão (MD)	112
6.3	A Função de Variância	112
6.4	Inferência	113
6.4.1	A "Deviance"	113
6.4.2	A Matriz de Informação de Fisher	115
6.4.3	Parâmetro de dispersão	117
6.4.4	Resultados teóricos para amostras grandes	120
6.5	Aplicação do Suporte Teórico Maior	122
6.5.1	O Modelo	122
6.5.2	O Algoritmo	123
7	OUTROS TÓPICOS	125
7.1	Uso Experimental: Tamanho de Amostra	125
7.2	Uso Biológico: Processo Estocástico	126
8	CONCLUSÕES	130
9	REFERÊNCIAS BIBLIOGRÁFICAS	132
10	APÊNDICES	138
10.1	Apêndice A	138
10.2	Apêndice B	143

LISTA DAS FIGURAS

1	Dispersão da mortalidade versus a idade das árvores de <i>E. grandis</i> por tratamento.	139
2	Dispersão da $V_{ij}(Z)$ versus $\bar{Z}_{ij}$, onde (k, l) representa à rotação k e região l , para $k, l = 1, 2, 3$	139
3	Dispersão do $\ln(\bar{Z}_{ij} - V_{ij}(Z)/n_{ij})$ versus $\ln(V_{ij}(Z) - \bar{Z}_{ij})$, onde (k, l) representa à rotação k e região l , para $k, l = 1, 2, 3$	140
4	Distribuição empírica e estimada, segundo o modelo Poisson, da mortalidade das árvores de <i>E. grandis</i>	140
5	Valores médios da mortalidade das árvores de <i>E. grandis</i> para as combinações de região e rotação.	141
6	Dispersão dos $\hat{z}_{ijk}$ versus $\hat{\eta}_{ijk}$ para a mortalidade das árvores de <i>E. grandis</i>	141
7	Valores estimados dos quocientes [”deviance” residual]/[graus de liberdade] versus θ , onde θ é tal que $V(Z) = \mu + \frac{1}{\theta}\mu$	142
8	Valores ótimos de tamanhos de amostras sugeridos pelos modelos Extra-Poisson e Binomial Negativo, para a obtenção de diferentes precisões na estimação da média da mortalidade de <i>E. grandis</i>	142
9	Curvas hipotéticas de sobrevivência segundo Raymond Pearl, em 1928.	143

LISTA DAS TABELAS

1	Valores médios e desvios padrões (d.p.) de mortalidade (%) das árvores de <i>Eucalyptus grandis</i> para os tratamentos	22
2	Valores médios e desvios padrões (d.p.) da idade (anos) das árvores de <i>Eucalyptus grandis</i> para os tratamentos.	26
3	Valores das estatísticas para exploração do modelo probabilístico da mortalidade das árvores de <i>E. grandis</i>	38
4	Análise da "deviance" para o modelo Extra-Poisson para a variável mortalidade das árvores de <i>Eucalyptus grandis</i> , no GLIM e no SPlus.	52
5	Estimativas dos parâmetros, e seus respectivos erros padrões e estatísticas de Wald, sob enfoque Extra-Poisson.	53
6	Estimativas dos parâmetros, e seus respectivos erros padrões e estatísticas de Wald, sob modelo Poisson (sem extra-variação).	56
7	Resultados para o ajuste máxima verossimilhança do modelo Binomial Negativo através das macros de Venables & Ripley (1994).	77
8	Valores do teste quociente de verossimilhança (L.R.) e do λ para verificar hipóteses sob modelos encaixados, usando as macros de Venables & Ripley (1994).	78
9	Valores estimados para os parâmetros do modelo Binomial Negativo através das macros de Venables & Ripley (1994).	79
10	Resultados para o ajuste de máxima verossimilhança do modelo Binomial Negativo através das macros de Hilbe (1994).	80
11	Valores estimados para os parâmetros do modelo de locação após ajuste máxima verossimilhança do modelo Binomial Negativo através das macros de Hilbe (1995).	81
12	Resultados para o ajuste máxima verossimilhança do modelo Binomial Negativo através das macros de Lawless (1987).	82
13	Valores estimados para os parâmetros do modelo de locação após ajuste de máxima verossimilhança através das macros de Lawless (1987).	83
14	Resultados para o ajuste de máxima verossimilhança do modelo Binomial Negativo para as funções de ligação canônica e raiz quadrada, através macros de Hinde (1996).	85

15	Valores do teste quociente de verossimilhança (L.R.) e do λ para verificar hipóteses sob modelos encaixados, usando as macros de Hinde (1996).	87
16	Valores estimados para os parâmetros do modelo de locação após ajuste máxima verossimilhança através das macros de Hinde (1996), usadas para duas funções de ligação.	88
17	Resultados para o ajuste máxima verossimilhança do modelo Binomial Negativo para observações positivas, através das macros de Hinde (1996).	90
18	Valores estimados para os parâmetros do modelo de locação somente para observações postivas através das macros de Hinde (1996).	91
19	Tamanhos ótimos de amostras segundo os enfoques Extra-Poisson (EP) e Binomial Negativo (BN).	126

MODELOS LINEARES GENERALIZADOS COM SUPERDISPERSÃO PARA MORTALIDADE DE ÁRVORES DE *Eucalyptus grandis*

Autora: María del Pilar Díaz

Orientador: Prof. Dr. Hilton Thadeu Zarate do Couto

RESUMO

Na área florestal, a estimativa da mortalidade das árvores de uma determinada espécie deve estar presente nos sistemas confiáveis de predição do crescimento da espécie. Porém, o estudo da mortalidade de indivíduos envolve um processo de classificação dicotômica, originando dados de contagem nas unidades experimentais. Assim, é fundamental, primeiramente, a identificação da distribuição das observações para o estudo do comportamento, estabelecendo um modelo que incorpore informação relacionada com uma adequada estratificação do material, como por exemplo, seleção da espécie, rotação, idade, espaçamento e local, entre os mais importantes.

No caso das observações serem independentes, um modelo básico para a mortalidade poderia ser o modelo Poisson. Porém, na prática é freqüente observar que a variância amostral excede á variância nominal, o que é chamado de superdispersão. Neste trabalho é identificado, através de técnicas formais e informais, o modelo probabilístico da mortalidade das árvores de *Eucalyptus grandis*.

O gênero *Eucalyptus* foi escolhido dada sua grande importância produtiva no Brasil. Dentro dele, a espécie *Eucalyptus grandis* mostra características silviculturais desejáveis, tais como a adaptação ambiental, rápido crescimento e boa forma, o que a torna a espécie mais utilizada para fins de produção de papel e celulose.

Sob o enfoque dos modelos lineares generalizados, estabeleceu-se um componente sistemático para o estudo dos efeitos dos fatores rotação da espécie e qualidade de sítio. O objetivo foi analisar duas metodologias para um modelo Poisson com superdispersão, isto é, o modelo Extra-Poisson, através de quase-verossimilhança, e o modelo Binomial Negativo, através de máxima verossimilhança. O parâmetro de dispersão para a mortalidade foi estimado em 2.18, indicando um

índice de agregação altamente significativo.

Dada complexidade do algoritmo para ajuste do modelo através do segundo enfoque, foram propostas e testadas metodologias alternativas visando uma simples e rápida implementação. As performances desses modelos, em termos da eficiência das estimativas foram estudadas para fins experimentais, indicando como a adequada à de máxima verossimilhança. Um modelo exponencial com dispersão discreto é apresentado como um suporte teórico maior, generalizando os modelos anteriores e respeitando a verdadeira estrutura probabilística dos dados. A partir disso, foi estabelecido um processo estocástico para a mortalidade das árvores, envolvendo, na avaliação do mesmo, as mudanças da rotação ao longo do tempo.

GENERALISED LINEAR MODELS WITH OVERDISPERSION FOR *Eucalyptus grandis* TREE MORTALITY

Author: María del Pilar Díaz

Adviser: Prof. Dr. Hilton Thadeu Zarate do Couto

SUMMARY

In the forestry area, a tree mortality estimation for a given species must be present in the reliable prediction systems for the species growth. However, the individual mortality study introduces a dichotomic process, arising to count data set for the experimental units. Thus, the identification of data distribution is the previous and important work to know the mortality behavior related to additional information, as the species selection, rotation, age, and site quality.

For independent observations mortality, the basic Poisson model could be proposed, nevertheless it is frequently to find that the sample variance is greater than the nominal one, which is called overdispersion. The present work identifies the probabilistic model of *Eucalyptus grandis* through formal and informal technics.

As the genus *Eucalyptus* is very important in the forestry brazilian production, it was chosen. The *E. grandis*, belonging to the genus, has desirable properties, like the enviroment adaptation, fast growth, so that it becomes one of the most useful species for the paper and cellulose national production.

Under generalized linear model approach, a sistematic component with two factors, rotation species and site quality. The aim was to study two methodologies for the overdispersion Poisson models, that is, the Extra-Poisson model, through maximum quase-likelihood, and the Negative Binomial model, through maximum likelihood method. The dispersion parameter was estimated by 2.18, which means that the mortality distribution has a hight aggregated index. Due to the fact that the maximum likelihood algorithm is complex, simple alternative methods were suggested. Their performances related to the efficiency were studied too. The results indicated that the proposed maximum likelihood method might be used. Finally, a discret exponential model was shown like a greater theoretical

support to generalize the model keeping the true probabilistic data structure. This generalization establishes the background to formulate a stochastic process for the tree mortality following the changes as the time goes by.

1 INTRODUÇÃO

Na Ciência Florestal e dentro da área de manejo para fins de produção de madeira e definição dos regimes de manejo, o estudo da combinação **rendimento-densidade** tornou-se o objetivo de intensas pesquisas durante as últimas décadas.

Em geral, os diversos enfoques que são abordados dessa ótica, dependem precisamente da natureza das características aleatórias consideradas embora, comumente a escolha seja feita sob as de comportamento contínuo (o que se relaciona com o processo de crescimento). As variáveis mais utilizadas são o volume cilíndrico ou sólido, área basal, diâmetro, etc. em função de alguns fatores controlados, como a qualidade do sítio (índice de sítio), espécie e procedência, idade do povoamento florestal, rotação, grau de intervenção (desbaste). Pouco se conhece do tratamento que deveria ser dado à outra variável, que diretamente está contribuindo na modelagem geral para a produção de uma floresta, em outras palavras, como se comportaria a mortalidade das árvores, por exemplo, em um período de tempo ou espaço determinado.

Os motivos desse desconhecimento no estudo dos processos mencionados não se relacionam com a parte metodológica da escolha da informação, visto que qualquer tipo de manejo florestal envolve a elaboração de inventários contínuos os quais contém as observações mencionadas. O problema está, então, centralizado na escolha da análise adequada, começando pela modelagem até as técnicas de inferências, sobre a definição da influência de fatores ambientais e antrópicos no povoamento.

Por exemplo, a densidade limite poderia ser reconstruída por diferentes métodos, dos quais o principal descreve-a através da sua distribuição mais natural, como o é a distribuição binomial; ou por outro lado, estudando a mortalidade como uma variável complementar e cuja distribuição aparece como pouco atraente para as pesquisas atuais. Essas definições poderiam fornecer um novo nível de consistência e precisão na descrição dos "stands" florestais. Conhecer a densidade ótima de um povoamento (o qual sem uma estimativa confiável do número de árvores

mortas é impossível) conduz à estimação apropriada da produção de biomassa.

Por outro lado, o estudo da mortalidade de indivíduos envolve um processo de classificação dicotômica, originando dados de contagem nas unidades experimentais nos ensaios da área de manejo florestal. Assim a identificação da distribuição das observações é fundamental para o desenvolvimento de um plano de amostragem confiável, e em consequência, a obtenção de estimações de parâmetros de fatores de interesse (Rogers, 1974).

A Estatística, como ciência de construção do conhecimento, deve fornecer modelos confiáveis e as ferramentas adequadas para sua validação em uma tentativa de descrever essas situações reais. Uma área de estudo muito importante no que diz respeito à modelação, principalmente nas duas últimas décadas, é a que compete aos modelos lineares generalizados (MLG). No enfoque desses modelos, a modelagem estabelece algumas **restrições** as quais caracterizam a distribuição assumida para os dados. Mas, um fenómeno que ocorre com frequência nos dados provenientes de contagem é a **superdispersão**, isto é, quando a $Var(Y_i)$, (variância empírica ou amostral) excede a variância nominal, $v(\mu_i)$ (variância esperada conforme um modelo probabilístico estabelecido). McCullagh & Nelder (1989) observaram que a superdispersão está quase sempre presente nos dados reais.

Uma das possíveis causas da superdispersão é basicamente experimental, por exemplo quando as observações não são independentes devido ao fato de que o delineamento amostral é **agrupado**. A omissão da identificação dos agrupamentos, ou a falha em modelar a correlação entre as respostas no mesmo grupo produz distribuições misturadas dentro das unidades experimentais, como por exemplo Aitkin *et al.* (1989) citam para o caso do modelo Binomial em tabelas de contingência.

Se o modelo proposto não é o apropriado, pode acontecer que os dados estejam superdispersos e, neste caso, a "deviance" (medida da variação utilizada nos MLG) supera o valor dos graus de liberdade associados, pelo qual um estudo prévio da estrutura de variação deve ser feito a fim de especificá-la no ajuste do modelo real.

Dentre as técnicas utilizadas para exploração da estrutura média-variância, estão, por exemplo, os gráficos de algumas estatísticas (ou funções) de posição e dispersão, o teste do quociente média-variância de Elliot (1979), ou o teste para verificar alguma relação pré-estabelecida entre a média e a variância, como o teste de Dean & Lawless (1989) para detectar superdispersão nos modelos Poisson, modelo básico usado para o ajuste para dados de contagem.

Se o fenômeno de superdispersão está presente nos dados, dois enfoques de análise são comuns na literatura. O primeiro deles, apresentado por Breslow (1984), e denotado por QV/M (na sequência, Quase-Verossimilhança-Momentos), é usar somente a estrutura da média-variância estabelecida, estimando os β 's pelo método de quase-verossimilhança e o parâmetro de dispersão pelo método dos momentos. O segundo enfoque, e menos usado é tratar as médias Poisson associadas com cada número (contagem) observado como sendo variáveis amostradas a partir de uma distribuição paramétrica determinada. A maioria dos autores tem considerado a distribuição Gama para o valor esperado, o que conduz a uma distribuição Binomial Negativa para as observações. Daí, as estimações para os parâmetros contidos no preditor linear e parâmetro de dispersão são obtidas por máxima verossimilhança.

O gênero *Eucalyptus* tem uma grande importância produtiva no Brasil. Dentro dele, a espécie *Eucalyptus grandis* mostra características silviculturais desejáveis, tais como a adaptação ambiental, rápido crescimento e boa forma. As plantações dessa espécie se distribuem praticamente por todas as regiões do país, sendo mais intensamente na região sudeste. Dentre as diversas finalidades do uso se destaca a fabricação de papel e celulose, painéis a base de madeira e carvão.

Visando garantir a produção requerida ao nível interno e externo (a qual requer um crescimento em torno de 2,5% ao ano até o fim do século), o setor de papel e celulose do Brasil vem expandindo sua capacidade desde a década passada, como também estudando a relação rendimento-densidade, para a qual a variável mortalidade merece previamente uma atenção destacável. Segundo Somers *et al.* (1980) o estudo da combinação rendimento-densidade, como processo, pode ser dividido em componentes interrelacionadas, das quais o crescimento real (sobrevivência) e a mortalidade são os mais importantes dada que a informação do crescimento final constitui a parte essencial da área do manejo florestal.

A predição da mortalidade deve estar presente nos sistemas confiáveis de predição do crescimento da espécie, isto é, o conhecimento da densidade ótima de um talhão conduz à estimações de produção de biomassa da população. Além disso, os sistemas atuais de inventário florestal estabelecem como necessária uma adequada estratificação do material (seleção da espécie, rotação, procedência, idade, produtividade, espaçamento e local) para a obtenção de regras gerais de manejo.

Para este trabalho foram propostos os seguintes objetivos:

1. Analisar a metodologia de análise e a performance dos modelos Extra-Poisson, através de quase-verossimilhança, e Binomial Negativo, através de máxima

verossimilhança, para o estudo dos efeitos dos fatores **rotação** e **local** na mortalidade das árvores de *Eucalyptus grandis*.

2. Propor metodologias alternativas para a análise do modelo linear generalizado, visando uma simples e rápida implementação, em função das dificuldades apresentadas pelos algoritmos fornecidos na literatura.
3. Apresentar um suporte teórico maior para esse modelo, generalizando os anteriores a fim de respeitar a verdadeira estrutura probabilística dos dados.
4. Estabelecer um processo estocástico para a mortalidade das árvores, envolvendo, na avaliação do mesmo, as mudanças ao longo do tempo, dos fatores estudados.

2 PROBLEMA

2.1 Material

Os dados de inventário florestal foram fornecidos pela Empresa Champion de Papel e Celulose Ltda., localizada no município de Mogi Guaçu (São Paulo-Brasil) e correspondem à espécie *Eucalyptus grandis*, classificados em três rotações e três locais ou qualidades de sítios.

A escolha deste gênero foi centrada na sua importância a nível nacional. O gênero *Eucalyptus* é amplamente utilizado para a implantação de maciços florestais pela maioria das empresas brasileiras dedicadas à produção de papel e celulose. A espécie *Eucalyptus grandis* mostra características silviculturais desejáveis, como adaptação ambiental, rápido crescimento e boa forma. Os talhões dessa espécie distribuem-se praticamente por todas as regiões do Brasil, sendo mais representativos na região sudeste. Os principais usos dela são a fabricação de papel e celulose, painéis de madeira e carvão vegetal.

A distribuição geográfica da produção de celulose e papel teve sempre no Estado de São Paulo o seu principal produtor, participando aproximadamente com 47% da produção nacional, da qual 83% do gênero *Eucalyptus*, e dentro dele, 43% para a espécie *E. grandis*.

Visando atingir os crescimentos da demanda interna e externa (esta última necessitando de um incremento 2,5% ao ano até fim do século), o setor de papel e celulose vem expandindo sua capacidade produtiva a partir da década passada. No ano 1991 foram implantados e reflorestados 75.000 hectares deste gênero, nas quais o investimento esteve concentrado na espécie *E. grandis*.

No setor de planejamento da empresa, as informações são obtidas através de sistemas de inventário florestal contínuo com parcelas permanentes e os dados para este estudo provieram de parcelas instaladas em três regiões geográficas do Estado de São Paulo. O referido sistema de inventário estabelece primeiramente uma estratificação do material, ou seja, agrupamentos dos maciços em estratos que podem ser considerados como unidades uniformes para fins de manejo e produção.

Os principais critérios considerados no processo citado foram: espécie, rotação, origem e/ou procedência, idade, produtividade, espaçamento e lotação (densidade populacional). As variações do solo são identificadas geralmente *a posteriori* à pré-amostragem. O local foi escolhido de forma a refletir a diversidade ambiental (solo-clima), isto é, três regiões das quais duas possuem condições edafo-climáticas, em princípio, semelhantes, e a outra, em condições extremas. Geograficamente, essas regiões formam um triângulo onde o vértice superior orientado para o norte corresponde à Região 3 e os outros dois, aproximadamente em uma linha reta leste-oeste, às Regiões 1 e 2.

As características edafo-climáticas e de localização das três regiões foram:

Região 1: Municípios de Mogi-Guaçu, Aguaí, Casa Branca e Espírito Santo de Pinhal. Estado de São Paulo. As precipitações durante o ano são regulares, com uma média de 1.295 mm e temperatura média de 20,85°C;

Região 2: Municípios de Brotas e Riberão Bonito. Estado de São Paulo. As precipitações durante o ano são regulares, com uma média de 1.511 mm e temperatura média de 21,04°C;

Região 3: Municípios de Altinópolis, São Simão, Santa Rita do Passa Quatro e Santa Rosa do Viterbo. Estado de São Paulo. As precipitações durante o ano são concentradas no verão, com uma média de 1.613 mm e temperatura média de 22,59°C.

Para a determinação da área total relacionada ao experimento e às técnicas de amostragem seguiram-se as normas de inventário florestal adotadas pela Empresa. A localização das parcelas foi feita em forma aleatória. Segundo estudos realizados, foi determinado que o uso de parcelas de tamanho 20,0 x 20,0 metros (400 m²) era o apropriado para as condições experimentais. O total de amostras foi obtido mediante diferentes intensidades em função do fator rotação, sendo a relação "área ocupada-área amostrada" de 0,32% para a primeira rotação e 0,64% para as restantes. Em cada parcela foi contado o número de árvores mortas e a idade média das árvores.

2.2 Método

2.2.1 Comentários Gerais

Segundo Jorgensen (1989), para um problema particular, o objetivo da análise estatística é encontrar um modelo estatístico adequado, testar a aderência

do modelo aos dados, estimar os parâmetros desconhecidos contidos nele, e testar hipóteses de interesse acerca desses parâmetros. Sob esse enunciado clássico motiva-se o que segue.

Para avaliar mudanças entre o número de indivíduos e seus tamanhos, existe uma lei muito conhecida, e por sinal muito criticada, a qual gerou, a partir de sua origem em 1963, o crescimento do conhecimento de mecanismos de importantes espécies arbóreas. Esta Lei da Potência $3/2$ de Yoda *et al.* (1963), relaciona o tamanho da árvore com a densidade populacional quando o processo de auto-desbaste está ocorrendo, tal que a população diminui tanto nos incrementos como no número de indivíduos, em função do tempo.

O desempenho desta relação tem recebido inumeráveis críticas devido às bases empíricas e teóricas, salientando-se, basicamente, a dependência das constantes envolvidas em função de alguns fatores biológicos, muitas vezes não controláveis.

As deficiências da lei não significam que as duas variáveis não estejam relacionadas, apenas que é impossível descrever tal relação com a ajuda de somente dois coeficientes, um dos quais se apresenta como "constante universal". Com isto, os caminhos de análise poderiam ser alterados, ou enfocados, de maneira diferente à usual seguida na literatura. Esta alteração divide o problema real considerando a natureza das duas variáveis, agora vistas como aleatórias e portanto tratáveis como tais no que diz respeito à sua modelagem e técnicas de inferências.

Visto que a informação do crescimento constitui uma parte essencial na área de manejo florestal, esse processo pode ser dividido em vários componentes inter-relacionados, dos quais o crescimento real (sobrevivência) e a mortalidade são os mais importantes. Mas, os processos de decisão no manejo florestal dependem de estimativas confiáveis do crescimento e produção do povoamento. Os valores dependem de características do indivíduo (a árvore), ou seja do volume real, da distribuição dos tamanhos e do número de sobreviventes. Um dos componentes mais importantes de um sistema de predição do crescimento e produção é a mortalidade.

Conhecida a história de um povoamento, é imperativo tentar modelar com precisão o processo de crescimento e da mortalidade e, conseqüentemente, prever o número de árvores que sobreviveram. O crescimento sempre foi o mais estudado, mediante o estudo das distribuições amostrais das variáveis que o refletem, ou dentro do sistema ecológico, mediante a modelagem dos valores médios, em função de alguns fatores controlados e de importância para a área de manejo florestal, como solo, clima, idade, espécie, etc.. Em ambos os casos, os métodos usuais de análise

estatística aplicam-se facilmente. Isto não acontece com o segundo processo, isto é, o número de plantas mortas, o qual é causado pelo incremento do tamanho da árvore do talhão, e devido à ação de fatores naturais que não estão sincronizados.

Assim, a predição da mortalidade é de essencial importância em qualquer sistema de predição do crescimento de um povoamento florestal. Infelizmente, essa característica é extremamente variável, o que dificulta o processo em si. Por outro lado, muitas técnicas que modelam o crescimento incorporam algum método para estimar a mortalidade de uma maneira explícita, tratando-a como modelos de processos estocásticos. Quando as unidades de avaliação geram espaços amostrais discretos, ou indivíduo vivo ou morto, esse processo torna-se de classificação dicotômica.

Por último, para a realização de uma boa modelagem também é necessário incluir informações que se refiram às condições do ecossistema em desenvolvimento, como por exemplo, diferentes qualidades de sítio.

2.2.2 Modelo Estatístico

Para certos tipos de dados existem modelos naturais ou padrões, os quais são usados, no geral, em uma primeira análise geral. Por exemplo, dados de contagens podem ser modelados como Binomiais (onde, entre outras coisas, o número de ensaios é assumido conhecido), ou como Poisson (se as contagens são feitas no tempo ou no espaço, prefixados de antemão). Assim, para observações provenientes de contagens há necessidade de utilizar modelos (e com isso significa estudar toda a estrutura da modelagem matemática-aleatória) que respeitem essa natureza original, como é o caso dos **Modelos Lineares Generalizados**.

Em termos gerais, um modelo linear generalizado (MLG) possui três componentes:

- a distribuição da variável resposta, ou componente aleatório ou estocástico, isto é, Z_i ($i = 1, \dots, n$) são variáveis aleatórias, com média μ_i e valores observados $E(z_i) = \mu_i$. Essas variáveis aleatórias constituem o *modelo probabilístico* para os dados;
- o preditor linear, isto é, A é um conjunto de p parâmetros conhecidos,

$$\beta = (\beta_1, \dots, \beta_p)^T$$

correspondente ao conjunto de variáveis explanatórias conhecidas $X_{n \times p} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$, e a matriz de delineamento ou matriz do modelo, definem o

modelo matemático:

$$\eta_i = \sum_{l=1}^n x_{il}\beta_l \quad (1)$$

ou, em forma matricial,

$$\mathbf{H} = \mathbf{X}\beta$$

Isto indica de que maneira a variável resposta muda de locação com respeito às variáveis explanatórias. Esse modelo matemático para os dados é o *componente sistemático*;

- a função de ligação, que faz a relação entre a média da i -ésima resposta e seu preditor linear. Ela é dada por funções monótonas e diferenciáveis sendo, pelo geral, a mesma para todas as observações.

$$\begin{aligned} \eta_i &= g_i(\mu_i) \\ &= \mathbf{x}_i^T \beta \end{aligned}$$

Dentro da especificação de um componente estocástico pertencente à família exponencial na forma canônica¹, a função de ligação que resulta a partir do parâmetro canônico θ conduz a importantes propriedades inferenciais, dado que todos os parâmetros desconhecidos da estrutura sistemática linear possuem estatísticas suficientes, caso o parâmetro de escala seja conhecido.

Combinando o modelo probabilístico com o modelo de locação obtém-se um modelo estatístico completo para descrever um conjunto determinado de observações. Nessa combinação está embutida a escolha de uma função de ligação apropriada para os dados. "Ela" depende de como são selecionados os dois componentes mencionados, principalmente o componente aleatório, e "dela" depende, um desses componentes, o componente sistemático na etapa de estimação. Isto será abordado em outra seção mas salienta-se que, no início, a função de ligação a propor para o modelo estatístico será a canônica.

Além disso, lembra-se que o modelo matemático é aquele estabelecido pela situação experimental incluindo o delineamento experimental (e não apenas o delineamento), isto é, fatores de tratamentos e/ou covariáveis que foram avaliadas e das quais se dispõe de informação. Como será apresentado na seção (3.3), essa não é a única maneira de abordar a estimação de um modelo linear generalizado.

¹O leitor pode consultar publicações básicas neste tópico, como por exemplo, Nelder & Wedderburn (1972); Cordeiro (1986), entre outros.

Mais precisamente, existem propostas de mudanças chaves para o componente sistemático, as quais não serão caso de estudo intensivo neste trabalho.

Por outro lado, e na estrutura dos modelos lineares generalizados, é dever fornecer uma forma para o componente aleatório do modelo, tal que permita pesquisar o componente sistemático livremente e com precisão. A escolha desse modelo, e em particular o componente aleatório, é a tarefa mais difícil na análise dos dados, descrita por Jorgensen (1989) como uma arte onde devem combinar o "background" teórico, a experiência e o erro próprio do tipo de ensaio que se está analisando.

O **método**, portanto, determina a conjunção de várias etapas de discussão e decisão, as quais envolvem, em uma segunda instância, todas as alternativas de análises em termos da estimação do modelo escolhido, as quais fornecerão os primeiros passos para a realização das inferências de interesse. A discussão acerca dos principais mecanismos envolvidos no sistema (biológico) em estudo é de vital importância, uma vez que dela podem ser extraídos novos caminhos para a construção da metodologia da análise em si, como assim descartadas algumas técnicas por não satisfazerem seus suportes teóricos. Salienta-se, portanto, que esse **método** não é definido em um sentido estático, pelo contrário, ele representa o "feedback" que todo pesquisador deveria considerar antes de tirar suas conclusões.

Os métodos que este material apresenta baseiam-se, principalmente, na teoria dos modelos lineares generalizados de McCullagh & Nelder (1989). Isto conduz à descrição de alguns modelos típicos encontrados na literatura. O estudo dos mesmos não significa, porém, apenas a apresentação dos algoritmos de estimação, senão também a construção de simples propostas que sirvam como alternativas possíveis para sua implementação. Algumas das seções deste trabalho abordam esses tópicos numa tentativa para contribuir para o uso desses métodos mostrados pela literatura como os mais adequados (em termos de obtenção de inferências confiáveis), já que alguns deles não são de compreensão, nem de uso simples, mesmo no caso de se dispor de recursos computacionais apropriados.

Os modelos lineares generalizados podem ser estimados usando como algoritmo uma generalização dos mínimos quadrados, os chamados mínimos quadrados ponderados iterativos, originado pelas equações de estimação do método da máxima verossimilhança. São bem conhecidas e demonstradas as boas propriedades que esse método possui, representando assim uma grande vantagem para a construção dos testes inferenciais. Além disso, os MLG podem ser estimados usando uma técnica **robusta**, como é o método da máxima quase-verossimilhança, a

qual utiliza apenas informação dos dois primeiros momentos da possível distribuição populacional (desconhecida) das observações. Essas, e outras metodologias de estimação, como as provenientes das equações generalizadas de estimação, da técnica de perfis, ou do enfoque de um modelo dois estágios ("two-stage model") para as observações, são abordadas neste material.

Às vezes, as dificuldades que apresenta a escolha do modelo propriamente dito, e sua correspondente estimação para dados de contagem, fazem com que se utilizem modelos não-paramétricos ou modelos semiparamétricos, os quais envolvem poucas pressuposições evitando a construção de pressuposições infundáveis nas conclusões. Não obstante, dado o grande universo de modelos a selecionar, com conhecimento da estrutura física do modelo para os dados, e um bom diagnóstico do modelo através de novas técnicas, as informações extraídas com um modelo paramétrico serão mais amplas do que com um a análise semi, ou não, paramétrica.

Neste material será apresentada também uma generalização do enfoque dos modelos lineares generalizados como uma proposta mais ampla para situações semelhantes à que está sendo estudada. Como se verá, a seção (6) apresenta as idéias e os conceitos mais importantes dessa generalização, em uma tentativa de estender o universo dos possíveis modelos para dados de contagem. Acredita-se que essa seção apresenta um novo enfoque de análise para alguns modelos que, embora possam ser abordados como lineares generalizados impondo alguma restrição no modelo probabilístico, podem ser tratados como modelos mais amplos respeitando sua verdadeira estrutura. Além disso, é a partir desse enfoque teórico diferente que será apresentada a construção do possível processo estocástico para a mortalidade das árvores de *E. grandis* permitindo a obtenção de resultados que poderão auxiliar no manejo das florestas de *Eucalyptus grandis* no Estado de São Paulo.

2.2.3 Modelo Matemático

A parte sistemática do modelo linear generalizado (McCullagh & Nelder; 1989), η , assumido para os dados de mortalidade de *E. grandis* foi um modelo da análise da covariância com estrutura fatorial de tratamentos e idade das árvores como covariável, isto é

$$\eta_{ijk} = \mu + \alpha_i + \gamma_j + \alpha\gamma_{ij} + \beta x_{ijk}$$

onde

- μ é uma constante inerente aos dados,

- α_i é o efeito da i -ésima região,
- γ_j é o efeito da j -ésima rotação, com $i, j = 1, \dots, 3$,
- $\alpha\gamma_{ij}$ é o efeito da interação entre os níveis i e j dos fatores região e rotação, e
- β é o coeficiente de regressão para a variável idade, x_{ijk} , com $k = 1, \dots, K(i, j)$.

A inclusão da variável explanatória idade foi decidida após o uso de ferramentas gráficas e testes apropriados para diagnósticos, como será apresentado na seção seguinte. A Figura 1 no Apêndice A ilustra a possível dependência entre a mortalidade e a idade das árvores para cada tratamento.

Em um modelo linear clássico as estimativas dos parâmetros envolvidos são obtidas, simplesmente, manipulando vetores e matrizes através da álgebra linear. Já para os modelos lineares generalizados, a função de ligação permite a não linearidade na relação valor esperado-preditor linear introduzindo assim um custo nas iterações requeridas da álgebra linear, por exemplo, com o uso do método de Newton-Raphson. Isto mostra mais uma vantagem dos MLG uma vez que podem ser escolhidas outras funções, e não somente a função identidade, para conectar os componentes que o formam.

Nos modelos lineares generalizados o método da máxima verossimilhança conduz a uma estimação bastante simples. O algoritmo para solução das equações de MV foi desenvolvido por Nelder & Wedderburn (1972) e equivale ao cálculo repetido de uma regressão linear ponderada. Além disso, e como foi mencionado, é similar ao processo iterativo de Newton-Raphson, sendo a característica principal o uso da matriz de valores esperados de derivadas segundas do logaritmo da verossimilhança (informação), em lugar da matriz correspondente de valores observados (CORDEIRO, 1986). Apresentam-se a seguir as idéias básicas para a obtenção do algoritmo de estimação nos MLG.

Seja o conjunto de variáveis aleatórias independentes Z_i , $i = 1, \dots, n$, com valores observados z_i . Um membro da família exponencial será tal que a forma *canônica* da função de probabilidade (ou densidade) para a variável Z_i , com parâmetro de *locação* ou parâmetro *canônico* θ é:

$$f(z_i; \theta) = \exp[z_i\theta_i - b(\theta_i) + c(z_i)]$$

Generalizando, para uma distribuição a dois parâmetros, com parâmetro de *escala* ou parâmetro de *dispersão* ϕ , pode ser escrita segundo:

$$f(z_i; \theta, \phi) = \exp\{[z_i\theta_i - b(\theta_i)]/a_i(\phi) + c(z_i, \phi)\}$$

Por exemplo, no caso $Z_i \sim P(\mu_i)$, tem-se

$$P(z_i; \mu_i) = \exp[z_i \log(\mu_i) - \mu_i - \log(z_i!)]$$

onde $\theta_i = \log(\mu_i)$, $b(\theta_i) = \exp(\theta_i)$, $a_i(\phi) = 1$, $c(z_i, \phi) = -\log(z_i!)$.

A importância nessa especificação é a relação média-variância. Para essa família, com parâmetro de dispersão, o logaritmo da função de verossimilhança, para uma única observação, é:

$$\log[L(\theta_i; \phi; z_i)] = [z_i \theta_i - b(\theta_i)]/a_i(\phi) + c(z_i, \phi)$$

e sua derivada primeira ou *score*,

$$U_i = [z_i - b'(\theta_i)]/a_i(\phi)$$

A partir disso resulta

$$\begin{aligned} E(Z_i) &= b'(\theta_i) \\ &= \mu_i \\ V(Z_i) &= b''(\theta_i) a_i(\phi) \\ a_i(\phi) &= \frac{\phi}{w_i} \end{aligned}$$

onde w_i são pesos *a priori* conhecidos, e $b''(\theta_i) = v(\mu_i)$ é chamada de *função de variância*.

Assim, estabelecido o modelo probabilístico, θ_i é o parâmetro de interesse, de modo que ϕ é usualmente considerado como parâmetro "*nuisance*". Com ϕ conhecido, z_i é uma estatística suficiente para θ_i , e a distribuição é completamente especificada, ou determinada, pela relação entre a média e a variância. As equações de estimação, ou *scores*, igualadas a zero e para o caso de independência das observações, produzem as estimativas de máxima verossimilhança para os parâmetros.

Em outras palavras, a obtenção das estimativas dos p parâmetros β_l , no modelo de locação, ou preditor linear $\eta_i = g(\mu_i) = \beta^T \mathbf{x}_i$, equivale a resolver o sistema:

$$\sum_{i=1}^n \sum_{k=1}^r \left[(z_{ik} - \mu_{ik}) \frac{\partial \theta_{ik}}{\partial \mu_{ik}} \frac{\partial \mu_{ik}}{\partial \eta_{ik}} x_{ilk} \right] = 0$$

Com r observações independentes na unidade i , esse sistema pode ser escrito em notação matricial como segue,

$$\sum_{i=1}^n \mathbf{X}_i \mathbf{D}_i \mathbf{d}_i = 0$$

onde $\mathbf{X}_i$ é uma matriz $p \times r$ de delineamento para a unidade experimental,

$$\mathbf{D}_i = \text{diag} \left[\frac{\partial \theta_{ik}}{\partial \mu_{ik}} \frac{\partial \mu_{ik}}{\partial \eta_{ik}} \right]$$

é uma matriz $r \times r$ e $\mathbf{d}_i = [\mathbf{z}_i - \mu_i]$ é um vetor $r \times 1$. Note-se que, dado que as observações na unidade experimental são assumidas independentes, a matriz $\mathbf{D}_i$ resulta diagonal.

O ajuste de um modelo linear generalizado pode ser feito no sistema GLIM de uma maneira muito simples. Deixando os detalhes para consulta do manual, a definição de um MLG requer 6 diretivas: YVARIATE (variável de resposta), ERROR (distribuição aleatória da variável de resposta), LINK (função de ligação), WEIGHT (pesos a priori), SCALE (parâmetro de escala ϕ), e OFFSET (valor fixo para uma parte linear conhecida do modelo). Além disso, o sistema GLIM pode ser usado para manipulação de dados, cálculos estatísticos, análise de resíduos e representação gráfica dos dados. Macros e algoritmos de resolução construídos em GLIM para o ajuste dos modelos apresentados neste material serão mostrados no Apêndice B.

2.2.4 Inferência e Verificação do Ajuste do Modelo

Os métodos de inferência nos modelos lineares generalizados baseiam-se, fundamentalmente, na teoria de máxima verossimilhança. De acordo com esta teoria, existem três estatísticas para testar hipóteses relativas aos parâmetros β 's, que são deduzidas de distribuições assintóticas, e de funções adequadas das estimativas dos β 's (Cordeiro, 1986). Sob a hipótese nula, as três estatísticas têm distribuição aproximada χ_p^2 . As expressões para hipóteses simples nos parâmetros são as seguintes:

- **Quociente de Máxima Verossimilhança:** é a única que estabelece um teste uniformemente mais poderoso para hipóteses relativas aos vários β 's:

$$R.V. = -2 \ln \lambda = 2[\ell(\hat{\beta}) - \ell(\beta_o)]$$

- **Estatística de Wald:** utiliza propriedades assintóticas dos estimadores de máxima verossimilhança dos β 's, como a aproximação normal:

$$W = (\hat{\beta} - \beta_o)^T \hat{j} (\hat{\beta} - \beta_o)$$

onde $\hat{j}$ é matriz de informação de Fisher avaliada através da estimativa da máxima verossimilhança de β .

- **Estatística Escore:** é obtida simplesmente da função escore apresentada no algoritmo de estimação:

$$E = U^T(\beta_o) J_o^{-1} U(\beta_o)$$

onde J_o é a matriz de informação de Fisher avaliada no valor de β estabelecido em H_o .

Maiores detalhes dos desenvolvimentos teóricos dessas estatísticas podem ser encontrados em Cox & Hinkley (1979); McCullagh & Nelder (1989), e para a distribuição assintótica do vetor $\hat{\beta}$ nos modelos lineares generalizados, Fahrmeir & Kaufmann (1985) apresentam as condições gerais, que garantem as propriedades de consistência e normalidade dessas estimativas.

Cordeiro (1986) apresenta resultados para a realização de inferências nos casos de hipóteses compostas do tipo $H_o : A\beta = 0$, com A uma matriz $q \times p$ de posto q com $q < p$, encontrando as expressões dessas estatísticas e comparando as performances assintóticas das mesmas. O autor apresenta também a estatística de Pearson generalizada

$$X^2 = \sum_{i=1}^n \frac{(z_i - \hat{\mu}_i)^2 w_i}{V(\mu_i)}$$

como uma medida da discrepância entre os dados z e os valores ajustados μ , a qual é usada para testar a adequação do modelo linear generalizado e definida, internamente no GLIM, após o ajustamento do modelo, pelo escalar do sistema %X2.

