



**MODELAGEM DA SUPERDISPERSÃO EM
DADOS BINOMIAIS POR UM MODELO
LINEAR GENERALIZADO MISTO**

JOSÉ AIRTON RODRIGUES NUNES

2003

JOSÉ AIRTON RODRIGUES NUNES

**MODELAGEM DA SUPERDISPERSÃO EM DADOS BINOMIAIS POR
UM MODELO LINEAR GENERALIZADO MISTO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

Orientador

Prof. Dr. Augusto Ramalho de Moraes

LAVRAS
MINAS GERAIS - BRASIL

2003

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Nunes, José Airton Rodrigues

Modelagem da superdispersão em dados binomiais por um modelo linear generalizado misto / José Airton Rodrigues Nunes. – Lavras : UFLA, 2003.

65 p. : il.

Orientador: Augusto Ramalho de Moraes.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Algoritmo. 2. Variação extrabinomial. 3. Máxima verossimilhança penalizada. 4. Deviance. 5. Modelo logístico-normal. 6. Ensaio de germinação.
I. Universidade Federal de Lavras. II. Título.

CDD-519.5

-630.20195

JOSÉ AIRTON RODRIGUES NUNES

**MODELAGEM DA SUPERDISPERSÃO EM DADOS BINOMIAIS POR
UM MODELO LINEAR GENERALIZADO MISTO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária.

APROVADA em sexta-feira, 14 de fevereiro de 2003.

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho UFLA

Prof. Dr. Daniel Furtado Ferreira UFLA

Prof. Dr. Eduardo Bearzoti UFLA



Prof. Dr. Augusto Ramalho de Moraes
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

A Deus,

pela infinita bondade e amor e por me proporcionar a realização desta importante conquista.

OFEREÇO

Ao meu sempre e querido pai, Manoel Raimundo Nunes Rodrigues
(*"in memorian"*), pela coragem, incentivo e amor;

À minha maravilhosa mãe, Maria Dedícia Rodrigues, pelos
ensinamentos de vida, amor e confiança;

Ao meu inesquecível avô, João Rodrigues de Barros (*"in*
***memorian"*), pela bondade e sabedoria.**

DEDICO

AGRADECIMENTOS

A Deus, por sua sabedoria plena e compaixão;

À Universidade Federal de Lavras pela oportunidade de concluir o mestrado em Agronomia – Estatística e Experimentação Agropecuária;

À Pró-Reitoria de Pesquisa e Pós-graduação da UFLA pelo incentivo e apoio no desenvolvimento de pesquisas;

Ao Departamento de Ciências Exatas pelo apoio estrutural durante o curso de mestrado;

Ao curso de Pós-graduação em Agronomia – Estatística e Experimentação Agropecuária, pelo apoio irrestrito na condução e desenvolvimento de pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela concessão de bolsa de pesquisa;

Ao coordenador do curso de mestrado em Agronomia – Estatística e Experimentação Agropecuária, Prof. Joel Augusto Muniz, pelo pleno critério e esforço frente ao programa e pela amizade e motivação durante todo o curso;

Ao Prof. Augusto Ramalho de Moraes pela amizade, confiança e pela orientação e apoio na elaboração deste trabalho;

Ao Prof. Júlio Sílvio Sousa Bueno Filho pela ajuda e colaboração essenciais para o andamento e conclusão deste trabalho;

Aos professores do Departamento de Ciências Exatas, Augusto, Daniel Furtado, Delly Veiga, Eduardo Bearzoti, Fortunato, Joel, Júlio, Lucas, Luis Henrique, Marcelo, Mário Vivanco, Paulo César e Telma, pelos ensinamentos fundamentais para minha formação profissional e humana;

Ao Prof. Raimundo de Pontes Nunes pelo incentivo, amizade, e por me iniciar na pesquisa científica.

Aos funcionários do Departamento de Exatas, Andréia, Íla, Maria e Maristela, pela simpatia e eficiência com que nos atenderam;

Aos colegas de curso, Alessandra, Ana, Carla, Cirilo, Élcio, Lúcia, Luis, João Marcos, José Ermelino, José Otaviano, Waldemar e Washington, pela amizade e ótimo convívio durante todo o curso;

Aos grandes amigos, Alexandre, João Luis, Leandro, Márcio José, Pedro, Vladimir e Zenaide Barbosa, pela amizade e incentivo, e em especial a Andréia pelo carinho, amor e cumplicidade;

Aos amigos e colegas de república, Agnaldo, Alex, Cristovão, Eduardo, Kaesel e Reinaldo, pela amizade, paciência e companheirismo;