Como teste de ajustamento do MLG, primeiramente observa-se a "deviance". Esta estatística foi definida por Nelder & Wedderburn (1972),

$$2(L_{(n)} - L_{(p)})$$

como a diferença entre os máximos dos logaritmo da função de verossimilhança para o modelo saturado (não informativo pois não sumariza os dados, mas simplesmente

os repete) e em investigação (modelo com p parâmetros, situado entre os modelos minimal e maximal).

Segundo Cordeiro (1986) para o teste, definem-se os graus de liberdade da "deviance" do modelo por $n - p$, os quais conjuntamente com o valor da "deviance" devem ser comparados com alguma distribuição de probabilidade de referência. Existem várias discussões a respeito, não havendo nenhuma referência acerca dessa distribuição, nem mesmo assintoticamente. Porém, Cordeiro (1986); Lindsey (1993), dentre outros, recomendam comparar a "deviance" com o valor crítico $\chi^2_{(n-p)}$, ou usar a estatística de Pearson, para a qual McCullagh (1985) mostrou resultados que avaliam a convergência assintótica a uma distribuição Qui-quadrado. Na realidade, é mostrado por Lawley (1956) que, através de expansões de Taylor para as equações de MV e $\hat{L}_p$, pode-se estabelecer uma ordem de convergência para a "deviance" tal que

$$2E(\hat{L}_p - L) = p + \epsilon_p + O(n^{-2})$$

onde o termo ϵ_p é de ordem n^{-1} , e de difícil obtenção já que depende de valores esperados de produtos de derivadas do logaritmo da função de verossimilhança e de derivadas desses valores esperados.

A partir de uma sequência de modelos (do tipo correspondente a hipóteses encaixadas) os efeitos dos fatores ou covariáveis incluídos, ou não, no componente sistemático, podem ser avaliados através da análise da "deviance", como pode ser visto em Demétrio (1993) e Cordeiro (1986). Essa análise utiliza a "deviance" de cada um dos modelos da sequência como uma medida da discrepância, gerando assim uma tabela de análise da "deviance", semelhante à tabela da análise da variância para um modelo normal do tipo Gauss-Markov. Em outras palavras, quando as hipóteses são do tipo $\Omega^n = H_o$ (hipótese saturada $\rightarrow n$ parâmetros, o que implica "deviance" nula) $\supseteq H_1 \supseteq H_2 \supseteq \dots$, cada uma dessas hipóteses representa uma restrição maior nos parâmetros do modelo, pelo qual o processo de inferência deve ser construído iterativamente. Isto é, através de uma análise dos resíduos é estudada a aderência do modelo H_1 , continuando-se com o teste de H_2 sob H_1 . Caso H_2 seja aceita, prossegue-se com H_3 sob H_2 , e assim em diante até que um resultado significativo é obtido, para o qual mantém-se o último modelo aceito como aquele modelo mais simples que explica os dados.

Como encerramento de uma primeira etapa do ajuste de um modelo a um conjunto de observações está a verificação do mesmo através de técnicas de diagnósticos, o qual gera novamente outro ciclo de ajustes visando à obtenção do modelo definitivo a propor para tal conjunto de dados. A literatura é vasta

neste tópico (Atkinson, 1982; Cordeiro, 1986; Atkinson *et al.*, 1989; Collett, 1991, dentre outros). Basicamente os autores concordam com a aplicação de métodos de covariáveis adicionais, análises de resíduos e diagnósticos e métodos intuitivos como os fornecidos por gráficos específicos. Neste material somente serão apresentadas as técnicas utilizadas para tal fim.

Para a verificação da parte funcional do modelo e função de ligação utilizou-se o método da covariável adicionada, o qual abrange a técnica de ajuste desse novo modelo e o gráfico dos resíduos de Pearson versus os resíduos da variável adicionada. Esse método consiste em aumentar a estrutura linear do modelo através de covariáveis escolhidas, segundo Cordeiro (1986), adequadamente para representar anomalias específicas no modelo linear generalizado usual. O componente sistemático aumentado resulta então $g(\mu) = X\beta + u\gamma$, onde u pode representar alguma função específica das covariáveis originais, ou uma matriz $n \times q$ (U) de variáveis adicionais.

A importância das covariáveis adicionais é dada pela diferença das "deviances" dos modelos $g(\mu) = X\beta$ e $g(\mu) = X\beta + u\gamma$, ou, em outras palavras, se a adição de alguma covariável u altera substancialmente o ajustamento, as anomalias em questão afetam, seriamente, o modelo original. De outra maneira, se o gráfico dos resíduos de Pearson versus os correspondentes ao novo modelo de locação - apresentar indícios de existência de relação linear, poder-se-ia considerar-se, então, essa covariável adicionada como necessária na modelagem na presença das outras restantes.

Essa metodologia também é usada para avaliar a performance da função de ligação e para identificar observações não explicadas pelo modelo. Segundo Cordeiro(1986) muitas vezes, para um conjunto particular de dados, pode ser difícil decidir qual a melhor função de ligação e, ainda, esta pode não pertencer à família potência senão a uma família maior (na generalização), como é, por exemplo, a família de Aranda-Ordaz. Uma estratégia frequentemente usada por sua simplicidade para verificar se uma função de ligação é adequada é computar a redução na "deviance" após a inclusão da covariável $\hat{\eta}^2$, que pode ser confirmada a partir de técnicas gráficas. Pregibon (1980) propõe indexar a família (potência) de ligações por um parâmetro λ e fazer um teste para o valor λ_0 . Um outro processo alternativo, chamado método do perfil de verossimilhança, para obter uma boa estimativa do λ é considerar λ fixado e pertencendo a um conjunto amplo de valores arbitrários, e então computar a "deviance"(λ) como função de λ , traçar o gráfico dessas funções versus λ e escolher $\tilde{\lambda}$ correspondente ao valor mínimo da "deviance".

Complementando, as análises de resíduos são importantes para detetar a presença de observações aberrantes que devem ser pesquisadas em detalhe. Um resíduo deve expressar, de alguma maneira conveniente e intuitiva, a discrepância entre a observação e o seu valor ajustado, isto é $r_k = h_k(z_k, \hat{\mu}_k)$, onde $h_k(., .)$ é selecionada em função de diferentes critérios. Em geral, essa função h_k é escolhida para estabilizar variâncias ou para induzir simetria na distribuição amostral de r_k . Uma vez estabelecidos os critérios de definição desses resíduos, esses conjuntos de valores são utilizados para a construção de importantes gráficos na exploração daquele modelo procurado como o adequado para as observações, como o são, os gráficos de índices, os gráficos dos r_k versus os $\hat{\mu}_k$, os gráficos dos r_k versus os níveis da variável ou fator correspondente ao parâmetro omitido, dentre os mais usados.

Para avaliar a presença de pontos discrepantes ou com ajuste pobre pelo modelo, foram usados os resíduos definidos como componentes da "deviance" estudentizados internamente (ou padronizados melhorados):

$$\begin{aligned} r'_{di} &= \frac{sg(z_i - \hat{\mu}_i) \sqrt{2\{z_i[\tilde{\theta}_i - \hat{\theta}_i] - b(\tilde{\theta}) + b(\hat{\theta})\}}}{\sqrt{\hat{\phi}(1 - h_i)}} \\ &= \frac{r_{di}}{\sqrt{\hat{\phi}(1 - h_i)}} \end{aligned} \quad (2)$$

onde h_i é um elemento da diagonal da matriz H ,

$$H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2}$$

Esse resíduo representa uma distância da observação ao seu valor ajustado, medida na escala da verossimilhança e padronizado por medidas de influência. Assim, um valor grande para o i -ésimo resíduo indica que o modelo ajusta mal à observação.

Pregibon (1979) mostrou que esses resíduos podem ser tratados, aproximadamente, como variáveis normais, e consequentemente seus quadrados como χ^2_1 . Esses componentes podem ser obtidos no GLIM mediante cálculos muito simples, já que o pacote fornece o valor do vetor de sistema %DI contendo os numeradores de (2). Esses resíduos são aproximadamente equivalentes aos resíduos de Anscombe (Cordeiro, 1986; Collett, 1991), embora suas formas, dados z_i 's e μ_i 's, sejam bem diferentes.

Neste trabalho não foram usados os resíduos de Pearson (componentes da estatística generalizada de Pearson) já que eles têm a desvantagem de que sua distribuição é bastante assimétrica para modelos não normais. Também

não foi usado o recurso gráfico de "envelope simulado", muito útil, principalmente, para ajustes de modelos com poucas observações. Esse gráfico é importante já que mostra como o modelo ajusta às observações comparando-as com o valor médio. Este recurso não foi possível de implementar visto que simula 19 conjuntos de dados, de tamanho igual à da amostra, para gerar modelos determinados no componente aleatório. Com o tamanho de amostra, $n = 2605$, que este material aborda foi impossível conseguir recurso computacional disponível para GLIM que suportasse as necessidades de memória.

3 EXPLORAÇÃO DO MODELO ESTOCÁSTICO

Os dados de contagem que este trabalho aborda possuem dois fatores de tratamentos com três níveis cada um. Considere-se então Z_{ijk} , a variável aleatória correspondente ao número de árvores mortas registradas na k -ésima parcela, representando uma repetição do tratamento ij , com $i, j = 1, 2, 3$. Pela natureza da variável aleatória, e tendo como base o desconhecimento de possíveis influências ou perturbações que as parcelas poderiam ter sofrido, pode-se ter basicamente, três tipos de distribuição, a saber:

1. distribuição **aleatória**, (modelo Poisson correto);
2. distribuição **regular**, (modelo Poisson com subdispersão);
3. distribuição de **contágio**, (modelo Poisson com superdispersão);

ou, em termos da relação média-variância:

1. $\sigma^2 = \mu \rightarrow$ Distribuição Poisson;
2. $\sigma^2 < \mu \rightarrow$ Distribuição Binomial (Positiva);
3. $\sigma^2 > \mu \rightarrow$ Distribuição Binomial Negativa.

Cabe assinalar que não significa que essas sejam as únicas distribuições para propor como modelo estocástico para a variável mortalidade, porém elas constituem uma primeira saída, básica e obrigatória, para a exploração de tal modelo.

Visando pesquisar a modelagem do componente aleatório do MLG a propor, e começando com uma exaustiva análise da relação média-variância da variável aleatória, a Tabela 1 apresenta os valores médios e os respectivos desvios padrões da mortalidade para cada tratamento (combinação dos níveis dos dois fatores), e a Figura 2, no Apêndice A, mostra a dispersão dos pares $(Var(Z), \bar{Z})$. Essa figura é utilizada para uma primeira inspeção do comportamento da variável aleatória.

A princípio, se os dados seguissem uma distribuição espacial aleatória, ou distribuição de Poisson, esse gráfico deveria mostrar uma relação 1-à-1 entre abscissa e ordenada, que tende a se aproximar de uma função de regressão linear com coeficiente angular igual a um. Por outro lado, nota-se que esse gráfico envolve apenas informação acerca dos primeiros dois momentos da distribuição hipotética das observações, o qual também não seria suficiente para a escolha do modelo estocástico da variável. Mas, essas ferramentas são muito úteis, e tornam-se necessárias em uma primeira instância da análise.

Outro método rápido, proposto por Elder (1995), para explorar uma provável agregação na distribuição dos dados, isto é, $\sigma^2 > \mu$, é baseado também no método dos momentos e consiste em plotar os $\ln(\bar{Z} - Var(Z)/n)$ versus $\ln(Var(Z) - \bar{Z})$, onde $\ln$ é o logaritmo neperiano. Nesse caso, se o modelo de agregação for o adequado para as observações, a tendência mostrada nesse gráfico deveria ser próxima a uma reta. A Figura 3 no Apêndice A ilustra essa situação, e indica como candidato um modelo que reflete magnitudes de dispersão maiores do que as de posição.

Dessas informações se deriva, a princípio, que o modelo probabilístico Poisson, assumido como hipótese preliminar para dados de contagem, não é o adequado, já que não existe uma relação linear do tipo $\bar{Z} = \alpha Var(Z)$, com α não significativamente diferente de 1. Aliás, o modelo que for proposto deverá relacionar os dois primeiros momentos na forma indicada no item 3, o que na literatura em geral (Elliot, 1979), é descrito como distribuição espacial agregada, ou de contágio.

Existe abundante bibliografia acerca da aplicação das distribuições de contágio a problemas de dispersão no espaço e no tempo. Segundo McGuire *et al.* (1957), o conceito de distribuição de contágio não implica necessariamente um verdadeiro contágio, apenas uma inflação da variância vis-a-vis àquela imposta pelo modelo Poisson, o que é chamado de superdispersão. Na seção que se segue serão expostos alguns argumentos que justificam esse fenômeno em conjuntos de dados, como também alguns caminhos possíveis para a implementação de técnicas de análises.

Às vezes, a superdispersão nos modelos de contagem, é estudada através de uma estrutura de extra-variação no modelo Poisson, chamado de modelo Extra-Poisson, e usado freqüentemente na literatura a partir do algoritmo de Breslow (1984) para ajustes de modelos lineares generalizados. Este trabalho destina uma seção para a apresentação e justificativa do mesmo. Mas, segundo Rabinovich (1978), quando o contágio, ou a superdispersão no modelo Poisson está

Tabela 1: Valores médios e desvios padrões (d.p.) de mortalidade (%) das árvores de *Eucalyptus grandis* para os tratamentos

Tratamento	Mortalidade (d.p.)
Rotação1 x Local 1	9,61(11,35)
Rotação1 x Local 2	9,24(8,22)
Rotação1 x Local 3	14,35(11,85)
Rotação2 x Local 1	24,46(22,11)
Rotação2 x Local 2	13,48(21,84)
Rotação2 x Local 3	31,92(26,82)
Rotação3 x Local 1	23,75(20,93)
Rotação3 x Local 2	28,68(25,55)
Rotação3 x Local 3	41,69(7,71)

presente, pode-se supor que os indivíduos são independentes entre si (não existe agrupamento), mas cada unidade de amostragem possui uma densidade variável de indivíduos, o qual, simula um contágio "aparente", e o padrão observado deve-se à heterogeneidade natural.

Obviamente, os gráficos são indicadores muito valiosos das características do sistema estocástico em estudo. Esses indicadores são chamados de ferramentas informais já que atuam e exploram o sistema em estudo apenas de maneira intuitiva. Precisa-se, então, de ferramentas formais, como os testes de hipóteses, os quais verificam, com um certo nível de significância, se o que a distribuição empírica mostra é realmente evidência para não rejeitar o modelo probabilístico proposto. Neste trabalho serão expostos dois procedimentos para verificação de hipóteses relacionadas com o modelo Poisson, dos quais o segundo é escassamente usado na literatura em geral. Nas seções (3.2) e (3.4) serão descritas essas ferramentas formais de diagnóstico do modelo estocástico. Não obstante, uma verificação simples e imediata poderia ser submeter os dados a uma análise da "deviance", com componente aleatório Poisson, e especificações para o componente sistemático e função de ligação conforme a seção (2.2.3).

3.1 Modelo Poisson

O modelo Poisson é um modelo probabilístico discreto muito especial dentre os modelos lineares generalizados, e como será mostrado na seção (6), ele é um modelo limite para os modelos exponenciais de dispersão discretos à semelhança do que acontece com o normal para os contínuos, como resultado do Teorema do Limite Central.

Seja Z uma variável aleatória com distribuição Poisson $P(m)$. Então, sua função de distribuição de probabilidade é:

$$\begin{aligned} p(z; m) &= \frac{m^z}{z!} e^{-m} \\ &= \frac{1}{z!} \exp \{ \tilde{\phi} z - e^{\tilde{\phi}} \}, \quad z \in N_o \end{aligned} \quad (3)$$

ou, na forma canônica da família exponencial uniparamétrica (McCullagh & Nelder, 1989; Demétrio, 1993),

$$p(z; m) = \exp \left\{ \frac{1}{a(\phi)} [z\theta - e^\theta] - \log z! \right\} \quad (4)$$

ou, ainda mais, definindo $m = \lambda e^\theta$

$$p(z; m) = \frac{\lambda^z}{z!} \exp \{ z\theta - \lambda e^\theta \} \quad z \in N_o \quad (5)$$

onde

$$\begin{aligned} \tilde{\phi} &= \log(m) \in \mathbb{R} \\ \theta &= \log m, \quad a(\phi) = \frac{\phi}{w} = 1 \end{aligned}$$

Expresso o modelo segundo (4), se satisfaz que $E(Z) = m = \mu = V(\mu)$ e o parâmetro de dispersão resulta conhecido e igual a 1, o que implica que $V(Z) = 1$. $V(\mu) = \mu$. Já no enfoque dos modelos exponenciais de dispersão¹ de Jorgensen (1989), como na forma (5), da qual a forma (3) é sua extensão, tem-se que $E(Z) = m = \lambda \mu$ e $V(Z) = \lambda V(\mu) = \lambda \mu$ pela qual o parâmetro de dispersão, λ , resulta desconhecido, e portanto há necessidade de estimá-lo.

¹Esse enfoque teórico será apresentado em uma seção separada pois a abordagem implica, entre outras coisas, introduzir o leitor no estudo de uma classe de modelos mais geral, relaxando algumas das restrições impostas pela família exponencial a um parâmetro, pelo qual, acredita-se então, que essa apresentação não respeita o contexto aplicado desta seção.

Para a forma (4), os termos para o algoritmo de estimação do modelo linear generalizado com preditor linear como em (2.2.2), desprezando os índices, são os seguintes:

$$\begin{aligned}
 \mu &= e^\eta \\
 V(\mu) &= \mu \\
 a(\phi) &= \frac{\phi}{w} \\
 &= 1 \\
 g'(\mu) &= \frac{1}{\mu} \\
 \zeta^{(m)} &= \eta^{(m)} + (z - \mu^{(m)}) \frac{1}{\mu^{(m)}} \\
 &= X\beta^{(m)} + H^{(m)}(z - \mu)^{(m)} \\
 W^{(m)} &= \text{diag}\{\mu^{(m)}\} \\
 H^{(m)} &= \text{diag}\left\{\frac{1}{\mu^{(m)}}\right\}
 \end{aligned}$$

Assim, a estimativa de máxima verossimilhança para os β 's do preditor linear obtém-se através do processo iterativo a seguir:

$$\Rightarrow \beta^{(m+1)} = \{[X'WX]^{-1}\}^{(m)}(X'W\zeta)^{(m)}$$

onde X é a matriz de delineamento, conhecida.

Os métodos de inferência neste modelo, como em todos os MLG, baseiam-se na teoria de máxima verossimilhança, como foi apresentado na seção (2.2.2). Como medidas do ajuste do modelo, ou estatísticas de aderência que medem as discrepâncias entre o observado e o estimado pelo modelo proposto, apresentam-se "deviance" e a estatística de Pearson, da forma:

$$\begin{aligned}
 D(z; \hat{\mu})/\phi &= \sum 2w\{z(\tilde{\theta} - \hat{\theta}) - b(\tilde{\theta}) + b(\hat{\theta})\}/\phi \\
 D(z; \hat{\mu}) &= 2 \sum \{z \log(z/\hat{\mu}) - (z - \hat{\mu})\}
 \end{aligned}$$

$$\begin{aligned}
 X^2 &= \sum (z - \hat{\mu})^2 / V(\hat{\mu}) \\
 &= \sum (z - \hat{\mu})^2 / \hat{\mu}
 \end{aligned}$$

onde $\tilde{\theta}$ e $\hat{\theta}$ são as estimativas sob os modelos saturado e corrente¹, respectivamente. No caso Poisson, e assumindo que o modelo é correto (H_0 verdadeira), essas estatísticas possuem distribuição χ^2 , apenas, assintoticamente, porém, esses resultados não podem ser relevantes quando os tamanhos de amostras não são suficientemente grandes. Não obstante, destaca-se que a "deviance" possui a grande vantagem de ser aditiva para modelos encaixados quando as estimativas são as de máxima verossimilhança, propriedade importante que a diferencia da estatística de Pearson generalizada.

O modelo Poisson também é importante no estudo dos processos estocásticos no tempo ou no espaço. Aplica-se ao número de eventos que ocorre em sucessivos intervalos pequenos, onde as ocorrências nesses intervalos são independentes, o qual é chamado de Processo Poisson e corresponde a intervalos entre eventos cuja distribuição é a exponencial.

Assim, com os recursos do GLIM foi ajustado o modelo linear generalizado para os dados de *E. grandis*, com erro Poisson, função de ligação canônica (log), parâmetro de escala unitário, componente sistemático seguindo um modelo de análise da covariância, declarando dois fatores, rotação e região com três níveis cada um. A covariável idade das árvores, em anos, foi escolhida pela mudança evidenciada nos seus valores médios ao longo dos tratamentos, o qual é ilustrado na Figura 1 do Apêndice A. A Tabela 2 apresenta os valores médios e desvios padrões (d.p.) para cada combinação da região e da rotação de *E. grandis*. O valor da "deviance" residual de 28.083,01, com seus respectivos 2.596 graus de liberdade, indicou um ajuste pobre do modelo.

Porém, já foi dito que, estabelecer um modelo para um conjunto de dados implica, entre outras coisas, escolher, por algum critério, as partes que o compõem, isto é, seu componente aleatório, seu componente sistemático, e sua função de ligação. O "feedback" da modelagem deve, sem dúvida, começar pelo que é mais conhecido, ou em outras palavras, por aquilo que pode ser pesquisado imediatamente, como é a distribuição das observações.

Tentando melhorar o ajuste, e analisando o grande ponto de massa que as observações nulas provocam na distribuição original da mortalidade, postulou-se ainda um modelo de dispersão aleatória apenas para as observações positivas. Assim, foi avaliado o modelo **Poisson Truncado** (David & Johnson, 1952), para o qual se assume que a probabilidade de um evento em um período fixo é independente

¹Modelos contendo um parâmetro para cada observação, e sob pesquisa, respectivamente.

Tabela 2: Valores médios e desvios padrões (d.p.) da idade (anos) das árvores de *Eucalyptus grandis* para os tratamentos.

Tratamento	Idade (d.p.)
Rotação1 x Local 1	4,99(2,17)
Rotação1 x Local 2	4,29(1,71)
Rotação1 x Local 3	4,03(1,69)
Rotação2 x Local 1	5,67(1,95)
Rotação2 x Local 2	5,74(2,21)
Rotação2 x Local 3	4,55(1,73)
Rotação3 x Local 1	4,95(1,71)
Rotação3 x Local 2	5,39(1,74)
Rotação3 x Local 3	2,34(0,12)

de que o evento tenha ocorrido previamente ou não. Para os dados de *E. grandis* essa afirmação poderia ser improvável, pelo qual existiriam dois processos em ação com duas taxas: uma para a primeira aparição de uma árvore morta e outra para a segunda, condicionada.

Assim, uma possibilidade de estudo seria assumir que, uma vez que ocorreu uma morte, o número de mortes é distribuído segundo um processo Poisson Truncado, da forma:

$$P(Z_i = z_i) = \frac{e^{-\lambda_i} \lambda_i^{z_i}}{z_i!(1 - e^{-\lambda_i})}, \quad i = 1, \dots, h$$

onde h é o número de parcelas ou unidades experimentais com, ao menos, uma morte, e $z_i \in N$. Assumindo que $\lambda_i = \exp(X'\beta)$, com as referências anteriores para o preditor linear, pode-se ajustar no GLIM o modelo linear generalizado segundo a macro apresentada por Ridout & Demétrio (1992) com os recursos do sistema de usuário que o pacote possui.

Os resultados obtidos da implementação dessas macros para os dados de mortalidade de *E. grandis* apuraram um valor da "deviance" residual de

13.313,05 versus os 1.658 graus de liberdade para essa fonte de variação. Isto indica que o ajuste é ainda muito pobre, e os gráficos exploratórios dos valores ajustados e suas estatísticas resumos forneceram indícios da presença de valores preditos muito pequenos, e portanto uma aproximação assintótica inadequada do χ^2 para a aderência do modelo. Assim, o modelo Poisson para as contagens positivas, com as mesmas especificações que no caso Poisson anterior, é inapropriado para os dados de mortalidade de *E. grandis*.

Uma outra possibilidade para ainda manter uma dispersão aleatória entre as observações positivas, mas contendo informação acerca das parcelas sem presença de mortes, é propor um modelo, chamado por Cohen (1960), citado por Macchiavelli (1987), de **Poisson com zeros**. Essa distribuição resulta de um processo de adicionar a classe ou categoria zero à uma Poisson truncada, como foi no caso anterior. A função de probabilidade é:

$$P(X = x) = \begin{cases} 1 - \theta & : z = 0 \\ & : \\ \frac{\theta e^{-\lambda} \lambda^z}{z!(1-e^{-\lambda})} & : z = 1, 2, \dots \end{cases}$$

Para esse modelo, os parâmetros são $\lambda > 0$ e $0 < \theta < 1$. As expressões para o valor esperado μ e a variância σ^2 podem ser obtidos, como em Macchiavelli (1987), através da sua função geradora de probabilidade:

$$\begin{aligned} G(s) &= (1 - \theta) + \sum \frac{\theta e^{-\lambda} \lambda^z}{z!(1 - e^{-\lambda})} s^z \\ &= (1 - \theta) + \frac{\theta e^{-\lambda}}{(1 - e^{-\lambda})} \sum \frac{(\lambda s)^z}{z!} \\ &= (1 - \theta) + \frac{\theta e^{-\lambda}}{(1 - e^{-\lambda})} (e^{\lambda s} - 1) \\ &= 1 - \theta \left\{ \frac{1 - \exp[\lambda(s - 1)]}{1 - \exp[-\lambda]} \right\} \end{aligned}$$

obtendo as derivadas de primeira e segunda ordem em relação ao s , isto é:

$$\begin{aligned} \mu &= \theta \lambda [1 - \exp(-\lambda)]^{-1} \\ \sigma^2 &= \mu [1 + \lambda - \mu] \end{aligned}$$

Observe-se que, o quociente média-variância será menor do que 1 se, e somente se, $\lambda > -\ln(1 - \theta)$. De fato, $e^{-\lambda} < 1 - \theta$ já que, pela definição que apresenta Cohen (1960), para que essa distribuição seja considerada como de contágio, ou agregada, a probabilidade da classe zero deveria ser maior que em uma distribuição de Poisson com o mesmo parâmetro λ . Então,

$$\begin{aligned}
 1 - e^{-\lambda} &> \theta \\
 1 &> \theta(1 - e^{-\lambda})^{-1} \\
 \lambda &> \lambda\theta(1 - e^{-\lambda})^{-1} \\
 \frac{\sigma^2}{\mu} &= 1 + \lambda - \lambda\theta(1 - e^{-\lambda})^{-1} \\
 &\Rightarrow \frac{\sigma^2}{\mu} > 1
 \end{aligned}$$

A interpretação do parâmetro λ é a mesma que para a distribuição Poisson; e θ representa a proporção populacional do número de parcelas com presença de mortes, ou em outras palavras, a quantidade de unidades experimentais incompletas observadas, visando o sucesso da implantação. Esse parâmetro funciona como "nuisance", uma vez que as equações de estimação pelo método da máxima verossimilhança são independentes. Dessas equações se deduz que o estimador de máxima verossimilhança para θ é $\hat{\theta} = n^*/n$, onde n^* indica o número de parcelas menos a frequência na classe zero, ou seja, $\hat{\theta} = 1 - n_0/n$.

Logo, o ajuste no GLIM do modelo linear generalizado, com η descrito em (2.2.3), resulta imediato uma vez que os parâmetros atuam de maneira independente. Isto é, no algoritmo utilizado para o ajuste anterior (Poisson truncado, através das macros de Ridout & Demétrio (1992)) somente resta multiplicar à equação contendo o λ pelo valor encontrado para θ e prosseguir com a estimação. Essa macro não será incluída no Apêndice B já que simplesmente adiciona um cálculo à já apresentada por esses autores para o caso Poisson Truncado.

Os resultados obtidos para os dados de mortalidade de *E. grandis* indicaram que o modelo Poisson com zeros é inapropriado para avaliar efeitos de interação entre os dois fatores, uma vez que "deviance" residual foi de 19.477,18 versus os 2.596 graus de liberdade para essa fonte de variação. Isto indica que o ajuste é ainda muito pobre, e mais precisamente, a aderência do modelo Poisson, condicionado ou não, para as observações não nulas (positivas) não é satisfatória.

3.2 Teste Quociente Média-Variância

Os testes de aderência no contexto da Estatística Aplicada são de vital importância, dado que constituem a ligação entre a realidade (amostra observada) e o modelo teórico proposto (distribuição da população considerada). Nesta seção se apresenta, e se justifica a escolha de um teste de aderência para um modelo estocástico, muito utilizado nas pesquisas provenientes das áreas de Ecologia e Entomologia.

Segundo Elliot (1979), o teste do **Quociente Média-Variância** é apropriado quando não se têm muitos elementos *a priori* para descrever a distribuição dos dados. Este teste é baseado na igualdade da variância e a média em séries Poisson. O Índice de Dispersão, I , se aproximará de 1 se houver concordância com uma série Poisson. Isto é:

$$I = \frac{\text{Variância Amostral}}{\text{Variância Teórica}} = \frac{S^2}{\bar{Z}} = \frac{\sum_{i=1}^n (Z - \bar{Z})^2}{\bar{Z}(n-1)} \quad (6)$$

onde $\bar{Z}$ é a média aritmética, n é o número de unidades amostrais.

A expressão $I(n-1)$ possui boa aproximação à χ^2 com $(n-1)$ graus de liberdade. Portanto, a estatística do teste, para a hipótese que existe uma série de Poisson, será:

$$\chi^2 = I(n-1) = \frac{S^2(n-1)}{\bar{Z}} = \frac{\sum_{i=1}^n (Z - \bar{Z})^2}{\bar{Z}}$$

Os diferentes valores de χ^2 a 5% de significância em função dos graus de liberdade podem ser obtidos da Tabela 8 de Pearson & Hartley (1966). Este teste aceita a hipótese nula ao nível $100(1-\alpha)\%$ de confiança. A Tabela 3 apresenta os valores dessa estatística indicando significância ao nível $\alpha = 0.01$, pelo que se conclui que o modelo não é o Poisson.

Existem outros testes de aderência de uma distribuição teórica à uma observada, dentre os quais podem-se mencionar os clássicos testes χ^2 , o teste de Kolmogorov-Smirnov, e o teste do quociente de verossimilhança. Mas, antes da seleção do teste a utilizar, deveriam ser consideradas três características fundamentais que, sem dúvida, condicionam a construção de um teste de aderência:

- a distribuição hipotética é discreta;
- a hipótese nula é composta (os parâmetros da distribuição são desconhecidos);

- não constituem eventos raros aqueles para os que correspondem frequências esperadas pequenas para algumas classes.

O teste de Kolmogorov-Smirnov apresenta problemas para o primeiro item, porém o de Qui-quadrado não. Além do mais, esse teste é mais conservador quando os parâmetros da distribuição são todos desconhecidos (Conover, 1971). Segundo Macchiavelli (1987), o teste do quociente média-variância foi estudado por Heltshe & Ritchey (1984)¹, os quais compararam a performance da estatística I com outra proposta por Stevens (1937)² para a mesma hipótese nula, concluindo que a aproximação da distribuição Qui-quadrado é boa ainda no caso em que n não seja suficientemente grande.

Segundo Elder (1995), testes de aderência como o Teste de Qui-Quadrado, Quociente de Verossimilhança, ou o Teste de Kolmogorov-Smirnov não superam as boas propriedades assintóticas do Teste de Dispersão (I é chamado de Índice) apresentado por Fisher em 1922. Ainda mais, em Frome *et al.* (1973) podem-se encontrar os fundamentos matemáticos e a construção dos algoritmos para a obtenção do teste exato.

3.3 Superdispersão

Nesta seção são discutidas algumas das possíveis causas da superdispersão no modelo Poisson, como também apresentadas diferentes metodologias para sua abordagem dentro da modelagem estatística.

É óbvio que, impor um modelo estocástico para um conjunto de observações, extraídas de uma família de distribuições caracterizada por um parâmetro implica em particular, forçar que a variância seja determinada univocamente pela média. Isto é, na seleção do modelo resulta implícita a relação matemática entre a variância e a média da variável aleatória, o que é consequência imediata de se assumir que μ_i é constante, como valor esperado da distribuição, e da **independência dos eventos**.

Esses argumentos possuem grande importância na hora de considerarmos a modelagem. Assim, conhecer o processo biológico de geração das observações é fundamental para o sucesso das análises subsequentes. Porém, um fenômeno muito frequente nos conjuntos de dados provenientes das Ciências Natu-

¹HELTSHE, J.F.; RITCHEY, T.A. Spatial pattern detection using quadrat samples. **Biometrics**, v. 40, p.877-886, 1984.

²STEVENS, W.L. Significance of grouping. **Annals of Human Genetics**, v. 8, p.57-69, 1937.

rais, é a presença de superdispersão, a qual conduz a uma falha da relação média-variância estabelecida.

Em termos gerais, para uma distribuição de dados determinada, o fato de possuir superdispersão na modelagem significa que a variância amostral excede a variância nominal. Os caminhos encontrados na literatura em geral para contornar esse problema coincidem em aplicar o método de quase-verossimilhança (QV), dada a simplicidade do algoritmo. Isto é devido ao fato que para definir a função de quase-verossimilhança somente se necessita especificar a relação entre os dois primeiros momentos, ou entre a média e a variância.

Se uma extra-varição estiver presente, significa que existe desconhecimento de informações, e assim, o que essa informação produz na estrutura aleatória dos dados. Então, a superdispersão está relacionada com alguma fonte aleatória adicional de variabilidade, ou também devido à correlação positiva entre as respostas dentro de um grupo, embora em princípio isto sugira a existência de mais variáveis explanatórias, cuja inclusão no modelo poderia reduzir a variância à mesma grandeza da estrutura estabelecida pelo modelo probabilístico adotado.

Segundo Lindsey (1993), a superdispersão em modelos Poisson pode originar-se, ao menos, por duas razões:

1. a probabilidade de um evento em uma unidade experimental pode ser a mesma para todas as unidades experimentais, porém, depende do número prévio de eventos ocorrendo naquela unidade. Em outras palavras, varia no tempo. Esse é um modelo de contágio.
2. a probabilidade de um evento em uma unidade experimental pode permanecer constante ao longo do tempo mas não necessariamente ser igual em todas as unidades. Esse é um modelo de heterogeneidade populacional de propensão ou debilidade.

Segundo Wedderburn (1974), uma situação geral é aquela onde existe uma relação entre ambas características possivelmente através de uma constante desconhecida. Se a variância for constante, a estimação é a de mínimos quadrados; quando outras relações média-variância estão presentes, a QV torna-se uma função de verossimilhança reconhecível. Esse é o caso das variáveis que tem coeficiente de variação constante, para as quais a função de QV é a mesma que a de verossimilhança obtida, de modo que se possa tratar as observações como se originassem de uma distribuição Gama.

Existem diversas consequências da superdispersão. Às vezes, pode-se estar interessado especificamente na superdispersão como um fenômeno em si,

em cujo caso sua magnitude necessita ser estimada. Em outras, há necessidade de realizar uma análise detalhada da sua estrutura para poder especificá-la, e logo, estimá-la. Relacionado com isso é frequente o uso de procedimentos bayesianos para completar as inferências acerca dos parâmetros. Na seção (4.4), esse tópico será abordado a nível introdutório. Mas, se a superdispersão é ignorada pode haver uma falta de eficiência nas estimativas pontuais, no sentido de que os erros padrões terão grandeza menor. Além disso, segundo Cox & Snell (1989), se uma quantidade de superdispersão é grande, pode ser conveniente concentrar primeiramente o estudo na forma real da variação. Esse é o caso do modelo Extra-Poisson, onde o componente $Var(Z_i) = \mu_i$ conforma a menor parte da variação real.

Aitkin & Francis (1995) comentam que o fato de as pressuposições do modelo proposto não serem satisfeitas, produz

- uma estimativa da "deviance" residual inflacionada,
- valores estimados dos resíduos muito grandes,
- subestimação dos erros padrões das estimativas dos parâmetros (e consequentemente, inferências errôneas para as variáveis explanatórias),
- severos vícios nas próprias estimativas no caso que algumas variáveis omitidas na modelagem serem correlacionadas com aquelas do modelo.

Em termos do ajuste do modelo, se o proposto não for o adequado, e descartando problemas da parte sistemática, a "deviance" residual, medida do ajuste para a maioria dos modelos lineares generalizados, supera os respectivos graus de liberdade.

Existem vários caminhos para abordar análises de dados sob modelos lineares generalizados com superdispersão, dentre os quais se destaca a abordagem através da composição do modelo de locação, amplamente usada na literatura. O ajuste desses modelos, em geral, é feito com algoritmo EM para estimação de máxima verossimilhança. Utilizado por Aitkin (1996), o algoritmo é obtido como uma forma de quadratura gaussiana assumindo uma distribuição de mistura não necessariamente normal, inclusive poderia ser usado também nos casos onde a distribuição é obtida por métodos não paramétricos. A idéia geral é baseada em Anderson & Hinde (1988), a qual é sintetizada como segue.

Seja $(Z_1, \dots, Z_n)$, uma amostra aleatória extraída de uma distribuição $f(z | \theta)$ pertencente à família exponencial com presença de extra-variação, parâmetro canônico θ , valor esperado μ e, variáveis explanatórias $X = (x_1, \dots, x_n)$

relacionadas com μ através da função de ligação $\eta_i = h(\mu_i)$, com η_i o preditor linear. Para esse preditor linear se propõe um efeito não observado u_i , para a i -ésima observação, com u_i assumindo, inicialmente, uma distribuição normal padrão. Assim, condicional à u_i , z_i é um modelo linear generalizado com preditor linear

$$\eta_i = \beta' x_i + \sigma u_i$$

Logo, a superdispersão é modelada na mesma escala que o preditor linear. Segundo Aitkin (1996), isso constitui uma escolha natural para modelagem no caso em que a superdispersão for originada pela heterogeneidade não controlada, e devida à omissão de importantes variáveis explanatórias. Estabelecido assim o modelo, calcula-se a função de verossimilhança, agora em função de (β, σ) , e estimam-se os parâmetros através do algoritmo EM. Esse modelo é usualmente chamado de Modelo Poisson Composto.

Segundo Fioravante (1993), o algoritmo EM, apresentado por Dempster *et al.*, (1977)¹ é um método iterativo para cálculo de estimativas de máxima verossimilhança quando as observações podem ser consideradas como dados incompletos. Cada iteração do algoritmo consiste em um passo de esperança (passo-E) seguido por um passo de maximização (passo-M). O processo EM é notável, em parte, devido à simplicidade e generalidade da teoria a ele associada, e em parte devido à grande amplitude de casos que recaem no seu âmbito de aplicação. Em termos gerais, seja $f(z | \phi)$ pertencente à família exponencial, escrita da forma:

$$f(z | \phi) = \frac{b(z)}{a(\phi)} \exp \{ \phi t(z)^T \} \quad (7)$$

onde ϕ denota um vetor de $1 \times p$ de parâmetros $\in \Omega$, e $t(z)$ denota um vetor de estatística suficientes dos dados completos. Assim, cada iteração pode ser escrita em dois passos:

1. Passo-E: estima-se $t(z)$ calculando

$$t^{(s)} = E[t(z) | z, \phi^{(s)}]$$

2. Passo-M: determina-se $\phi^{(s+1)}$ como o valor do $\phi \in \Omega$ que maximiza

$$-\log a(\phi) + \phi t^{(s)T}$$

¹DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, v. 39, n. 1, p.1-38, 1977.

Aitkin & Francis (1995) propuseram a obtenção dos erros padrões corretos das estimativas usando a propriedade que, para amostras grandes provenientes de um modelo com função de verossimilhança quadrática nos parâmetros, o teste do quociente de verossimilhança e o teste de Wald para a significância de um parâmetro na reta real são equivalentes, implicando que a mudança na "deviance", caso a variável seja omitida, é igual ao quadrado da estatística t (estimativa do parâmetro/desvio padrão da estimativa).