Ao meu querido pai (*"in memorian"*) pelo esforço, pela coragem, pela vontade de vencer e pelo amor e confiança que depositou em mim;

À minha querida mãe pela forma sábia e dedicada como lançou à vida, pelo infinito amor e pelo anseio que nos guia e contagia;

Aos meus valiosos avôs, Josefa, João Rodrigues (*"in memorian"*), José Rodrigues e Júlia, pela vontade de realizar, pela plenitude na sabedoria e pelo carinho e ensinamentos tão especiais;

Aos meus queridos irmãos, Denise, Francisco, João e Luiz, pela união e pelo carinho;

À minha maravilhosa sobrinha, Leticia, pela simpatia e pela alegria que nos proporciona;

Aos familiares e demais pessoas que direta ou indiretamente estiveram presentes durante essa conquista.

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	ii
RESUMO	iii
ABSTRACT	iv
1 INTRODUÇÃO.....	01
2 REFERENCIAL TEÓRICO.....	03
2.1 Modelos lineares generalizados com efeitos fixos	03
2.1.1 Introdução.....	03
2.1.2 Componentes dos modelos lineares generalizados	04
2.1.3 Estimação de máxima verossimilhança	06
2.1.4 “Deviance”	10
2.1.4.1 Análise de “deviance”	12
2.2 Modelo binomial	13
2.3 Superdispersão em dados binomiais	16
2.3.1 Introdução.....	16
2.3.2 Modelos.....	17
2.3.3 Quase-verossimilhança	20
2.4 Modelos lineares generalizados mistos	21
2.4.1 Introdução.....	21
2.4.2 Estimação	23
3 MATERIAL E MÉTODOS.....	26
3.1 Material	26
3.1.1 Exemplo 1: Ensaio de germinação	26
3.1.2 Exemplo 2: Ensaio de sobrevivência	27
3.2 Métodos	29

3.2.1 Modelo linear generalizado binomial	29
3.2.1.1 Exemplo 1: Ensaio de germinação.....	30
3.2.1.2 Exemplo 2: Ensaio de sobrevivência	31
3.2.2 Verificação de ajuste.....	31
3.2.3 Modelo linear generalizado binomial misto	33
3.2.3.1 Exemplo 1: Ensaio de germinação.....	38
3.2.3.2 Exemplo 2: Ensaio de sobrevivência	38
3.2.4 Verificação de ajuste.....	39
4 RESULTADOS E DISCUSSÃO	40
4.1 Exemplo 1: Ensaio de germinação	40
4.2 Exemplo 2: Ensaio de sobrevivência	47
5 CONCLUSÕES	53
REFERÊNCIAS BIBLIOGRÁFICAS	54
ANEXOS	57

LISTA DE TABELAS

TABELA 1	Distribuições de probabilidade e respectivas funções de ligação	06
TABELA 2	Distribuições de probabilidade e respectivas “deviances”	11
TABELA 3	Exemplo de um esquema de análise de “deviance” para um experimento de dois fatores, contendo a níveis do fator A, b níveis do fator B e repetidos r vezes	13
TABELA 4	Estimativas dos parâmetros para os modelos de regressão logística descritos em (1.19) por três algoritmos para MLG ..	41
TABELA 5	Estimativas dos parâmetros para os MLGM’s descritos em (1.31) por três algoritmos para MLGM	44
TABELA 6	Estimativas dos parâmetros para o modelo binomial padrão (1.20) por três algoritmos para MLG	48
TABELA 7	Estimativas dos parâmetros para os MLGM’s descritos em (1.32) por três algoritmos para MLGM	50

LISTA DE FIGURAS

- FIGURA 1 Representação gráfica do experimento sobre germinação de sementes 27
- FIGURA 2 Representação gráfica do experimento sobre sobrevivência de ovos de truta 28
- FIGURA 3 Representação gráfica da proporção média de sementes germinadas em função das variedades *O. aegyptiaco* 73 e 75 e dos extratos de raízes de feijão e pepino 41
- FIGURA 4 Representação gráfica das proporções de sementes germinadas observadas e ajustadas pelos modelos binomiais padrão de efeitos principais “M1” e com interação “M2” 43
- FIGURA 5 Proporções de germinação ajustadas pelos modelos mistos binomiais de efeitos principais “M1” e com interação “M2” por meio do glimmix SAS e algoritmo MLGM (a) e pelo glmmPQL R (b)..... 46
- FIGURA 6 Proporções de ovos de truta sobreviventes observados e ajustados pelo modelo (1.20) para as 20 parcelas 49
- FIGURA 7 Proporções de ovos de truta sobreviventes observados e ajustados pelo modelo (1.32) para as 20 parcelas 51

1 INTRODUÇÃO

Na pesquisa agrônômica, o pesquisador se depara, não raramente, com situações nas quais os dados obtidos apresentam erros binomiais. Proporções são comuns, por exemplo, em ensaios toxicológicos ou relacionados à germinação de sementes.