Na realidade foi Hinde (1982) quem apresentou primeiramente um modelo desse tipo, o modelo Poisson Composto, para dados de contagens com extra-variação, o mesmo que Lindsey (1993) chamou de modelo composto normal, dado que consiste simplesmente em supor uma distribuição normal para os efeitos aleatórios do modelo. Isto é, se Z_{ij} tem distribuição Poisson (ou Binomial), então um modelo de efeitos aleatórios é:

$$g(\mu_{ij}) = \lambda_j + \sum_{l=1}^p \beta_{il} x_{il}$$

com

$$\lambda_j \sim N(0, \delta)$$

e a distribuição composta resulta:

$$P(z_{ij}) = \int P(z_{ij} \mid \lambda_j) \phi\left(\frac{\lambda_j}{\sqrt{\delta}}\right) d\lambda_j \quad (8)$$

onde $\phi(\cdot)$ é a função de densidade para a distribuição normal padrão. Note-se que a variância da distribuição das contagens, τ_j^2 , e assim a correlação intra-classe, $\rho_j = \delta / (\tau_j^2 + \delta)$, será uma função do valor esperado μ_{ij} . O autor propôs o método de estimação permitindo que seja incorporada essa variável aleatória (independente) no modelo linear para a média de uma distribuição Poisson, e afirma que esse método refere-se principalmente ao fenômeno de superdispersão ocasionado pelo fato de não possuir informação acerca de importantes variáveis explanatórias que deveriam ter sido incluídas na modelagem. Porém, o método poderia ser utilizado em outras situações, como as surgidas nos casos onde o mecanismo não pode ser determinado sem informação adicional.

Assim, a utilização desse algoritmo para dados de contagem com superdispersão se justifica no fato de que são dados incompletos, uma vez que as observações não trazem informação a respeito da variável explanatória introduzida no modelo. A aplicação do algoritmo permite que o ajuste do modelo composto seja feito através de métodos normalmente utilizados para dados completos, já que divide

o processo de estimação em dois passos, estimando primeiramente a informação dada pela variável não observada, e em seguida, os parâmetros com base na estimativa calculada no primeiro passo.

Cabe assinalar que, sob este enfoque, não é possível obter uma expressão explícita para a distribuição fornecida pela integral, o qual difere daqueles modelos, também chamados de compostos, que serão apresentados na seção (5.1). O resultado disso é uma função de verossimilhança complexa, a qual pode ser avaliada apenas por integração numérica. Lindsey (1993) sugere usar outra distribuição para compor (embora a distribuição normal represente uma alternativa em muitos casos, robusta), para cuja escolha precisa de um "background" teórico do problema no que diz respeito ao comportamento da resposta média ao longo da população. Jorgensen *et al.* (1991) usaram uma outra possibilidade, compor com uma distribuição Inversa Gaussiana.

Ainda mais, Aitkin & Francis (1995) propõem o ajuste de modelos lineares generalizados com superdispersão através do enfoque de máxima verossimilhança **não paramétrica**. Essa técnica consiste em adicionar ao componente sistemático um efeito aleatório não observado, modelado por uma *distribuição de mistura* estabelecida como discreta sobre um número finito de pontos de massa. Esse enfoque não será abordado neste material, mas cabe assinalar que as macros para a obtenção das estimativas dos parâmetros são aquelas apresentadas por Hinde & Wood (1987), contendo apenas informação adicional acerca da estrutura de superdispersão para a função de variância.

Um outro caminho para abordar análises de dados de contagem sob modelos lineares generalizados com superdispersão, caso se conheça a estrutura de superdispersão como sendo uma função quadrática da média, é compor o modelo probabilístico, e não o preditor linear. Um modelo muito usado é aquele onde:

1.

$$\begin{aligned}
 i - Z_{ij} : P(z_{ij} \mid \mu_{ij}) &= \frac{e^{-\mu_{ij}} \mu_{ij}^{z_{ij}}}{z_{ij}!} \\
 ii - \mu_{ij} : f(\mu_{ij}) &= \frac{\nu_j^{-\kappa_j} \mu_{ij}^{\kappa_j-1} e^{-\frac{\mu_{ij}}{\nu_j}}}{\Gamma(\kappa_j)}
 \end{aligned}$$

2. logo, integrando em relação a μ_{ij} obtém-se a distribuição marginal do Z_{ij} , isto

é

$$\begin{aligned}
 P(z_{ij}) &= \frac{\nu_j^{\kappa_j} \int e^{-\mu_{ij}} \mu_{ij}^{z_{ij}} \mu_{ij}^{\kappa_j-1} e^{-\frac{\mu_{ij}}{\nu_j}} d\mu_{ij}}{\Gamma(\kappa_j) z_{ij}!} \\
 &= \frac{\Gamma(z_{ij} + \kappa_j)}{z_{ij}! \Gamma(\kappa_j)} \left(\frac{1}{1 + \nu_j} \right)^{\kappa_j} \left(\frac{\nu_j}{1 + \nu_j} \right)^{z_{ij}}
 \end{aligned}$$

o qual conduz ao modelo Binomial Negativo para as contagens, para o qual a seção (5.1) apresenta alguns dos caminhos possíveis para sua estimação.

3.4 Teste para Detectar Superdispersão no Modelo Poisson

Esta seção apresenta um teste para detectar a variação Extra-Poisson em modelos de regressão. O teste foi introduzido por Dean & Lawless (1989), unificando e estendendo os trabalhos de Collings & Margolin (1985) para testar esse fenômeno em alguns modelos de Poisson arbitrários. Em termos gerais, a construção do teste é como segue.

Os modelos Poisson estabelecem que, para variáveis de resposta independentes Z_i e vetores de covariáveis associados X_i , para $i = 1, \dots, n$, a distribuição do Z_i dado X_i é Poisson com média $\mu_i = \mu_i(X_i; \beta)$, onde β é um vetor p -dimensional de parâmetros desconhecidos. Por convenção, denotaremos $Z_i \sim P(\mu_i)$, (a partir deste, é amplamente usado o modelo log-linear, $\mu_i = \exp(X_i' \beta)$).

Para considerar a possível variação extra-Poisson estende-se o anterior para modelos de efeitos aleatórios: sejam $\nu_1, \dots, \nu_n$ variáveis aleatórias contínuas, *i.i.d.* com valores positivos tais que, dado X_i e ν_i , $Z_i \sim P(\nu_i \mu_i)$. Assume-se que os ν_i 's têm os dois primeiros momentos finitos, e sem perda da generalidade, escolhe-se $E(\nu_i) = 1$ e $Var(\nu_i) = \tau$.

Para este caso, e como será provado neste trabalho, tem-se que

$$Var(Z_i | X_i) = \mu_i + \phi \mu_i^2$$

pelo que pode-se testar o **modelo Poisson** versus o **modelo Extra-Poisson** estabelecendo, $H_0 : \phi = 0$ versus $H_1 : \phi > 0$.

Dean & Lawless (1989) estenderam este teste a situações de regressão mais gerais, obtendo a seguinte estatística de teste:

$$T = \frac{1}{2} \sum_{i=1}^n (Z_i - \hat{\mu}_i)^2 - Z_i \quad (9)$$

onde $\hat{\mu}_i = \mu_i(X_i; \hat{\beta})$, sendo $\hat{\beta}$ o estimador de máxima verossimilhança do β sob o modelo Poisson. Os autores apresentam as distribuições limites para versões padronizadas e apropriadas de T e as aproximações para os cálculos dos níveis de significância. Note-se que valores positivos grandes desta estatística indicam superdispersão; valores negativos grandes indicam subdispersão, relativo ao modelo Poisson (isto é, sob H_0).

Quando n é fixado e os $\mu_i \rightarrow \infty$, uma extensão direta do argumento do artigo original mostra que o teste para $\phi = 0$ baseado em (9) é aquele baseado em

$$T_2 = \sum_{i=1}^n (Z_i - \hat{\mu}_i)^2 / Z,$$

onde $\bar{Z} = \sum Z_i / n$, são assintoticamente equivalentes.

Além da exploração fornecida pela Figura 4 no Apêndice A, a qual expressa claramente o desvio que existe entre as distribuições Poisson (a originalmente postulada, sem nenhum conhecimento prévio) e a observada para os dados de mortalidade de *Eucalyptus grandis*. O Apêndice B apresenta a macro no GLIM4.0 para o cálculo da estatística.

Os valores obtidos para as estatísticas são mostrados na Tabela 3; indicando que a distribuição em estudo não é a Poisson. Ainda mais, o teste de Dean & Lawless (1989) indicou que existe uma extra-variação, caso seja adotado o modelo probabilístico Poisson, ou um outro modelo com distribuição agregada. Neste caso um outro modelo a ser testado poderia ser o **Binomial Negativo**.

Para populações que seguem esta distribuição, o parâmetro ϕ , agora denotado por $1/\lambda$, na estrutura $(\mu + \frac{1}{\lambda}\mu^2)$, representa o índice de agregação que permanece inalterado quando se produzem mudanças na densidade devidas às mortes aleatórias. Na medida que o λ for pequeno, o padrão é mais agregado, e quando $\lambda \rightarrow \infty$, mais aleatório. Esse comportamento foi definido por Macchiavelli (1987) como **processo de composição** e será apresentado mais completamente na seção (5.1). Esse modelo é adequado para os casos em que as observações não são independentes (estão correlacionadas), ou também quando existem variáveis explanatórias importantes que não foram avaliadas no ensaio e, conseqüentemente, excluídas na relação proposta no modelo matemático.

Smith & Heitjan (1993) forneceram um teste, baseado na estatística *score*, para detectar desvios da dispersão nominal nos modelos lineares generalizados, e apresentaram também um método para ajustar a dispersão nominal caso a superdispersão esteja presente. A idéia é relativamente simples, porém, a proposta analítica resulta complicada de ser implementada. O estabelecimento do MLG é

Tabela 3: Valores das estatísticas para exploração do modelo probabilístico da mortalidade das árvores de *E. grandis*.

Estatística Usada	Valor obtido
Teste de Dispersão	
Região1xRotação1	8.996,58
Região1xRotação2	2.063,51
Região1xRotação3	1.534,52
Região2xRotação1	1.615,68
Região2xRotação2	3.329,85
Região2xRotação3	463,77
Região3xRotação1	3.870,08
Região3xRotação2	6.215,09
Região3xRotação3	0,89
Teste de Extra-Poisson	
Região1xRotação1	16.955,13
Região1xRotação2	1.763,42
Região1xRotação3	865,02
Região2xRotação1	1.435,48
Região2xRotação2	2.452,97
Região2xRotação3	234,71
Região3xRotação1	2.730,14
Região3xRotação2	4.622,08
Região3xRotação3	0,69

semelhante ao de McCullagh & Nelder (1989) até a definição do preditor linear, ou seja, $\eta_i = g(\mu_i)$ onde $g(\mu_i) = \tilde{\beta}_i^T x_i$, com $i = 1, \dots, n$ e $\tilde{\beta}_i$ é o vetor de coeficientes gerado por uma distribuição F desconhecida, com média β e variância igual à $\text{diag}(\tau)$, $\tau = (\tau_1, \dots, \tau_p)^T$. Logo, se $\tau_j = 0$, $\forall j = 1, \dots, p$, então η corresponde a um modelo linear generalizado sem superdispersão, caso contrário, quando $\tau_j > 0$, a variância de z é acrescentada pela variabilidade no $\tilde{\beta}_i$, resultando um modelo com superdispersão. A estatística do teste proposta segue assintoticamente uma distribuição χ^2 , com p graus de liberdade, e utiliza fortemente as condições de regularidade do modelo e o fato de as estimativas serem as de máxima verossimilhança sob $H_o : \tau = 0$.

4 MODELO EXTRA-POISSON DE BRESLOW

4.1 Introdução

O enfoque de Breslow (1984), baseado nos trabalhos de Williams (1982) e Margolin *et al.* (1981) para o caso extra-binomial, tornou-se o mais usado na literatura, em geral, para ajustes de modelos lineares generalizados para dados de contagem que apresentem superdispersão. Esse tipo de modelagem utiliza-se, principalmente, na análise de experimentos chamados de *Bioensaios*, freqüentes na área de Toxicidade Genética, onde teve sua origem, expandindo-se depois a áreas como a Entomologia, Horticultura, Epidemiologia, entre outras.

O modelo Extra-Poisson, como assim foi chamado pelo autor, nada mais é que um modelo Poisson que incorpora uma estrutura extra de variação para as observações. Como será visto nesta seção, o algoritmo de estimação incorpora uma nova metodologia e a implementação computacional foi realizada no GLIM. Esse algoritmo não requer muito trabalho computacional (não mais do que um MLG usual), razão pela qual é o preferido na literatura em geral, como Wagner (1994)¹; Elder (1995); Bennett (1989); Morgan & Smith (1992); Smith & Morgan (1989); Breslow (1990), dentre outros.

Originalmente, Breslow (1984) apresentou esse modelo, assumindo as Z_{ij} como distribuídas por uma Poisson, com média μ_i , onde μ_i é a realização de uma variável aleatória Λ_i obedecendo à relação $\ln(\Lambda_i) = x_i\beta + \epsilon_i$, sendo β o vetor dos parâmetros ou coeficientes de regressão, e ϵ_i o vetor dos erros *i.i.d.*, com média 0 e variância comum σ^2 . A distribuição marginal de Z foi estabelecida de modo que o valor esperado $\mu = \mu(x, \beta)$ e a variância $V(\mu; \phi) = \mu(x, \beta) + \phi\mu^2(x, \beta)$, onde ϕ é o parâmetro de heterogeneidade ou parâmetro de variância. A estimação foi feita através do algoritmo de mínimos quadrados ponderados iterativos usando os recursos fornecidos pelo GLIM.

Esse tipo de estrutura de média-variância, ou melhor, da variân-

¹Bruce Wagner. US Department of Agriculture, Rochester (USA). Comunicação Pessoal. Maio de 1994.

cia como uma $f(\mu)$, possui uma interpretação valiosa, principalmente na área das Ciências Biológicas, a qual é comentada na seção (5.1). Além disso, os modelos probabilísticos que apresentam essa relação quadrática entre os dois primeiros momentos possuem propriedades interessantes para a realização de inferências quando o enfoque de análise é um modelo exponencial de dispersão, como será exposto na seção (6).

Considerando a existência de extra-variação em um modelo Poisson, Moore (1986), independentemente do apresentado por Breslow, mostra que pode ser estabelecido e ajustado univocamente um modelo de regressão no caso de se dispor somente de informação acerca dos dois primeiros momentos de Z , sendo essa informação a relação média-variância.

Margolin *et al.* (1981) discute algumas das possíveis causas do fenômeno de superdispersão para dados de natureza Binomial Positiva, dando ênfase à presença de correlação entre as unidades experimentais. Smith & Morgan (1989) apresentam macros em GLIM para ajustar um modelo linear generalizado para dados de respostas a doses, segundo um modelo Poisson, ilustrando com uma aplicação na área de Entomologia, ou seja o problema de Wadley acerca de moscas de frutos. Os autores construíram as macros para a estimação do modelo usando a equação do Breslow (1984) para a obtenção da estimativa do parâmetro de superdispersão. Morgan (1992) dedica um capítulo completo para o tratamento de dados binários e de contagem com superdispersão, e apresenta as diferentes escolhas de análises para a estimação de parâmetros.

Por outro lado, outras especificações da função de variância foram apresentadas na literatura, como, por exemplo, $V(\mu, \sigma) = \phi^2 \mu^b$ (Breslow, 1990), ou $V(\mu, \phi) = \mu/\phi$ (McCullagh & Nelder, 1989). Lembrando então, que a função de variância descreve o **tipo de superdispersão**, onde esta pode ser pensada como uma função que assumi o valor nulo. Esse fenômeno também pode ser especificado diretamente assumindo uma distribuição específica para o parâmetro, como será visto na seção seguinte que, para a primeira,

$$V(\mu, \sigma) = \phi^2 \mu^b$$

corresponde a um modelo Gama*Poisson, e para a segunda,

$$V(\mu, \phi) = \mu/\phi$$

a um modelo Poisson*Poisson.

O enfoque de Breslow (1984), comumente, é apresentado como QL/M (máxima quase-verossimilhança/método dos momentos), já que usa esses

métodos de estimação combinando-os, respeitando unicamente a estrutura média-variância da variável aleatória. Dessa maneira, os parâmetros do modelo de locação são estimados por quase-verossimilhança e o parâmetro da dispersão ou de variação, em princípio, pelo método dos momentos. Díaz (1994) descreve alguns exemplos clássicos e apresenta também o enfoque para amostras grandes usando aproximação normal para quando é conhecido o número total de indivíduos na amostragem. Segundo Morgan (1992), o método de quase-verossimilhança fornece uma ferramenta poderosa e de fácil implementação para os usuários de pacotes tais como o GLIM ou GENSTAT.

A função de quase-verossimilhança foi introduzida por Wedderburn (1974) com o objetivo de estender o tratamento dos modelos lineares generalizados permitindo que as pressuposições estocásticas necessárias na especificação do componente aleatório possam ser substituídas por outras, muito menos exigentes, definidas simplesmente pelo primeiro e segundo momento distribucional. Segundo Nelder & Pregibon (1987), Wedderburn fez semelhante à façanha de Gauss, quando substituiu o suposto de normalidade nos modelos lineares clássicos pela homogeneidade de variâncias. Nelder (1985) afirma que a QV ganhou o interesse na literatura principalmente porque permitiu que os MLG fossem especificados usando somente a função de ligação e a função de variância, em vez da função de distribuição.

Em termos intuitivos, a estimação de QV pode considerar-se como uma estimação **robusta**. Como será apresentado na seção (5.4), embora o método produza estimadores dos $\beta's$ assintoticamente eficientes para o modelo verdadeiro Binomial Negativo restrito a um parâmetro de forma constante, com o método dos momentos há alguma perda da eficiência na estimação da variância.

A seguir se aborda o modelo de quase-verossimilhança propriamente dito e as respectivas equações de estimação para o caso Extra-Poisson.

4.2 Quase-Verossimilhança

Continuando com a notação, seja Z_{ijk} o número de árvores mortas para o grupo ij , com $i, j = 1, 2, 3$ e $k = 1, \dots, n_{ij}$. Assume-se que os Z_{ijk} são independentes, e pode-se supor que Z_{ijk} tem, aproximadamente, uma distribuição Poisson. Portanto, no contexto dos modelos lineares generalizados de McCullagh & Nelder (1989), a função de ligação natural é a logarítmica. Assim, estendendo as idéias de Breslow (1984) a um modelo de locação de análise da variância, considera-se o componente sistemático η relacionado com o valor esperado na forma logarítmica,

isto é,

$$\begin{aligned} E(Z_{ijk}) &= \mu_{ij} \\ &\cong \exp\{(X\beta)\} \\ &\cong \exp\{\mu + \alpha_i + \gamma_j + \alpha\gamma_{ij} + \beta x_{ijk}\} \end{aligned}$$

e também (e somente isso), a expressão para a variância,

$$V(Z_{ijk}) \cong \mu_{ij} + \phi\mu_{ij}^2$$

Isso implica que se está impondo uma estrutura de variação para a variável Z condicionada, e para seu valor esperado, condicionado também, ou que

$$\begin{aligned} E\{V(Z_{ijk} | m_{ij})\} &= \mu_{ij} \\ V\{E(Z_{ijk} | m_{ij})\} &= \phi\mu_{ij}^2 \end{aligned}$$

onde m_{ij} são tais que $P(m_{ij})$. De fato:

$$\begin{aligned} E\{V(Z_{ijk} | m_{ij})\} &= E\{E(Z_{ijk}^2 | m_{ij})\} - E\{[E(Z_{ijk} | m_{ij})]^2\} \\ &= V(Z_{ijk}) - E\{[E(Z_{ijk} | m_{ij})]^2\} + E\{[E(Z_{ijk} | m_{ij})]^2\} \\ &= V(Z_{ijk}) - V\{E(Z_{ijk} | m_{ij})\} \end{aligned}$$

$$\begin{aligned} \Rightarrow V(Z_{ijk}) &= E\{V(Z_{ijk} | m_{ij})\} + V\{E(Z_{ijk} | m_{ij})\} \\ &= \mu_{ij} + \phi\mu_{ij}^2 \end{aligned}$$

Note-se que, se $\phi = 0$, o valor esperado de Z_{ijk} , dado m_{ij} , têm variância 0, resultando em um modelo com a mesma variância que o modelo Poisson. Ou seja, têm-se uma variável aleatória Z_{ijk} cuja distribuição não é completamente especificada.

4.2.1 Método de Estimação

Em termos gerais, suponha que z_{ijk} , $k = 1, \dots, n_{ij}$ são observações independentes com esperanças μ_{ij} e variâncias $V(\mu_{ij})$, onde V é alguma função conhecida. Além disso, e para cada observação, μ_{ij} é alguma função conhecida de um conjunto de parâmetros $\beta_1, \dots, \beta_p$. Então, para cada observação define-se a função de quase-verossimilhança, $K(z_{ijk}, \mu_{ij})$, pela relação

$$\frac{\partial K(z_{ijk}, \mu_{ij})}{\partial \mu_{ij}} = \frac{z_{ijk} - \mu_{ij}}{V(\mu_{ij})} \quad (10)$$

ou, equivalentemente

$$K(z_{ijk}, \mu_{ij}) = \int_0^{\mu_{ij}} \frac{z_{ijk} - \mu_{ij}}{V(\mu_{ij})} d\mu_{ij} + f(z_{ijk})$$

o qual conforme COX (1983) pode-se expressar como,

$$K(z_{ijk}, \mu_{ij}) = z_{ijk}\theta_{ijk} - b(\theta_{ijk}) + c(z_{ijk}; \phi)$$

onde se satisfaz:

$$\begin{aligned} \theta_{ijk} &= \int V(\mu_{ij})^{-1} d\mu_{ij} \\ b'(\theta_{ijk}) &= \mu_{ij} \\ b''(\theta_{ijk}) &= d\mu_{ij}/d\theta_{ijk} \\ &= V(\mu_{ij}) \end{aligned}$$

Portanto, a função de probabilidade de Z_i pode ser escrita na forma da família exponencial de distribuições com um parâmetro. Logo, dado que a recíproca é verdadeira, admitir que as observações têm distribuição na família exponencial a um parâmetro, equivale, simplesmente, supor uma relação média-variância para os dados.

Em outras palavras, Cox (1983) afirma que, formular uma família exponencial de distribuições é a *mais fraca suposição distribucional que pode ser feita para os dados*. Pode ser difícil decidir qual deve ser a distribuição populacional, mas a forma da relação média-variância é muito mais simples de ser postulada, o que torna a quase-verossimilhança de grande utilidade. Aliás, Wedderburn (1974) postula o modelo de quase-verossimilhança como aquele que especifica a função de variância **a menos de uma constante multiplicativa**. Com a definição (10), podem ser demonstradas várias propriedades desta função, algumas das quais são comuns com as da função log-verossimilhança; como por exemplo, no caso em que a distribuição de Z^1 pertence à família exponencial uniparamétrica, K é a função logaritmo da função de verossimilhança.

Uma outra propriedade é a seguinte: se as observações são não correlacionadas então a *função de covariância* $V(\mu)$, tal que

$$Cov(Z) = \varphi^{-1}V(\mu)$$

¹Para alguns enunciados de propriedades da função de quase-verossimilhança, sem perda da generalidade, se considerará z_{ijk} como z .

φ constante de proporcionalidade, $V(\mu)$ semi-definida positiva, é da forma,

$$V(\mu) = \text{diag}\{V(\mu_1), \dots, V(\mu_n)\}$$

e conclui-se que o logaritmo da função de quase-verossimilhança para as

$$n = \sum_{i,j} \sum_{k=1}^{n_{ij}} n_{ijk}$$

observações iguala à soma das n contribuições individuais,

$$K(\mu; z) = \sum_{i=1}^n K(\mu_i; z_i)$$

Entretanto, isso, em geral, não é satisfeito, já que superdispersão está frequentemente presente nos dados.

O seguinte *Teorema* resume outras características da função de quase-verossimilhança, imediatas a partir da sua definição.

Teorema 4.1 *Sejam Z e K definidas como em (10), e suponha que μ é função dos parâmetros $\beta_1, \dots, \beta_p$. Então K tem as seguintes propriedades:*

- $E\left(\frac{\partial K}{\partial \mu}\right) = 0,$
- $E\left(\frac{\partial K}{\partial \beta_i}\right) = 0,$
- $E\left(\frac{\partial K}{\partial \mu}\right)^2 = -E\left(\frac{\partial^2 K}{\partial \mu^2}\right) = \frac{1}{V(\mu)},$
- $E\left(\frac{\partial K}{\partial \beta_i} \cdot \frac{\partial K}{\partial \beta_j}\right) = -E\left(\frac{\partial^2 K}{\partial \beta_i \partial \beta_j}\right) = \frac{1}{V(\mu)} \frac{\partial \mu}{\partial \beta_i} \frac{\partial \mu}{\partial \beta_j}$

Quando $V(\mu) = 1$, a estimação de MQV se reduz a mínimos quadrados. Se $V(\mu_i) = \mu_i$, ou, $V(\mu_i) = \mu_i + \mu_i^2/\lambda$, obtêm-se, integrando (10), os respectivos logaritmos da função de quase-verossimilhança:

$$z_i \log \mu_i - \mu_i$$

$$z_i \log[\mu_i/(\phi + \mu_i)] + \phi \log[\phi/(\phi + \mu_i)]$$

os quais correspondem, a menos de constante, aos modelos Poisson e Binomial Negativo, respectivamente; e se $V(\mu_i) = \varphi \mu_i$,

$$K(z_i, \mu_i) = \varphi^{-1} [z_i \log \mu_i - \mu_i]$$

Neste último caso, se φ for igual a um, então a função de QV coincide com a função de verossimilhança, e segundo Jorgensen (1983), se $\varphi \neq 1$ a função de QV não é função de verossimilhança de nenhuma família de distribuições.

A precisão dos estimadores de máxima QV pode ser estimada a partir dos valores esperados das segundas derivadas de K de forma semelhante ao que acontece com os estimadores de máxima verossimilhança em relação à logaritmo da função de verossimilhança. Para cada observação, seja u o vetor com componentes $\partial K / \partial \beta_i$. Então, usando as propriedades do Teorema (6.3) tem-se que u tem média zero e matriz de dispersão com elementos

$$-E\left(\frac{\partial^2 K}{\partial \beta_i \partial \beta_j}\right)$$

satisfazendo a

$$-E\left(\frac{\partial^2 K}{\partial \beta_i \partial \beta_j}\right) \leq -E\left(\frac{\partial^2 L}{\partial \beta_i \partial \beta_j}\right)$$

como consequência imediata da desigualdade de Cramér-Rao (Wedderburn; 1974). Esta expressão torna-se uma igualdade, quando a L é a logaritmo da função de verossimilhança da família exponencial de distribuições. O membro esquerdo dessa fórmula é uma medida da informação, quando se conhece apenas a relação média-variância dos dados, enquanto que o direito, é a informação obtida pelo conhecimento da distribuição dos dados.

Portanto, dada a estrutura média-variância

$$v_i^* = \text{var}(Z_i) = \mu_i + \phi \mu_i^2$$

o modelo descrito por:

$$\begin{aligned} E(Z) &= \mu(\beta) \\ \text{Cov}(Z) &= \varphi V(\mu) \end{aligned}$$

é chamado **Modelo de Quase-Verossimilhança**, para o qual as funções $v_i^*(.)$ são conhecidas.

Note-se que φ não é constante, isto é, φ_i é igual a $1 + \phi \mu_i$. Muitas estruturas de extra-variação envolvem a estimação de parâmetros de dispersão independente dos parâmetros do preditor linear, de tal forma que seus processos de estimação também funcionam independentemente. Isso não acontece no caso da

estrutura de variação em estudo $(\mu + \phi\mu^2)$, já que existe uma função quadrática envolvendo os valores preditos.

Como deveríamos generalizar o método de mínimos quadrados para obter a estimação do β , quando $V = V(\mu)$?. Em termos matriciais, o método equivalente é da forma:

$$(Z - \mu)'V^{-1}(\mu)D = 0' \quad (11)$$

Note-se que, como em mínimos quadrados, as equações a serem resolvidas *não dependem* do parâmetro de superdispersão φ .

A expressão (11) foi interpretada por Firth (1990) como segue: **o método de quase-verossimilhança generaliza o dos mínimos quadrados como um método geométrico (mais que algébrico)**. Assim, e semelhante ao que ocorre em um modelo normal de Gauss-Markov, o estimador de QV do vetor β é uma projeção ortogonal dos dados no modelo definido.

Assim, e fazendo algumas passagens de caráter algébrico, as equações de estimação possuem a seguinte expressão:

$$f_{ij}(\beta, \phi) = \sum_{k=1}^{n_{ij}} \left\{ \frac{(Z_k - \mu_k \mathbf{X}_k' \beta)}{v * k} \right\} \Delta_k X_{kij} = 0, \quad (12)$$

onde $\Delta_k = \partial\mu(\eta)/\partial\eta$ e $\mu(\eta) = \exp\{\eta\}$ expressa a média de Z como uma função das covariáveis. No caso de assumir função de ligação *log* (canônica para o modelo Poisson), tem-se

$$\Delta_k = \mu_k$$

e (12) se simplifica para

$$\sum_{k=1}^{n_{ij}} \frac{(z_k - \mu_k(X' \beta))}{1 + \phi \mu_k} x_{kij} = 0 \quad (13)$$

Assim, se ϕ for conhecido, a solução do vetor β baseada na QV, seria obtida através das seguintes equações de estimação:

$$U(\beta; \phi) = \sum_{i=1}^n \frac{z_i - \mu_i}{V(\mu_i; \phi)} \frac{\partial \mu_i}{\partial \beta} = 0$$

Observe-se que, dizer "se ϕ for conhecido" significa que ϕ funciona como constante estimada.

Porque se estima ϕ separadamente?. A matriz de variâncias-covariâncias assintótica dos estimadores dos parâmetros de regressão, obtida da equação (11), depende somente do primeiro e segundo momentos. Portanto, os cálculos dos erros padrões são diretos e não requerem pressuposições adicionais (o que é muito importante na prática), a saber,

$$Cov(\hat{\beta}) \approx \phi(X'V^{-1}X)^{-1}$$

Dado que X e V dependem do β , estimam-se os desvios padrões, d.p. (β_{ij}) , substituindo $\beta = \hat{\beta}$, e assim se estima ϕ separadamente.

Como segunda etapa, na sequência, a **equação dos momentos** adicional para ϕ será:

$$\begin{aligned} U^*(\beta; \phi) &= \sum_{i=1}^n \left[\frac{(z_i - \mu_i)^2}{V(\mu_i; \phi)} - 1 \right] \\ &= \sum_{i=1}^n \left[\frac{(z_i - \mu_i)^2}{\mu_i + \phi \mu_i^2} - 1 \right] \\ &= \sum_{i=1}^n \left[\frac{(z_i - \mu_i)^2 - \mu_i - \phi \mu_i^2}{\mu_i + \phi \mu_i^2} \right] \\ &= \sum_{i=1}^n \left[\frac{z_i^2 + (1 - \phi) \mu_i^2 - (2z_i + 1) \mu_i}{\mu_i + \phi \mu_i^2} \right] \end{aligned}$$

Na prática, segundo Breslow (1984) e Morgan (1992), é importante corrigir para o "viés" quando p , a quantidade de parâmetros no modelo para o preditor linear, é muito grande comparado com o n , segundo

$$U^*(\beta; \phi) = \sum_{i=1}^n \left[\frac{(z_i - \mu_i)^2}{V(\mu_i; \phi)} - \frac{n - p}{n} \right] = 0$$

o qual não é de importância para os dados que este material aborda.

Dado que ϕ é não negativo, sua estimativa será igual a zero quando a equação de estimação leve a valores negativos. Cabe assinalar que essas mudanças não afetam as propriedades assintóticas desenvolvidas em Breslow (1990).

Mostradas essas equações para a obtenção das estimativas, independentemente, resulta o algoritmo geral de estimação. Em outras palavras, se ϕ for conhecido, a solução de máxima QV para os β 's pode ser facilmente obtida no GLIM usando função do erro Poisson (\$ERROR P\$); função de ligação natural logarítmica, e definindo os pesos *a priori* $w_i = (1 + \phi \hat{\mu}_i)^{-1}$ (\$CAL W=1/(1+PH*%FV)\$). Ainda no caso em que os pesos fossem usados erroneamente, um valor novo do ϕ pode ser outra vez estimado estabelecendo o critério do chi-quadrado (i.é, usando QV/M na sequência descrita na seção (4.2) como sendo igual aos seus correspondentes graus de liberdade, i.e.

$$\sum \frac{(z_i - \hat{\mu}_i)^2}{\{\hat{\mu}_i(1 + \phi \hat{\mu}_i)\}} = n - p$$

ou

$$\phi = (N - p)^{-1} \sum_{i=1}^n \left\{ \frac{(z_i - \hat{\mu}_i)^2}{\hat{\mu}_i(\hat{\mu}_i + \phi)} \right\} \quad (14)$$

Isto conduz ao chamado por Breslow (1984) **Procedimento II**:

1. Ajustar o modelo Poisson conforme a estrutura do modelo linear generalizado proposto. Se a "deviance" residual está próxima dos seus graus de liberdade, finaliza o processo e conclui-se que a variação residual está explicada satisfatoriamente.
2. Resolver (14) recursivamente para ϕ .
3. Definir pesos novos $w_i = (1 + \phi \hat{\mu}_i)^{-1}$ e voltar ao passo [1].

O Apêndice B contém o conjunto de macros em GLIM para a implementação do algoritmo de BRESLOW (1984).

Por outro lado, pode-se melhorar a velocidade de convergência do algoritmo propondo uma estimativa inicial do ϕ da forma:

$$\hat{\phi}_0 = \frac{\sum_{i=1}^n (z_i - \hat{\mu}_i)^2}{\hat{\mu}_i} - \frac{(n - p)}{\sum_{i=1}^n \hat{\mu}_i(1 - \hat{\mu}_i q_i)} \quad (15)$$

obtida de maneira análoga à (14) considerando o problema associado de mínimos quadrados. A matriz $\hat{W}$ dos pesos iterativos possui na diagonal elementos $\hat{\mu}_i$, e a matriz $X(X^T \hat{W} X)^{-1} X^T$, os q_i . De forma alternativa, o denominador de (15) pode ser aproximado por $(\sum \hat{\mu}_i)(n - p)/n$.

Para a obtenção da (14) construiu-se, no SPlus, o algoritmo de estimação de acordo ao método de Newton-Raphson, para o qual se tem:

$$\begin{aligned}
 f(\phi) &= \phi - (n - p)^{-1} \sum_{i=1}^n \left\{ \frac{(z_i - \hat{\mu}_i)^2}{\hat{\mu}_i(\hat{\mu}_i + \phi^{-1})} \right\} \\
 f'(\phi) &= 1 - (n - p)^{-1} \sum_{i=1}^n \left\{ \frac{(z_i - \hat{\mu}_i)^2 \phi^{-2}}{\hat{\mu}_i(\hat{\mu}_i + \phi^{-1})^2} \right\} \\
 \frac{f(\phi)}{f'(\phi)} &= \frac{\phi - (n - p)^{-1} \sum_{i=1}^n \left\{ \frac{(z_i - \hat{\mu}_i)^2}{\hat{\mu}_i(\hat{\mu}_i + \phi)} \right\}}{1 - (n - p)^{-1} \sum_{i=1}^n \left\{ \frac{(z_i - \hat{\mu}_i)^2 \phi^{-2}}{\hat{\mu}_i(\hat{\mu}_i + \phi)^2} \right\}}
 \end{aligned}$$

Assim, o processo iterativo tem a seguinte expressão:

$$\phi_{h+1} = \phi_h - \frac{\phi_h - (n - p)^{-1} \sum_{i=1}^n \left\{ \frac{(z_i - \hat{\mu}_i)^2}{\hat{\mu}_i(\hat{\mu}_i + \phi_h)} \right\}}{1 - (n - p)^{-1} \sum_{i=1}^n \left\{ \frac{(z_i - \hat{\mu}_i)^2 \phi_h^{-2}}{\hat{\mu}_i(\hat{\mu}_i + \phi_h)^2} \right\}}$$

O Apêndice B contém a "function" (macro) em SPlus para o algoritmo completo de estimação.

A função "deviance", já falou, é obtida analogamente ao feito para a função de verossimilhança, isto é, ela é definida como 2 vezes a diferença entre o máximo do logaritmo da função de quase-verossimilhança do modelo saturado e o máximo do logaritmo da função de quase-verossimilhança do modelo corrente, ou

$$\begin{aligned}
 D(z, \mu) &= -2\{K(z, \mu) - K(z, z)\} \\
 &= -2 \int_z^\mu \frac{z - \mu}{V(\mu)} d\mu
 \end{aligned}$$

e não pode ser utilizada como uma estatística para observar a aderência do mesmo. Para outros modelos, onde a expressão para os pesos não contenha justamente o valor estimado do parâmetro de dispersão, essa medida será a primeira inspeção a realizarmos acerca da aderência do modelo.

De fato, esse ajuste propõe estimativas de máxima quase-verossimilhança (MQV) para os parâmetros envolvidos no preditor linear (efeitos de interação e coeficiente linear) e estimativas do parâmetro de dispersão equacionando a estatística χ^2 aos seus respectivos graus de liberdade, para assim coloca-las como pesos no algoritmo de estimação dos β 's. Ao longo das iterações a "deviance" iguala-se aos graus de liberdade, o que faz que essa estatística não possa ser usada como diagnóstico do modelo maximal.

As inferências acerca dos termos do modelo (isto é, efeitos de interação ou efeitos principais, coeficiente de regressão, etc.) são feitas comparando modelos encaixados (hipóteses do tipo $H_1 \supseteq H_2 \supseteq \dots H_k$) usando uma estimativa comum obtida do χ^2 generalizado do modelo maximal. Ainda mais, se H_A e H_B são duas hipóteses encaixadas de dimensão $A < B$, então, sob H_A , a mudança na "deviance",

$$D(\hat{\mu}_B, \hat{\mu}_A) = D(z, \hat{\mu}_A) - D(z, \hat{\mu}_B)$$

tem uma distribuição assintótica χ^2_{B-A} .

Além disso, já foi mencionado que a distribuição dessa estatística, para a fonte de variação residual, não é conhecida mas, na prática, isso será usado no enfoque da seção (5.1), usa-se testar o ajuste de um modelo comparando-se sua "deviance" residual com o valor de uma estatística $\chi^2_{(n-p),\alpha}$, onde p é o número de parâmetros do modelo sob pesquisa, e α é o nível de significância adotado. Se o ajuste for bom, o valor da "deviance" residual observado deveria ser menor ou igual ao valor da $\chi^2_{(n-p),\alpha}$. Esse critério foi o usado para a rejeição do primeiro modelo, aquele sem superdispersão, representado pelo erro Poisson.

Assim, os dados de mortalidade de árvores de *E. grandis* foram submetidos a esses algoritmos de ajustes do modelo Extra-Poisson, no GLIM e no SPlus. A Tabela 4 apresenta os resultados das análises em GLIM e SPlus, constando os valores para as "deviances" iniciais (Dev. Ini.), isto é sem ajustar a superdispersão, ou seja com componente aleatório segundo um modelo Poisson, as "deviances" em cada iteração, ajustando a variação extra, com os respectivos graus de liberdade (g.l.), e as estimativas do parâmetro de superdispersão. A Tabela 5 apresenta as estimativas dos parâmetros, seus respectivos erros padrões, e estatísticas de Wald para cada um deles¹, obtidos no ajuste final, para o modelo de locação estabelecido como uma análise de covariância.

Para esse enfoque foram observados os valores estimados para os componentes de "deviance" padronizados e estudentizados externamente para a

¹Neste caso o valor da estatística de Wald coincide com o valor de t.

Tabela 4: Análise da "deviance" para o modelo Extra-Poisson para a variável mortalidade das árvores de *Eucalyptus grandis*, no GLIM e no SPlus.