Neste caso, as pressuposições básicas requeridas para aplicação da metodologia da análise de variância associada ao teste F, técnica normalmente utilizada, são violadas. Negligenciando estas restrições, o pesquisador incorrerá em elevadas taxas de erro para os testes de hipóteses realizados e em inferências pouco confiáveis.

Para tornar válida a realização deste método estatístico, os pesquisadores têm adotado a mudança adequada da escala da variável aleatória por meio de transformações nestes dados (Bartlett, 1947). Porém, em muitas situações práticas, a escolha de uma escala que produza, simultaneamente, efeito aditivo da parte sistemática do modelo e variância constante para a parte aleatória se torna uma tarefa difícil.

Com a introdução dos modelos lineares generalizados, problemas com escalas foram grandemente reduzidos (McCullagh & Nelder, 1989). Trata-se de uma extensão dos modelos lineares, desenvolvida por Nelder & Wedderburn (1972), para dados não normalmente distribuídos. Esta metodologia tem como base o fato de que os efeitos sistemáticos são linearizados por uma transformação adequada dos valores esperados, permitindo que os valores ajustados variem dentro da amplitude real das respostas.

Não obstante o uso deste método estatístico, dados binomiais apresentam, em muitas ocasiões práticas, uma variância nas respostas superior à variância nominal da distribuição binomial comportada pelo modelo,

denominada de variação extrabinomial ou superdispersão. Vários autores têm mencionado a importância de considerar a presença deste fenômeno na modelagem e, com isso, têm sugerido várias formas de lidar com este problema prático.

A incorporação de efeitos aleatórios no preditor linear, os modelos lineares generalizados mistos (MLGM's), têm se mostrado numa técnica de grande utilidade e aplicabilidade na área das ciências biológicas para acomodação da superdispersão. Esta se fundamenta numa extensão da teoria dos modelos mistos para dados com distribuições pertencentes à família exponencial, assumindo-se uma distribuição particular para os efeitos aleatórios.

O objetivo principal deste trabalho foi a aplicação de modelos lineares generalizados mistos com base numa quase-logverossimilhança penalizada, em dois ensaios envolvendo dados binomiais superdispersos, com o intuito de acomodar de forma satisfatória a variabilidade extra presente, permitindo uma inferência adequada e segura. Além disso, implementar um algoritmo em ambiente R para executar um MLGM com função de ligação logística para estimação de parâmetros inerentes ao modelo proposto pela maximização da função de quase-logverossimilhança penalizada conjunta.

2 REFERENCIAL TEÓRICO

2.1 Modelos lineares generalizados com efeitos fixos

2.1.1 Introdução

A unidade de muitos métodos estatísticos envolvendo combinações lineares de parâmetros foi demonstrada por Nelder & Wedderburn (1972) usando a idéia de um modelo linear generalizado (MLG).

Os modelos lineares clássicos requerem pressuposições de independência, normalidade e variância constante das observações (Graybill, 1976). Estas duas últimas são abrandadas pela teoria dos MLG's, pois neste caso admite-se qualquer distribuição de probabilidade pertencente à família exponencial de distribuições para as observações, incluindo distribuições úteis como a normal, binomial, Poisson, multinomial, gama, binomial negativa, dentre outras; quanto à variância, a exigência é que esta tenha relação conhecida com a média. A aditividade dos efeitos continua sendo uma pressuposição importante nos MLG's, como uma propriedade das respostas esperadas (McCullagh & Nelder, 1989).

Em virtude da amplitude das distribuições abrangidas pelos MLG's, e pela menor exigência de pressupostos, esta metodologia tem tido grande aplicabilidade na resolução de problemas estatísticos em muitas áreas, principalmente nas Ciências Biológicas.

2.1.2 Componentes dos modelos lineares generalizados

Os MLG's utilizados neste trabalho são abordados extensivamente em McCullagh & Nelder (1989) e Dobson (1990). Um MLG é definido por três componentes:

i. Componente aleatório. As variáveis aleatórias independentes Y_1, Y_2, \dots, Y_n são pertencentes a uma mesma distribuição da família exponencial na forma canônica com médias $\mu_1, \mu_2, \dots, \mu_n$. A função densidade de probabilidade de Y_i é dada por

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i; \phi) \right\}, \quad (1.1)$$

na qual θ_i e ϕ são parâmetros e $a_i(\phi)$, $b(\theta_i)$ e $c(y_i; \phi)$, funções conhecidas. Em geral, $a_i(\phi) = \phi/w_i$, sendo w_i pesos a priori.

O parâmetro θ_i é o parâmetro natural ou canônico da distribuição e $\phi > 0$ é o parâmetro de escala. Além disso, pode-se mostrar que se Y_i tem distribuição na família exponencial, então a média e a variância são dadas por

$$\begin{aligned} \mu_i &= E(Y_i) = b'(\theta_i) \\ \sigma_i^2 &= Var(Y_i) = a_i(\phi) b''(\theta_i) \end{aligned} \quad (1.2)$$

sendo $b'(\theta_i)$ e $b''(\theta_i)$ as derivadas primeira e segunda de $b(\theta_i)$ em relação a θ_i .

Quando $a_i(\phi) = \phi/w_i$, a variância tem a forma mais simples expressa por

$$Var(Y_i) = \sigma_i^2 = \frac{\phi b''(\theta_i)}{w_i}.$$

A partir desta equação, evidencia-se que a variância de Y_i é o produto de duas funções: $b''(\theta_i)$, que depende do parâmetro canônico θ_i , e conseqüentemente de μ_i , chamada *função de variância*; e $a_i(\phi)$, que depende somente de ϕ , o parâmetro de escala ou parâmetro de dispersão. A função de variância, considerada como função de μ_i , é denotada por $V(\mu_i)$, ou seja,

$$V(\mu_i) = b''(\theta_i). \quad (1.3)$$

As distribuições da família exponencial, portanto, têm uma relação conhecida entre a média e variância dada por

$$Var(Y_i) = V(\mu_i) \frac{\phi}{w_i},$$

sendo $\mu_i = E(Y_i)$.

ii. Componente sistemático. As variáveis explicativas ou preditoras entram na forma de uma soma linear de seus efeitos

$$\eta = X\beta,$$

sendo $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ a matriz de incidência do modelo; $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ o vetor de parâmetros desconhecidos e $\eta = (\eta_1, \eta_2, \dots, \eta_n)^T$ o vetor de preditores lineares.

iii. **Função de ligação.** Uma função que liga o componente aleatório ao componente sistemático, ou seja, relaciona a média ao preditor linear, isto é,

$$\eta_i = g(\mu_i) \Rightarrow \mu_i = g^{-1}(\eta_i), \quad (1.4)$$

sendo $g(\cdot)$ uma função monótona e diferenciável chamada de função de ligação e μ_i o valor esperado ajustado para Y_i . A Tabela 1 mostra exemplos de funções de ligação para uma variedade de distribuições de probabilidade.

TABELA 1 Distribuições de probabilidade e respectivas funções de ligação.

<i>Distribuição de probabilidade</i>	<i>Função de ligação canônica</i>
Normal	$\eta_i = \mu_i$
Binomial	$\eta_i = \log\left(\frac{\mu_i}{1 - \mu_i}\right)$
Poisson	$\eta_i = \log \mu_i$
Gama	$\eta_i = \mu_i^{-1}$
Inversa Normal	$\eta_i = \mu_i^{-2}$

Fonte: McCullagh & Nelder (1989).

2.1.3 Estimação de máxima verossimilhança

Os estimadores dos parâmetros β 's para os MLG's são obtidos pelo método da máxima verossimilhança, maximizando a verossimilhança ou a logverossimilhança dos dados observados (Mood et al., 1974). Se a distribuição

de probabilidade é especificada por (1.1) e baseada em um conjunto de observações independentes $\mathbf{y} = (y_1, y_2, \dots, y_n)'$, então a função de logverossimilhança conjunta é dada por

$$l(\theta, \phi; \mathbf{y}) = \sum_{i=1}^n \log f_i(y_i; \theta_i) = \sum_{i=1}^n \left(\frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} \right) + \sum_{i=1}^n c(y_i, \phi), \quad (1.5)$$

sendo $E(Y_i) = \mu_i$, $\eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$.

Uma propriedade da família exponencial de distribuições é a que seus elementos satisfazem condições de regularidade suficientes para assegurar que o máximo global do logaritmo da função de verossimilhança conjunta $l(\theta, \phi; \mathbf{y})$

(1.5) seja dado unicamente pela solução do sistema de equações $\mathbf{U}_\theta = \frac{\partial l}{\partial \theta} = \mathbf{0}$

ou, equivalentemente, pelas soluções de $\mathbf{U}_\beta = \frac{\partial l}{\partial \beta} = \mathbf{0}$ (Dobson, 1990). Tem-se

que as funções escore são dadas por

$$U_j = \sum_{i=1}^n \frac{\partial l(\theta_i, \phi; y_i)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_j}.$$

Em geral, as equações $U_j = 0$, $j = 1, 2, \dots, p$ não são lineares e têm que ser resolvidas numericamente por processos iterativos do tipo Newton – Raphson. O método iterativo de Newton – Raphson para a solução de uma equação $f(x) = 0$ é baseado na aproximação de Taylor para a função $f(x)$ nas vizinhanças do ponto \mathbf{x}_0 .