Resultados obtidos no GLIM				
Estatísticas	Ajuste Inicial	1a. Iter.	2a. Iter.	3a. Iter.(g.l.)
"Deviance" Residual	33.083	2.578,03	2.597,54	2.596 (2.596)
$\hat{\lambda}$	–	0,8693	0,9299	0,9302
Resultados obtidos no SPlus				
Estatísticas	Ajuste Inicial	1a. Iter.	2a. Iter.	3a. Iter.(g.l.)
"Deviance" Residual	33.939	2.678	2.598,21	2.596 (2.596)
$\hat{\lambda}$	–	0,8793	0,9245	0,9298

aderência do modelo, dado que a simples observação da "deviance" não é o apropriado. No GLIM essas informações podem obter-se facilmente operando com os vetores do sistema %di (componentes da "deviance"), %rs (resíduos de Pearson), %le ("leverage"), %cd (distância de Cook), entre outros. Os resíduos componentes de "deviance" padronizados estiveram contidos, basicamente, no intervalo [–3, 54; 2, 90], embora tenham aparecido observações com valores de -7,53 e 10,40 para as quais seus baixos valores de "leverage" indicaram que não eram observação influentes para o modelo. Dado o tamanho de amostra que foi analisada ($n = 2.605$), a listagem desses resíduos, assim como os valores preditos não foram incluídos neste material², somente informa-se um resumo das suas características. O valor da mediana desses resíduos padronizados estimados foi –0,0142 e a distribuição mostrou assimetria à esquerda, indicando que o modelo superestima com frequência.

Porém, dada a estrutura e os fins do modelo estabelecido, isto é, um modelo de análise da covariância e não um modelo de regressão, o importante da análise é conseguir avaliar os efeitos de interação e/ou efeitos principais, controlando o erro experimental através da inclusão da variável idade das árvores. Ou seja, com este modelo não cabe fazer predições a nível do número de indivíduos mortos para uma determinada combinação de região e rotação, apenas inferir acerca da possível dependência entre os fatores. Portanto, e dada que a performance do modelo Extra-

²Essas listagens podem ser consultadas com a autora.

Tabela 5: Estimativas dos parâmetros, e seus respectivos erros padrões e estatísticas de Wald, sob enfoque Extra-Poisson.

Parâmetro	Estimativa	Desvio Padrão	Valor de t
μ	2,0200	0,0362	5,5801
β	0,1845	0,0096	19,2187
α_2	-0,1757	0,0658	2,6702
α_3	0,2567	0,0551	4,6588
γ_2	0,7369	0,0881	8,3644
γ_3	0,8145	0,1003	8,1206
$\alpha\gamma_{(2,2)}$	-0,2646	0,1352	-1,9571
$\alpha\gamma_{(2,3)}$	0,3552	0,2047	1,7352
$\alpha\gamma_{(3,2)}$	0,0815	0,1093	0,7456
$\alpha\gamma_{(3,3)}$	0,2232	0,7036	0,3174

Poisson em geral, poderse-ia considerar como o adequado para tais fins.

Assim, pode-se concluir que a mortalidade de árvores de *E. grandis* segue um modelo Poisson com extra-variação, e que existe interação, ao nível 0,10 de significância, entre os fatores região e rotação. Essa dependência estaria evidenciada na combinação da região compreendida pelo município de Brotas e Riberão Bonito e a rotação 2 da espécie. Embora não haja independência entre os fatores (e portanto conclusões isoladas acerca dos fatores se tornam incorretas) pode observar-se nas estatísticas de Wald que existiu um marcado efeito principal de rotação. A Figura 5 do Apêndice A mostra esse comportamento da variável mortalidade.

O índice de agregação foi de 1,0753, indicando que a extra-variação não é desprezível. Não obstante, da análise de Breslow, pouco se pode inferir acerca desse parâmetro ϕ , devido ao fato de que o algoritmo não fornece mecanismo nenhum para a obtenção de alguma medida da precisão da estimativa, como seria seu erro padrão.

Se, pelo menos, essas estimativas fossem obtidas mediante o método do perfil de verossimilhança, as propriedades assintóticas da função logaritmo da função de verossimilhança forneceriam valores para o erro padrão. Daí, a construção de intervalos de confiança permitiria a verificação de hipóteses acerca do índice de

agregação. Ou, relaxando ainda mais, se a regra de parada fosse construída no GLIM segundo o critério da diminuição do quociente de verossimilhança, e não segundo o decréscimo da "deviance", essa estimativa para o λ poderia ter sido considerada como a de máxima verossimilhança, facilitando a realização de inferências.

Porém, é claro que o enfoque de Breslow tenta colocar resposta aos problemas de estimação dos parâmetros embutidos no preditor linear, usando como ferramenta uma estimativa que diz-se ser de máxima verossimilhança para o parâmetro de superdispersão, e não inferir sob a ferramenta propriamente dita. Aliás, o autor propôs duas formas para se obter a estimativa, pelo método dos momentos e por máxima verossimilhança, mas como será apresentado na seção (5.4), para este modelo, a primeira delas não mereceu sequer ser implementada.

Observando os valores de superdispersão que encontrados na literatura para dados provenientes de diferentes áreas de pesquisa, isto é, em um intervalo $[0, 52, 0]$, pode-se arriscar significância para o valor do $\phi = 1,0753$ para os dados de mortalidade de *E. grandis*.

Por outro lado, o modelo Extra-Poisson indica, entre outras coisas, uma real falta de conhecimento do processo de atribuição de chances. Ou seja, embora ele seja um artifício totalmente novo e prático para lidar com problemas do tipo "variação extra" no caso Poisson, esse modelo não estabelece a verdadeira distribuição dos dados. Esse enfoque será exposto na seção (5.1) a seguir. Portanto, acredita-se que a construção de testes do tipo quociente de verossimilhança para, por exemplo, testar hipóteses acerca de modelos encaixados (com e sem interação, com e sem idade, etc) poderia ser inconsistente. Isto implica que a base para a obtenção de conclusões do modelo (performance e inferências sob os parâmetros) somente é regida pelas análises de diagnóstico.

Já para o caso dos modelos extra-binomial, Haseman & Kupper (1979) alertam para o uso dos testes quociente de verossimilhança como ferramenta para decisões e apresentam uma revisão junto com um estudo de simulação que mostra valores dos erros do **tipo I** muito superiores aos correspondentes valores nominais, no caso de duas amostras. Shirley & Kickling (1981) também apresentam estudos da potência desses testes e concluíram que nos casos extra-binomial, ou extra-Poisson, quando for possível, deveriam ser usados *testes não paramétricos* ou *transformações de dados*.

Questionando o modelo, existem fatos importantes a serem considerados. Por exemplo, e isto se relaciona com o diagnóstico propriamente dito de um MLG segundo o enfoque de Breslow, da definição da função de quase-verossimilhança

se deduz que a função de variância **determina** a unidade de medida para a $D(z, \hat{\mu})$ e X^2 , razão pela qual não é possível avaliar diferentes funções de variâncias.

A escolha ou seleção da função de variância torna-se uma etapa muito importante na construção do modelo estatístico (praticamente define a regra de atribuição de eventos da variável resposta). Portanto, precisa-se de alguma ferramenta de análise que vise testar a aderência da mesma. Para logr-lo h necessidade de estender a definio da funo de quase-verossimilhan de quase-verossimilhan estendida, proposta por Nelder & Pregibon (1987). Neste trabalho no ser abordado esse enfoque de anlise, apenas sua apresentao para conhecimento do leitor de uma famlia mais ampla no que diz respeito s restrios impostas na distribuio de uma varivel aleatria qualquer. Para uma s observo z , com mdia μ e varincia $\phi V(\mu)$, se define a **funo de quase-verossimilhan estendida** pela expresso:

$$K^+(z, \mu) = -\frac{1}{2} \log\{2\pi\phi V(z)\} - \frac{1}{2} D(z, \mu)/\phi$$

onde $D(z, \mu)$ a "deviance" definida anteriormente. Note-se, que K^+ (como o K) no precisa do conhecimento da distribuio da varivel aleatria, apenas os dois primeiros momentos. Sob o enfoque dos modelos exponenciais de disperso (seo (6) essa funo torna-se importante na estimao do parmetro de disperso, dado que $\exp(K^+)$ uma das aproximaos utilizadas (chamada de "saddlepoint")¹.

Um outro fato interessante em termos das interpretaos destes modelos a ser resgatado o seguinte: o que aconteceria com o desempenho das estatsticas dos testes para os parmetros (e no somente para eles) se o modelo no levasse em considerao a superdisperso?. Ou, em outras palavras, como seria a dependncia que essas estatsticas tem com os valores reais de agregao na distribuio das observaos?. Lembre-se que foi mencionado que, quando ϕ for nulo, as estimativas de MV e de MQV coincidem. Obviamente que se o modelo escolhido fosse o Poisson, os intervalos de confiana para parmetros seriam mais estreitos, ou, as estimativas sob esse modelo, menos eficientes. Ou seja, se considerarmos o modelo Poisson, sem declarar a extra-variao, os erros padres das estimativas resultam menor valor absoluto. Isto pode ser observado na Tabela 6.

Ou seja, incorporar a extra-variao implica eficincia na estimao. Mas, a eficincia, como uma das propriedades que torna mais valiosa qualquer es-

¹Para maiores detalhes a respeito, o leitor pode consultar Nelder & Pregibon (1987), ou Jorgensen (1989; 1992).

Tabela 6: Estimativas dos parâmetros, e seus respectivos erros padrões e estatísticas de Wald, sob modelo Poisson (sem extra-variação).

Parâmetro	Estimativa	Desvio Padrão	Valor de t
μ	2,0200	0,0128	157,8125
β	0,1959	0,0046	42,5869
α_2	-0,1757	0,0246	-7,1423
α_3	0,2567	0,0183	14,0273
γ_2	0,7369	0,0240	30,7042
γ_3	0,8145	0,0261	31,2069
$\alpha\gamma_{(2,2)}$	-0,2646	0,0416	-6,3606
$\alpha\gamma_{(2,3)}$	0,3552	0,0507	7,0059
$\alpha\gamma_{(3,2)}$	0,0815	0,0294	2,7721
$\alpha\gamma_{(3,3)}$	0,2232	0,1380	0,6174

timativa, significa, em termos práticos, possuir menor variância, o que sob este enfoque se traduz, para o caso de observações com superdispersão, ao inverso da sua definição.

Na realidade existe uma pequena sutileza na literatura em geral que confunde o leitor que tenta interpretar esse fenômeno. Isto é, a superdispersão deve ser entendida em todo o processo de análise, sem deixá-la de lado na hora de inferir acerca dos parâmetros. E é nessa hora que deve ser lembrado o já foi discutido na seção (3.3), acerca dos caminhos plausíveis que podem levar um conjunto de dados a apresentar extra-variação. Assim, se a superdispersão (estudado e descartado qualquer outro motivo que possa ser sanado na modelagem) pode ser causada pela existência de agrupamentos, ou correlação entre as realizações da variável que se pensaram como independentes, é lógico então que esses desvios padrões (e portanto os testes de hipóteses simples acerca dos parâmetros) apresentem grandeza maior que aqueles obtidos sem incorporar a extra-variação. Qual é o motivo então?: um só, **observam-se realmente um número menor de dados. Ou, observações correlacionadas aportam, em síntese, menos graus de liberdade**, e portanto não se pode considerar n como o número real de observações. Dessa maneira, a definição de eficiência continua sendo coerente aos argumentos básicos da Teoria

Estatística.

Cox (1983) mostrou que para quantias pequenas de superdispersão, a estimação pelo método MQV (estimação dos β 's para um ϕ dado) é altamente eficiente, como por exemplo cita o autor caso extra-binomial com $\phi < 0,2$. No nosso caso a estimativa dos momentos foi $\phi = 1/\hat{\lambda} = 1/0,93 = 1,0753$, pelo qual a alerta mencionada deveria ser considerada. Na seção (5.4) é apresentado um estudo da eficiência na estimação baseado na matriz de informação de Fisher. Esse estudo também foi motivado pelo que se segue.

O parâmetro de superdispersão sempre é escolhido tal que a estatística χ^2 esteja próximo dos seus graus de liberdade, uma vez ajustado o modelo maximal. Em tais circunstâncias, essa estatística consiste de dois componentes: um devido à variação em torno da média para um tratamento particular, e outro devido às discrepâncias entre as médias observadas em cada tratamento e seus valores estimados a partir do modelo ajustado. Esses dois componentes são análogas ao erro puro e ao termo falta de ajuste no modelo usual de regressão. Usá-las para estimar ϕ pode ser considerado como incorreto dado que **superestima** o parâmetro, conduz a erros padrões inflacionados dos parâmetros contidos no preditor linear e produz, então, testes de significância muito conservadores.

Por último e em uma tentativa de aportar elementos para se tomarem decisões, apresenta-se a seguir a construção de um teste **aproximado** da razão de verossimilhança, conforme Brooks (1984), o poderia ser uma saída simples e adequada para a realização das inferências. Isto é, assumindo um modelo discreto para as observações, com parâmetros desconhecidos contidos no preditor linear do MLG estabelecido, seja $L(\beta; \phi)$ a função de verossimilhança admitida como função dos parâmetros (avaliada na amostra). O método para a construção do teste é o seguinte:

- encontrar L , máximo sob os valores de ϕ de $L(\hat{\beta}; \phi)$, onde $\hat{\beta}$ é o estimador de máxima verossimilhança do β (sob o modelo maximal) para um valor dado de $\hat{\phi}$ na forma do algoritmo de Breslow (1984) e apresentado como **Procedimento II**, isto é, estabelecer um modelo discreto sem superdispersão e utilizar um processo iterativo para o ajuste ponderando com pesos iguais aos inversos do componente adicionado de variação,
- obter os diferentes valores de L conforme seja o modelo reduzido para o preditor linear,
- calcular os respectivos $-2\log(r.v.)$, onde $r.v.$ é o quociente de verossimilhança.

Observe-se que essa abordagem para um teste de verossimilhança, também permite testar a **homogeneidade de ϕ** ao longo dos grupos (tratamentos), o qual não é permitido com o uso daqueles testes sugeridos por Williams (1982) e Breslow (1984). Isto é, a maximização do L é aproximada calculando-se o somatório dos L obtidos para os grupos quando são analisados separadamente. Assim, os valores de $-2\log(r.v.)$ são usados para testar ϕ constante, e se ϕ for assumido como um valor arbitrário conhecido de antemão, para comparar modelos não saturados adequados e estudar sua aderência.

Essa abordagem, construída mediante um caminho diferente de simples interpretação, será apresentada na seção (5.3.2).

4.3 Equações Generalizadas de Estimação

Outro enfoque para abordar modelos para variáveis respostas que apresentam superdispersão, é converter a informação univariada a uma situação multivariada através das **Equações Generalizadas de Estimação (EGE)** (Lian & Zeger, 1986). A grande mudança reside em que o ajuste é obtido diretamente dos algoritmos de estimação e não dos modelos propostos propriamente ditos. A idéia geral é introduzir no algoritmo uma dependência estocástica na forma de uma matriz de variâncias e covariâncias.

Para apresentar a construção geral dessas EGE, lembre-se das equações de estimação para um modelo linear generalizado, chamadas equações dos escores, para observações independentes obtidas de uma distribuição hipotética pertencente à família exponencial, com um parâmetro de dispersão, para cada observação,

$$U_i = [z_i - b'(\theta_i)]/a_i(\phi)$$

onde $b'(\theta_i) = \mu_i$. No caso do modelo Poisson, tem-se

$$U_i = [z_i - e^{\theta_i}]/1 = z_i - \mu_i$$

Assim, quando a soma desses escores, ao longo das observações, igualava-se a zero, obtinham-se as estimativas de máxima verossimilhança dos parâmetros contidos no modelo de locação. Ou seja, substituindo adequadamente conforme for o problema, as soluções de

$$\sum_{i(j)=1}^3 \sum_{k=1}^{n_{ij}} \left[(z_{ijk} - \mu_{ijk}) \frac{\partial \theta_{ijk}}{\partial \mu_{ijk}} \frac{\partial \mu_{ijk}}{\partial \eta_{ijk}} \zeta_{ijk} \right] = 0$$

tornavam-se as estimativas procuradas para os β 's. Ou, no caso Poisson,

$$\sum_{i(j)=1}^3 \sum_{k=1}^{n_{ij}} [(z_{ijk} - \mu_{ijk}) \zeta_{ijk}] = 0$$

Mas, se existem n_{ij} observações independentes na (i, j) -ésima, unidade o sistema pode ser escrito como,

$$\sum_{i=1}^n X_i D_i d_i = 0$$

onde X_i é a matriz de delineamento,

$$D_i = \text{diag} \left[\frac{\partial \theta_{ijk}}{\partial \mu_{ijk}} \frac{\partial \mu_{ijk}}{\partial \eta_{ijk}} \right]$$

e

$$d_i = z_i - \mu_i$$

Note-se que a matriz D é diagonal devido ao fato de as observações serem modeladas como independentes. É aí que pode ser introduzida uma dependência estocástica, isto é, fazendo D não diagonal. Mas, uma primeira questão é a seguinte: como escolher a matriz D ? Que informação poderia ser refletida por essa matriz? Intuitivamente, dado que as observações apresentam superdispersão, essa matriz pode ser construída em termos das correlações. De fato, da expressão da distribuição de probabilidade conforme um modelo pertencente à família exponencial,

$$V(Z) = a(\phi)V(\mu)$$

ou,

$$V(Z) = a(\phi)\tau^2$$

onde τ^2 é a função de variância, ou

$$V(Z) = \frac{\phi\tau^2}{w}$$

define-se

$$U_i = \text{diag}(\phi\tau^2/w)$$

Assim, a matriz R_i de variâncias e covariâncias procurada, pode ser obtida segundo,

$$\Sigma_i = U_i^{1/2} R_i U_i^{1/2}$$

e as equações de estimação resultam

$$\sum_{i=1}^n C_i \Sigma_i^{-1} d_i = 0$$

onde

$$C_i = \frac{\partial \mu_{ijk}}{\partial \beta_{\tau}} = X_i \text{diag} \left[\frac{\partial \mu_{ijk}}{\partial \eta_{ijk}} \right]$$

Note-se que se R_i é a matriz identidade, recupera-se o modelo original de independência uma vez que $b''(\theta) = \partial \mu / \partial \theta = \tau^2 / w$. O algoritmo de estimação assim construído não pode ser avaliado a respeito da sua performance através de "deviances", mas pela estatística de Wald (Lindsey, 1993).

Esse enfoque não foi utilizado neste trabalho, apenas introduzido com o objetivo de proporcionar ao leitor metodologias que estão disponíveis na literatura e que podem ser consideradas novas quanto a sua abordagem de interpretação. Sem dúvida, o caminho que o estudo dos problemas da modelagem da superdispersão deveria adotar é através da especificação desse fenômeno naquelas partes do algoritmo que envolvem informação relevante do processo de geração dos dados, como o é sua matriz de variâncias e covariâncias.

Presença da superdispersão nas análises dos dados mediante modelos lineares generalizados pode assim ser incorporada, e portanto estimada, de várias maneiras, cada uma das quais tenta incorporar informações diferentes, ou, o que é mais provável, com fins distintos.

4.4 Enfoque Bayesiano

As idéias que brevemente serão expostas nesta seção se baseiam no argumento seguinte: existe um parâmetro de interesse que está relacionado com a realização de uma variável aleatória e que possui uma distribuição de frequências *a priori*, a qual pode ser estimada a partir de experiências anteriores relevantes. Isto é, além dos dados, cuja distribuição depende diretamente do parâmetro de interesse (ou seja, indexada por ele), existe outro conjunto de dados dependendo de parâmetros provavelmente mais próximos do parâmetro de interesse imediato. Estabelecido isto, seguem-se as noções introduzidas por Cox & Snell (1989) para a versão estritamente paramétrica, as quais Díaz (1994) descreveu em uma primeira instância.

Suponha que os dados disponíveis dependem de um parâmetro geral Φ e de outros parâmetros $\gamma_1, \dots, \gamma_k$, onde os γ 's são representados por variáveis aleatórias *i.i.d.* com densidade $p(\gamma; \kappa)$, onde κ é um parâmetro desconhecido. Suponha que o interesse está centralizado num γ particular, γ_1 , o qual representa alguma propriedade do primeiro grupo de observações; alternativamente o parâmetro de interesse poderia ser uma constante, como por exemplo $\gamma_1 - \gamma_1$. Se os γ 's são tratados por separado, obtém-se imediatamente uma estimativa para γ_1 . A geração aleatória dos $\gamma_2, \dots, \gamma_k$ significa que pode-se "emprestar potência" da distribuição dos $\gamma_2, \dots, \gamma_k$ para melhorar a estimativa do γ_1 . Isto comumente envolve a estimativa direta do γ_1 seguindo a média geral de todos os γ 's.

Suponha, primeiramente, k amostras de tamanho r^1 de populações normais com médias $\mu_1, \dots, \mu_k$ e variância σ^2 . Logo, as médias amostrais correspondentes $\bar{Z}_1, \dots, \bar{Z}_k$ são normais com variância σ^2/r . Ainda mais, suponha agora que as médias populacionais são normais *i.i.d.* com média ν e variância τ^2 , e que o interesse imediato está, sem perda da generalidade, no μ_1 . Pelo Teorema de Bayes (Box & Tiao, 1973), a distribuição a posteriori de μ_1 será proporcional à:

$$\frac{1}{\sqrt{2\pi}\sigma/\sqrt{r}} \exp \left\{ -\frac{(\bar{z}_1 - \mu_1)^2}{2\sigma^2/r} \right\} \frac{1}{\sqrt{2\pi}\tau} \exp \left\{ -\frac{(\mu_1 - \nu)^2}{2\tau^2} \right\}$$

Completando quadrados na exponencial como uma função de μ_1 , tem-se que a densidade a posteriori para μ_1 é uma normal com

$$\text{Média} = \frac{\bar{z}_1/(\sigma^2/r) + \nu/\tau^2}{1/(\sigma^2/r) + 1/\tau^2}$$

e

$$\text{Variância} = \frac{1}{1/(\sigma^2/r) + 1/\tau^2}$$

Note-se que a expressão da média representa uma ponderação em $\bar{z}_1$, a média das observações, e em ν , a média a priori, representando assim ambas informações, a do passado e a do presente. Esta expressão pode ser re-escrita como

$$\bar{z}_1 + \frac{1/\tau^2}{1/(\sigma^2/r) + 1/\tau^2}(\nu - \bar{z}_1) \quad (16)$$

indicando que $\bar{z}_1$ se modifica através de um termo que reflete sua dispersão em torno do ν .

¹Na notação utilizada nas outras seções, $k = ij$ e $r = n_{ij}$

Para o uso de (16) devem ser estimados ν , τ^2 e σ^2 . Por outro lado a teoria Normal-Padrão do modelo de componentes de variância da análise da variância estabelece que os quadrados médios esperados *entre* e *dentro* de grupos são, respectivamente, $\sigma^2 + r\tau^2$, σ^2 . De fato, a distinção entre a formulação empírica Bayesiana e a de componentes de variância se baseia num único argumento: no **passado** o interesse está em μ_1 (por exemplo), e no **futuro**, em (ν, τ^2, σ^2) . Se k não for muito pequeno, i.e., $k > 10$, o efeito em (16) da estimação de ν e τ^2/σ^2 é, obviamente, relativamente pequeno.

Particularizando o tipo de dados que este trabalho aborda, suponha k grupos de observações de contagens, dos quais o interesse centraliza-se na estimação de um dos μ_i , por exemplo, μ_1 . Segundo o mencionado anteriormente, existem duas formas de ação. A primeira delas, usa as propriedades do primeiro e segundo momento da distribuição, enquanto que a segunda depende explicitamente do estabelecimento de um modelo do tipo "passado-presente". Para o primeiro dos enfoques, têm-se as estimativas independentes de μ_1 , $\bar{z}_1$ e μ_θ , as quais combinadas na ponderação com os inversos das variâncias originam a seguinte expressão para a média

$$\frac{\bar{z}_1 \sigma_{\bar{z}_1}^{-2} + \mu_\theta \sigma_\theta^{-2}}{\sigma_{\bar{z}_1}^{-2} + \sigma_\theta^{-2}} = \bar{z}_1 + \frac{\sigma_\theta^{-2}}{\sigma_{\bar{z}_1}^{-2} + \sigma_\theta^{-2}} (\mu_\theta - \bar{z}_1)$$

O segundo dos enfoques é chamado de modelo **Gama-Poisson** pois utiliza a distribuição Gama para o parâmetro que indexa a distribuição dos dados (como seria a distribuição Poisson). Mostra-se que se a distribuição *a priori* ($\pi(\theta)$) do μ_i é uma **Gama** de parâmetros α e β , a distribuição a posteriori ($\pi(\theta | z)$) para μ_i é também uma Gama, agora com parâmetros

$$\alpha' = (\alpha + n\bar{z})$$

e

$$\beta' = (n + \beta)$$

De fato, para θ variável aleatória, com $0 < \theta < \infty$, a distribuição *a priori*

$$\pi(\theta) = \frac{\beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha, \beta)}$$

é combinada com a função de verossimilhança para θ (função de θ dadas as ob-

servações, cuja expressão é a da função de distribuição conjunta avaliada no parâmetro) do modelo Poisson, origina

$$\begin{aligned}
 \pi(\theta | z) &= \frac{\pi(\theta)l(\theta)}{\int_0^\infty \pi(\theta)l(\theta)d\theta} \\
 &= \frac{\theta^{(\alpha + \sum_{i=1}^n z_i - 1)} e^{(n+\beta)\theta}}{\int_0^\infty \theta^{(\alpha + \sum_{i=1}^n z_i - 1)} e^{(n+\beta)\theta} d\theta} \\
 &= \frac{(n + \beta)^{\alpha + n\bar{z}} \theta^{\alpha + n\bar{z} - 1} e^{(n+\beta)\theta}}{\Gamma(\alpha + n\bar{z})} \quad (17)
 \end{aligned}$$

Uma distribuição *a priori* com essa propriedade, isto é, a distribuição *a posteriori* resultante é também um membro dessa família de distribuições, é dita de **função a priori conjugada**.

Como utilizar esta informação para a realização de inferências?. O relacionado com o parâmetro está contido nessa expressão. De fato, conhecer a distribuição do parâmetro é a situação ideal dado que se tem em mão a lei do comportamento assintótico, para qualquer evento de interesse. Além do mais, podem ser construídos intervalos de credibilidade para medidas de posição ou de dispersão, como também estimativas para o valor de máxima verossimilhança variável aleatória (ou seja, do parâmetro), o qual coincide com a moda da distribuição *a posteriori*.

Dado que a distribuição *a priori* é uma Gama, então, *a posteriori* também o é. Porém, para o conhecimento real dessa distribuição há necessidade de proporcionar informação completa do passado, ou seja, os valores dos parâmetros α e β para assim combiná-los com a informação do presente. Se, na realidade, não se conhece de antemão os valores dessas constantes (situação mais usual), uma maneira plausível de se proceder é estabelecer, por experiências prévias, ou por uma revisão bibliográfica ampla e consistente, valores para esses parâmetros a partir da solução do sistema de equações gerado pelas expressões para os dois primeiros momentos dessa distribuição. Isto é, se $\theta \sim G(\alpha, \beta)$ então,

$$\begin{aligned}
 E(\theta) &= \frac{\alpha}{\beta} \\
 V(\theta) &= \frac{\alpha}{\beta^2}
 \end{aligned}$$

e portanto,

$$\frac{\alpha}{\beta} = \lambda_o \Rightarrow$$

$$\begin{aligned}
\alpha &= \lambda_o \beta \Rightarrow \\
v_o &= \frac{\beta \lambda_o}{\beta^2} \\
&= \frac{\lambda_o}{\beta} \Rightarrow \\
\beta &= \frac{\lambda_o}{v_o} \Rightarrow \\
\alpha &= \frac{\lambda_o^2}{v_o}
\end{aligned}$$

ou, por exemplo, substituindo uma dessas equações pela corresponde à moda,

$$M_o = \frac{\alpha - 1}{\beta}, \quad \forall \alpha \geq 1$$

Ou também, através do coeficiente de variação, $C.V. = \alpha^{-1/2}$, informação estatística que pode, às vezes, ser conhecida facilmente por pesquisas anteriores, e da qual se sabe que um intervalo de credibilidade para $C.V. = \sigma/\mu$ é da forma (Patel *et al.*, 1976):

$$P \left[\left(\frac{2n\zeta}{c_1 \chi_{[1-(\alpha/2); n-1]}^2} \right)^{1/2} \leq \frac{\sigma}{\mu} \leq \left(\frac{2n\zeta}{c_2 \chi_{[\alpha/2; n-1]}^2} \right)^{1/2} \right] \approx 1 - \alpha$$

onde, $\zeta = \log \bar{Z} - [\sum (\log z_i / n)]$, e

$$\begin{aligned}
c_1 &= \frac{1}{2} \left[1 + \left(1 + \frac{4(n+1)\zeta}{3\chi_{[1-(\alpha/2); n-1]}^2} \right)^{1/2} \right] \\
c_2 &= \frac{1}{2} \left[1 + \left(1 + \frac{4(n+1)\zeta}{3\chi_{[\alpha/2; n-1]}^2} \right)^{1/2} \right]
\end{aligned}$$

Assim, com a solução desse sistema, por exemplo, pode-se obter a expressão unívoca da distribuição *a posteriori* para o parâmetro que indexa a distribuição dos dados.

5 MODELO BINOMIAL NEGATIVO

5.1 Descrição do Modelo

Uma variável aleatória é dita Binomial Negativa se a distribuição de probabilidade é da forma:

$$P(Z = z) = \binom{\lambda + z - 1}{z} (1 - p)^z p^\lambda; \quad z = 0, 1, 2, \dots \quad 0 < p \leq 1$$

ou, parametrizada segundo:

$$P(Z = z) = \binom{\lambda + z - 1}{z} \left(1 + \frac{\mu}{\lambda}\right)^{-\lambda} \left(\frac{\mu}{\mu + \lambda}\right)^z; \quad z = 0, 1, 2, \dots, \quad \lambda > 0$$

para a qual se satisfaz que, o valor esperado é

$$\mu = \frac{\lambda(1 - p)}{p}$$

e a variância

$$\sigma^2 = \frac{\lambda(1 - p)}{p^2} = \mu + \frac{1}{\lambda} \mu^2$$

Alguns autores (Macchiavelli, 1987; Johnson & Kotz, 1969, dentre outros) preferem obter a função de probabilidade a partir da expansão de $(q - p)^{-\lambda}$, com $p > 0$, $q - p = 1$ e $\lambda > 0$, ou seja:

$$P(Z = z) = \frac{\Gamma(\lambda + z)}{z! \Gamma(\lambda)} p^z (1 + p)^{-\lambda - z} \quad z = 0, 1, 2, \dots, \quad (18)$$

As características dessa distribuição são:

- o coeficiente de variação, $CV = \frac{1}{\sqrt{\lambda q}}$,
- o coeficiente de assimetria, $CA = \frac{1 + q}{\sqrt{\lambda q}}$,

- a expressão geral dos r -ésimos momentos populacionais, $\mu_{r+1} = q(\frac{\partial \mu_r}{\partial q} + \frac{r\lambda}{p^2} \mu_{r-1})$,
- função geradora de momentos, $\phi = p^\lambda(1 - qe^t)^{-\lambda}$,

dentre outras (Patel *et al.*; 1976).

Por outro lado, a nível de família de distribuições:

- Se λ for conhecido, $p(z)$ é completa,
- Se λ for conhecido, a estatística $T = \sum_i Z_i$ é completa e suficiente para p ,
- Se $\lambda = 1$, $\frac{(n-1)}{(n+T-1)}$ é UMVUE para p ,

$\frac{T}{n\lambda}$ é UMVUE para $\mu = \frac{\lambda(1-p)}{p}$ (λ conhecido),

$\frac{(n+t)T}{(n+1)n}$ é UMVUE para $\sigma^2 = \frac{(1-p)}{p^2}$, ($\lambda = 1$) (Patel *et al.*; 1976)

Algumas propriedades importantes da distribuição Binomial Negativa são as seguintes:

Propriedade 1: Se $Z_1, Z_2, \dots, Z_n$ são variáveis aleatórias independentes, e identicamente distribuídas, com funções de probabilidade $p(z; \lambda_i, p)$, Binomial Negativa, $i = 1, \dots, n$, respectivamente, então a variável aleatória $Y = \sum_{i=1}^n Z_i$ também tem distribuição Binomial Negativa, com função de probabilidade $p(z; \sum_{i=1}^n \lambda_i, p)$ (Patel *et al.*; 1976).

Propriedade 2: A distribuição Binomial Negativa para $\lambda = 1$ é conhecida como a distribuição geométrica (Mood *et al.*; 1974).

Em termos intuitivos, a distribuição Binomial Negativa pode ser gerada a partir de diferentes situações experimentais partindo de um modelo probabilístico aleatório. Por exemplo, uma forma ideal de construir um modelo que represente dependência entre as respostas (realizações da variável aleatória) é diretamente através de uma estrutura de dependência estocástica, como por exemplo é o caso do modelo de componentes de variância (modelo de locação aleatório ou misto). Uma maneira indireta de tratar esse problema, caso as respostas possuam uma dependência constante, é imaginar um efeito aleatório adicional, o que produz uma distribuição composta.

Como foi mencionado, e relacionando com a estrutura de um modelo linear generalizado, o parâmetro de locação μ descreve algumas características

básicas da população à qual pertence, e assim esse modelo de locação $g(\mu)$ é usado para descrever como essas características variam sistematicamente, de forma observável, para alguns subconjuntos da população de modo que as condições experimentais mudam. Mas, em geral, essas condições não são observáveis, o que complica o processo de estimação, dado que o parâmetro de locação pode ser pensado agora como tendo alguma distribuição aleatória, independentemente da distribuição das respostas. Esse *processo de composição* é chamado de distribuição composta, e o parâmetro de locação é usado para criar a dependência estocástica (Lindsey, 1993).

A distribuição do parâmetro de locação pode quantificar a variação aleatória ou heterogeneidade entre os indivíduos. Portanto, há necessidade de obter a distribuição conjunta, das observações e do parâmetro aleatório, a partir da qual (por integração), resulta a distribuição (marginal) dos dados, conhecida como *distribuição composta*. Isto é, se $P(z_{ijk} | \mu_{ij})$ é a função de probabilidade das respostas, dado o parâmetro aleatório, e $p_j(\mu_{ij})$ é a distribuição que compõe (isto é, do parâmetro), então a distribuição conjunta é dada por:

$$P(z_{ij1}, \dots, z_{ijk}) = \int \left[\prod_{k=1}^{n_{ij}} P(z_{ijk} | \mu_{ij}) \right] p_j(\mu_{ij}) d\mu_{ij}$$

onde k é o índice que indexa dentro das unidades experimentais.

Então, o primeiro problema é a escolha correta da distribuição que compõe a distribuição dos dados. Essa distribuição pode ser procurada tendo em conta o seguinte: o valor esperado pertence a um espaço paramétrico definido e restrito, um conjunto restrito de valores plausíveis, o qual origina, dentro das distribuições da família exponencial. A literatura (Lindsey; 1993) denomina essa distribuição de *conjugada*. Seja $f(.,.)$ uma distribuição qualquer da forma:

$$\begin{aligned} f(z; \phi | \mu) &= \exp\{[y\theta - b(\theta)]/a(\phi) + c(z, \phi)\} \\ &= \exp\{[yg(\mu) - b(g(\mu))]/a(\phi) + c(z, \phi)\} \end{aligned}$$

onde $g(.)$ é a função de ligação canônica e λ é agora o parâmetro que representa o valor esperado aleatório. Se a distribuição a propor é uma distribuição *conjugada*, então será da forma:

$$p(\mu; \kappa, \nu) = \exp\{\kappa g(\mu) - b(g(\mu))/\nu + d[\kappa, \nu]\}$$

onde $d[.,.]$ é a função que independe do parâmetro. Portanto, a distribuição da

variável de resposta é

$$f(z; \kappa, \nu, \phi) = \exp\{d[\kappa, \nu] + c(z, \phi) - d[\kappa + z/a(\phi), \nu a(\phi)/(\nu + a(\phi))]\}$$

Para enxergar a regra da função de ligação, e para o tipo de distribuições que este trabalho aborda, suponha-se que exista uma distribuição Poisson para as observações:

$$P(z | \mu) = \exp[z \log(\mu) - \mu - \log(z!)]$$

logo, a distribuição conjugada é:

$$p(\mu; \kappa, \nu) = \exp\{\kappa \log(\mu) - \mu/\nu + d[\kappa, \nu]\}$$

e tal que μ tem distribuição Gama com,

$$d(\kappa, \nu) = -(\kappa + 1)\log(\nu) - \log[\Gamma(\kappa + 1)]$$

da qual resulta que a distribuição marginal é:

$$\begin{aligned} P(z | \kappa, \nu) &= \exp\{d(\kappa, \nu) - \log(z!) - d[\kappa + z, \nu/(\nu + 1)]\} \\ &= \frac{\Gamma(\kappa + z + 1)\nu^z}{\Gamma(\kappa + 1)z!(\nu + 1)^{\kappa+z+1}} \end{aligned}$$

isto é, Binomial Negativa, onde o parâmetro de forma κ é igual a $(n - 1)$ para uma reparametrização padrão como é a de Lindsey (1993). Como

$$E(Z) = \kappa\nu = \mu$$

resulta que um modelo log-linear, com função de ligação $\log(\cdot)$ pode ser usado para avaliar efeitos e/ou coeficientes em um preditor linear. Observe-se, mais uma vez, outra das vantagens da teoria geral dos modelos lineares generalizados, isto é, a função de ligação $g(\cdot)$, que relaciona os componentes aleatório e sistemático, pode ser combinada livremente, e portanto escolhida, como neste caso, visando uma simples interpretação. Obviamente, isso se contrasta com a teoria geral dos modelos lineares clássicos, para os quais a função *Identidade* é a única possibilidade para relacionar ambos componentes.

Continuando com a idéia do modelo de composição, note-se que, se $P(\mu) = f(\mu)$ é uma Gama, tem-se que $E(\mu_{ij}) = \kappa_j \nu_j$ e $V(\mu_{ij}) = \kappa_j \nu_j^2$, e assim

$$E(Z_{ij}) = \mu_j = \kappa_j \nu_j$$

e

$$V(Z_{ij}) = \kappa_j \nu_j (1 + \nu_j)$$

Uma forma simples de interpretar esse parâmetro de forma, κ_j , é mantendo à correlação $\rho_j = 1/\kappa_j$ como constante, isto é:

$$\begin{aligned} V(Z_{ij}) &= \kappa_j \nu_j (1 + \nu_j) \\ &= \mu_j (1 + \mu_j / \kappa) \\ &= \mu_j (1 + \rho \mu_j) \end{aligned}$$

note-se que a variância é uma função quadrática da média, a qual, segundo Jorgensen (1989), possui interessantes propriedades sob o enfoque de um modelo exponencial de dispersão.

Embora a distribuição Gama seja a mais usada para a distribuição conjugada para contagens (assim como a Beta é para o caso de uma distribuição Binomial Positiva para a variável resposta), existem outras que também conduzem a distribuições marginais manipuláveis. Porém, o significado ou interpretação intuitiva na construção não é mantida.

Por outro lado, se λ é um número natural, então essa função de probabilidade corresponde a uma variável aleatória com distribuição de Pascal, e se $\lambda = 1$, à distribuição geométrica. Em estudos de dispersão espacial o parâmetro λ interpreta-se como um indicador do grau de agregação ou contágio na população (Rojas, 1964¹). Mas, segundo Elliot (1979), o uso desse tipo de índice possui a desvantagem de não ser independente do número de unidades de amostragem consideradas (n), nem do tamanho das mesmas.

Foi mencionado anteriormente que a distribuição Binomial Negativa pode originar-se através de um processo de composição, isto é, os indivíduos são independentes e a distribuição das contagens dos mesmos é uma Poisson, condicionada a um valor do parâmetro, para o qual existe uma distribuição Gama como lei para atribuição das probabilidades dos eventos. Existe outro processo, o de *generalização*, descrito por Macchiavelli (1987), que também conduz a uma distribuição Binomial Negativa. Nesse processo, supõe-se que:

¹Citado por Macchiavelli, 1987.

- grupos de indivíduos constituem os itens que possuem um padrão aleatório (distribuição Poisson) e,
- o número de indivíduos por grupo é uma variável aleatória com distribuição logarítmica.

Isto é facilmente demonstrado obtendo-se, primeiramente, a função geradora de probabilidade para o processo e, depois, compondo com a função geradora de probabilidade da variável logarítmica. Dessa composição resulta uma função geradora de probabilidade $G(s) = (1 + p - ps)^{-\lambda}$, correspondente à distribuição Binomial Negativa.