Considerando que se deseja obter a solução do sistema de equações

$$\mathbf{U}_p = \frac{\partial l}{\partial \beta} = \mathbf{0} \text{ e utilizando da versão multivariada do método de Newton -}$$

Raphson, obtém-se

$$\beta^{(m+1)} = \beta^{(m)} + (I_0^{-1})^{(m)} \mathbf{U}^{(m)},$$

sendo $\beta^{(m)}$ e $\beta^{(m+1)}$ os vetores de parâmetros estimados nos passos (m) e (m+1),

$\mathbf{U}^{(m)}$ o vetor escore com elementos $\frac{\partial l}{\partial \beta_j}$, avaliado no passo m, e $(I_0^{-1})^{(m)}$ a

inversa da negativa da matriz de derivadas parciais de 2ª ordem de $f(x)$, com

elementos $-\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}$, avaliada no passo (m).

Um procedimento alternativo que algumas vezes é mais simples do que o método Newton - Raphson é chamado método de Escore de Fisher. Ele envolve a substituição da matriz de derivadas de segunda ordem pela matriz de valores esperados das derivadas parciais, isto é, a substituição da matriz de informação observada, I_0 , pela matriz de informação esperada de Fisher, \mathfrak{I} . Portanto,

$$\beta^{(m+1)} = \beta^{(m)} + (\mathfrak{I}^{-1})^{(m)} \mathbf{U}^{(m)},$$

sendo que \mathfrak{I} tem elementos dados por $\mathfrak{I}_{jk} = E \left[\frac{-\partial^2 l}{\partial \beta_j \partial \beta_k} \right] = E \left[\frac{\partial l}{\partial \beta_j} \frac{\partial l}{\partial \beta_k} \right]$,

que é a matriz de covariâncias dos U_j 's.

Através do rearranjo dos termos, tem-se

$$\mathfrak{I} = \frac{1}{\phi} X^T W X$$

$$U_j = \sum_{i=1}^n \frac{w_i}{\phi} \frac{(y_i - \mu_i)}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} = \sum_{i=1}^n \frac{1}{\phi} x_{ij} W_i \frac{d\eta_i}{d\mu_i} (y_i - \mu_i) = \frac{1}{\phi} X^T W \Delta (\mathbf{y} - \boldsymbol{\mu})$$

$$W = \text{diag}(W_1, W_2, \dots, W_n), \text{ em que } W_i = \frac{w_i}{V(\mu_i) [g'(\mu_i)]^2}$$

$$\Delta = \text{diag} \left\{ \frac{d\eta_1}{d\mu_1}, \frac{d\eta_2}{d\mu_2}, \dots, \frac{d\eta_n}{d\mu_n} \right\} = \text{diag} \{ g'(\mu_1), g'(\mu_2), \dots, g'(\mu_n) \};$$

e fazendo-se

$$\mathbf{z}^{(m)} = X \boldsymbol{\beta}^{(m)} + \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)} = \boldsymbol{\eta}^{(m)} + \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)},$$

o vetor de variáveis dependentes ajustadas, tem-se que a equação iterativa para o método de escore pode ser escrita como

$$\begin{aligned} X^T W^{(m)} X \hat{\boldsymbol{\beta}}^{(m+1)} &= X^T W^{(m)} \mathbf{z}^{(m)} \\ \hat{\boldsymbol{\beta}}^{(m+1)} &= (X^T W^{(m)} X)^{-1} X^T W^{(m)} \mathbf{z}^{(m)}, \end{aligned} \quad (1.6)$$

que tem a forma da solução das equações normais para um modelo linear obtido pelo método dos quadrados mínimos ponderados, exceto que, nesse caso, a solução $\hat{\boldsymbol{\beta}}$ é obtida por processo numérico iterativo.

2.1.4 “Deviance”

O processo de ajuste de um modelo a um conjunto de dados pode ser considerado como uma maneira de substituição dos dados y por um conjunto de valores ajustados $\hat{\mu}$ para um modelo envolvendo um número relativamente pequeno de parâmetros. Em geral, os $\hat{\mu}$'s não serão iguais aos y 's exatamente, e a questão é em quanto eles diferem (McCullagh & Nelder, 1989).

Dadas n observações, a elas podem ser ajustadas modelos contendo até n parâmetros. O modelo mais simples é o **modelo nulo** que tem um único parâmetro, representado por um valor comum a todos os dados. Esse modelo atribui toda variação entre os y 's ao componente aleatório. No outro extremo está o **modelo saturado** ou **completo** que tem n parâmetros, um para cada observação. Este atribui toda variação ao componente sistemático e, portanto, ajusta-se perfeitamente (Demétrio, 2001).

Na prática, o modelo nulo é simples demais e o modelo completo não é informativo, pois não resumiriam os dados. Entretanto, o modelo completo nos dá uma base para medir a discrepância em relação a um modelo intermediário com p parâmetros. O modelo contendo o menor número de parâmetros para o ajuste é chamado **modelo minimal**; por outro lado, o modelo contendo o maior número de termos que podem ser considerados é chamado de **modelo maximal** (Demétrio, 2001).

Nelder & Wedderburn (1972) propuseram, como uma medida de discrepância, a “deviance”, representada por

$$D(y; \hat{\mu}) = 2[l(y; y) - l(\hat{\mu}; y)], \quad (1.7)$$

sendo que $l(y; \theta)$ é a função de máxima logverossimilhança realizável para um ajuste exato para o qual θ é expresso com uma função de y (modelo saturado), e $l(\hat{\mu}; y)$ é a função de máxima logverossimilhança calculada para os parâmetros estimados para o qual θ é expresso como uma função de $\hat{\mu}$ (modelo corrente).

TABELA 2 Distribuições de probabilidade e respectivas “deviances”.

<i>Distribuição de Probabilidade</i>	<i>“Deviance”</i>
Normal	$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2$
Poisson	$2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right]$
Binomial	$2 \sum_{i=1}^n \left\{ y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) + (m_i - y_i) \log \left[\frac{(m_i - y_i)}{(m_i - \hat{\mu}_i)} \right] \right\}$
Gama	$2 \sum_{i=1}^n \left[-\log \left(\frac{y_i}{\hat{\mu}_i} \right) + \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} \right]$
Inversa Normal	$\sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2 y_i}$

Fonte: McCullagh & Nelder (1989).

¹¹ A “deviance” é uma generalização da soma de quadrados dos erros na análise de variância e da razão de verossimilhanças χ^2 em tabelas de contingência, sendo equivalente, respectivamente, para erros normais e Poisson (McCullagh & Nelder, 1989).

A “deviance” pode ser utilizada para propósitos de avaliação de ajustamento de modelos e para testar hipóteses. Nas distribuições de

probabilidade da família exponencial a aproximação de χ^2 é bastante acurada para diferença de “deviance” (idêntica à estatística razão de verossimilhanças) com graus de liberdade equivalentes à diferença entre o número de parâmetros nos dois modelos. As formas das “deviances” para algumas distribuições são apresentadas na Tabela 2.

Associada a cada modelo está uma quantidade p de graus de liberdade que é dado pelo posto da matriz do modelo X ou equivalentemente ao número de parâmetros linearmente independentes a serem estimados. Para uma amostra de n observações independentes, a “deviance” para modelo em pesquisa tem graus de liberdade residual dado por $(n-p)$. A adequação do modelo linear generalizado é testada por meio do valor calculado para S_p (com $n-p$ graus de liberdade), a ser comparado com o percentil de alguma distribuição de probabilidade de referência a um nível de significância igual a α

$$S_p = \frac{D(\mathbf{y}; \hat{\boldsymbol{\mu}})}{\phi} \sim \chi_{n-p, \alpha}^2 \text{ (aproximada)}, \quad (1.8)$$

sendo que S_p denota “scaled deviance” e ϕ , o parâmetro de escala.

2.1.4.1 Análise de “deviance”

A análise de “deviance” é uma generalização da análise de variância usual para os modelos lineares generalizados com finalidade de se obterem, a partir de modelos seqüenciais, os efeitos dos fatores, covariáveis e suas interações. Dado esta seqüência de modelos hierarquizados, utiliza-se a “deviance” como uma medida de discrepância generalizada para testar hipóteses por meio de tabela de diferença de “deviances” (Cordeiro, 1986).

Seja $M_{p_1}, M_{p_2}, \dots, M_{p_r}$ uma seqüência de modelos hierárquicos de dimensões p_1, p_2, \dots, p_r , com matrizes dos modelos $X_{p_1}, X_{p_2}, \dots, X_{p_r}$ e “deviances” $D_{p_1}, D_{p_2}, \dots, D_{p_r}$, tendo os modelos a mesma distribuição e a mesma função de ligação. Assim, para o caso de um ensaio fatorial inteiramente casualizado com r repetições, com a níveis do fator A e b níveis do fator B, obtêm-se os resultados mostrados na Tabela 3.