Um outro enfoque de obtenção de um modelo probabilístico Binomial Negativo é pensar que as observações são produzidas por um processo estocástico Poisson agrupado, onde cada evento contribui para uma quantidade aleatória total (Firth, 1990). Em outras palavras, observa-se que

$$Z = Y_1 + Y_2 + \dots + Y_N$$

onde Y_i 's são independentes e identicamente distribuídas, e N segue uma distribuição Poisson independente do Y_i , $\forall i = 1, \dots, N$. Então:

$$\begin{aligned} E(Z) &= \sum_{i=1}^N E(Y_i) \\ &\simeq NE(Y) \\ &= E(N)E(Y) \end{aligned}$$

$$\begin{aligned} V(Z) &= E[Z - E(Z)]^2 \\ &= E(N)E(Y)^2 \end{aligned}$$

pois os Y 's são *i.i.d.* Ou seja,

$$V(Z) > E(Z) \text{ se e somente se } E(Y^2) > E(Y)$$

e assim,

$$V(Y) \propto [E(Y)]^2$$

i.e., $Y \sim \Gamma(\mu, \lambda)$ resultando o modelo Binomial Negativo.

5.2 Estimação por Máxima Verossimilhança

Segundo Bliss & Fisher (1953), considerando amostragem aleatória, os estimadores de máxima verossimilhança para os parâmetros da Binomial Negativa são eficientes. Esses estimadores são obtidos do sistema de equações seguinte:

$$\begin{aligned} \bar{Z} - \hat{\lambda}\hat{p} &= 0 \\ \sum_{r=0}^n f_r \left[\sum_{i=1}^r \frac{1}{\hat{\lambda} + i - 1} - \ln(1 + \hat{p}) \right] &= 0 \end{aligned}$$

onde $\bar{Z}$ é a média amostral, e f_r é a frequência observada na classe r -ésima.

Dado que existe um modelo de locação para uma função do valor esperado, isto é, o componente sistemático, e supondo que o parâmetro de dispersão é conhecido, tem-se o algoritmo de estimação para os parâmetros envolvidos nesse modelo, a partir da expressão da função de probabilidade como membro da família exponencial (McCullagh & Nelder, 1989). De fato,

$$\begin{aligned} p(z; \mu, \lambda) &= \exp \left\{ \ln[\Gamma(z + \lambda)] + \lambda \ln\left(\frac{\lambda}{\mu}\right) - \ln(z!) - \ln[\Gamma(\lambda)] - (z + \lambda) \ln\left(1 + \frac{\lambda}{\mu}\right) \right\} \\ &= \exp \left\{ (-z) \ln\left(\frac{\mu + \lambda}{\mu}\right) - \lambda \left[\ln\left(\frac{\mu + \lambda}{\mu}\right) - \ln\left(\frac{\lambda}{\mu}\right) \right] + \ln[\Gamma(z + \lambda)] - \ln(z!) - \ln[\Gamma(\lambda)] \right\} \\ p(z; \mu, \lambda) &= \exp \left\{ z \ln\left(\frac{\mu}{\mu + \lambda}\right) - \lambda \ln\left(\frac{\mu + \lambda}{\lambda}\right) + \ln \left[\frac{\Gamma(z + \lambda)}{z! \Gamma(\lambda)} \right] \right\}, \quad z = 0, 1, \dots \end{aligned}$$

Assim,

$$\begin{aligned} a(\phi) &= 1 \\ \theta &= \ln \left(\frac{\mu}{\mu + \lambda} \right) \\ b(\theta) &= \lambda \ln \left(\frac{\mu + \lambda}{\lambda} \right) \\ &= -\lambda \ln(1 - e^\theta) \\ c(z; \phi) &= \ln \left[\frac{\Gamma(z + \lambda)}{z! \Gamma(\lambda)} \right] \end{aligned}$$

Portanto, e desprezando os índices, para λ **fixo**, têm-se os termos do algoritmo de estimação do MLG:

$$\begin{aligned}
 V(\mu) &= \mu \left(\frac{\mu}{\lambda} + 1 \right) \\
 w &= 1 \\
 g'(\mu) &= \frac{\lambda}{\mu(\mu + \lambda)} \\
 \eta^{(o)} &= \ln \left(\frac{z}{z + \lambda} \right) \\
 Z^{(m)} &= \eta^{(m)} + (z - \mu^{(m)}) \frac{\lambda}{\mu^{(m)}(\lambda + \mu^{(m)})} \\
 w^{(m)} &= \mu^{(m)}(\mu^{(m)} + \lambda) \\
 W^{(m)} &= \text{diag}\{\mu^{(m)}(\mu^{(m)} + \lambda)\} \\
 H^{(m)} &= \text{diag}\left\{ \frac{\lambda}{\mu^{(m)}(\mu^{(m)} + \lambda)} \right\}
 \end{aligned}$$

ou seja, as estimativas de máxima verossimilhança para os β 's contidos no preditor linear são obtidas através de um processo iterativo segundo a expressão:

$$\Rightarrow \beta^{(m+1)} = \{[X'W X]^{-1}\}^{(m)}(X'W Z)^{(m)}$$

onde X é a matriz de delineamento, de constantes, conhecida.

Como foi mencionado nas seções anteriores, a "deviance" é uma das estatísticas usadas para a aderência do modelo,

$$D = 2 \sum_{i=1}^n \left\{ z_i \log\left(\frac{z_i}{\mu_i}\right) - (z_i + \lambda) \log\left[\frac{z_i + \lambda}{\mu_i + \lambda}\right] \right\}$$

Lembre-se a idéia de Breslow (1984), de estimar os parâmetros do preditor linear como em um MLG usando, em cada passo, a atualização do vetor dos pesos contendo a estimativa do λ , obtida pelo método dos momentos. É óbvio que se forneça um valor estimado desse parâmetro de dispersão, uma vez que a estimativa para o vetor de parâmetros β é obtida mediante um algoritmo, que envolve, em todos seus componentes, o valor de λ .

O usuário pode fornecer esse valor, colocando um valor arbitrário, obtido, por exemplo, de ferramentas gráficas (no "plot" dos pares *médias-variâncias* dos tratamentos esse valor pode corresponder ao inverso do coeficiente de regressão), ou por indícios do material experimental, tratado anteriormente na literatura, etc.

Embora a estimativa desse parâmetro pareça não importar muito, algum critério de estimação deve-se seguir para continuar com a realização de inferências completas e consistentes. Por outro lado, os estimadores correspondentes aos β 's (os parâmetros realmente de interesse) são os de máxima verossimilhança (MV).

O método de MV, introduzido por Fisher em 1922, citado por Leite & Singer (1990), é certamente uma das técnicas mais usadas na estimação paramétrica, dado que produz equações bastante convenientes do ponto de vista computacional, e estimadores com boas propriedades assintóticas. Essas boas propriedades permitem que sejam estabelecidas estatísticas de testes para hipóteses simples, ou compostas, acerca dos β 's de maneira muito simples. Ainda mais, o Teorema (5.2) que se segue enuncia um resultado clássico da literatura, expreso neste material para variáveis aleatórias discretas, que permite a construção de intervalos de confiança e estatísticas com distribuição assintótica conhecida para a verificação de hipóteses do tipo $H : A'\beta = 0$ (ou $A'\beta = \delta$).

Teorema 5.2.1 *Sejam $Z_1, Z_2, \dots, Z_n$ variáveis aleatórias independentes e identicamente distribuídas com função de probabilidade $p(z; \mu)$, $\mu \in \Theta \subset \mathbb{R}$, satisfazendo às seguintes condições:*

1. $\frac{\partial}{\partial \mu} p(z; \mu)$ e $\frac{\partial^2}{\partial \mu^2} p(z; \mu)$ existem em quase toda parte, e são tais que $|\frac{\partial}{\partial \mu} p(z; \mu)| \leq H_1(z)$ e $|\frac{\partial^2}{\partial \mu^2} p(z; \mu)| \leq H_2(z)$, onde $\int_{\mathbb{R}} H_j(z) dz < \infty$, $j = 1, 2$.
2. $\frac{\partial}{\partial \mu} \log p(z; \mu)$ e $\frac{\partial^2}{\partial \mu^2} \log p(z; \mu)$ existem em quase toda parte, e são tais que:
 - $0 < I_\mu = E_\mu \left\{ \frac{\partial}{\partial \mu} \log p(Z_1; \mu) \right\}^2 = \int_{\mathbb{R}} \left\{ \frac{\partial}{\partial \mu} p(z; \mu) \right\}^2 \{p(z; \mu)\}^{-1} dz < \infty$, ou seja, Z_1 tem informação de Fisher finita;
 - $E_\mu \left\{ \sup_{\{h: |h| \leq \delta\}} \left| \frac{\partial^2}{\partial \mu^2} \log p(Z_1; \mu + h) - \frac{\partial^2}{\partial \mu^2} \log p(Z_1; \mu) \right| \right\} = \Psi \rightarrow 0$ quando $\delta \rightarrow \infty$.

Então, o estimador de MV de μ , $\hat{\mu}_n$, é tal que $\sqrt{n}(\hat{\mu}_n - \mu) \rightarrow_D N(0, I_\mu^{-1})$.

Observe-se que o Teorema (5.2) garante esse resultado de convergência assintótica através do seguinte fato: a distribuição de probabilidade é um membro da família exponencial a um parâmetro. Satisfeita essa propriedade, as condições de regularidade necessárias para provar a convergência em distribuição resultam certas. Então, o que acontece com o outro parâmetro, visto, talvez, como "nuisance" nesse sistema de estimação? A literatura em geral não explora o recurso que o modelo probabilístico "two-stage" Binomial Negativo oferece em si, isto é,

os parâmetros μ e λ (ou ϕ como seu recíproco) são independentes e podem, portanto, ser estimados separadamente. O usual é encontrar valores arbitrários dos λ colocados nas equações de estimação dos β 's.

Por outro lado, no modelo de locação se verifica que $\mu = \mu(\beta)$, então, se ϕ for desconhecido mas constante, as equações de estimação para β serão, em geral, diferentes daquelas obtidas por mínimos quadrados ponderados. Não obstante, podem ser comparados os dois conjuntos de estimativas, um deles baseado na verossimilhança Binomial Negativa e o outro, baseado na verossimilhança Poisson, como duas "seqüências de variáveis aleatórias". Assim, McCullagh & Nelder (1989) informaram que essas duas seqüências diferem por um termo da ordem de magnitude ϕ^{-2} , ou em $O_p(\phi^{-2})$. Isto implica que para quantidades pequenas de superdispersão essa diferença é desprezível, o qual é conveniente na prática pela simplicidade do algoritmo de estimação.

Pensada à distribuição Binomial Negativa agora como membro da família exponencial a dois parâmetros, a estimação por máxima verossimilhança é menos utilizada do que os enfoques anteriores, e apresenta muitas discussões devido ao fato que uma das equações obtidas da derivada parcial em relação ao parâmetro de interesse não é de imediata resolução. Isto é, a equação correspondente ao $\lambda = 1/\phi$ além de envolver um processo de Newton-Raphson para sua resolução, dado que ela é uma equação não linear, precisa dos recursos de aproximação de funções não explícitas para alguns dos seus termos.

De fato, lembrando a expressão para a função de probabilidade de uma variável aleatória Binomial Negativa, a equação que surge da derivada do logaritmo da função de verossimilhança em relação a ϕ envolve funções $\partial\Gamma/\partial\phi$, definidas como funções **Digama**, e denotadas por $\Gamma' = \Psi$. Isto é,

$$\ell(p, \lambda; z_i) = \sum_{i=1}^n \log \Gamma(\lambda + z_i) - \Gamma(\lambda)^n \sum_{i=1}^n \log z_i! + \sum_{i=1}^n \log(1-p)^{z_i} + \log(p^{n\lambda})$$

e assim,

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} &= \sum_{i=1}^n \frac{\Gamma'(\lambda + z_i)}{\Gamma(\lambda + z_i)} - n[\Gamma(\lambda)]^{n-1} \Gamma'(\lambda) \sum_{i=1}^n \log z_i! + \frac{1}{p^{n\lambda}} [np^{n\lambda} \log(n\lambda)] \\ &= \sum_{i=1}^n \frac{\Psi(\lambda + z_i)}{\Gamma(\lambda + z_i)} - n[\Gamma(\lambda)]^{n-1} \Psi(\lambda) \sum_{i=1}^n \log z_i! + n \log(n\lambda) \end{aligned}$$

É óbvio que o problema não se acaba resolvendo essa expressão, mas necessita-se,

ainda, da segunda derivada da função Gama, ou seja, uma função **Trigama** avaliada na estimativa, para corroborar que seja um máximo local da função logaritmo da função de verossimilhança.

Por exemplo, é comum encontrarmos trabalhos como de Schukken *et al.* (1991), onde foi estudado o número de casos de mastite clínica registrados em lotes de gados de algumas fazendas do Estado de Nova York, usando a distribuição Binomial Negativa. Os autores justificaram a escolha no fato de que o número de casos por fazenda segue uma distribuição Poisson e as médias dessas distribuições uma distribuição Gama. Mas, dado que o algoritmo de estimação de máxima verossimilhança envolve um sistema de equações de resolução bastante trabalhosa, os autores optaram por usar uma aproximação razoável mediante um modelo Poisson com um parâmetro extra para a superdispersão, usando para isso o método da máxima quase-verossimilhança, como foi apresentado na seção anterior.

Até a data da edição deste material, somente dois pacotes de uso mundial, pela comunidade estatística, incorporaram trabalhos científicos de autores em forma de macros disponíveis em uma biblioteca, chamada de "library", comum para o uso de algoritmos já implementados. Esses pacotes são o SPlus e o SAS (Statistical Analysis System). Embora essas macros estejam incorporadas como ferramenta do pacote, elas são super extensas, muito trabalhosas, e não são de compreensão rápida. Neste trabalho foram estudadas ambas rotinas, consultadas com seus respectivos autores, e utilizadas para os dados de mortalidade de *E. grandis*.

A primeira delas foi construída por Venables & Ripley (1994). Os autores apresentaram uma macro (chamada de "glm.nb() function") em SPLUS, para a geração da família Binomial Negativa com o parâmetro de forma fixo, mas essa macro é estendida para o caso sem restrição. Explicita-se claramente a notação do modelo "two-stage" para uma variável aleatória, e a macro que produz as estimativas de máxima verossimilhança para os parâmetros (μ, ϕ) , permitindo o ajuste de um modelo linear generalizado para μ , o qual no nosso caso é um modelo de análise da covariância. Para evitar o uso da função Ψ contendo o parâmetro como argumento, o autor utiliza o recurso de funções Log-Gama para a aproximação da função Digama.

Como saída da análise são fornecidas estimativas para a verificação da aderência do modelo, valores estimados, e erros padrões, dos parâmetros e testes do quociente de verossimilhança para hipóteses encaixadas. Essa macro¹, aplicada para aos dados de mortalidade de árvores, forneceu os resultados mostrados na

¹A macro não é apresentada neste material e está disponível no Cap. 9 do Venables & Ripley (1994).

Tabela 7: Resultados para o ajuste máxima verossimilhança do modelo Binomial Negativo através das macros de Venables & Ripley (1994).

Estatísticas	Valor(d.p.)	Estatísticas	Valor
$\hat{\lambda}$	0,4.579(0,0158)	$\hat{\phi}$	2,1838
"Deviance"	2.903,4760	"Deviance"/g.l.	1,1188
X^2 de Pearson	2.605,8800	X^2 /g.l.	1,0041
2xLog-verossimilhança	11.4052,5725 (**)		
[Min r_{d_i} , Max r_{d_i}]	[-1,9683; 2,1662]	Med r_{d_i}	-0,0035

Tabela 7. Essa tabela também contém o resumo das estatísticas, e seus desvios padrões (d.p.), e dos resíduos obtidos dos componentes da "deviance" do último modelo ajustado, e donotados por r_{d_i} .

Cabe assinalar que no modelo de análise da variância com estrutura fatorial de tratamentos para o preditor linear foi adicionada a variável idade média das árvores como sugeriu a Figura 1 do Apêndice A. Além disso, esse modelo foi submetido a análises de diagnóstico, principalmente ao teste para a variável adicionada (Atkinson, 1982) para determinar a inclusão ou não de potências dessa covariável. A contribuição dessa nova variável (função quadrática da idade) foi avaliada através da estatística X^2 . Para um modelo de médias, isto é,

$$\eta_{ij} = g(\mu_{ij}) = m_{ij}$$

o valor dessa estatística foi de 461, 2371, dos quais, 386, 0286 corresponderam à fonte de variação estabelecida pela covariável (x_{ij}) e, 2, 1029, à potência da covariável (x_{ij}^2). Isto implica que a explicação da mortalidade em termos dos fatores estudados se vê fortalecida pela inclusão de apenas idade como covariável adicional.

Salienta-se que para comparar modelos encaixados com o parâmetro de dispersão fixado de antemão, a diferença de "deviances" é uma medida apropriada como diagnóstico do ajuste. Porém, para comparar modelos provenientes de hipóteses encaixadas com ϕ estimado, como é o caso analisado, devem ser usadas as logaritmo da função de verossimilhança do modelo Binomial Negativo.

A Tabela 8 apresenta o valor do teste do quociente de verossimilhança (L.R.) para modelos encaixados (com ou sem interação dos fatores, com ou sem a idade das árvores) e os respectivos valores de λ (tal que $var(Y) = \mu + \frac{1}{\lambda}\mu^2$),

Tabela 8: Valores do teste quociente de verossimilhança (L.R.) e do λ para verificar hipóteses sob modelos encaixados, usando as macros de Venables & Ripley (1994).

Modelo	λ (d.p.)	$2 * \ell(\mu, \lambda; z_i)$	L.R.(Pr(Chi))
rot+reg+rot*loc	0,4143(0,0138)	113.870,5100	-
ida+rot+reg+rot*loc	0,4579(0,0158)	114.052,5725	182,1199(0,000)
ida+rot+reg	0,4548(0,0156)	8.532,0041	-
ida+rot+reg+rot*loc	0,4579(0,0158)	11.4052,5725	105.520,5684(0,000)

com $1/\lambda=\phi$ índice de agregação da distribuição, com seu desvio padrão, para o ajuste Binomial Negativo.

Segundo Hinde (1995)¹, essa macro fornece a estimativa de máxima verossimilhança para ϕ atualizandó-lo em cada iteração do ajuste do modelo linear generalizado. A obtenção em cada passo é através de um algoritmo do tipo de Newton-Raphson, embora em termos gerais isso não seja percebido. Como será mencionado depois, o autor implementou uma macro semelhante à feita no SPlus, na nova versão do GLIM, ou seja usando a aproximação Log-Gama para a função Digama.

Assim, o teste do quociente de verossimilhança para hipóteses encaixadas mostrou que o modelo incluindo a covariável idade das árvores é o adequado para a mortalidade. O coeficiente e desvio padrão para a variável explanatória foi 0,2028(0,0156).

A Tabela 9 apresenta as estimativas dos parâmetros do modelo completo com os valores das estatísticas de Wald para significância dos mesmos.

Pode-se observar a mudança apresentada nas estimativas dos índices de agregação para ambos os modelos: sem covariável, 2,414, com covariável, 2,184, ambos significativos.

Os resultados indicaram que a Região 2 correspondente aos Municípios de Brotas e Riberão Bonito com a Rotação 2 é a que proporciona efeito negativo para o valor médio da mortalidade das árvores. A Figura 5 no Apêndice A

¹HINDE, J. Department of Mathematical Statistics and Operational Research. University of Exeter (UK). Comunicação pessoal. Dezembro de 1995.

Tabela 9: Valores estimados para os parâmetros do modelo Binomial Negativo através das macros de Venables & Ripley (1994).

Parâmetro	Estimativa	Erro padrões	Valor de t
Constante	1,5940	0,1436	11,1014
Idade	0,2029	0,0157	12,9573
Região1	-0,0854	0,0594	-1,4385
Região2	0,2636	0,1195	2,2065
Rotação1	0,2989	0,0376	7,9304
Rotação2	0,2396	0,1222	1,9609
Reg1xRot1	-0,1588	0,0509	-3,1179
Reg2xRot1	0,0283	0,0234	1,2109
Reg1xRot2	0,0571	0,0517	1,1035
Reg2xRot2	0,0606	0,1185	0,5114

ilustra a natureza da interação entre os fatores estudados.

A semelhança do SPlus, o pacote SAS possui uma macro (Sistemas SAS, versão 6.08), que permite o ajuste de um modelo linear generalizado com uma função de ligação logarítmica para o valor esperado.

Na realidade, o algoritmo de estimação por máxima verossimilhança pode ser feito através de duas macros, "hilbenb" e "stata" (Versão 1.0, Technical Report 26. SAS Sample Library) ambas construídas por Hilbe (1995)¹, as quais utilizam uma aproximação para a função Digama semelhante à utilizada pelo pacote MATHEMATICA, chamada de PolyGamma. Um desenvolvimento em série dessa função e algumas cotas e propriedades são apresentadas em Wilks (1987).

Fornecidas pelo autor, essas macros foram usadas para o ajuste do modelo de análise da covariância, na estrutura de um MLG, para os dados de mortalidade de *E. grandis*. No processo de ajuste é usado o procedimento específico do SAS para o ajuste de modelos lineares generalizados, o PROC GENMOD, em-

¹A partir de comunicações pessoais do Prof. Barry Ames (Department of Political Science. Washington University. USA, Junho-Dezembro de 1994) foi contactado o Prof. Joe Hilbe (Graduate College Committee on Statistics, Arizona State University. USA, Dezembro de 1995) para consultas a respeito.

Tabela 10: Resultados para o ajuste de máxima verossimilhança do modelo Binomial Negativo através das macros de Hilbe (1994).

Estatísticas	Valor	Estatísticas	Valor
No. Iterações	8		
$\hat{\lambda}$	0,8597	$\hat{\phi}$	1,1632
"Deviance"	5.116,6292	"Deviance"/g.l.	1,9687
X^2 de Pearson	2.599,0638	X^2 /g.l.	1,0000
Log-verossimilhança	-8642,9456		

butido em uma série de macros que modificam o ajuste log-binomial negativo. A "deviance" utilizada contém alguns recursos interessantes do sistema como, por exemplo, aproximações tal que um dos componentes da "deviance" resulta:

$$\begin{aligned} d_i &= -2(z_i \log(\frac{\mu_i}{\lambda}) - (z_i + \lambda) \log(1 + \frac{\mu_i}{\lambda})) - \\ &= -2(\text{lgamma}(z_i + \lambda) - \text{lgamma}(z_i + 1) - \text{lgamma}(\lambda)) \end{aligned}$$

A Tabela 10 apresenta esses resultados, mostrando os valores estimados para: $1/\phi = \lambda$ após o ajuste final, a "deviance", a estatística de Pearson (X^2) e a log-verossimilhança. A Tabela 11 mostra as estimativas para os parâmetros do modelo de locação com seus respectivos erros padrões, X^2 e "p-value" encontrado. Segundo Hilbe (1994), o ajuste do modelo apenas foi adequado devido ao fato de existirem muitos valores nulos que poderiam estar afetando tanto o ajuste Poisson como o Binomial Negativo. Serão expostos ao final desta seção ajustes com truncamento.

Os resultados indicaram que a idade das árvores contribui significativamente para o mecanismo da mortalidade das árvores em função de um forte efeito de interação entre as qualidades de sítio e as rotações ensaiadas. Em termos experimentais propriamente ditos, esses resultados coincidem com os obtidos pelas macros de Venables & Ripley em SPlus, no que diz respeito a falta de independência entre a rotação da espécie e a região onde foi implantada no Estado de São Paulo.

A grande diferença entre essas duas ferramentas de ajuste está no valor estimado do parâmetro de dispersão, ou, índice de agregação da distribuição

Tabela 11: Valores estimados para os parâmetros do modelo de locação após ajuste máxima verossimilhança do modelo Binomial Negativo através das macros de Hilbe (1995).

Parâmetro	g.l.	Estimativa	Erro Padrão	X^2	"p-value"
Constante	1	2,2625	0,1054	461,2371	0,0001
Região1	1	-0,5284	0,0486	118,1613	0,0001
Região2	1	-0,6328	0,0577	120,1691	0,0001
Região3	0	0,0000	0,0000	-	-
Rotação1	1	-0,8216	0,0879	87,3841	0,0001
Rotação2	1	-0,1625	0,0950	2,9270	0,0871
Rotação3	0	0,0000	0,0000	-	-
Idade	1	0,1997	0,0102	386,0286	0,0001
Escala	0	1,0000	0,0000	-	-

espacial, provavelmente porque provêm de aproximações diferentes. Esse valor estimado não pertence ao intervalo de confiança $[0,4264; 0,4894]$ para λ , com um $(1 - \alpha) = 95\%$, obtido segundo as macros em SPlus.

Por outro lado, e como conclusão das discussões mantidas com o Prof. Hilbe, quando o λ varia ao longo das parcelas (isto é, pode ser considerado como uma variável aleatória) devido à agrupamentos ou autocorrelação, os modelos lineares generalizados não tem boa performance.

Ainda mais, se a fonte de superdispersão é o agrupamento, chamado de "clustering", então deviam ser usadas as *Equações Generalizadas de Estimação* (comumente chamadas de GEE na literatura) apresentadas por Liang & Zeger, citado por Lindsey(1993), para as que já existem as macros em SAS (ainda não publicadas); ou também o novo enfoque de Nelder, o MLG Duplo que estende a distribuição Gaussiana para o efeito de "clustering", e para o qual se precisa do pacote GENSTAT como base para o "input" de dados e resolução de algoritmos de busca. Uma saída simples para este problema, e no caso do modelo de regressão, é verificar a colinearidade entre as variáveis explanatórias antes de impor a estrutura do erro no modelo linear generalizado.

Em Lawless (1987) se apresentam, pela primeira vez na literatura e sem ser incorporado no pacote GLIM (na versão para essa data, GLIM3.77), as

Tabela 12: Resultados para o ajuste máxima verossimilhança do modelo Binomial Negativo através das macros de Lawless (1987).

Modelo Ajustado	"Deviance" Residual(g.l.)	"Deviance" Média	$\hat{\lambda}$
reg+rot	5.704,8965(2,600)	2,1942	0,9123
reg+rot+reg*rot	5.618,3487(2,596)	2,1651	0,9305
ida+reg+rot+reg*rot	3.241,3255(2,595)	1,2491	0,7420

macros para ajuste de modelos de regressão, sob o enfoque dos modelos lineares generalizados, com erro Binomial Negativo, usando estimativas de máxima verossimilhança para ambos parâmetros. Usando os recursos da nova versão do GLIM, o GLIM4.08, Hinde (1996) apresentou uma macro que também produz estimativas de máxima verossimilhança para os dois parâmetros. A Tabela 12 apresenta um resumo dos resultados obtidos da implementação das macros de Lawless (1987), e a Tabela 13 mostra os valores estimados dos parâmetros com seus respectivos erros padrões e valor de t para os testes de hipóteses simples construídos para cada um deles.

E é nesse ponto que há necessidade de fazer um comentário geral, o qual acredita-se pode contribuir na compreensão do leitor. Salienta-se que dizer "estimativas de máxima verossimilhança" não significa que essas estimativas tenham sido obtidas a partir do sistema de equações original para estimação de máxima verossimilhança, isto é, o sistema de duas equações com duas incógnitas ((μ, λ) , ou (p, λ) , segundo a notação utilizada), ou o que é comumente chamado de "perfil de verossimilhança". Ou seja, a estimação torna-se uma estimação atualizada em forma iterativa devido ao fato de ter as duas incógnitas nas duas equações. No caso de Lawless (1987) apenas é usada a equação da logaritmo da função de verossimilhança para um deles, o μ , dado que contém os parâmetros de interesse contidos no preditor linear, e a estimativa do parâmetro restante é obtida de alguma outra equação atualizando os valores ajustados para μ recém obtidos, como por exemplo a igualdade estabelecida entre o valor esperado da estatística do Qui-quadrado e os seus graus de liberdade respectivos. É óbvio que, na prática, essa função deveria respeitar o fato de ser *injetora*, em todo ponto da reta real positiva para garantir a invariância na estimação.

Essencialmente, o algoritmo de estimação de Lawless (1987) baseia-

Tabela 13: Valores estimados para os parâmetros do modelo de locação após ajuste de máxima verossimilhança através das macros de Lawless (1987).

Parâmetro	Estimativa	Erro Padrão	Valor de t
Constante	1,1560	0,0605	19,1074
Idade	0,1843	0,0096	19,1974
Região2	1,1420	0,0653	17,4805
Região3	1,5890	0,0545	29,1559
Rotação2	0,6927	0,0815	8,4994
Rotação3	0,6834	0,0914	7,4770
Região2xRotação2	-0,5245	0,1243	-4,2196
Região2xRotação3	0,0746	0,1855	0,4022
Região3xRotação2	-0,0344	0,1003	-0,3429
Região3xRotação3	0,6486	0,6304	1,0289

se no enfoque de Breslow (1984), isto é, a forma iterativa de encontrar as estimativas do parâmetro de dispersão é através da estatística do Qui-quadrado, usando os vetores do sistema GLIM3.77. Hinde (1996) atualiza esse algoritmo construindo um "subfile" chamado de "negbin", e permitindo que o parâmetro de superdispersão seja fixado de antemão, ou estimado usando um processo iterativo dentro do processo de ajuste do modelo linear generalizado propriamente dito. Nesse passo é permitido também o uso de diferentes funções de ligação: identidade; logarítmica e raiz quadrada, o qual não é possível nas macros dos outros autores anteriormente mencionados.

Os métodos plausíveis de serem usados na "negbin" são: um valor fixo do λ fornecido pelo usuário ou estimação por máxima verossimilhança, estimação qui-quadrado média, estimação "deviance" média. Em uma tentativa de estabelecer relações com os outros enfoques apresentados neste material, salienta-se que as últimas duas macros que conformam a "negbin" são as que utilizam a aproximação Log-Gama para a função Digama, correspondente a uma das equações obtidas ao derivar a logaritmo da função de verossimilhança em relação ao λ .

Como resultados, "negbin" fornece as estimativas dos parâmetros

embutidos no modelo de locação, do λ (quando é escolhida a estimação), com os seus respectivos erros padrões, a "deviances" (usada para a aderência do modelo com o último valor encontrado para λ), e os valores para as log-verossimilhança (para a verificação de hipóteses encaixadas). As observações de mortalidade das árvores de *E. grandis* foram submetidas à "negbin", para duas funções de ligação, a logarítmica e a raiz quadrada, e os primeiros resultados são mostrados na Tabela 14.

Cabe assinalar que foram construídas em GLIM4.0, e anexadas à "negbin", rotinas para cálculo dos resíduos padronizados e estudentizados obtidos dos componentes da "deviance", como também a obtenção dos gráficos de índices para observar o padrão de dispersão e a possível banda de confiança para os mesmos, com uma confiança de 95% (aproximadamente os valores deveriam pertencer ao intervalo $[-2, 2]$, Collett, 1991).

Na Tabela 14 se apresentam os valores das estatísticas para a aderência do modelo, como a "deviance" residual, o resumo para os resíduos dos componentes da "deviance" padronizados (mínimo, máximo e mediana para $r(d_i)$), a logaritmo da função de verossimilhança para o modelo de análise da covariância, e o desvio padrão da estimativa do parâmetro de superdispersão, para as duas ligações que tiveram melhor performance.

Embora os dois modelos ajustados para o valor esperado através das funções de ligação logarítmica e raiz quadrada, hajam fornecidos resultados semelhantes, como por exemplo a "deviance", foi escolhido o modelo não linear **exponencial** para o μ .

O anterior foi devido, entre outras coisas, ao fato de:

- encontrarmos diferenças no comportamento dos resíduos, onde para a função de ligação raiz quadrada, o gráfico de índices mostrou tendência;
- os intervalos de confiança para cada um dos parâmetros de interesse (coeficiente de regressão e efeitos dos tratamentos) foram menos precisos no caso da função de ligação raiz quadrada (Ver Tabela 16).

Além disso, e dada a impossibilidade de construir gráficos do tipo envelope simulado (ver seção (2.2.3), foi construído para essa ligação o gráfico dos $\hat{z}_{ijk}$ versus $\hat{\eta}_{ijk}$, o qual não mostrou uma tendência linear (Collett, 1991). A Figura 6 do Apêndice A apresenta esse fato. Como foi mencionado na seção (2.2.3), existem outros métodos formais para verificar a adequação de uma função de ligação, como é o teste para a adição da variável construída $\hat{\eta}^2$ na estrutura do preditor linear, porém neste caso, acredita-se que não era necessário, dada a evidência dos métodos não formais anteriores.

Tabela 14: Resultados para o ajuste de máxima verossimilhança do modelo Binomial Negativo para as funções de ligação canônica e raiz quadrada, através macros de Hinde (1996).

Ligação Logarítmica			
Estatísticas	Valor(d.p.)	Estatísticas	Valor
$\hat{\lambda}$	0,4579(0,0158)	$\hat{\phi}$	2,1838
"Deviance"	2.903,5600	"Deviance" /g.l.	1,1200
2xLog-Veros. (L.V.)	11.4053 (**)	g.l	2.595
2x"Full" L.V	-16651.	Escala	1,0000
$[Minr_{(d_i)}, Maxr_{(d_i)}]$	$[-2, 6760; 2, 0662]$	$Medr_{(d_i)}$	-0,0002
Ligação Raiz Quadrada			
Estatísticas	Valor(d.p.)	Estatísticas	Valor
$\hat{\lambda}$	0,4547(0,0156)	$\hat{\phi}$	2,1996
"Deviance"	2.900,9000	"Deviance" /g.l.	1,1178
2xLog-Veros. (L.V.)	114.040. (**)	g.l	2.595
2x"Full" L.V	-16.663.	Escala	1,0000
$[Minr_{(d_i)}, Maxr_{(d_i)}]$	$[-4, 9683; 6, 0315]$	$Medr_{(d_i)}$	-0,0135

Uma outra justificativa da escolha da função de ligação logarítmica foi a fácil interpretação do modelo para o μ , o qual não se deve esquecer em termos de aplicação da pesquisa em si. A etapa de seleção da função de ligação é muito importante e deve ser pesquisada integralmente com os outros dois componentes que conforma o modelo linear generalizado, o sistemático e o estocástico, como é apresentado neste material e fornecido pelos métodos formais e informais aplicados a cada uma das duas combinações para esses componentes. Por outro lado, acredita-se que para o sistema biológico em estudo, não se corre o risco de que a escala da variável explanatória possa afetar a adequação da função de ligação, como alerta Demétrio (1993).

A Tabela 15 mostra os valores das estatísticas, com os seus respectivos p-"values", para os testes do quociente de verossimilhança obtidos para as hipóteses apropriadas, correspondentes à modelos encaixados intermediários, ou seja, com e sem a covariável, e com ou sem efeito da interação entre os fatores região e rotação. Os resultados desses testes foram, em termos gerais, semelhantes aos obtidos nos casos anteriores (SPlus e SAS), mostrando que o modelo da análise da variância com estrutura fatorial pode ser melhorado, significativamente, incluindo a informação fornecida pela variável idade das árvores. Para essa variável explanatória, a estimativa (e desvio padrão entre parêntese) do coeficiente de regressão foi 0,2038(0,0156), com um p-"value" para a estatística de Wald de 0,0001.

Hinde (1996) indicou que os erros padrões obtidos dessas macros são sub-estimados (em uma quantidade desprezível), dado que não consideram essa estimação conjunta.

A Figura 1, do Apêndice A, apresenta o comportamento dessa variável desdobrado para cada nível de tratamento, podendo se observar que essa relação de dependência no modelo geral poderia dever-se à influência da idade nas rotações 2 e 3 do *E. grandis*. Porém, e dado que o teste do L.R. indicou a inclusão dessa variável no preditor linear, em uma tentativa de melhorar esse modelo de locação (e assim descartar o estudo e diagnóstico do componente sistemático como uma das etapas requeridas em toda pesquisa), incluiu-se outra variável adicional do tipo $u = f(x)$, onde $f(.)$ é a função quadrática, e X representa a idade das árvore. Para esse novo modelo o valor da estatística, e o correspondente p-"value" para o teste do quociente de verossimilhança foi igual a 1,0239(0,7805). Implica que o modelo da análise da variância com apenas uma covariável é o adequado para os dados de mortalidade de *E. grandis*.

A Tabela 16 apresenta as estimativas dos β 's, seus erros padrões e

Tabela 15: Valores do teste quociente de verossimilhança (L.R.) e do λ para verificar hipóteses sob modelos encaixados, usando as macros de Hinde (1996).

Modelo	λ (d.p.)	$2 * \ell(\mu, \lambda; z_i)$	L.R.(Pr(Chi))
rot+reg+rot*loc	0,4143(0,0139)	113.873	-
ida+rot+reg+rot*loc	0,4579(0,0158)	114.053.	180(0,000)
ida+rot+reg	0,4544(0,0146)	90.560,02	-
ida+rot+reg+rot*loc	0,4579(0,0158)	114.053	23.492,98(0,000)

os correspondentes valores das estatísticas de Wald (ou valores do teste t de Student) para cada um deles. Utilizando esses valores, conclui-se que existe interação entre região e rotação (Figura 5 do Apêndice A). Ainda mais, a combinação da região do município de Brotas e Riberão Bonito com a rotação 2 foi a responsável dessa dependência região-rotação, proporcionando efeito significativo para o valor médio da mortalidade das árvores. O menor número de falhas na rotação 2 deve-se ao fato de que as árvores submetidas ao corte na primeira rotação estavam vivas, com o sistema radicular completo. Após o corte, o sistema radicular continua inalterado fazendo com que a planta suporte o estresse do corte. Na primeira rotação, principalmente nos primeiros anos de vida, a planta é mais sensível a variações climáticas.

A mudança apresentada no valor estimado do índice de agregação, $\hat{\phi}$, foi a seguinte: para o modelo sem covariável, 2,411, e com covariável, 2,184, ambos significativos ($p < 0,05$). Salienta-se que esse valor de $\hat{\phi}$ para o modelo sem idade não pertence ao intervalo de confiança obtido para esse coeficiente quando é incorporada a covariável no modelo, com $\alpha = 0,05$.

Os recursos expostos acima são válidos se a geração e avaliação do modelo Binomial Negativo é o seguinte. Existe um modelo probabilístico do tipo "two-stage", para o qual o valor esperado do número de árvores mortas segue uma distribuição Gama, originando uma distribuição condicional Binomial Negativa (BN) para Z_i . A função de densidade de uma variável aleatória distribuída segundo uma Binomial Negativa, reparametrizada segundo Jorgensen (1989) conduz à relação média-variância

$$Var(Z) = \mu + \mu^2/\lambda \Rightarrow \lambda = \frac{\mu}{\sigma^2 - \mu}$$

Tabela 16: Valores estimados para os parâmetros do modelo de locação após ajuste máxima verossimilhança através das macros de Hinde (1996), usadas para duas funções de ligação.

Ligação Logarítmica			
Parâmetro	Estimativa	Erro Padrão	Valor de t
Constante	0,8608	0,0966	8,9109
Idade	0,2038	0,0156	13,0641
Região2	0,0326	0,0981	0,3323
Região3	0,5403	0,0838	6,4475
Rotação2	0,8566	0,1324	6,4698
Rotação3	0,7976	0,1513	5,2716
Região2xRotação2	-0,6373	0,2036	-3,1302
Região2xRotação3	0,0253	0,3097	0,0817
Região3xRotação2	-0,1464	0,1646	-0,8894
Região3xRotação3	0,6420	1,0670	0,6017
Ligação Raiz Quadrada			
Parâmetro	Estimativa	Erro Padrão	t-"value"
Constante	1,213	0,1171	10,3587
Idade	0,2798	0,0234	11,9573
Região2	0,0862	0,1169	6,3913
Região3	0,7318	0,1145	5,6193
Rotação2	1,3610	0,2322	5,8613
Rotação3	1,1540	0,2705	4,2662
Região2xRotação2	-0,8748	0,3304	-2,6477
Região2xRotação3	0,0966	0,5922	0,1631
Região3xRotação2	0,0148	0,3101	0,0492
Região3xRotação3	1,4960	2,789	0,5364

e portanto para um λ fixo, o modelo BN pertence à família exponencial e então é um modelo linear generalizado. Os valores do λ considerados como fixos são aqueles obtidos em cada iteração, e colocados dentro do algoritmo de estimação para os β 's apresentado na seção (2.2.3). Se a intenção for reforçar a idéia descrita em um parágrafo anterior, onde se alerta de como são obtidas essas estimativas de máxima verossimilhança (e onde também se introduz o conceito de "perfil de verossimilhança), esse modelo BN, ou "two stage" é um modelo exponencial de dispersão, como será descrito na seção (6).

Por outro lado, e levando em conta a sugestão do Hilbe (1995), foi considerado o modelo binomial negativo somente para as observações positivas, isto é, somente foram avaliadas as parcelas que mostraram falhas ou mortalidade das árvores. Salienta-se que, no ensaio de *E. grandis*, aproximadamente 36 % dos dados coletados tinham valores nulos (sem mortalidade). A Tabela 17 apresenta os valores das estatísticas para avaliar a aderência do modelo, e a estimativa para o índice de agregação da distribuição da variável número de indivíduos mortos.

Pelo resumo das estatísticas dos resíduos componentes de "deviance" padronizados externamente, existiriam alguns valores aberrantes dentre as 1.668 realizações da variável, mas dada a boa performance da "deviance" e do teste do R.L. (quociente de verossimilhança), esse modelo pode ser considerado como adequado também para dados somente positivos. Ainda mais, poderia ser pesquisado um modelo para a variável mortalidade tal que condicionado às observações não nulas seja um binomial negativo.

No caso Poisson, encontrar a função de probabilidade condicional ao truncamento não é tarefa complicada, no caso da Binomial Negativo já não é tão simples. Hurrell (1993) apresenta a Binomial Negativa Truncada, baseado em Rolph *et al.* (1981)¹, e comenta que, para o ajuste de um MLG com erro Binomial Negativo Truncado, há necessidade de utilizar-se o algoritmo EM (seção (3.3) dada a dificuldade de expressar a função de verossimilhança como membro da família exponencial. Esse modelo torna-se útil quando λ tende a zero, e assim a distribuição Binomial Negativa Truncada tende à distribuição das Séries Logarítmicas, muito usadas na áreas de Ecologia e Entomologia.

A Tabela 18 apresenta as estimativas dos β 's, seus erros padrões e os correspondentes valores das estatísticas de Wald para cada um deles. Observe-se

¹ROLPH, J.E; CHAIKEN, J.M; HOUCHENS, R.L. Methods for estimating crime rates of individuals. Report R-2730-NIJ. Santa Monica: Rand Corporation, 1981. 24p.

Tabela 17: Resultados para o ajuste máxima verossimilhança do modelo Binomial Negativo para observações positivas, através das macros de Hinde (1996).

Para Observações Positivas - Ligação Logarítmica			
Estatísticas	Valor(d.p.)	Estatísticas	Valor
$\hat{\lambda}$	9,2800(0,5384)	$\hat{\phi}$	0,1078
"Deviance"	1.760,9	"Deviance"/g.l.	1,0621
2xLog-Veros. (L.V.)	11.9832 (**)	g.l	1.658
2x"Full"L.V	-10.871	Escala	1,0000
$[Minr_{(d_i)}, Maxr_{(d_i)}]$	$[-3, 1456; 5, 8205]$	$Medr_{(d_i)}$	-0,1027

que para este caso que também existem evidências de dependência entre os fatores região e rotação, sendo a combinação da região do município de Brotas e Ribeirão Bonito (região 2), e a região dos municípios de Altinópolis, São Simão, Cajurú, Santa Rita do Passa Quatro e Santa Rosa de Viterbo (região 3), com a rotação 2 a responsável por essa interação. Não obstante existir dependência entre os fatores, podem-se observar quão marcantes são os efeitos principais, basicamente aquele correspondente à rotação.

Em termos de comportamento de ajuste, e visando à estimação dos parâmetros envolvidos no preditor linear, as macros fornecidas pelos autores mencionados nesta seção (e estendidas para aplicação de técnicas de diagnósticos) tiveram performance semelhante. A grande diferença foi o valor estimado para o $\lambda = 1/\phi$ no SAS, de 0,8595, não contido no intervalo de confiança para esse parâmetro segundo os outros dois enfoques (GLIM e SPlus). Porém, isso não significa que os cálculos e as respectivas aproximações para a função Digama, feitas por Hilbe (1995) nessa macro estejam errados. Pelo contrário, acredita-se que a grande semelhança entre os resultados de Venables & Ripley (1994) e Hinde (1996) foi simplesmente devido ao fato que o segundo autor construiu a macro em GLIM adaptando à já existente em SPlus.

O que há necessidade de resgatar é que nenhuma dessas estimativas forneceu indícios que sustentem um modelo Poisson usual para a variável de resposta, já que todas elas foram diferentes do valor 1. Isto foi verificado pelas estatísticas avaliadas para a aderência do modelo proposto. Cabe assinalar que a ferramenta

Tabela 18: Valores estimados para os parâmetros do modelo de locação somente para observações positivas através das macros de Hinde (1996).

Parâmetro	Estimativa	Erro Padrão	Valor de t
Constante	4,3710	0,3775	11,5788
Idade	1,2851	0,0689	18,6517
Região2	-0,9582	0,3353	-2,8577
Região3	4,0083	0,3531	11,3517
Rotação2	14,3464	1,0215	14,0144
Rotação3	16,6985	1,3333	12,5242
Região2xRotação2	4,4345	1,9427	2,2826
Região2xRotação3	4,3640	2,9754	1,4667
Região3xRotação2	6,3412	1,345	4,7146
Região3xRotação3	-0,5457	7,5147	0,0762

fornecida por Hinde (1996) em GLIM4.0 resulta a mais útil, uma vez que ela abrange várias possibilidades para lidar com o parâmetro de superdispersão, e facilita a manipulação, e portanto a aderência, do modelo permitindo a opção de escolher diferentes funções de ligação.

Portanto, a variável aleatória mortalidade de árvores de *E. grandis* seguiu um modelo Poisson com extra variação, onde essa estrutura de variação mostrou ser corretamente modelada segundo a estrutura $V(Z) = \mu + \phi\mu^2$, o que resulta em um modelo probabilístico condicional Binomial Negativo. Ou, em outras palavras, o valor esperado do número de árvores mortas em florestas de três rotações de *E. grandis* implantadas em três regiões do Estado de São Paulo, pensado como variável aleatória, teve comportamento estocástico conforme uma distribuição Gama, o que implicou que a distribuição do número de mortes era agregada, do tipo Binomial Negativa.

5.3 Metodologias Alternativas

5.3.1 Iterando a variância

Uma variável aleatória Z com distribuição Binomial Negativa satisfaz à relação $Var(Z) = \mu + \phi\mu^2$. Dado então que o ajuste do modelo linear generalizado no GLIM4.0 precisa de valores dados para o λ (onde $\lambda = 1/\phi$), o algoritmo proposto nesta seção é um processo iterativo extraindo o λ dessa expressão. Lembre-se que os μ_i são fornecidos pelo sistema como os valores ajustados, %fv, os quais são calculados na macro para a opção do modelo do usuário no comando `link`.

O processo iterativo começa com uma "semente" para λ , e depende fortemente da proximidade desse valor à estimativa definitiva. Esse valor pode ser escolhido, por exemplo, como a estimativa obtida pelo método dos momentos, ou como o recíproco do coeficiente de regressão estimado na relação entre as X =médias e as Y =variâncias amostrais, para os nove tratamentos. Para os dados de mortalidade de árvores de *E. grandis*, o valor estimado para $\lambda = 1/\beta_1$, onde β_1 é o coeficiente de regressão, foi igual a 0,6961 com um desvio padrão de 0,1621, indicando significância do valor do parâmetro. Com isso, o algoritmo calcula as estimativas $\hat{\mu}_i$'s através do ajuste do modelo linear generalizado, com componente sistemático igual ao modelo de análise da covariância, função de ligação logarítmica, e erro Binomial Negativo, com λ fixo. O processo prossegue com o cálculo de λ segundo:

$$\lambda = \frac{\hat{\mu}_i^2}{Var(Z) - \hat{\mu}_i} \quad (19)$$

A macro para sua implementação foi denominada "var.bn", e é apresentada no Apêndice B. A regra de parada foi estabelecida em função da diferença de dois valores consecutivos do λ ser menor ou igual à $[10E - 4]$.

Salienta-se que esta proposta poderia ser uma saída simples e rápida de ajuste de qualquer modelo linear generalizado com erro Binomial Negativo. Isto é, continua na estrutura de um MLG, ou seja, o componente aleatório é estabelecido como um membro da família exponencial a um parâmetro.

Quando é proposto um método de estimação, essa proposta não resulta *convvincente*, e portanto ela não é *completa*, se não apresenta ou descreve a performance das estimativas, isto é, grau de precisão e comportamento assintótico das mesmas. Em outras palavras, há necessidade de que essas estimativas sejam caracterizadas, no sentido de incluí-las em uma família comum, conhecida caso seja possível, ou não.

O que acontece com as estimativas propostas nesta seção? Essas

estimativas são as de **máxima verossimilhança**. De fato, as $\hat{\mu}_i$ são as obtidas pelo algoritmo de estimação sob o estabelecimento do modelo linear generalizado (contemplando as estruturas para os componentes sistemáticos e aleatórios conjuntamente com a função de ligação) e portanto são as de MV; de modo que para λ , utiliza-se o conceito de invariância como uma das boas propriedades dos estimadores de máxima verossimilhança para distribuições uniparamétricas. O conceito é lembrado a seguir.

Considere-se o modelo $f_Z(z; \theta)$, com $\theta \in \Omega$. Suponha-se que para todas as transformações g em um grupo $\mathfrak{N}$, existe um único valor do parâmetro transformado $\xi = g^*\theta$ tal que a distribuição da variável aleatória transformada é dada por $f_Z(z; \xi)$ e $g^*\Omega = \Omega$, então qualquer estimativa $t(z)$ de θ satisfaz que $t(gz) = g^*t(z)$ (Cox & Hinkley; 1974).

Muitas das transformações usadas na Estatística, em geral, originam-se de argumentos físicos, como a escala de medida, ou porque na realidade são expressões para representar quantidades físicas propriamente ditas. Embora, às vezes, as transformações consideradas provêm simplesmente de uma estrutura matemática do problema e não possuem tanta significância, o que faz que a força do princípio de invariância seja fraca.

Essa idéia foi estendida neste material para o caso onde os domínios de definição são subconjuntos A contidos no plano. Isto é, dado que λ é obtido da expressão (19) como uma função h a duas variáveis, os argumentos a seguir tornam-se uma demonstração da propriedade de invariância agora estendida. Seja

$$h : \mathfrak{R}^+ \times \mathfrak{R}^+ \rightarrow \mathfrak{R}$$

uma função tal que

$$\lambda(\mu, \sigma_{(n)}^2) = \frac{\mu^2}{\sigma_{(n)}^2 - \mu}$$

onde $\sigma_{(n)}^2$ representa a variância populacional da variável aleatória que depois será estimada pelo método da máxima verossimilhança, e seja

$$A = \{(\mu, \sigma_{(n)}^2) \in \mathfrak{R}^+ \times \mathfrak{R}^+ : \sigma_{(n)}^2 > \mu\}$$

isto é,

$$A = \{(\mu, \sigma_{(n)}^2) \in \mathfrak{R}^+ \times \mathfrak{R}^+ : 0 < h(\mu, \sigma_{(n)}^2) < 1\}$$

com $h : A \rightarrow \mathfrak{R}$. Sejam

$$(\mu_1, \sigma_{1(n)}^2) \neq (\mu_2, \sigma_{2(n)}^2) \in A$$

logo,

$$h(\mu_1, \sigma_{1(n)}^2) = h(\mu_2, \sigma_{2(n)}^2) \iff$$

$$\frac{\mu_1^2}{\sigma_{1(n)}^2 - \mu_1} = \frac{\mu_2^2}{\sigma_{2(n)}^2 - \mu_2} \iff$$

$$\mu_1^2(\sigma_{o(n)}^2 - \mu_1) = \mu_2^2(\sigma_{o(n)}^2 - \mu_2) \iff$$

$$\mu_1^2\sigma_{o(n)}^2 - \mu_2^2\sigma_{o(n)}^2 = \mu_1^2\mu_2 - \mu_2^2\mu_1 \iff$$

$$\sigma_{o(n)}^2 = \frac{\mu_1\mu_2}{\mu_1 + \mu_2}$$

o qual é absurdo dada a definição do subconjunto $A \subseteq \mathfrak{R}^+ \times \mathfrak{R}^+$. Além disso,

$$\begin{aligned} \frac{\partial h(\mu, \sigma_{(n)}^2)}{\partial \mu} &= \frac{2\mu(\sigma_{(n)}^2 - \mu) - (-1)\mu^2}{(\sigma_{(n)}^2 - \mu)^2} \\ &= \frac{2\mu\sigma_{(n)}^2 - \mu^2}{(\sigma_{(n)}^2 - \mu)^2} \\ &= \mu \frac{(2\sigma_{(n)}^2 - \mu)}{(\sigma_{(n)}^2 - \mu)^2} \end{aligned}$$

mas, dado que $\mu > 0$, por definição, $\sigma_{(n)}^2 > \mu$ e $(\sigma_{(n)}^2 - \mu)^2 > 0$, $\forall (\mu, \sigma_{(n)}^2)$ tem-se que a função $h(\mu, \sigma_{(n)}^2)$ é monótona crescente $\Rightarrow h(.,.)$ é 1-á-1. E, para a outra variável, se tem:

$$\frac{\partial h(\mu, \sigma_{(n)}^2)}{\partial \sigma_{(n)}^2} = \frac{-\mu^2}{(\sigma_{(n)}^2 - \mu)^2} < 0, \quad \forall \sigma_{(n)}^2$$

$\Rightarrow h$ é injetora na reta real positiva.

Assim, dado que essa função h proposta é monótona em ambas variáveis (e portanto marginalmente 1-a-1), substituindo os μ_i e $\sigma_{(n)}^2$ pelas respectivas estimativas de máxima verossimilhança, $\hat{\mu}_i$ e $\hat{\sigma}_{(n)}^2$, obtém-se também uma **estimativa $\hat{\lambda}$ de MV** para λ .

Contudo, os dados de mortalidade de *E. grandis* foram submetidos à macro "var.bn", obtendo-se, após cinco iterações, um valor de λ igual a 0,4201, semelhante ao obtido nas macros do GLIM e do SPlus.

Porém, para a realização de inferências há necessidade de conhecer a distribuição amostral do λ , ou pelo menos algumas características do seu comportamento. Essa tarefa é uma das mais trabalhosas na pesquisa da Teoria Estatística. A solução clássica para contornar esse problema é o uso de **aproximações assintóticas**. A prática de utilizar aproximações assintóticas é bastante comum em toda a Estatística.

Sabe-se que, sob determinadas condições de regularidade, os estimadores de máxima verossimilhança (e de mínimos quadrados também), têm distribuição assintótica normal, independentemente da distribuição original dos erros. Isto é, para um tamanho de amostra suficientemente grande, a distribuição desses estimadores estará próxima da distribuição normal. Para o caso em estudo¹, Jorgensen (1989) mostrou² que se $\mu = \mu(\beta)$ e $\sigma^2 = 1/\lambda$, então se satisfaz que $(\hat{\beta}, \hat{\sigma}^2)$ é assintoticamente normal, com matriz de variâncias dada pela inversa da matriz de informação,

$$\begin{pmatrix} i(\mu(\beta) | \sigma^2(\lambda))/\sigma^2(\lambda) & 0 \\ 0 & i(\sigma^2(\lambda) | \mu(\beta)) \end{pmatrix}$$

onde, para a estrutura de um modelo linear generalizado com hipóteses do tipo $H_1 : \mu = \mu(\beta)$, com $\beta = (\beta_1, \dots, \beta_p)^T$,

$$i(\beta | \sigma) = \mathbf{X}(\beta)^T \mathbf{W} i(\mu(\beta) | \sigma) \mathbf{X}(\beta)$$

com $\mathbf{X}(\beta) = \partial \mu / \partial \beta^T$, e

¹Na realidade, esse resultado foi provado para qualquer modelo exponencial de dispersão, dos quais, os MLG são uma sub-classe.

²Introduz-se notação nova com o objetivo de familiarizar o leitor com as expressões utilizadas nos modelos exponenciais de dispersão, os quais serão apresentados em uma próxima seção.

$$i(\sigma^2 | \mu) = E_{\mu, \sigma^2} \{-c''(Z, \sigma^2) + D(Z, \mu)/\sigma^6\}$$

com D definida como a "deviance", e $c(z, \sigma^2)$ uma função arbitrária tal que a logaritmo da função de verossimilhança satisfaça:

$$L(\mu, \sigma^2) = c(z, \sigma^2) - D(z, \mu)/(2\sigma^2)$$

Note-se que os zeros fora da diagonal ocorrem pela independência das observações e pela ortogonalidade dos parâmetros. Assim, para o caso Binomial Negativo, a matriz resulta:

$$\begin{pmatrix} 1/\mu(\beta)(1 + \mu(\beta)) & 0 \\ 0 & i(\sigma^2(\lambda) | \mu(\beta)) \end{pmatrix}$$

onde se mostra a independência do $i(\beta | \sigma)$ em relação a σ .

Em particular, $\hat{\beta}$ e $\hat{\sigma}^2$ são assintoticamente independentes, e a distribuição normal para o $\hat{\sigma}^2$ é dada por:

$$\hat{\sigma}^2 \sim N(\sigma^2, i^{-1}(\sigma^2 | \mu(\beta)))$$

Jorgensen (1989) mostrou que, para tamanhos grandes de amostra, se satisfaz $i(\sigma^2 | \mu) \cong n/(2\sigma^4)$. Assim,

$$\hat{\sigma}^2 \sim N(\sigma^2, (2\sigma^4)/n)$$

Por exemplo, no caso da estimativa obtida pelas macros de Hinde (1996), o

$$\lambda = 1/\sigma^2 = 0,4579$$

e

$$\begin{aligned} \frac{(2\hat{\sigma}^4)}{n} &= \frac{2}{(2.605 * 0,4579^4)} \\ &= 0,0174 \end{aligned}$$

então, $\sqrt{0,0174} = 0,1321$, seria o correspondente erro padrão estimado pela aproximação normal. Obviamente que esse valor é mais grosseiro que aquele obtido por

Hinde (1996), através de aproximações em séries de Taylor para a expressão do erro da estimativa¹ e cujo valor era 0,0158. Embora, acredita-se que, intervalos de confiança a partir dele podem ser de valiosa utilidade na pesquisa aplicada, onde às vezes, não se dispõe dos recursos computacionais necessários para implementar tais macros.

Para a estimativa do parâmetro de dispersão, obtida pelo método apresentado nesta seção e definido como "iterando a variância", i.é., $\hat{\lambda} = 0,4201$, o erro padrão correspondente à aproximação normal seria igual a:

$$\begin{aligned}\frac{(2\hat{\sigma}^4)}{n} &= \frac{2}{(2.605 * 0,4201^4)} \\ &= 0,0246\end{aligned}$$

então, $\sqrt{0,0246} = 0,1570$. Portanto, pode-se inferir que, a estimativa de máxima verossimilhança do índice de agregação ϕ é 2,3804, altamente significativa, já que a estatística de Wald para o $\hat{\lambda}$ o é.

Por outro lado, se considerarmos o modelo Binomial Negativo para observações positivas (isto é, desprezando as parcelas que não tiveram falhas), foi mostrado que o ajuste do modelo Binomial Negativo é adequado, conforme tinha recomendado o Prof. Hilbe. Para esses dados, esta metodologia apurou um valor para a estimativa do parâmetro de dispersão igual a:

$$\begin{aligned}\frac{(2\hat{\sigma}^4)}{n} &= \frac{2}{(1.658 * 8,01^4)} \\ &= 0,00000029\end{aligned}$$

então, $\sqrt{0,00000029} = 0,00054$. Cabe assinalar que, dado o valor estimado para o $\phi = 0,1248$, muito próximo do 0, estar-se-ia praticamente trabalhando com outro modelo, o modelo Poisson, uma vez que $\mu + \phi\mu^2$ com $\phi \cong 0 \Rightarrow V(Z) = \mu$. Portanto, a maneira exposta para obter os erros padrões assintóticos carece de validade.

Para a tomada de decisões, é importante lembrar que este enfoque fornece estimativas com erros assintóticos e estatísticas com distribuições aproximadas. Além disso, é freqüente, nas aplicações propriamente ditas, encontrarmos o problema de não saber quão grande deve ser o n para se ter uma "boa" aproximação. Para um tamanho amostral que não seja suficientemente grande a validade

¹O autor utiliza para esses cálculos os recursos do sistema GLIM, como o são os %pw, componentes do vetor de pesos "a priori".

das aproximações poderá ficar bastante comprometida. No nosso caso, acredita-se que o n igual a 2.605 pode ser considerado como "suficientemente grande". Além do mais, as condições de regularidade estão satisfeitas¹, pelo que se optou pelo uso da distribuição assintótica do estimador de λ para fornecer uma medida de precisão para a mesma.

Portanto, esse algoritmo apresenta uma saída muito simples e rápida ao problema de estimação, por máxima verossimilhança, dos parâmetros contidos no modelo de locação, e portanto de interesse principal, e do parâmetro de dispersão de uma distribuição Binomial Negativa (ou, parâmetro de superdispersão no caso de um modelo probabilístico Poisson).

Um outro caminho a seguir para o conhecimento do comportamento do estimador do λ seria o uso dos chamados *métodos modernos em Estatística*. Nas duas últimas décadas, vem-se desenvolvendo juntamente com o advento da computação de alta velocidade, com grandes capacidades de armazenamento de dados a custos mais acessíveis, uma nova metodologia para análise estatística conhecida como *computacionalmente intensiva*. Segundo Silva, N. (1995), os métodos computacionalmente intensivos começaram a se tornar ferramentas atraentes, pois, eles visam substituir as complexidades analíticas ou uso de aproximações assintóticas de um determinado problema por computação maciça. Entre os métodos mais importantes, pode-se destacar o "*Jackknife*" e o "*Bootstrap*" (Efron, 1979).

Suponha-se uma amostra aleatória $z_1, z_2, \dots, z_n$ extraída independentemente de um membro da família uniparamétrica $\{F_\theta \mid \theta \in \Theta\}$ de distribuições. No nosso caso, os μ_i estarão fixos dado que os parâmetros da Binomial Negativa podem considerar-se como sendo independentes (e portanto as suas estimativas), e serão iguais às estimativas de máxima verossimilhança. Seja então $\hat{\lambda} = T(z)$ uma função da amostra, usada como a estatística do parâmetro de dispersão λ .

O procedimento sugerido para ser utilizado é o "*bootstrap*". Basicamente, o método é um conjunto de técnicas de reamostragem para se obterem informações das características da distribuição de alguma variável aleatória, que não podem ser facilmente avaliadas por métodos analíticos tradicionais ou cuja aproximação existente tenha suposições questionáveis em alguma situação.

Em termos práticos, o método consiste em extrair m amostras da amostra $\{z_i\}$ com reposição e calcular os $\hat{\lambda}_j$ para cada uma delas. A idéia é avaliar a variabilidade do $\hat{\lambda}$ em torno do parâmetro desconhecido $1/\phi = \lambda$ pela variabilidade

¹Uma demonstração, bastante matemática por sinal, dessas condições pode ser encontrada em Jorgensen (1989), págs. 112-116.

do $\hat{\lambda}_j$ em torno do $\hat{\lambda}$. Em particular, o viés do $\hat{\lambda}$ pode ser estimado pela média de $(\hat{\lambda}_j - \hat{\lambda})$. Por exemplo, pode-se estar interessado na distribuição do $1/\phi$, para o qual a teoria assintótica normal estabelece que sua distribuição aproximada, para tamanhos de amostras grandes, é a normal.

Mas, para não introduzir um tópico novo nesta seção, o leitor que quiser se aprofundar na construção propriamente dita das estimativas "bootstrap" das variâncias do $\hat{\lambda}$, pode consultar Silva N. (1995); Efron (1979); Efron & Tibshirani (1986), dentre outros. O que será usado neste material é o recurso que a simulação Monte Carlo fornece ao método de reamostragem "bootstrap". Ou seja, felizmente sempre é possível obter uma aproximação numérica para $\hat{\sigma}_{boot}^2(\hat{\lambda})$ através de simulação de Monte Carlo, seguindo-se o seguinte algoritmo. Este algoritmo não converge para o $\hat{\sigma}_{boot}^2(\hat{\lambda})$ se o tamanho da amostra "bootstrap" diferir do n original. O algoritmo resume-se nos seguintes passos:

1. usando um gerador de números aleatórios, extrair B amostras "bootstrap" independentes, $\mathbf{z}^*(1), \mathbf{z}^*(2), \dots, \mathbf{z}^*(B)$, isto é, B amostras aleatórias simples de tamanho n , extraídas com reposição da amostra atual $\{z_1, z_2, \dots, z_n\}$;
2. para cada amostra $\mathbf{z}^*(b)$, calcular a estatística de interesse, $\hat{\lambda}^*(b) = t(\mathbf{z}^*(b))$, com $b = 1, \dots, B$;
3. calcular a variância amostral dos B valores $\hat{\lambda}^*(1), \hat{\lambda}^*(2), \dots, \hat{\lambda}^*(B)$, isto é,

$$\hat{\sigma}_B^2(\hat{\lambda}) = \frac{\sum_{b=1}^B (\hat{\lambda}^*(b) - \hat{\lambda}^*(.))^2}{B - 1}$$

onde

$$\hat{\lambda}^*(.) = \frac{\sum_{b=1}^B \hat{\lambda}^*(b)}{B}$$

Mas, como parte do processo de inferência estatística, não seria suficiente somente avaliar essas determinadas características com o procedimento de estimação pontual. Deseja-se recorrer a um mecanismo de estimação por intervalo, que possibilite avaliar o erro que se comete nessa estimação pontual. A avaliação desse erro pode ser feita com a construção de um intervalo de confiança para o parâmetro de interesse, isto é, um intervalo do tipo $[l_1(z), l_2(z)]$, tal que, a probabilidade deste intervalo conter o verdadeiro valor do parâmetro, λ , seja igual a um valor pré-fixado $1 - 2\alpha$, ou seja,

$$Pr_P\{l_1(Z) \leq \lambda \leq l_2(Z)\} = 1 - 2\alpha$$

O lado esquerdo é denominado *probabilidade de cobertura*, e, o valor $1 - 2\alpha$ ($0 < \alpha < 0.5$) é o *nível de confiança* do intervalo. O valor (2α) é denominado *nível de significância* do intervalo.

Esse intervalo é construído sob a suposição de normalidade aproximada da estatística padronizada. Isto é, seja $\hat{\sigma}_{boot}(\hat{\lambda})$ a estimativa "bootstrap" de $\sigma = \{V(\hat{\lambda})\}^{1/2}$, definida no parágrafo anterior. O intervalo de confiança *Bootstrap padrão* para o parâmetro λ , com probabilidade de cobertura de aproximadamente $1 - 2\alpha$, é dado por:

$$[\hat{\lambda} + z_{\alpha}\hat{\sigma}_{boot}(\hat{\lambda}), \hat{\lambda} - z_{\alpha}\hat{\sigma}_{boot}(\hat{\lambda})]$$

onde z_{α} é o α -ésimo percentil da distribuição normal padrão, isto é, $z_{\alpha} = \Phi^{-1}(\alpha)$.

A construção desse intervalo baseia-se na aproximação assintótica,

$$T = \frac{\hat{\lambda} - \lambda}{\hat{\sigma}_{boot}(\hat{\lambda})} \sim N(0, 1)$$

Caso a distribuição de T seja perfeitamente normal padrão, o intervalo é exato, caso contrário, é somente aproximado.

Existem outras maneiras de se construírem intervalos "bootstrap", os quais se relacionam com outras aproximações. Porém, a principal vantagem do intervalo apresentado é a sua facilidade de construção. Mesmo quando a estimativa "bootstrap" for complicada de ser calculada analiticamente, pode-se utilizar em seu lugar, a aproximação Monte Carlo, como foi feito neste caso. A grande desvantagem tem a ver com sua exatidão. Para tamanhos de amostras que não sejam suficientemente grandes para a validade da aproximação da T , ou que, características como assimetria, tendenciosidade, etc, que podem estar presentes na distribuição de T , podem prejudicar o desempenho dos intervalos "bootstrap" no sentido da sua exatidão.

Por outro lado, e para fornecer estimativas mais precisas (e assim independentes do tamanho da amostra) o SPLUS fornece ferramentas para a implementação de algoritmos de buscas da distribuição do $\hat{\lambda}$. No Apêndice B apresenta-se a macro para sua implementação. Essa macro leva apenas um minuto de cálculo. O gráfico que esse algoritmo apresenta representa a distribuição "bootstrap" do $\hat{\lambda}_j$ em torno do $\hat{\lambda}$, o qual informará se a aproximação foi, ou não, adequada. Isto está

relacionado com os métodos de "smoothing" de estimação não paramétrica.

5.3.2 Enfoque Intuitivo

Nesta seção é apresentado um método alternativo de busca do λ . Ele é alternativo aos já descritos pois os valores do λ mudam dentro de um intervalo plausível de valores, sugerido por exemplo pela Figura 2 no Apêndice A acerca da relação média-variância. Isto é, a regressão linear entre as médias e os desvios padrão das três regiões combinadas com três rotações, mostrou significância ($p < 0,001$) com um coeficiente linear estimado ($1/\lambda = \beta_1$, onde β_1 é o coeficiente de regressão) igual a 1,4366, indicando como intervalo plausível, por exemplo, o intervalo $[0, 2; 1, 5]$.

O algoritmo propõe o ajuste do modelo linear generalizado para todos os valores do λ dentro do intervalo escolhido selecionando como estimativa desse parâmetro aquele valor que deixa a "deviance" residual mais próxima dos graus de liberdade para essa fonte de variação.

É óbvio que este método de procura não oferece mais do que um valor isolado do λ , útil para o futuro ajuste dos parâmetros de interesse embutidos no preditor linear, e que essa busca é feita de maneira **independente** deles. Portanto, nada poderia ser dito acerca das propriedades assintóticas dessa "estimação", nem do seu erro padrão. Além disso, é importante salientar que essa metodologia pode ser usada, *sem muita complexidade*, como alternativa prática para ajustes de modelos com observações apresentando extra-variação.

Por outro lado, o fato do λ ser selecionado sob esse critério não indica que se obtenha algum valor da "deviance" residual próxima dos seus graus de liberdade, ou ainda, que o modelo seja o adequado para o conjunto de dados em estudo. Pelo contrário, independentemente disso, haveria a necessidade de submeter ao modelo as devidas técnicas de diagnóstico para tal fim, como as propostas na seção (2.2.3).

A macro para sua implementação é apresentada no Apêndice B, e a Figura 7, do Apêndice A, mostra o valor do λ que deixa o quociente ["deviance" residual]/[graus de liberdade] próximo a um. Observe-se que esse valor foi de 0,3801, semelhante à estimativa de máxima verossimilhança. Já que esse método possui a grande vantagem da sua simplicidade de implementação, poderia construir-se uma medida do erro padrão. Isto é, da amostra original podem ser extraídas com reposição, g amostras aleatórias, as quais serão submetidas ao algoritmo. Desse procedimento surgirão os g valores de λ que fornecerão uma estimativa do erro

padrão, conforme foi exposto na seção anterior através do método "*bootstrap*" de reamostragem.

5.4 Eficiência na Estimação

5.4.1 Alguns Comentários Gerais

As estimativas de máxima verossimilhança MV de $\beta(\lambda)$ são obtidas maximizando a logaritmo da função de verossimilhança $\ell(\beta, \lambda)$ em relação a β para valores selecionados do λ . Se essas estimativas são colocadas na equação da logaritmo da função de verossimilhança para λ , o procedimento de estimação é chamado de Perfil de Verossimilhança, o qual em forma iterativa conduz às estimativas $\hat{\beta}(\hat{\lambda})$ e $\hat{\lambda}(\hat{\beta})$. Esse foi o enfoque utilizado na seção (5.2), isto é, as estimativas surgem do método de máxima verossimilhança propriamente dito pois o $\ell(\beta, \lambda)$ é maximizado em relação aos dois argumentos, supondo a função de verossimilhança como função do (β, λ) dada a amostra $(z_1, \dots, z_n)$ variáveis. Já foi mencionada a complexidade do cálculo sob este enfoque, dado que, no caso Binomial Negativo, as aproximações disponíveis para a função Digama não são de fácil implementação.

Os outros enfoques de busca do λ não implementa perfis de verossimilhança, apenas procura o λ em função de dois critérios diferentes. O primeiro deles fornece estimativas com boas propriedades, uma vez que elas podem ser expressas como funções um-a-um de estimativas de máxima verossimilhança, garantindo portanto a propriedade da invariância do método de estimação. O outro algoritmo de busca do λ é totalmente independente das estimativas dos β 's, pelo qual estudos de simulação e reamostragem são necessários na hora de informar a precisão de tais estimativas. Acerca do comportamento assintótico dos β 's na seção (6) foram mencionadas as distribuições limites.

Por outro lado, já foi dito que quando o λ é conhecido, o modelo probabilístico Binomial Negativo é um modelo linear generalizado. Ainda mais, as equações de MV, $\partial \ell / \partial \beta_r, r = 1, \dots, p$ são também as equações de máxima QV, e as equações de mínimos quadrados ponderados (Breslow, 1984).

Para considerar o λ fixo, outros enfoques usam estimadores dos momentos. Segundo Moore (1986), o estimador dos momentos do λ é considerado como mais robusto que o estimador de MV, mas é menos eficiente quando o modelo binomial negativo é o correto. Usando Lawless (1987) pode ser provado que sob condições de regularidade, β e λ são assintoticamente independentes e normalmente distribuídos. Isto indica que se se procede segundo Breslow (1984), o estimador de

β obtido é assintoticamente equivalente ao de máxima verossimilhança, e sua matriz de variâncias e covariâncias assintótica é consistentemente estimada pela matriz de informação de Fisher. Porém, existe alguma perda da eficiência na estimação de λ pelo método dos momentos.

Lawless (1987) apresenta um estudo básico realizado por Ascombe (1950) acerca da eficiência do λ relativo ao caso em que há apenas um modelo de médias para o valor esperado da variável, $\mu_i = \mu \forall i = 1, \dots, n$. O autor encontrou que a eficiência relativa RE aproxima-se de 1, quando $\lambda \rightarrow 0$, e nas situações mais reais, RE alcança 0,90, não sendo válido esse resultado quando existem covariáveis no modelo para o valor esperado, como é o nosso caso em que existe uma covariável, a idade das árvores.

5.4.2 Caso Binomial Negativo

Mudando a notação, sabe-se que a binomial negativa pode ser escrita como expansão $(q - p)^{-\lambda}$, onde $q = 1 + p$, λ positivo, ou segundo a expressão:

$$p^z \frac{1}{(1 + p)^{\lambda+z}}$$

Na área experimental para o caso em que o expoente for desconhecido, essa distribuição é obtida de uma extensão simples das condições que apresentam as séries Poisson. As séries Poisson são obtidas quando amostras iguais são extraídas de material perfeitamente homogêneo, e é completamente determinada pela média μ , ou o número esperado de ocorrências por amostra. No caso de amostras desiguais, ou material não homogêneo, o valor de μ poderia variar de amostra para amostra.

A expressão para os estimadores pelo método dos momentos dos parâmetros é equivalente à da Binomial Positiva, mudando o sinal de p e lembrando que λ corresponde a $(-n)$ e $q = 1 + p$. Em consequência, para amostras grandes, as equações de estimação são:

$$\hat{p} = \frac{Var(Z) - \bar{Z}}{\bar{Z}}, \quad \hat{\lambda} = \frac{\bar{Z}}{Var(Z) - \bar{Z}}$$

onde $\bar{Z}$ é a média aritmética e $Var(Z)$ a variância amostral. No nosso caso, $N = 2.605$, $\bar{Z} = 11,307$ e $Var(Z) = 166,290$, e assim $p = 13,707$ e $\lambda = 0,825$.

Para examinar a eficiência do método precisa-se do determinante da matriz de variâncias e covariâncias das estimativas do p e λ , denotada por C . A matriz C de $\bar{Z}$ e $Var(Z)$ para N grande é,

$$\frac{1}{N} \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & (\mu_4 - \mu_2^2) \end{pmatrix}$$

onde $\mu_2 = pq\lambda$, $\mu_3 = pq(p+q)$, $\mu_4 - 3\mu_2^2 = pq(1+6pq)$. Substituindo pelo p e λ resulta:

$$\frac{1}{N^2} \begin{pmatrix} pq\lambda & pq(p+q)\lambda \\ pq(p+q)\lambda & pq(1+6pq)\lambda + 2p^2q^2\lambda^2 \end{pmatrix}$$

cujo determinante é

$$\frac{2}{N^2} p^3 q^3 \lambda^2 (\lambda + 1)$$

Para se obter o determinante da matriz de variâncias e covariâncias das estimativas do p e λ , precisa-se apenas multiplicar pelo quadrado do jacobiano:

$$\frac{\partial(p, \lambda)}{\partial(\bar{Z}, Var(Z))}$$

O jacobiano é:

$$\begin{pmatrix} -\frac{Var(Z)}{\bar{Z}} & \frac{1}{\bar{Z}} \\ \frac{\bar{Z}(2Var(Z) - \bar{Z})}{(Var(Z) - \bar{Z})^2} & \frac{\bar{Z}}{(Var(Z) - \bar{Z})^2} \end{pmatrix} = \frac{-1}{Var(Z) - \bar{Z}} = \frac{-1}{p^2\lambda}$$

Portanto, o determinante da matriz de p e λ estimados pelos dois primeiros momentos é:

$$\frac{2p^2(\lambda + 1)}{N^2p}$$

o qual para nosso caso é $694,842/93.014,687 = 0,0000074$. Assim, pode-se comparar isto com o correspondente determinante de qualquer outro método de estimação. A forma mais conveniente é obter a matriz de informação, a qual será a recíproca da matriz de variâncias e covariâncias para a estimação proposta como eficiente. Mostra-se que o determinante da matriz de informação é:

$$\frac{N^2}{p\lambda} \sum_{z=2}^{\infty} \frac{1}{z} \frac{p^z}{q^z} \frac{(z-1)! \lambda!}{(\lambda+z-1)!}$$

Se o determinante da matriz de covariância correspondente à qualquer método de estimação é multiplicado pela expressão acima, tem-se a recíproca

da eficiência. Para o método dos momentos:

$$\frac{1}{E} = 1 + \frac{4}{3} \frac{p}{q(\lambda + 2)} + 3 \frac{p^2}{q^2(\lambda + 2)(\lambda + 3)} + \dots + \dots$$

Note-se que $E < 1$ será próximo de 1 quando

$$\frac{p}{q(\lambda + 2)} = \frac{\bar{Z}}{(\bar{Z} + \lambda)(\lambda + 2)}$$

for pequeno. No caso da mortalidade de *E. grandis* essa igualdade teve um valor de 0,326.

Por outro lado, quando $\bar{Z}$ é pequeno e $\lambda \approx 1$, a eficiência é alta e a expressão acima tem como limite $1/33$; se $\lambda \rightarrow 0$, a eficiência é baixa; se $\lambda > 9\bar{Z}$, então o limite é $1/20$. Nenhum desses casos caracteriza as observações de mortalidade das árvores em estudo. Quando a média está próxima de 10, λ deve ser 9 para obter-se esse limite esperado. Se o produto $(1 + \frac{1}{p})(\lambda + 2) > 20$, a eficiência está garantida. Para *E. grandis* esse produto foi 3,057 o que leva a suspeitar de uma eficiência muito baixa para a estimação do método dos momentos.

Calculando, tem-se:

$$\frac{1}{E} = 1 + \frac{4}{3} \frac{p}{q(\lambda + 2)} + 3 \frac{p^2}{q^2(\lambda + 2)(\lambda + 3)} + \dots + \dots$$

$$\begin{aligned} \frac{1}{E} &= 1 + 0,4399 + 0,2412 + 0,1490 + 0,0993 + 0,0697 + \\ &= +0,0393 + 0,0081 + 0,0012 + 0,0005 + \dots \end{aligned}$$

$$\frac{1}{E} = 2,1211 \Rightarrow E = 0,4715 \Rightarrow E \approx 50\%$$

Assim, esse processo de estimação é ineficiente para as observações de mortalidade de *E. grandis* e portanto descartado. Isto justifica a adoção do método da máxima verossimilhança para o ajuste do modelo proposto. Johnson & Kotz (1969) apresentam as variâncias assintóticas para os estimadores do λ , e discutem que o estimador do método dos momentos somente será eficiente quando λ for pequeno.