TABELA 3 Exemplo de um esquema de análise de “deviance” para um experimento de dois fatores, contendo a níveis do fator A, b níveis do fator B e repetido r vezes.

<i>Modelo</i>	<i>G.L.</i>	<i>“deviance”</i>	<i>Dif. “deviance”</i>	<i>Dif. G.L.</i>	<i>Significado</i>
Nulo	$rab-1$	D_1			
A	$A(rb-1)$	D_A	$D_1 - D_A$	$a-1$	A ignorando B
A+B	$A(rb-1)-(b-1)$	D_{A+B}	$D_A - D_{A+B}$	$b-1$	B incluindo A
A+B+AB	$ab(r-1)$	D_{AB}	$D_{A+B} - D_{AB}$	$(a-1)(b-1)$	Interação AB Incluído A e B
Completo	0	0	D_{AB}	$ab(r-1)$	Resíduo

Fonte: Cordeiro (1986)

2.2 Modelo Binomial

A distribuição binomial é de grande importância e aplicação na Estatística Experimental, pois modela um grande número de fenômenos observáveis, neste caso, dados de proporções.

Uma variável aleatória Y_i , número de sucessos em amostras de tamanho m_i , $i = 1, \dots, n$, é definida ter distribuição binomial se sua função de densidade discreta é dada por (MOOD et al., 1974)

$$f_{Y_i}(y_i) = \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i} I_{\{0,1,\dots,m_i\}}(y_i) \quad 0 \leq \pi_i \leq 1,$$

que pode ser reescrita na forma (1.1) como

$$f_{Y_i}(y_i) = \exp \left\{ y_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + m_i \log(1 - \pi_i) + \log \binom{m_i}{y_i} \right\},$$

em que

$$\theta_i = \log \left(\frac{\pi_i}{1 + \pi_i} \right); \quad a_i(\phi) = 1; \quad b(\theta_i) = -m_i \log(1 - \pi_i); \quad c(y_i; \phi) = \log \binom{m_i}{y_i}.$$

A variável aleatória $R_i = Y_i/m_i = \text{Bin}(m_i; \pi_i) / m_i$, ou seja, a proporção de sucessos em amostras de tamanho m_i , tem distribuição binomial e, portanto, pertencente à família exponencial, com função de densidade discreta dada por

$$f_{R_i}(r_i) = \binom{m_i}{m_i r_i} \pi_i^{m_i r_i} (1 - \pi_i)^{m_i(1 - r_i)} I_{\left\{ \frac{0}{m_i}, \frac{1}{m_i}, \dots, \frac{m_i}{m_i} \right\}}(r_i) \quad 0 \leq \pi_i \leq 1,$$

que pode ser reescrita na forma (1.1) como

$$f_{R_i}(r_i) = \exp \left\{ m_i \left[r_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + \log(1 - \pi_i) \right] + \log \binom{m_i}{m_i r_i} \right\}, \quad (1.9)$$

em que

$$\theta_i = \log\left(\frac{\pi_i}{1 + \pi_i}\right); a_i(\phi) = \frac{1}{m_i}; b(\theta_i) = -\log(1 - \pi_i); c(r_i; \phi) = \log\left(\frac{m_i}{m_i r_i}\right),$$

tendo-se o valor esperado e a variância de R_i dados por

$$E(R_i) = b'(\theta_i) = \pi_i$$

$$Var(R_i) = a_i(\phi_i) b''(\theta_i) = \frac{\pi_i(1 - \pi_i)}{m_i}. \quad (1.10)$$

É possível modelar as proporções esperadas π_i em termos de fatores qualitativos ou variáveis explicatórias (x_1, \dots, x_p) por meio da combinação linear

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

para parâmetros desconhecidos β_1, \dots, β_p .

As funções de ligação $g(\pi)$ comumente utilizadas na prática são:

- A logit ou função logística

$$g_1(\pi) = \log\left[\frac{\pi}{1 - \pi}\right];$$

- A probit ou função normal inversa

$$g_2(\pi) = \Phi^{-1}(\pi);$$

- A função complemento log-log

$$g_3(\pi) = \log[-\log(1 - \pi)];$$

- A função log-log

$$g_4(\pi) = -\log[-\log(\pi)].$$

A aplicação da distribuição binomial em dados superdispersos é limitada pelo fato de a variância ser automaticamente determinada estando a média especificada (Hinde & Demétrio, 1998).