6 UM SUPORTE TEÓRICO MAIOR

6.1 Introdução

É amplamente conhecido que os modelos lineares generalizados (MLG) foram introduzidos por Nelder & Wedderburn (1972) como uma maneira de unificar um número de modelos estatísticos clássicos tais como os modelos da teoria normal e análise da variância, regressão logística, regressão Poisson, modelos log-lineares para tabelas de contingência. A generalização dos métodos de estimação e de inferência também formou parte desse trabalho. O método de estimação é baseado nos mínimos quadrados ponderados iterativos e foi implementado no pacote estatístico GLIM, juntamente com uma das ferramentas mais importantes para a realização de inferências, a chamada análise da "deviance".

Porém, os modelos lineares generalizados se restringem às distribuições pertencentes à família exponencial a um parâmetro. Uma generalização da distribuição do erro pode ser encaminhada através dos *modelos com dispersão*, e dentro deles, os *Modelos Exponenciais com Dispersão*, apresentados por Jorgensen (1986, 1987).

No artigo original, Nelder & Wedderburn (1972), enfatizaram o valor pedagógico dos modelos lineares generalizados, uma idéia que já foi confirmada pela experiência, e ainda mais pelo sucesso que esses modelos tiveram como tópico de estudo e pesquisa. Segundo Jorgensen (1992), a generalização das distribuições para o componente aleatório nos MLG é feita de maneira compreensível em uma classe de interesse, e de fácil análise. No entanto, o entendimento da série completa dos modelos requer conhecimento de uma variedade de tópicos da teoria estatística, e da teoria da probabilidade, tais como famílias exponenciais, processos estocásticos e convergência assintótica.

Os modelos exponenciais com dispersão possuem boas propriedades, generalizando às dos modelos lineares generalizados, em particular, a análise da "deviance" construída para os primeiros, conduz a versões assintóticas das famílias dos testes χ^2 , t e F , cujas estatísticas se constróem de forma semelhante que para os

MLG. Foi Jorgensen (1984), quem estudou as propriedades gerais conjuntamente com a geração de alguns exemplos interessantes nesta nova classe de modelos, antes descartados por não pertencer à família exponencial a um parâmetro.

Embora a grande importância que esses tópicos possuem na área de Modelação Estatística propriamente dita, acredita-se que o grande sucesso dessa generalização se baseia na interpretação, daqueles modelos apresentados na literatura como estáticos, em termos de processos estocásticos com incrementos estacionários e independentes. Serão apresentados os conceitos mais importantes desses modelos apenas para o caso discreto.

6.2 Definições

6.2.1 Modelo Exponencial com Dispersão

Foi Tweedie (1947), citado por Jorgensen (1986) quem apresentou as idéias básicas dos modelos exponenciais com dispersão¹, inclusive as demonstrações de algumas importantes propriedades e classes de MED, porém, outros autores (Morris, 1982; McCullagh, 1983; Jorgensen, 1986, 1987; dentre outros) independentemente, foram fornecendo conhecimento para a construção da teoria geral. Segundo Jorgensen (1992) um modelo exponencial com dispersão é definido pela função de densidade de probabilidade:

$$p(y|\theta, \lambda) = a(\lambda, y) \exp [\lambda(\theta y - K(\theta))], \quad y \in \mathcal{R},$$

onde a e k são funções e λ e θ são parâmetros definidos num subespaço contido no plano real. A variância desse modelo será $V(Y) = \sigma^2 V(\mu)$, com $V(\mu)$ sendo a função de variância. Da mesma forma, um modelo exponencial com dispersão discreto define-se segundo:

$$p(z, \theta, \lambda) = a^*(\lambda, z) \exp [\theta z - \lambda K(\theta)], \quad z = 0, 1, \dots \quad (20)$$

para funções conhecidas $a^*(\cdot, \cdot)$ e $K(\cdot)$ e parâmetros θ e λ tais que $(\theta, \lambda) \in \Theta \times \Lambda \subseteq \mathcal{R} \times \mathcal{R}$.

Observe-se que uma conexão simples entre (20) e a correspondente versão para o caso contínuo obtém-se estabelecendo $1/\lambda = \phi^2 = \sigma^2$, onde σ^2 é o parâmetro de dispersão para o caso contínuo.

¹No texto, em geral, serão denotados por MED.

Denota-se então

$$Z \sim ED^*(\theta, \lambda)$$

com $E(Z) = \lambda\mu$ e $Var(Z) = \lambda V(\mu)$, onde $\mu = \tau(\theta)$ e $Var(\mu) = K''(\tau^{-1}(\mu))$, onde $V(\mu)$ é a função de variância com suporte $\Omega = \tau(int\Theta)$, definida como na família exponencial.

Note-se que um modelo linear generalizado pode, então, ser definido como o seguinte modelo estatístico: sejam as variáveis aleatórias $Z_1, \dots, Z_n$, independentes e identicamente distribuídas, tal que $Z_i \sim ED(\theta_i, \lambda)$ (isto é, pode ser expresso conforme (20), e para o qual o valor esperado satisfaz

$$g(\mu_i) = \sum_{j=1}^k x_{ij}\beta_j = g(\tau(\theta_i)), \quad i = 1, \dots, n$$

com $g(\cdot)$ uma função um-a-um dos β 's (ou dos μ 's sua inversa), chamada função de ligação, o que é compatível com os MED.

Um dos modelos exponenciais com dispersão discretos mais importantes é a distribuição **Binomial Negativa**. Isto é, Z_i têm função de distribuição de probabilidade dada por

$$\begin{aligned} p(z; \phi, \lambda) &= \binom{\lambda + z - 1}{z} \phi^z (1 - \phi)^\lambda \\ &= \binom{\lambda + z - 1}{z} \exp\{\theta z + \lambda \log(1 - \exp \theta)\} \end{aligned}$$

então, e segundo (20),

$$a^*(\lambda, z) = \binom{\lambda + z - 1}{z}, \quad K(\theta) = -\log(1 - \exp \theta)$$

com $\theta = \log \phi < 0$. Logo, o modelo Binomial Negativo será dado por

$$Z \sim BN(\lambda, \mu)$$

com

$$E(Y) = \lambda\mu, \quad \mu = \exp \theta / (1 - \exp \theta)$$

o qual, para valores inteiros do λ , conduz à distribuição de Pascal, interpretada

como o tempo de espera até a falha λ -ésima em uma sequência de ensaios Bernoulli independentes. Por outro lado, e dado que $V(\mu) = \mu(1 + \mu)$, resulta que

$$Var(Z) = \lambda V(\mu) = \lambda\mu(1 + \mu)$$

É neste exemplo que se percebe claramente como se reduz (no sentido de relaxar) o problema da restrição imposta pelos modelos lineares generalizados (McCullagh & Nelder, 1989; Cordeiro, 1986) as distribuições pertencentes unicamente à família exponencial uniparamétrica. Isto é, a variância amostral resulta um produto da variância nominal e o parâmetro de dispersão, agora este considerado parâmetro é originado pela expressão do MED (dentro da família exponencial natural, a família mais ampla que contém a exponencial com um parâmetro usada pelos autores mencionados).

Assim, e como será mostrado nesta seção, os MED formam uma classe de modelos estatísticos de interesse devido a dois fatos principais:

- possuem um número grande de propriedades matemáticas, por exemplo, a geração de um processo estocástico através do conceito de convolução (seção seguinte), e
- incluem distribuições muito utilizadas na pesquisa biológica não contempladas na modelagem geral dos MLG, apenas restringidas considerando o resto dos seus parâmetros como parâmetros "nuisance", e abrangem, como casos muito especiais, um número de importantes distribuições agrupando-as de forma conveniente para técnicas comuns de análises.

6.2.2 Convolução

Este conceito é muito importante na teoria dos MED e generaliza a propriedade de convolução da distribuição normal. Segundo Jorgensen(1987), a fórmula de convolução

$$ED * (\theta, \lambda_1) * \dots * ED * (\theta, \lambda_k) = ED * (\theta, \lambda_1 + \lambda_2 + \dots + \lambda_k)$$

aplica-se a qualquer modelo exponencial com dispersão discreto. Isto pode ser provado utilizando a função geradora de momentos. Assim, podem ser construídos os processos estocásticos com incrementos estacionários e independentes para qualquer MED, isto é:

$$\Delta Z_i = Z(t_i) - Z(t_{i-1})$$

para $t_o < \dots < t_k$.

A expressão da convolução mostra que pode ser definido o processo estocástico $\{Z(t) : \rho \in \Lambda\}$ através de

$$\Delta Z_i \sim ED^*(\theta, \rho(t_i - t_{i-1})), \quad i = 1, \dots, k.$$

$\forall \text{ MED}, \forall (\theta, \rho) \in \Theta \mathfrak{R}$. Do anterior se deduz que (Jorgensen, 1992): a distribuição de $[Z(t_2) - Z(t_o)]$ será

$$ED^*(\theta, \rho(t_2 - t_1)) * ED^*(\theta, \rho(t_1 - t_o)) = ED^*(\theta, \rho(t_2 - t_o))$$

observe-se que este processo assim definido tem incrementos estacionários e independentes.

Portanto, denota-se $Z(t) \sim ED^*(\theta, \rho t)$ à distribuição do processo estocástico, tem-se que $E(Z(t)) = \rho \mu t$, onde o produto $(\rho \mu)$ é considerado como uma taxa do processo. Como exemplo, o processo $Z(t) \sim BN(\rho t, \mu)$ fornecido pelas observações provenientes de um MED Binomial Negativo, e definido segundo

$$\Delta Z_i \sim BN(\rho(t_i - t_{i-1}), \mu)$$

fornece o **processo Binomial Negativo**.

Jorgensen (1989) apresenta um exemplo de construção de um processo aleatório para o caso discreto, o qual foi a base para um possível processo estocástico para a mortalidade das árvores de *E. grandis* em função da informação adicional relacionada com um dos dois fatores estudados como causa de variação. No caso do exemplo do citado autor, a variável aleatória Z_{ij} , número de acidentes em rodovias, foi classificada em função de dois fatores, o tipo da rodovia e a estação do ano avaliada. A propriedade de convolução dos MED permitiu interpretar o modelo em termos de um processo estocástico Poisson no tempo. É interessante a forma na qual o autor constrói este processo, valendo-se fortemente da intuição nas etapas de geração e interpretação, e da teoria dos modelos exponenciais com dispersão na etapa de análises e verificação de hipóteses. Na seção 7.2 será mostrado o processo construído para a mortalidade das árvores.

Na realidade, o caso Poisson, como o exemplo anterior, é um caso especial de MED discreto dado que, pode ser demonstrado, é um modelo limite dos MED discretos. Nesse resultado usa-se o fato que, parametrizando a distribuição de Poisson da forma $E(Z) = \lambda \exp(\theta) = \lambda \mu$ e $Var(Z) = \lambda V(\mu) = \lambda \mu$, o parâmetro (θ, λ) resulta não identificável, somente o é μ . Para maiores detalhes desses argumentos ver Jorgensen (1989), págs. 23-26.

6.2.3 Modelos com Dispersão (MD)

As idéias anteriores foram baseadas pelo próprio autor em uma família maior de modelos, os *Modelos com Dispersão*. Em Jorgensen (1983) são definidos esses modelos apenas para o caso contínuo, e salientada a importância da generalização já que nesta família podem ser incluídas funções de densidades que não pertencem aos MED, como são os casos da distribuição de von Mises-Fisher e a classe de distribuições Simplex, utilizadas para análises de sobrevivência.

6.3 A Função de Variância

A função de variância desempenha um papel muito importante na teoria dos MED, dado que possui a propriedade fundamental de caracterizar a família exponencial natural da qual provêm as observações. O seguinte teorema o expressa formalmente.

Teorema 6.3.1 *A função de variância V com domínio Ω caracteriza $\wp$ dentro da classe de todas as famílias exponenciais naturais [Jorgensen, 1990].*

Note-se que V é determinada a partir de $\wp$ somente até uma transformação de escala do λ , pelo que esse teorema corresponderia à família $\lambda V(m/\lambda)$ onde $m = E(Z)$.

As funções de variância mais estudadas correspondem a três classes de funções, a saber, funções quadráticas, potência e exponencial. Dada a natureza do nosso estudo, serão somente abordados os dois primeiros casos. Define-se função de variância quadrática à

$$V(\mu) = a\mu^2 + b\mu + c$$

para $\mu \in \Omega$. Exemplos imediatos são as distribuições Poisson, Binomial Positiva, Binomial Negativa, Normal e Gama. Todas essas distribuições junto com a secante hiperbólica, e salvo transformações lineares das mesmas, caracterizam todos os MED com função de variância quadrática (Morris, 1982). Se existiram raízes reais desse polinômio, e se o coeficiente a for não nulo, pode ser definido o ponto de estacionalidade $d = -b/2a$, e portanto, $V(d)$ denotará o valor da função de variância no ponto de estacionalidade. No caso em estudo, o da Binomial Negativa, o a é sempre positivo existindo então três possibilidades: $V(d) < 0$, $V(d) = 0$ ou $V(d) > 0$.

Por outro lado, a classe de funções de variância potência, definidas segundo $V(\mu) = \mu^p$, $\forall \mu \in \Omega$, onde $\Omega = \mathbb{R}$ para $p = 0$, e $\Omega = \mathbb{R}$ para $p \neq 0$. Note-se que assim definida, várias distribuições na classe anterior também pertencem à classe potência. Jorgensen (1986) provou que para $0 < p < 1$ não corresponde

nenhum modelo exponencial com dispersão, e para $1 < p < 2$ corresponde à Poisson Composta (ver Jorgensen, 1989, págs. 47-50). Na seção 3.3 foi introduzida essa distribuição com a forma explícita para a parte de continuidade da função de distribuição.

Souza & Jorgensen (1992) discutiram a estimação e inferência dos parâmetros envolvidos no processo Poisson Composto de Tweedie para dados de apólices de seguros na cidade do Rio de Janeiro, e Jorgensen (1992) apresentou as macros em GLIM4.0 para sua obtenção. Essa classe de modelos exponenciais com dispersão foi apresentada por Tweedie (1987), citado por Jorgensen (1987), e satisfaz que $V(\mu) = \mu^p$, para algum p no conjunto $(-\infty, 0] \cup [1, \infty)$. O modelo Tweedie, é usado, basicamente, para dados Poisson com superdispersão, ou Poisson Composto para $0 < p < 1$, e para o resto, para dados seguindo um processo Poisson misturado com uma distribuição Gama.

6.4 Inferência

6.4.1 A "Deviance"

Análoga à análise da variância para os modelos lineares Normal, existe uma técnica de análise (conhecida pelos usuários dos MLG) para fazer inferências nos modelos com dispersão.

Para hipóteses encaixadas dos μ 's, que são funções dos β 's, do tipo

- $H_0 : \mu \in \Omega^n$, com $\dim(\mu) = n$,
- $H_1 : \mu = \mu(\beta)$, com $\dim(\beta) = k_1$,
- $H_2 : \mu = \mu(\beta)$ e $\beta = \beta(\gamma)$, com $\dim(\gamma) = k_2$

os conceitos de inferência desses modelos são os baseados na "deviance", $D(z, \mu)$, (generalização da soma de quadrados residual da teoria normal), para o parâmetro μ dado o vetor de observações z . Isto é, se

$$L(\mu, \phi^2) = \sum_{i=1}^n \log a(\phi^{-1} w_i; z_i) + \phi^{-2} \sum_{i=1}^n t(z_i, \mu_i) w_i$$

é a função de logaritmo da função de verossimilhança para uma amostra aleatória de tamanho n , com $t(., .)$ uma função um-a-um e real, então, a "deviance" é definida segundo:

$$D(z, \mu) = 2 \sum_{i=1}^n \left\{ \sup_{\mu \in \Omega} t(z_i; \mu_i) - t(z_i, \mu_i) \right\} w_i$$

$$= 2 \sum_{i=1}^n \{t(z_i; \tilde{\mu}_i) - t(z_i, \mu_i)\}$$

a qual, que para o caso normal, resulta na conhecida expressão $D(z; \mu) = \sum (z_i - \mu_i)^2$. Para o caso da Binomial Negativa resulta:

$$D(z; \mu) = 2 \sum_{i=1}^n \left\{ z_i \log\left(\frac{z_i}{\mu_i}\right) - (z_i + \lambda) \log\left[\frac{z_i + \lambda}{\mu_i + \lambda}\right] \right\}$$

Lembra-se que o mínimo valor da "deviance" sob H_j ($j = 1, 2$) será $D_j = D(z, \hat{\mu}^{(j)})$, onde $\hat{\mu}^{(j)}$ é a estimativa de máxima verossimilhança de μ sob H_j . Assim as "deviances" $D_j > D_{j-1}$ são as correspondentes às hipóteses $H_j \subseteq H_{j-1}$, respectivamente. No caso de λ ser conhecido, a estatística do quociente de verossimilhança é mais usada, isto é para testar H_2 sob H_1 ,

$$\Delta D_2 = (D_2 - D_1)/\phi^2$$

a qual tem distribuição aproximada $\chi^2_{(k_1 - k_2)}$ sob H_2 , para n grande.

Este resultado pode ser estendido ainda para algumas situações que incorporaram elementos do modelo proposto. Em outras palavras, pode ser garantida a convergência a uma χ^2 ainda no caso em que o $\min(w_1, w_2, \dots, w_n) \rightarrow \infty$ (pesos "a priori" conhecidos e estabelecidos pelo MED). Esse resultado estende o domínio da teoria padrão para amostras grandes, o qual é muito útil dado que, às vezes, resulta impraticável trabalhar com amostras grandes e, em troca, não é difícil propor um modelo que seja com dispersão finita.

Quando as hipóteses são do tipo $\Omega^n = H_o$ (hipótese saturada= n parâmetros, o que implica "deviance" nula) $\supseteq H_1 \supseteq H_2 \supseteq \dots$, cada uma dessas hipóteses representa uma restrição maior nos parâmetros do modelo, pelo qual o processo de inferência deve ser construído iterativamente, como foi exposto na seção 2.2.3.

É conveniente salientar a opinião de Jorgensen (1989) em relação ao primeiro parágrafo desta seção. O autor afirma que o paralelismo que existe entre as análises da variância e da "deviance" para o caso de dados normais também se dá nos MED, como classe dentro dos MD. Isto é baseado no fato que os modelos são parametrizados tal que a matriz de informação é bloco-diagonal, e em particular fazendo com que β e ϕ^2 sejam ortogonais. Assim, esta idéia de generalização da análise da variância surgida no caso contínuo poderia ser aplicada para o caso dos MED discretos sempre que for possível garantir uma parametrização adequada para tais fins.

Por outro lado, para entender o papel da "deviance" como medida do ajuste poderia ser útil plotar um componente da "deviance", como uma função de z , para um valor dado do μ . Caso a distribuição fosse a Normal, esses gráficos seriam parábolas, o que não é usual nas restantes distribuições (Jorgensen, 1989).

6.4.2 A Matriz de Informação de Fisher

Uma outra forma de fazer inferências nos modelos exponenciais com dispersão é através da matriz de informação de Fisher. Lembre-se que essa matriz reflete a estrutura do processo inferencial propriamente dito, em particular se os parâmetros são, ou não, ortogonais. Para um MED, $Z_i \sim ED(\mu_i, \phi^2/w_i)$, com $i = 1, \dots, n$, regular (ou seja, um modelo que satisfaça às condições de regularidade da teoria assintótica), pode ser facilmente demonstrado que:

$$t(z, \mu) = z\tau^{-1}(\mu) - k(\tau^{-1}(\mu)), \quad \mu \in \Omega$$

$$t'(z, \mu) = (z - \mu)/V(\mu)$$

$$t''(z, \mu) = \frac{-V(\mu) - (z - \mu)/V(\mu)}{V(\mu)^2}$$

onde, $\tau(\theta) = k'(\theta)$, ou $E(Z) = \mu = \tau(\theta)$, e $V(\mu) = \tau'(\tau^{-1}(\mu))$. Portanto

$$\frac{\partial^2 L(\mu, \phi^2)}{\partial \sigma^2 \partial \mu_i} = -t'(z_i, \mu_i)w_i/\sigma^4$$

tem valor esperado nulo, o que implica que os parâmetros são ortogonais. Lembre-se que, dada a expressão do modelo com dispersão para o caso contínuo, o

$$\phi^2 = \sigma^2 = 1/\lambda$$

onde λ era o recíproco do índice de agregação na convenção da distribuição de probabilidade do modelo BN. Essas expressões para o caso Binomial Negativo resultam:

$$t(z, \mu) = z\tau^{-1}(\mu) - k(\tau^{-1}(\mu)), \quad \mu \in \Omega$$

$$t'(z, \mu) = (z - \mu)/\mu(1 + \mu)$$

$$t''(z, \mu) = \frac{\mu(1 + \mu) - \frac{z - \mu}{\mu(1 + \mu)}}{\mu^2(1 + \mu)^2}$$

$$= \frac{\mu^2(1 + \mu)^2 - (z - \mu)}{\mu(1 + \mu)}$$

Assim a matriz de informação de Fisher para $\mu(\beta)$ e $\phi^2(\lambda)$ resulta bloco-diagonal, da forma:

$$\begin{pmatrix} \mathbf{i}(\mu(\beta) | \phi^2(\lambda)) / \phi^2(\lambda) & 0 \\ 0 & i(\phi^2(\lambda) | \mu(\beta)) \end{pmatrix}$$

onde, para a estrutura de um modelo linear generalizado com hipóteses do tipo $H_1 : \mu = \mu(\beta)$, com $\beta = (\beta_1, \dots, \beta_p)^T$,

$$\mathbf{i}(\beta | \phi) = \mathbf{X}(\beta)^T \mathbf{W}_{\mathbf{i}}(\mu(\beta) | \phi) \mathbf{X}(\beta)$$

com $\mathbf{X}(\beta) = \partial \mu / \partial \beta^T$, e

$$i(\phi^2 | \mu) = E_{\mu, \phi^2} \{-c''(Z, \phi^2) + D(Z, \mu) / \phi^6\}$$

com D definida como a "deviance", e $c(z, \phi^2)$ tal que a log-verossimilhança satisfaz:

$$L(\mu, \phi^2) = c(z, \phi^2) - D(z, \mu) / (2\phi^2)$$

Esse desenvolvimento já foi apresentado na seção 5.3.1, onde também havia se notado que os zeros fora da diagonal indicavam a independência das observações e a ortogonalidade dos parâmetros. Assim, para o caso Binomial Negativo, essa matriz resultava:

$$\begin{pmatrix} 1/\mu(\beta)(1 + \mu(\beta)) & 0 \\ 0 & i(\phi^2(\lambda) | \mu(\beta)) \end{pmatrix}$$

As inferências podem ser feitas lembrando que por teoria assintótica o estimador do (β, ϕ^2) é assintoticamente normal com matriz de variância dada pela inversa da matriz de informação de Fisher. Pela ortogonalidade de β e ϕ^2 , os dois blocos da matriz podem ser invertidos separadamente. Em particular $\hat{\beta}$ e $\hat{\phi}^2$ são assintoticamente independentes e resulta:

$$\hat{\beta} \sim N(\beta, \phi^2 \mathbf{i}^{-1}(\beta | \phi^2))$$

$$\hat{\phi}^2 \sim N(\phi^2, i^{-1}(\phi^2 \mid \mu(\beta)))$$

Esses resultados se baseiam nas condições gerais de regularidade e no fato que a sub-matriz de informação dos β 's tende a infinito, ou dito de outra maneira, para cada j , com $l = 1, \dots, p$, p número de parâmetros, muitas são as observações que contribuem na estimação de cada β_l .

Jorgensen (1989) mostrou, quando $w \rightarrow \infty$,

$$i(\phi^2 \mid \mu) \simeq n/(2\phi^4)$$

notando que essa expressão não depende do modelo sob consideração. Em particular,

$$i(\phi^2 \mid \mu) = n/(2\phi^4)$$

para as distribuições normal e inversa gaussiana. Para um modelo linear, o estimador do σ^2 é proporcional a uma variável χ^2 , o que sugere que a aproximação χ^2 poderia ser melhor do que a aproximação normal para a distribuição do estimador $\hat{\sigma}^2$ nos casos com dispersão pequena.

Por outro lado, e dada a expressão do erro padrão de cada β_l , há necessidade, caso ϕ^2 for desconhecido, de estimar o parâmetro de dispersão para ambos casos, contínuos ou discretos. Muitas referências podem ser citadas a respeito, e muitas discussões acerca da escolha do método de estimação desse parâmetro fazem com que esse tópico seja abordado em uma outra seção.

6.4.3 Parâmetro de dispersão

Como foi mencionado, o parâmetro de dispersão deve ser estimado para poder obter boas estimativas dos erros padrões assintóticos dos $\hat{\beta}_l$. Para manter as propriedades desejadas desses erros, e em consequência das suas estimativas, pode ser usado o estimador de máxima verossimilhança, ou qualquer outro estimador **consistente**. A consistência é uma propriedade assintótica importante de um estimador. Lembra-se que: a sequência $\{T_n\}$, $T_n = t(Z_1, Z_2, \dots, Z_n)$, de estimadores de θ é dita de sequência *consistente* de estimadores se $\lim_{n \rightarrow \infty} E(T_n - \theta)^2 = 0$, $\forall \theta \in \Omega$ (Mood *et al.*, 1974). Apresentam-se a seguir os estimadores propostos por Jorgensen (1989) para o parâmetro de dispersão:

- O estimador de **máxima verossimilhança** do $\phi^2 = 1/\lambda$ é da forma:

$$\hat{\phi}^2 = \frac{1}{n-p} \sum_{i=1}^n w_i (z_i - \hat{\mu}_i)^2 \quad (21)$$

cabe assinalar que esse estimador possui boas propriedades, como consistência e eficiência, mas ele é tendencioso.

- O estimador do **Perfil de Verossimilhança Modificado** é definido por aquele valor de ϕ^2 , denotado por ϕ_o^2 , que maximize, sob H_1 :

$$L^o(\phi^2) = \frac{k_1}{2} \log(\phi^2) + L(\mu(\hat{\beta}), \phi^2)$$

note-se que o segundo termo dessa expressão é o perfil do logaritmo da função de verossimilhança usual. Se o modelo tiver dispersão pequena, isto é $w \rightarrow \infty$, então

$$\partial L^o / \partial \phi^2 \simeq \frac{k_1 - n}{2\phi^2} + \frac{D(z, \mu(\hat{\beta}))}{2\phi^4}$$

ou seja, aproximadamente

$$\hat{\phi}_o^2 \simeq \frac{D(z, \mu(\hat{\beta}))}{n - k_1}, \quad w \rightarrow \infty \quad (22)$$

o qual para n grande, é equivalente ao estimador de máxima verossimilhança. Assim, para amostras grandes esse estimador resulta **consistente**.

- O estimador **baseado na "Deviance"** é definido como o valor assintótico de (22), o que dá por estimativa:

$$\tilde{\phi}^2 \simeq \frac{D(z, \mu(\hat{\beta}))}{n - k_1}, \quad w \rightarrow \infty \quad (23)$$

Para a distribuição normal esse estimador é equivalente à soma de quadrados residual usual, o que outorga um valor intuitivo a este novo estimador. Porém, esse estimador não é recomendável devido ao fato de não possuir, para grandes amostras, a propriedade de consistência.

- O estimador de **Pearson**, $\bar{\phi}^2$ pode ser obtido resolvendo a equação:

$$\phi^2 = \frac{1}{(n - k_1)} X^2(\mu(\hat{\beta}), \phi^2) \quad (24)$$

onde a estatística X^2 (generalizada) de Pearson é definida segundo

$$X^2(\mu, \phi^2) = \sum_{i=1}^n t(z_i, \mu_i)^2 w_i / \mathbf{i}_{ii}(\mu \mid \phi^2)$$

Observe-se que para um MED, e pelo descrito na seção 6.4.2, esse estimador resulta da seguinte forma, uma vez que foram estimados os β 's no modelo estabelecido:

$$\bar{\phi}^2 = \frac{1}{(n - k_1)} \sum_{i=1}^n (z_i - \hat{\mu}_i)^2 w_i / V(\hat{\mu}_i)$$

Jorgensen (1989) prova a consistência para amostras grandes e para dispersão finita, reescrevendo-o como função dos $\hat{\beta}$'s, os quais são consistentes por serem estimativas de máxima verossimilhança.

Para comparar esses estimadores no nosso caso de um modelo Binomial Negativo, sejam as expressões seguintes dos estimadores propostos:

1.

$$\hat{\phi}^2 = \frac{1}{(n - p)} \sum_{i=1}^n (z_i - \hat{\mu}_i)^2 \quad (25)$$

onde p é a quantidade de parâmetros envolvidos no preditor linear (η), e $\hat{\mu}_i = g^{-1}(\eta_i)$. Os valores desse estimador podem ser obtidos facilmente no GLIM4.0 com os recursos simples do sistema, uma vez que é ajustado o modelo. Isto é:

...

\$scal p=10\$

\$scal %E1=1/(%cu(1)-p)*%cu(%yv-%fv)**2\$

...

2. Os estimadores obtidos através do Perfil de Verossimilhança modificado e baseado na "deviance" possuem a mesma expressão em termos da sua aproximação.

$$\phi^2 = \frac{D(z, \mu(\hat{\beta}))}{n - p} = \frac{2 \sum_{i=1}^n \{z_i (\log(\frac{z_i}{\mu_i}) - (z_i + \lambda) \log[\frac{z_i + \lambda}{\mu_i + \lambda}])\}}{n - p} \quad (26)$$

onde p é a quantidade de parâmetros envolvidos no η . Esse estimador também pode ser obtido no GLIM4.0 da seguinte maneira:

...

$$\text{\$cal \%E2}=\text{\%dv}/(\text{\%cu}(1)-p)$$

...

Quando o tamanho da amostra é suficientemente grande, esses dois estimadores coincidem com o de máxima verossimilhança, assegurando assim a consistência.

3. Os valores do estimador de Pearson são os propostos no enfoque de BRESLOW (1984) e extraídos da equação do valor esperado para a estatística X^2 (igual aos seus respectivos graus de liberdade), ou o que é equivalente:

$$\phi^2 = (n - p)^{-1} \sum_{i=1}^n \left\{ \frac{(z_i - \hat{\mu}_i)^2}{\hat{\mu}_i(\hat{\mu}_i + \phi^2)} \right\} \quad (27)$$

Visto o processo iterativo de resolução, o estimador $\bar{\phi}^2$ terá como estimativa o valor ϕ_{k+1}^2 obtido de (27) que satisfaça alguma regra de parada. Esse estimador é consistente por ser função dos β 's, e assim de máxima verossimilhança. A obtenção no GLIM4.0 requer um processo iterativo de busca, uma macro, a qual é apresentada no Apêndice B.

Mudando o enfoque, existem outras alternativas de busca para o estimador do parâmetro de dispersão; elas foram apresentadas na seção 5.3.1.

6.4.4 Resultados teóricos para amostras grandes

Em Jorgensen (1986) são apresentados os resultados assintóticos para os modelos exponenciais com dispersão, inclusive a prova das condições de regularidade, necessárias para obter as distribuições limites dos estimadores dos parâmetros.

Os três resultados básicos são:

$$\hat{\beta} \sim_{H_1} N(\beta, \phi^2 \mathbf{i}^{-1}(\beta, \phi^2))$$

para $n, w, \rightarrow \infty$

$$D_1/\phi^2 \sim_{H_1} \chi^2(n - p_1)$$

para $w \rightarrow \infty$

$$(D_2 - D_1)/\phi^2 \sim_{H_2} \chi^2(p_2 - p_1) \quad (28)$$

para $n, w \rightarrow \infty$, onde $D_i, \forall i$, são as "deviances" associadas com as hipóteses H_i . Pode-se provar também que $\hat{\beta}$ e D_1 são assintoticamente independentes sob H_1 e que D_1 e $D_2 - D_1$ também o são, sempre para dispersão assintótica pequena.

A partir disso, surgem os resultados mais usados na análise da "deviance":

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\phi}_{\alpha} \mathbf{j}^{jj}(\hat{\beta})^{1/2}} \sim_{H_1} t(n - p) \quad (29)$$

para $n, w \rightarrow \infty$, e

$$\hat{\phi}_{\alpha} \mathbf{j}^{jj}(\hat{\beta})^{1/2}$$

é o erro padrão do $\hat{\beta}_j$ baseado na informação observada $\mathbf{j}_{ii} = -t''(z_i, \mu_i)$;

$$(n - p_1) \hat{\phi}_o^2 / \phi^2 \sim_{H_1} \chi^2(n - p_1)$$

para $w \rightarrow \infty$ e,

$$\frac{D_2 - D_1}{\hat{\phi}_o^2(p_1 - p_2)} \sim_{H_2} F(p_1 - p_2, n - p_1) \quad (30)$$

para $n, w \rightarrow \infty$

O mais importante na análise da "deviance" é que a estatística F pode ser usada para testar H_2 sob H_1 . Isto é obtido analisando a relação entre o teste F e o teste do quociente de verossimilhança (LR). Segundo Jorgensen (1989) a estatística (30) é o teste do quociente do logaritmo da função de verossimilhança quando ϕ^2 é conhecido. Porém, quando ele é desconhecido, o teste do quociente do logaritmo da função de verossimilhança é uma função mais complicada das observações e então precisa-se estudar essa relação entre ambos testes. O autor utiliza aproximações do tipo ponto de sela para a verossimilhança e obtem que a expressão para esse teste é a seguinte:

$$\begin{aligned} LR_{1,2} &\simeq n \log(D_2/D_1) \\ &= n \log\{1 + (D_2 - D_1)/D_1\} \end{aligned}$$

o que é simplesmente uma função monótona do

$$\frac{(D_2 - D_1)/(p_1 - p_2)}{D_1/(n - p_1)}$$

e portanto F e LR são assintoticamente equivalentes. Assim o teste F é aproximadamente ótimo para testar H_2 sob H_1 . A partir disso resulta que o teste t em

(29) é também ótimo para testar hipóteses acerca dos β 's, já que se pode pensar como uma versão específica do teste F . Cabe assinalar que todas essas propriedades de optimalidade somente são válidas quando as condições de regularidade do modelo exponencial com dispersão são satisfeitas, para maiores detalhes, ver Jorgensen (1989), págs. 112-115.

6.5 Aplicação do Suporte Teórico Maior

6.5.1 O Modelo

Explorada a função de distribuição candidata para as observações, pode estabelecer-se o componente aleatório de um modelo linear generalizado, sempre e quando essa função pertença à família exponencial a um parâmetro. Se for o caso, ou ainda mais, se essa lei de probabilidades não for um membro de tal família, pode se propor um modelo exponencial com dispersão, segundo o descrito nas seções anteriores.

Para os dados de mortalidade de *E. grandis*, na seção (3.4, concluiu-se que a candidata com mais chance era a função de distribuição Binomial Negativa. Esse modelo probabilístico não pertence à família exponencial, salvo que um dos parâmetros seja conhecido, ou fixado de antemão. Dado que um dos parâmetros está ligado a uma função linear dos parâmetros de interesse, os quais informam sob os dois fatores ensaiados, a única escolha para restringir o modelo, e assim incluí-lo em tal família (e portanto na definição dos MLG, segundo McCullagh & Nelder, 1989), é ao nível do parâmetro de dispersão, $\sigma^2 = 1/\lambda$, na estrutura $V(Z) = \mu + \mu^2/\lambda$. Assim,

$$\begin{aligned} Z_i &\sim ED^*(\mu_i, \lambda) \\ Z_i &\sim BN^*(\mu_i, \lambda) \end{aligned} \tag{31}$$

define o MED e o caminho de estimação. O vetor de parâmetros (μ, λ) assume a estrutura de regressão,

$$\eta_i = \log(\mu_i) = \sum_{l=1}^p x_{il}\beta_l \tag{32}$$

onde $X = x_{il}$ denota a matriz de delineamento do modelo. Para ele serão obtidas as estimativas de máxima verossimilhança. Já foi discutido o critério para a escolha para a função de ligação como a $\log$, por isso não será abordado novamente.'

Logo, a função da log-verossimilhança para $(\mu(\beta), \lambda) = (\beta, \lambda)$ é dada por

$$\ell(\beta, \lambda) = \sum \{ \log[\Gamma(\lambda + z)] - \log[z! \Gamma(\lambda)] + z \log(\mu) + \lambda \log(1 - \mu) \}$$

a partir da qual serão realizadas as etapas inferenciais: a estimação e as provas de hipóteses.

6.5.2 O Algoritmo

É útil primeiramente considerar a estimação dos β 's para um λ fixo, pois isto fornece a essência do procedimento geral de estimação. Já que o modelo (31) e (32) é um modelo linear generalizado, pode ser usado então o algoritmo iterativo de mínimos quadrados ponderados implementado no GLIM, simplesmente com os recursos do modelo definido pelo usuário. Salienta-se que se estima β , para valores dados do λ .

Como montar o algoritmo para a estimação do conjunto "full" (maximal) de parâmetros?. A forma básica do algoritmo é um processo do tipo Newton-Raphson, e sua geração recai nas propriedades que essa função da logaritmo da função de verossimilhança possui. Pode ser demonstrado que os perfis dessa função, para cada um dos dois parâmetros, tem máximos locais nos seus domínios de definição ($\mathbb{R}^+$), e, o que é mais importante, que os parâmetros (β, λ) são ortogonais. Esse fato se deriva da estrutura da família exponencial natural, como uma generalização também da família exponencial usual, e torna a matriz de informação bloco diagonal, com um bloco $p \times p$ correspondendo ao vetor β , e outro bloco 1×1 para o λ . Assim, a inversa dessa matriz de informação também é bloco diagonal, permitindo a inversão de cada bloco individualmente.

Essa propriedade valiosa de função do logaritmo da função de verossimilhança **separa** as equações de Newton-Raphson em duas equações, uma para cada um dos parâmetros. Isto é, a atualização de, por exemplo, a estimativa MV do β , pode assim ser realizada como se o outro parâmetro fosse conhecido (uma estimativa de MV também sem atualizar), e vice-versa. O algoritmo altera permanentemente sua busca através das expressões para as estimativas, obtidas de cada perfil de verossimilhança, isto é, $\partial \ell(\beta, \lambda_o) / \partial \beta = 0 \Rightarrow \hat{\beta}$, até obter convergência. O critério também deve ser cuidadoso na hora da seleção dos valores iniciais, os quais podem ser escolhidos como, por exemplo, a estimativa pelo método dos momentos, ou por meio de gráficos convenientes.

Por outro lado, a regra de parada que o GLIM utiliza para fornecer o ajuste do modelo é baseada na diminuição relativa do valor da "deviance". Já para o proposto nesta seção, deveria observar-se a diminuição da verossimilhança. O sistema GLIM fornece as estimativas para os parâmetros β 's do preditor linear com os seus correspondentes erros padrões.

Para o cálculo do outro erro padrão aplica-se a propriedade de ortogonalidade dos parâmetros. Essa ortogonalidade implica que o erro padrão para a estimativa de MV do λ pode ser calculada na informação esperada para λ como se os β 's fossem conhecidos, aliás, acontece com os β 's que a matriz de informação esperada o tem ao λ ingresada como um fator de proporcionalidade. Ainda mais, o GLIM estima o $\sigma^2 = 1/\lambda$ como $D/g.l.$, onde D é a "deviance" para o modelo proposto, pelo qual os erros padrões dos β 's apurados no "output" devem ser multiplicados por uma correção do tipo $\hat{\sigma}/\sqrt{D/g.l.}$. Para ambos casos, esses erros padrões são aproximados, eles seriam verdadeiramente exatos no caso em que o bloco cruzado da matriz de informação fosse zero, fato que na prática nunca acontece, apenas aproximadamente.

Para a verificação de hipóteses, note-se que se o enfoque de geração do modelo foi um Poisson Composto segundo Firth (1990), estar-se-ia modelando simultaneamente N (pensado como variável aleatória agora) e Z . Dessa maneira, a análise da "deviance" resulta ineficiente, já que não considera ao N propriamente dito no cálculo das discrepâncias entre o que foi observado e o que foi ajustado. Logo, as inferências acerca de hipóteses encaixadas devem ser realizadas através do teste do quociente de verossimilhança (LR).

Mas, segundo Jorgensen (1992), os valores dos p-"values" para o teste do quociente de verossimilhança e para o valor da estatística de Pearson são bastante semelhantes, e ambos bem diferentes do obtido mediante o valor do teste F baseado na "deviance". Logo, o cálculo do teste LR poderia ser dispensado dado que o GLIM fornece diretamente com o escalar %X2 o valor da estatística de Pearson. Isto confirma o fato que o estimador baseado na "deviance" para λ é tendencioso, ou seja, produz testes viesados também. Não obstante esse argumento, esse tópico merece ainda um pouco mais de atenção e estudo para o caso específico do modelo Binomial Negativo.

7 OUTROS TÓPICOS

7.1 Uso Experimental: Tamanho de Amostra

Segundo Rogers (1974), o conhecimento da dispersão dos indivíduos é fundamental para o desenvolvimento de planos de amostragem confiáveis, e, a obtenção de estimativas de parâmetros de fatores de interesse. Além disso, conhecendo a distribuição amostral podem-se estabelecer relações úteis dessas contagens com alguma teoria de crescimento ou de dispersão populacional. Se, se conhece a distribuição amostral das observações (e portanto se infere que essa é a distribuição populacional das contagens) é possível estimar o tamanho amostral requerido para obter uma precisão dada na estimação do parâmetro de locação (avaliada, por exemplo, como o coeficiente de variação da média amostral).

Quando a amostragem corresponde a um modelo de dispersão aleatória, ou Poisson, esse cálculo é imediato, uma vez que o coeficiente de variação da média é $D = \sigma/(\sqrt{n}\mu)$, e portanto $n = \sigma^2/D^2\mu^2$. Se a distribuição for Binomial Negativa, Rojas (1964) citado por Rogers (1974) apresenta o valor de n colocando as respectivas média e variância na equação anterior, o que resulta:

$$n = \frac{(\frac{1}{\mu} + \frac{1}{\theta})}{D^2}$$

Assim, e substituindo pelos seus respectivos estimadores, a expressão para o valor estimado do tamanho de amostra resulta:

$$n = \frac{(\frac{1}{\hat{\mu}} + \frac{1}{\hat{\theta}})}{D^2}$$

Essa equação pressupõe conhecer a média amostral, mas para relaxar essa dificuldade pode-se plotar n em função da $\hat{\mu}$ para vários níveis de precisão desejados e, também, D em função de $\hat{\mu}$ para diferentes valores de n pré-selecionados.

Seguindo esse enfoque, foram calculados os tamanhos ótimos de amostra para diferentes precisões, expressados em termos da porcentagem da média

Tabela 19: Tamanhos ótimos de amostras segundo os enfoques Extra-Poisson (EP) e Binomial Negativo (BN).

Precisão (% da média)	Tamanho n [EP]	Tamanho n [BN]
1	63	95
1,5	30	42
2	16	24
2,5	10	15
3	9	11
3,5	8	8
4	7	6
4,5	6	5
5	3	3

amostral, para os valores de θ obtidos segundo os modelos Extra-Poisson (EP), segundo estimação máxima quase-verossimilhança, e Binomial Negativo (BN), segundo estimação máxima verossimilhança. A Tabela 19 apresenta os resultados dessas estimações para precisões usuais na pesquisa florestal, os quais são ilustrados na Figura 8 do Apêndice A, onde são mostradas as curvas para a busca do tamanho ideal de amostra conforme a situação.

Podemos observar-se que para precisões menores na estimação os tamanhos de amostras são semelhantes sob ambos enfoques, resultando mais rigoroso o modelo Binomial Negativo a medida que a exigência se incrementa. Isto fornece elementos de decisão para os futuros planos de manejo florestal, uma vez que o delineamento contempla certos níveis de exigência desejados no que diz respeito as estimativas de posição. Salienta-se que, esta metodologia para a determinação dos tamanhos amostrais, é independente da teoria de testes de hipóteses, requisito sempre usado na obtenção das expressões para os tamanhos procurados.

7.2 Uso Biológico: Processo Estocástico

A mortalidade é um dos parâmetros chaves que determinam mudanças nas populações. Isto implica que há necessidade de propor uma técnica que resuma, de alguma forma, como se produz tal fenômeno. Existe toda uma linha

de pesquisa a respeito, a qual explora os padrões da mortalidade de uma determinada espécie através de tabelas de vida. Essas tabelas de vida foram construídas para estudos de demografia, principalmente para companhias de seguros, mas pouco foram adaptadas para espécies animais e, menos ainda, para espécies vegetais. Não obstante, o objetivo da construção duma tabela de vida independe da espécie em estudo, já que elas podem ser definidas como um resumo, específico para cada idade, das taxas de mortalidade que operam sobre uma população. Nota-se que, para compor uma tabela de vida, há necessidade de determinar os intervalos de idade nos quais agrupar os dados, ou seja, estabelecer os "pontos" da amostragem (ou censo, caso for possível) onde coletar a informação. Quanto menor esses intervalos, maior o detalhe a obter da imagem da mortalidade.

Na seção 6 foram introduzidas as idéias de Jorgensen (1989) acerca dos modelos exponenciais de dispersão. Um conceito apresentado, como propriedade interessante e muito útil desses modelos, foi o de convolução. Apresenta-se a seguir uma proposta de linha de estudo para os dados de mortalidade de *E. grandis* que visa à recuperação de informação adicional relevante, já não em forma de covariável para controlar o erro experimental do modelo de locação selecionado, senão em um sentido dinâmico, de maneira que contribua ao conhecimento do processo biológico em si.

Seja z_{ijk} , a k -ésima observação, correspondente à i -ésima região e j -ésima rotação, $i, j = 1, 2, 3$ e $k = 1, \dots, n_{i,j}$. Lembra-se que, para essa observação k -ésima, se dispunha também de informação adicional, como a idade das árvores. Suponha-se que, para cada uma das rotações do ensaio, são extraídas informações para determinados intervalos ou "comprimentos" t_j (em alguma unidade de tempo, por exemplo anos, embora não haja necessidade de coincidir com a idade real das árvores) das mesmas.

Na realidade, informações desse tipo, que reflitam os verdadeiros intervalos de tempo, onde as árvores apresentam mudanças marcantes a nível de crescimento, merecem ser estudadas. Mas, essa informação somente poderá ser fornecida se se dispõe de um conhecimento profundo da espécie, e da sua adaptação, como no caso do *E. grandis*, a qual pode originar-se de estudos, de alguns componentes intrínsecos de pesquisa, como o são os planos de manejo técnico-florestal, ou ainda mais de etapas de observação propriamente ditas, como as observações preliminares de comportamento da sobrevivência da espécie introduzida. Uma forma muito prática de obter informações a respeito é através das tabelas de vida.

Com as tabelas de vida se constróem as famosas curvas de sobre-

vivência, que expressam o comportamento do número de sobreviventes em relação à idade, avaliada em alguma medida de tempo. Pearl em 1921, citado por Krebs (1986), distinguiu três tipos gerais de curvas de sobrevivência. No Apêndice A se ilustram essas tendências na Figura 9. Observe-se que as curvas **tipo I** pertencem a populações com poucas falhas (perdas) entre as primeiras idades e depois, muitas para os indivíduos mais velhos; as curvas diagonais (**tipo II**) implicam uma taxa constante de mortalidade independentemente da idade; e as curvas **tipo III** indicam alta mortalidade nas primeiras idades, seguida por um período de perdas constantes e de grandeza inferior.

Dada a importância desta espécie no Brasil, porque não tentar estudar a curva típica para ela, com o objetivo de recuperar seu processo de crescimento com perfil ecológico?. Uma vez reconstruída essa informação, poderia ter certeza dos intervalos corretos nos quais deve ser amostrada a população, ou seja, os t_j . Por exemplo, se a curva para *E. grandis* for a do **tipo II**, é óbvio que não há necessidade de realizar amostragem específica, nem esforços na frequência delas dentro de um período do tempo determinado, já que, sob esse comportamento da sobrevivência, não existe um período crítico para a mortalidade das árvores. Logo, podem ser tomados os intervalos equidistantes, ou, ainda mais, coincidindo com a idade propriamente dita das árvores em relação a data de plantio. Já, e por exemplo, se essa curva for do **tipo III**, o esforço para a coleta da informação será maior e deverá estar concentrada no período primeiro de vida, de forma a reproduzir o comportamento do número de mortes no início da etapa de crescimento.

Em outras áreas de experimentação, esses comprimentos, ou pontos de amostragem, podem ser de natureza física, como apresenta Jorgensen (1992) em um processo industrial textil para a determinação da permeabilidade de uma fibra de tecido, ou como poderiam ser abordados os problemas de rendimento na área de manejo de solos em relação a respostas sob um determinado gradiente de fertilidade.

Por outro lado, e isto é preliminar à incorporação dessa nova informação, foi provado nas seções anteriores que um modelo apropriado para refletir o comportamento aleatório da variável mortalidade das árvores de *E. grandis*, é o modelo Binomial Negativo. Assim, pelo exposto na seção 6, o modelo exponencial de dispersão associado é:

$$Z_{ijk} \sim ED^*(\theta_{ijk},) \quad (33)$$

onde se satisfaz que a $E(Z_{ijk}) = \lambda t_j$. Lembra-se que, μ_{ijk} é o número esperado de falhas por unidade de tempo (ou comprimento de tempo) na rotação j , implantada

na região i . Visando relacionar com a teoria geral dos MLG, se estabelece que μ_{ijk} é tal que $g(\mu_{ijk}) = \eta_{ijk}$, o preditor linear que determina o modelo de locação da análise da covariância. É importante salientar que em qualquer metodologia de análise, sempre será possível a avaliação de um modelo para o $g(\mu_{ijk})$.

Estabelecido dessa maneira o modelo, a propriedade de convolução permite a interpretação do modelo em termos de um processo estocástico subjacente. Isto é, sejam divididas as rotações em sub-intervalos de comprimento t_{jl} , $l = 1, \dots, L$, onde L é a quantidade de intervalos convenientemente escolhidos de forma tal que possa refletir possíveis mudanças na mortalidade. Logo, assumindo que os Z_{ijlk} são independentes, e que

$$Z_{ijlk} \sim ED^*(\theta_{ijk}, \lambda t_{ijlk})$$

interpreta-se Z_{ijlk} como o número de falhas, ou a mortalidade, para o l -ésimo intervalo da j -ésima rotação na região i , e o que é mais importante, a

$$Z_{ijk} = \sum_l Z_{ijlk}$$

segue (33). Em outras palavras,

$$Z_{ijk} = \sum_l Z_{ijlk} \sim BN(\mu_{ijk}, \lambda t_j)$$

interpreta-se como um modelo BN usual só que ele provém de um processo estocástico trazendo informação acerca da taxa de crescimento. Muita informação é considerada no estabelecimento desse modelo, e muito é também o proveito que pode ser tirado com o uso das equações estimadas que resultam do processo de estimação. Lembra-se que o objetivo central em um modelo exponencial de dispersão é a modelação do valor esperado μ , respeitando a estrutura probabilística original da variável de resposta. Por isso, muitas vantagens em termos do manejo da espécie poderiam ser encontradas num modelo como o exposto acima, para o qual é incorporada também a informação temporal para esse valor esperado, implicando, ainda mais, a obtenção de inferências acerca dos fatores envolvidos no experimento ao longo da etapa de crescimento.

8 CONCLUSÕES

Do presente trabalho foram obtidas as conclusões que seguem:

- (a) Dado que o fenômeno de superdispersão aparece freqüentemente nos conjuntos de dados de contagens o modelo Poisson deve ser utilizado com cuidado. Primeiramente é necessário estabelecer o tipo de relação **média-variância** através de técnicas empíricas e inferenciais. Feito isto, a estimação do modelo assumido para o valor esperado da variável de resposta pode ser feita no GLIM ou SPLUS usando um dos dois enfoques para a modelagem: o **modelo linear generalizado** ou, o **modelo exponencial com dispersão**.
- (b) Para a variável aleatória mortalidade das árvores de *Eucalyptus grandis*, no Estado de São Paulo obteve-se:
 - (i) um componente extra de variação ao modelo probabilístico Poisson, indicando assim um processo ecológico de composição no *stand* florestal. Foi selecionada e verificada como candidata para a aproximação dessa distribuição agregada a distribuição Binomial Negativa.
 - (ii) razoável o estabelecimento do modelo de locação igual a um modelo da análise da covariância, incluindo assim a informação adicional da idade das árvores para o controle do erro experimental como fonte de variação.
 - (iii) um índice de agregação, $\phi = 1/\lambda$, altamente significativo, com valor estimado igual a 2,18, o qual mostrou que a mortalidade das árvores desta espécie segue uma distribuição com estrutura **média-variância** quadrática, e pode portanto ser estudada a partir de um processo estocástico.
- (c) O modelo exponencial com dispersão Binomial Negativo relaxa a restrição imposta no modelo linear generalizado para o parâmetro de dispersão, e pode ser ajustado facilmente no GLIM utilizando os recursos do modelo do usuário desse sistema. Ele é o preferido para a obtenção de inferências biparamétricas gerais.

- (d) Dentro do enfoque do modelo linear generalizado, o modelo Extra-Poisson de BRESLOW (1984) através de quase-verossimilhança foi implementado facilmente, embora tenha mostrado baixa eficiência da estimativa do parâmetro extra de variação em relação à obtida pelo método da máxima verossimilhança. O método de busca dessa estimativa definido através da variância nominal pode ser recomendado como alternativa metodológica pela sua simples e rápida implementação.
- (e) Existiu interação entre os fatores região e rotação da espécie, a qual mostrou que nos municípios de Brotas e Riberão Bonito as plantas suportam melhor o estresse do primeiro corte.
- (f) O processo estocástico Binomial Negativo para a mortalidade das árvores pode ser construído incorporando informação temporal para o valor esperado em função da rotação da espécie. Inferências acerca dos fatores ao longo da etapa de crescimento podem ser obtidas.

9 REFERÊNCIAS BIBLIOGRÁFICAS

- AITKIN, M. A general maximum likelihood analysis of overdispersion in generalized linear models. **Statistics and Computing**, v. 5, p.342-351, 1996.
- AITKIN, M.; FRANCIS, B. Fitting overdispersed generalized linear models by nonparametric maximum likelihood. **GLIM Newsletter**, v. 25, p.37-45, 1996.
- AITKIN, M.; ANDERSON, D.; FRANCIS, B.; HINDE, J. **Statistical modelling in GLIM**. Oxford: Oxford Statistical Sciences Series-Claresdon Press, 1989. 583p.
- ANDERSON, D.A.; HINDE, J. Random effects in generalized linear models and the EM algorithm. **Communication in Statistics**, v. 17, p.3847- 3856, 1988.
- ATKINSON, A.C. Regression diagnostics, transformations and constructed variables **Journal of the Royal Statistics Society, B**, v. 44, n. 1, p.1-36, 1982.
- ATKINSON, A.C.; DAVISON, A.C; NELDER, J.A; O'BRIEN, C.M. **Model checking**. London: Imperial College, 1989. 235p.
- BENNETT, S. An extention of Williams' method for overdispersion models. **GLIM Newsletter**, v. 17, p.12-18, 1989.
- BLISS, C.I.; FISHER, R.A. Fitting the negative binomial distribution to biological data. **Biometrics**, v. 9, p.176-200, 1953.
- BOX, G.E.P; TIAO, G.C. **Bayesian inference in analysis**. New York: Addison Wesley, 1973, 588p.
- BRESLOW, N.E. Extra-Poisson variation in log-linear models. **Applied Statistics**, v. 33, p.38-44, 1984.
- BRESLOW, N.E. Test of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. **Journal of the American Statistical Association**, v. 85, p.565-571, 1990.
- BROOKS, R.J. Approximate likelihood ratio tests in the analysis of Beta-Binomial data. **Applied Statistics**, v. 33, n. 3, p.285-289, 1984.
- COLLETT, D. **Modelling binary data**. London: Chapman and Hall, 1991. 369p.
- COLLINGS, B.J.; MARGOLIN, B.H. Testing goodness of fit for the Poisson assumption when observations are not identical. **Journal of the American Statistical Association**, v. 84, p.411-418, 1985.

- CONOVER, W.J. **Practical nonparametric statistics**. New York: John Wiley and Sons, 1971. 548p.
- CORDEIRO, G.M. **Modelos lineares generalizados**. Campinas: VII SINAPE, 1986. 286p.
- COX, D.R. Some remarks on overdispersion. **Biometrika**, v. 70, p. 269-274, 1983.
- COX, D.R.; HINKLEY, D.V. **Theoretical statistics**. London: Chapman and Hall, 1979. 342p.
- COX, D.R.; SNELL, E.J. **Analysis of binary data**. 2.ed., London: Chapman and Hall, 1989. 376p.
- DAVID, F.N.; JOHNSON, N.L. The truncated Poisson. **Biometrics**, v. 8, p.275-285, 1952.
- DEAN, C.; LAWLESS, J.F. Tests for detecting overdispersion in Poisson regression models. **Journal of the American Statistical Association**, v. 84, p. 406-415, 1989.
- DEMÉTRIO, C.G.B. **Modelos lineares generalizados na experimentação agrônômica**. Porto Alegre: 5o. SEAGRO, 1993. 125p.
- DÍAZ, M.P. Ajustes de modelos lineares Generalizados para dados de contagem com superdispersão. Piracicaba, 1994. 48p. Seminário. Escola Superior de Agricultura "Luiz de Queiroz". Universidade de São Paulo.
- EFRON, B. Bootstrap methods: another look at the jackknife. **The Annals of Statistics**, v. 21, p.460-480, 1979.
- EFRON, B.; TIBSHIRANI, R. **An introduction to the bootstrap**. New York: Chapman and Hall, 1986. 421p.
- ELDER, L. F. **Statistical analysis of genotoxicity assays**. Santa Marta: SIMEST, 1995. 52p.
- ELLIOT, J.M. **Some methods for the statistical analysis of samples of benthic invertebrates**. 2.ed. Warehan: Freshwater Biological Association, 1979. 285p.
- FAHRMEIR, L.; KAUFMAN, H. Consistency and asymptotic normality on the maximum likelihood estimator in generalized linear models. **Annals of Statistics**, v. 13, p.342-368, 1985.
- FIORAVANTE, A.M. Uma aplicação do algoritmo EM. Piracicaba, 1993. 45p. Seminário. Escola Superior de Agricultura "Luiz de Queiroz". Universidade de São Paulo.
- FIRTH, D. Overdispersion. Piracicaba, 1990. 25p. Minicurso. Escola Superior de Agricultura "Luiz de Queiroz". Universidade de São Paulo.
- FRANCIS, B.; GREEN, M.; PAYNE, C. **The GLIM system**. Oxford: Clarendon Press, 1993. 821p.

- FROME, E.L.; KUTNER, M.H.; BEANCHAMP, J.J. Regression analysis of Poisson-distribution data. **Journal of American Statistical Association**, v. 68, p. 935-940, 1973.
- HASEMAN, J.K.; KUPPER, L.L. Analysis of dichotomous response data from certain toxicological experiments. **Biometrics**, v. 35, p.281-303, 1979.
- HINDE, J. Compound Poisson regression models. In: INTERNATIONAL CONFERENCE OF GENERALIZED LINEAR MODELS, 1982. **Proceedings**. New York: Springer-Verlag, p.109-121, 1982.
- HINDE, J. Macros for fitting overdispersion models. **GLIM Newsletter**, v. 25, p.10-19, 1996.
- HINKLEY, D.V. Likelihood inference for a simple growth model. **Biometrika**, v. 66, p.659-662, 1979.
- HURREL, K. T. Modelling the relationship between crime count and observation period in prison inmates self-report data. **Applied Statistics**, v. 42, n. 2, p. 355-367, 1993.
- JOHNSON, N.L.; KOTZ, S. **Distributions in statistics**. New York: Houghton Mifflin, v.1, 1969. 436p.
- JØRGENSEN, B. Maximum likelihood estimation and large-sample inference for generalized linear and non-linear regression models. **Biometrika**, v. 70,n. 1, p.19-28, 1983.
- JØRGENSEN, B. The delta algorithm and GLIM. **International Statistical Review**, v. 52, p.283-300, 1984.
- JØRGENSEN, B. Some properties of exponential dispersion models. **Scandinavian Journal of Statistics**, v. 13, p.187-198, 1986.
- JØRGENSEN, B. Exponential dispersion models. **Journal of the Royal Statistical Society, B**, v. 49, p.127-162, 1987.
- JØRGENSEN, B. **The theory of exponential dispersion models and analysis of deviance**. São Paulo: Ia. Escola de Modelos Lineares, 1989. 129p.
- JØRGENSEN, B. Exponential dispersion models and extentions: a review. **International Statistical Review**, v. 60, p.5-20, 1992.
- JØRGENSEN, B.; MARTÍNEZ, J.R. **On the asymptotic behaviour of the variance function**. Rio de Janeiro: Informes de Matemática B-059, IMPA. 1990.
- JØRGENSEN, B.; SESHADRI, V.; WHITMORE, G.A. On the mixture of the inverse Gaussian distribution with its complementary reciprocal. **Scandinavian Journal of Statistics**, v. 18, p.77-89, 1991.
- KREBS, C.J. **Ecología**. Madrid: Ediciones Pirámide, 1986. 782p.
- LAWLESS, J.F. Negative binomial and mixed Poisson regression. **Canadian Jour-**

- nal of Statistics, v. 15, p.209-225, 1987.
- LAWLEY, D.N. A general method for approximating to the distribution of likelihood ratio criteria. **Biometrika**, v. 43, p.295-303, 1956.
- LEITE, J.G.; SINGER, J.M. **Métodos Assintóticos em Estatística**. São Paulo: 9o. Simpósio Brasileiro de Probabilidade e Estatística, 1990. 72p.
- LIANG, K. Y.; ZEGER, S. L. Longitudinal data analysis using generalized linear models. **Biometrika**, v. 73, p.13-22, 1986.
- LINDSEY, J.K. **Models for repeated measurements**. Oxford: Oxford Science Publications, 1993. 413p.
- MACCHIAVELLI, R.E. Fundamentos y aplicaciones del análisis estadístico en muestreo de poblaciones de insectos. Buenos Aires, 1987. 328p. Tese (Mestrado). Universidade de Buenos Aires.
- MARGOLIN, B.; KAPLAN, N.; ZEIGER, E. Statistical analysis of the ames *salmonella*. **National Academic Science**, v. 76, p.3779-3783, 1981.
- McCULLAGH, P. Quase-likelihood functions. **Annals of Statistics**, v. 11, p.59-67, 1983.
- McCULLAGH, P. On the asymptotic distribution of Pearson's statistic in linear exponential-family models. **International Statistical Review**, v. 53, n. 1, p.61-67, 1985.
- McCULLAGH, P.; NELDER, J. A. **Generalized linear models**. 2.ed. London: Chapman and Hall, 1989. 511p.
- McGUIRE, J.V.; BRINDLEY, T.A.; BANCROFT, T.A. The distribution of european corn borer larvae (*Pyrausta nubilalis* HBN) in field corn. **Biometrics**, v. 13, p.65-78, 1957.
- MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the theory of statistics**. 3.ed. Auckland: McGraw-Hill, 1974. 598p.
- MOORE, D.F. Asymptotic properties of moment estimators for overdispersed counts and proportions. **Biometrika**, v. 73, p.583-588, 1986.
- MORGAN, B.J.T. **Analysis of quantal response data**. London: Chapman and Hall, 1992. 511p.
- MORGAN, B.J.T.; SMITH, D.M. A note on Wadley's problem with overdispersion **Applied Statistics**, v. 41, n. 2, p.349-354, 1992.
- MORRIS, C. Natural exponential families with quadratic variance functions. **The Annals of Statistics**, v. 10, n. 1, p.65-80, 1987.
- NELDER, J.A. Quasi-likelihood and GLIM. In: INTERNATIONAL CONFERENCE OF GENERALIZED LINEAR MODELS, 1985. **Proceedings**. New York: Springer-Verlag, p.127, 1985.

- NELDER, J.A.; PREGIBON, D. An extended quasi-likelihood function. **Biometrika**, v. 74, p.221-232, 1987.
- NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistics Society, A**, v. 135, p.370-384, 1972.
- PATEL, J.K; KAPADIA, C.H.; OWEN, D.B. **Handbook of Statistical Distributions**. New York: Marcel Dekker, Inc, 1976. 302p.
- PEARSON, E.S.; HARTLEY, H.O. **Biometrika Tables for Statisticians**. 3.ed., Cambridge: Biometrika, 1966. 24p.
- PREGIBON, D. Data analytic methods for generalized linear models. Toronto, 1979. Tese (Ph.D.). Universidade de Toronto.
- PREGIBON, D. Goodness of link test for generalized linear models. **Applied Statistics**, v. 29, n. 1, p.15-24, 1980.
- RABINOVICH, J.E. **Ecologia de Poblaciones Animales**. Washington: Serie de Biología, 21, Organización de Estados Americanos (OEA), 1978. 178p.
- RIDOUT, M.S.; DEMÉTRIO, C.G.B. Generalized linear models for positive count data. **Revista de Matemática e Estatística**, v. 10, p.139-148, 1992.
- ROGERS, A. **Statistical analysis of spatial dispersion**. London: Pion, 1974. 543p.
- SCHUKKEN, Y.H.; CASELLA, G.; VAN DEN BROEK, J. Overdispersion in clinical mastitis data from dairy herds : a negative binomial approach. **Preventive Veterinary Medicine**, v. 10, n. 3, p.239-249, 1991.
- SILVA, N. O método bootstrap e aplicações à regressão múltipla. Campinas, 1995. 158p. Tese (Mestrado). Universidade de Campinas.
- SHIRLEY, E.A.; KICKLING, H. An evaluation of some statistical methods for analysis numbers of abnormalities found amongst litters in teratology studies. **Biometrics**, v. 37, p.819-829, 1981.
- SMITH, P.J.; HEITJAN, D.F. Testing and adjusting for departures from nominal dispersion in generalized linear models. **Applied Statistics**, v. 42, n. 1, p.31-41, 1993.
- SMITH, D.M.; MORGAN, B.J.T. Extended models for Wadley's problem. **GLIM Newsletter**, v. 18, p.21-35, 1989.
- SOMERS, R.G; ODERWALD, W.; LANGDON, O.G. Predicting mortality with a Weibull distribution. **Forest Science**, v. 26, p.291-300, 1980.
- SOUZA, P.; JORGENSEN, B. Fitting Tweedie's compound Poisson model to insurance claims data. **unpublished**, 1992.
- VENABLES, W.N.; RIPLEY, B.D. **Modern applied statistics with S-Plus**. New York: Springer-Verlag, Inc, 1994. 462p.

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

WILKS, S. **Mathematical Statistics**. New York: John Wiley and Sons, 1987. 277p.

WILLIAMS, D. A. Extra-binomial variation in logistic linear models. **Applied statistics**, v. 31, n. 2, p.144-148, 1982.

YODA, R.T; KIRA, T.; OGAEA, H. ; OZUMI, K. Self-thinning in verycrowded pine stands under cultivated and natural conditions. **Journal of Biology**, v. 14, p.107-129, 1963.

10 APÊNDICES

FIGURAS

PROGRAMAS E MACROS

10.1 Apêndice A

Figura 1: Dispersão da mortalidade versus a idade das árvores de *E. grandis* por tratamento.

Figura 2: Dispersão da $V_{ij}(Z)$ versus $\bar{Z}_{ij}$, onde (k, l) representa à rotação k e região l , para $k, l = 1, 2, 3$.

Figura 3: Dispersão do $\ln(\bar{Z}_{ij} - V_{ij}(Z)/n_{ij})$ versus $\ln(V_{ij}(Z) - \bar{Z}_{ij})$, onde (k, l) representa à rotação k e região l , para $k, l = 1, 2, 3$.

Figura 4: Distribuição empírica e estimada, segundo o modelo Poisson, da mortalidade das árvores de *E. grandis*.

Figura 5: Valores médios da mortalidade das árvores de *E. grandis* para as combinações de região e rotação.

Figura 6: Dispersão dos $\hat{z}_{ijk}$ versus $\hat{\eta}_{ijk}$ para a mortalidade das árvores de *E. grandis*.

Figura 7: Valores estimados dos quocientes ["deviance" residual]/[graus de liberdade] versus θ , onde θ é tal que $V(Z) = \mu + \frac{1}{\theta}\mu$.

Figura 8: Valores ótimos de tamanhos de amostras sugeridos pelos modelos Extra-Poisson e Binomial Negativo, para a obtenção de diferentes precisões na estimação da média da mortalidade de *E. grandis*.

Figura 9: Curvas hipotéticas de sobrevivência segundo Raymond Pearl, em 1928.

10.2 Apêndice B

Programa utilizado no **GLIM4.0** para o cálculo da estatística do teste de superdispersão do modelo Poisson.

```
$data 2605 reg rot ida mor$
$dinp 'a:morta.dad' reg rot ida mor$
$factor reg 3 rot 3$
$yvar mor $error p$
$fit reg+rot+reg.rot+ida$
$cal T=0.5*(%cu((%yv-%fv)**2-%yv))$
$look T $stop
```

Macros utilizadas no **GLIM4.0** para o ajuste do modelo Extra-Poisson através de máxima quase-verossimilhança. **Procedimento II.**

```
$mac chis $prin 'Qui-quadrado=%X2' on '*i%df' D.F.':$
$endmac
!
$mac brs2 !k e alpha para o modelo Poisson de Breslow
$cal %k=(%x2-%cu((1-qt)*%pw))/(%cu((1-qt)*%pw*%fv))$
$cal a=1+%k*%fv$
$endmac
!
$mac ovr !ajusta o modelo de superdispersao com k=escala
$use %1 $prin 'escala=' %sc$
$cal w=1/a $out 0 $tra $fit. $out %poc $tra i o f h
$use chis$
$endmac
!
$mac over !ajusta o modelo de superdispersao com k=parametro extra
```

```

$ext %vl $scal qt=%vl*%wt*%pw/%sc
$use %l $prin 'k=%k$
$scal w=1/a $out 0 $tra $fit. $out %poc $tra i o f h
$use chis$
$endmac
!
$data 2605 reg rot ida mor$
$dinp 'a:morta.dad' reg rot ida mor$
$fac reg 3 rot 3$
$warm $tra i o f h
$yvar mor $
$erro p$
$link l$
$fit reg+rot+reg.rot+ida $disp me$
$use chis$
$scal w=1 $weig w$
$sca %sc $fit. $use chis $fit reg+rot+reg.rot+ida$
$use chis !$ajuste sub-modelo$
$yvar mor $erro p$
$fit reg+rot+reg.rot+ida $disp me $use chis$
$scal w=1 $weig w$
$arg over brs2 $use over $sus !use over ate convergencia
$fit reg+rot+reg.rot+ida$
$dis me$ $use chis$
$stop

```

Macro utilizada no **SPLUS** para o ajuste do **modelo Extra-Poisson** através e máxima quase-verossimilhança.

```

function{mor, reg, rot, ida, par, m, k}
{
  wpeso<-rep(1, noobs)
  j<-1
  i<-1
  for(j in 1:m){
    modelo1.df<-glm(mor reg+rot+reg*rot+ida,
                     family=poisson, weight=peso)
    dev<- deviance(modelo1.df)
    if(dev/(noobs-par)>=1.5){
      model2.fit<- fitted(modelo1.df)
      sigma1<- sum((mor-model2.fit) 2/model2.fit)-
        (sum(model2.fit)*(noobs-par))/noobs
      if(i<=1){
        sigma2<- sigma1-(sigma1-1/(noobs-par)*sum(mor-model2.fit)
          2/model2.fit*(model2.fit+sigma1 (-1)))/(1+1/
          (noobs-par)*sum((mor-model2.fit) 2/(model2.fit*(sigma1*
          model2.fit+1) 2)))
        if(sigma1 !=sigma2) {

```

```

        sigma1<-sigma2
        i<- i+1
        print("interacoes para sigma")
        print(i)
    }
}
else {
    i<- 1000000
    print("o processo convergiu")
}
wpeso<-1/(1+sigma2*model2.fit)
print("interacoes para o modelo")
print(j)
j<- j+1
}
else j<- 1000000
}
print("sigma")
print("sigma2")
print(summary(model2.fit))
print(summary(modelo1.df))
}

```

Macros utilizadas no **GLIM4.0** para o ajuste do modelo **Binomial Negativo** através de máxima verossimilhança.

```

$mac bpp!
!
!a funcao de ligacao sera %m=1 log
!
$yvar mor $switch %m$
$cal i2=1-i1=%eq(x,0)$
$own m1x m2x m3 m4 $sca 1$
$cal %lp=%log(%yv+0.01):%p=0$
$fac reg 3 rot 3 $
$fit reg+rot+reg.rot+ida$
$use nb $cal %z=%p$
$yvar mor$
$own m1 m2 m3 m4 $sca 1$
$cal %lp=m=0:%p=0$
$prin 'primeiro modelo com erro Poisson'$
$fit m-%gm+reg+rot+reg.rot+ida $disp mer$
$prin 'segundo modelo com erro binomial negativo'
        fator de dispersao '%z $prin "$
!
$cal %lp=%z$
$fit m=%gm+reg+rot+reg.rot+ida $disp me$
$endmac

```



```

!
$mac m1x $scal %fv=%exp(%lp)$
$endmac
$mac m2x $scal %dr=1/%fv$
$endmac
!
$mac m1 $scal %s=%ne(%pl,0)+1 $switch %s init next$
$warm $
$scal ta=i1+i2*(%eq(%m,1)/(1+%exp(lp)))
      :tb=%n*(%exp(lp)*ta*ta*%eq(%m,1))
      :m=i1+i2*(-ta/tb)
      :%fv=%n*ta
      :%lp=lp+%n*m$
$warm $
$endmac
!
$mac m2 $warm $scal %dr=i1+i2*(-1/tb) $warm $
$endmac
$mac m3 $warm $scal %va=%fv*(1+%lp*%fv) $warm $
$endmac
!
$mac m4 $warm $scal %di=2*(%yv*log(%yv/%fv)-(%yv-%fv)*%eq(%p, 0)
      -%log((1+%p*%yv)/(1+%p*%fv))*(1+%p*%yv)/%p*(%ne(%p,0)))$
$warm$
$endmac
$mac init $sort ta %yv $warm $
$scal %n=ta*(%nu)*1.1$
      :lp=i2*(%log(%n/%yv-1)*%eq(%m,1)
      :%x=1/%sqrt(2*%pi)$
$warm$
$endmac
!
$mac BN!
!
!Erro Binomial Negativo
!Estimativa inicial do sigma2
$prin 'para a estimacao inicial do sigma2'
!
$scal %p=%nu*(%x2-%df)/(%cu(%fv)*%df): %i=5$
$while %i BN1$
$endmac
!
$mac BN1!
$prin 'parte iterativa da Binomial Negativa, iterando em sigma2'!
!
$scal %q=%p : %p=%cu(((%yv-%fv)**2)/(%fv*(%fv+1/%q)))/%df$
!
$prin 'iteracao'%i: 'sigma2=' %p$
$fit reg+rot+reg.rot+ida $fit-%gm+reg+rot+reg.rot+ida$
$disp mecv$ $scal %q=(%q-%p)/%p : %f=%ge(%q, -0.001)*%le(%q, 0.001)$
$prin 'maximo numero de iteracoes em macor BN1'
$endmac
$data 2605 reg rot ida mor$
$dinp 'a:morta.dad' reg rot ida mor$

```

```
$fac reg 3 rot 3$
$cal %m=1$ $use bpp $prin 'muda o modelo'$
$delete model $use bpp$
$delete model$
$stop
```

Programa utilizado no **GLIM4.0** para o ajuste do modelo **Binomial Negativo** através de máxima verossimilhança iterando a variância nominal¹.

```
$SUBFILE VAR.BN $
*****
!Autora: Maria del Pilar Diaz. Departamento de Matematica e
! Estatistica. ESALQ- Universidade de Sao Paulo (BR).
! pilar@tambaqui.esalq.usp.br
!
! VAR.BN ajusta um modelo linear generalizado com erro
! Binomial Negativo iterando a variancia nominal.
!
! [Os valores utilizados como "input" para o conjunto de dados
! somente são fornecidos como exemplo ilustrativo].
! *****
$mac teste
$cal %i=%i-1$
$prin %i %k$
$use modelo$
$cal me=(%cu(%fv))/2605
: nu=((%cu(%fv-me))**2)/2605
: %k=(me**2)/(nu=me)$
$endmac
$mac modelo!
!Ajuste do modelo
$yvar mor$
$fac reg 3 rot 3$
$link o fvdr$
$cal %lp=%log(yv/(yv+%k))$
$erro o vadi$
$sca 1$
$terms (ida+reg+rot+reg.rot)$
$endmac
$mac fvdr!
!Modo "own" para o valor ajustado e
! a funcao derivada
$cal %fv=%k*%exp(%lp)/(1-%exp(%lp))
: %fv=%fv+0.001
: %dr=%k/(%fv*(%fv+%k))$
$endmac
$mac vadi!
```

¹Lembra-se que o parâmetro de dispersão $k = \theta = 1/\lambda$.

```

! Modo "own" para a funcao de variancia e
! componente de "deviance"
$cal %fv=%if(%lt(%fv,0),0.01,%fv)$
      :%va=(%fv+0.001)*(((%fv+0.001)/%k)+1)
      :%di=2*(%yv*%log(yv/%fv)+(%k+%yv)*
      %log((%fv+%k)/(yv+%k)))$
$endmac
$data 2605 reg rot ida mor$
$dinp 'a:morta.dad' reg rot ida mor$
$yvar mor$
$fac reg 3 rot 3$
$cal yv=%if(%eq(%yv,0),0.01,%yv)
! Colocar semente segundo criterio
      :%k=0.95
      :%i=10$
$link o fvdr$
$cal %lp=%log(yv/(yv+%k))$
$erro o vadi$
$sca 1$
! Declare modelo de locacao
$terms (ida+reg+rot+reg.rot)$
$while %i teste$
$disp m e c$
$finish

```

Programa utilizado no **GLIM4.0** para o ajuste do modelo **Binomial Negativo** através do enfoque intuitivo, o qual somente estima os β 's¹.

```

$mac busca
! uma grade para o theta
! e contador para "deviance"
$var 200 th dev qu$
$cal th=%gl(200,1)
      :th=th+1
      :th=th/120
! define contador de busca
      :%b=1$
$while %b principal$
$mac principal
! para o ajuste do modelo com o recurso
      "own" para os componentes
$cal %th=%th(%b)$
$fac reg 3 rot 3$
$link o fvdr$
$cal yv=%if(%eq(%yv,0),0.001,%yv)
      :%lp=%log(yv/(yv+%th))$
$erro o vadi$

```

¹Lembra-se que o parâmetro de dispersão $k = \theta = 1/\lambda$.

```

$scs 1$
! modelo de locacao
$terms (ida+reg+rot+reg.rot)$
$cal dev(%b)=%dv
! calculo do quociente "deviance" /gl
      :%qu=dev(%b)/2595$
$look dev(%b) %b %qu$
$cal %b=%b+1
! regra de parada
      :%b=%if(0.90;%quo;1.5, 0, %b)$
$prin %th dev(%b)$
$endmac
$mac fvdr
! modo "own" para o valor ajustado e
! a funcao derivada
$cal %fv=%th*%exp(%lp)/(1-%exp(%lp))
      :%fv=%fv+0.001
      :%dr=%th/(%fv*(%fv+%th))$
$endmac
$mac vadi!
! Modo "own" para a funcao de variancia e
! componente de "deviance"
$cal %fv=%if(%lt(%fv,0),0.01,%fv)$
      :%va=(%fv+0.001)*(((%fv+0.001)/%th)+1)
      :%di=2*(%yv*%log(yv/%fv)+( %th+%yv)*
      %log((%fv+%th)/(yv+%th)))$
$endmac
$data 2605 reg rot ida mor$
$dinp 'a:morta.dad' reg rot ida mor$
$use busca
$output resul.dad'$
$look (style=-1) %th dev(%b) %qu$
$output %poc$
$plot(t='Busca do theta otimo') %qu %th '*'$
$finish

```

Macro no **SPLUS** para implementação do método "**bootstrap**" para a obtenção do erro da estimativa de máxima verossimilhança do k , no enfoque **Binomial Negativo**.

```

attach(nome_arq)
sample(mort, 1000, replace=T)
...
(ó usuário deverá usar a "function" neg.bin)
...
density(kas, n=1000)
set.seed(101):m <- 1000
res <- numeric(m)
for (i in 1:m) res[i] <- mean(sample(mort, 1000, replace=T))

```

```
mean(res - mean(mort))
sqrt(var(res))
...
...
hist.FD(res, prob=T)
lines(density(res, n=1000, width=bandwidth.nrd(res)))
c(ucv(res, 0), bcv(res, 0), width.SJ(res, 0))
quantile(res, c(0.025, 0.975))
```