2.3 Superdispersão em dados binomiais

2.3.1 Introdução

O fenômeno da superdispersão tem sido largamente considerado na literatura, particularmente em relação à distribuição binomial, em trabalhos como o de Williams (1982). A superdispersão em dados binomiais ou variação extra-binomial constitui um problema prático comum em dados de proporções provenientes de experimentos em vários campos de aplicação. A superdispersão está presente quando os dados mostram maior variabilidade do que é previsto pelo modelo amostral assumido, ou seja, a variância de Y excede a variância nominal; neste caso a variância binomial nominal (1.10) (McCullagh & Nelder, 1989).

Falhas na modelagem de dados sem considerar tal fenômeno podem conduzir a sérios problemas de subestimação dos erros padrão e a uma inferência mal conduzida para os parâmetros do modelo (Hinde & Demétrio, 1998).

McCullagh & Nelder (1989) comentam que a superdispersão pode ocorrer de várias formas. A forma mais comum é o agrupamento de dados, por exemplo, de famílias, colônias de ocorrência natural ou leitegadas, descrita em vários trabalhos.

Fitzmaurice (1997) relata que a superdispersão pode ser representada pela correlação positiva entre as variáveis respostas, o que invalidaria a pressuposição de independência que é usada na construção das estatísticas de ajustamento de razão de verossimilhança e erros padrão nominais.

Alguns autores têm sugerido testes ou diagnósticos para verificar a presença de superdispersão em dados binomiais. Dean (1992) desenvolveu um método para obtenção de testes para superdispersão em modelos de regressão binomial e Poisson, derivado a partir da construção de uma família de modelos superdispersos e, por conseguinte, da construção de testes de escore para testar a hipótese de adequação do modelo natural (sem inclusão de parâmetro de superdispersão). Ganio & Schafer (1992) apresentam diagnósticos para variação extrabinomial e extrapoisson, por meio de inclusão de parâmetros de dispersão no modelo natural, fazendo o teste de adequação a partir da razão de verossimilhança e testes escore para estes parâmetros extra usando a verossimilhança da família exponencial dupla. Shall (1991) sugere que a estimativa do componente extra de dispersão ($\hat{\sigma}^2$) proveria uma checagem para superdispersão ou subdispersão presente, a qual seria, respectivamente, indicada pelos valores $\hat{\sigma}^2$ significativamente maiores ou menores que 1, para modelos lineares generalizados com efeitos aleatórios.

2.3.2 Modelos

Vários modelos têm sido utilizados para modelar a superdispersão em dados de proporções. Williams (1975) analisa dados binários envolvendo reprodução e teratogênese a partir de ensaios toxicológicos completamente aleatorizados, nos quais as unidades experimentais são constituídas por leitegadas animais por um modelo beta-binomial. Hines & Lawless (1993)

modelam a superdispersão em dados de mortalidade em experimentos de toxicidade com os tratamentos aplicados em grupos de animais no tempo, por modelos apresentados por Williams (1982).

Brooks et al. (1997) utilizaram o modelo beta-binomial, modelos de misturas de binomial com beta-binomial e misturas de binomiais e outros em seis conjuntos de dados de controle de mortalidade fetal de ninhadas de ratos. Os autores observaram, em cinco desses conjuntos, que a mistura de um modelo beta-binomial com a distribuição binomial proporcionou melhor descrição dos dados. Vieira (1998) comparou seis modelos para modelar a superdispersão em dados de um ensaio de controle biológico para *Diatraea saccharalis*, em que encontrou que o modelo binomial truncado superdisperso apresentou melhor ajuste aos dados.

Existem vários modelos específicos para a superdispersão em dados binomiais que podem ser organizados em grupos de acordo com a abordagem utilizada (Hinde & Demétrio, 1998; Lee & Nelder, 2000):

i. Assumindo uma forma mais geral para a função de variância, possivelmente incluindo parâmetros adicionais. O método de estimação, neste caso, é realizado por meio da quase-verossimilhança (Wedderburn, 1974);

ii. Assumindo um modelo de duplo estágio para a resposta, ou seja, o parâmetro do modelo básico da resposta tem alguma distribuição. Neste caso, o modelo beta binomial, incluso nos modelos Williams tipo II, é o mais utilizado;

iii. Assumindo um modelo de efeito aleatório, ou seja, uma forma de modelagem da superdispersão constituída numa natural extensão dos MLG's pela incorporação de efeitos aleatórios, os modelos lineares generalizados mistos (MLGM's). Os parâmetros podem ser estimados pelo método da

quase-verossimilhança (Gilmour, Anderson & Rae, 1985; IM & Gianola, 1988; Shall, 1991; Breslow & Clayton, 1993; McCulloch, 1997; Hall, 2000).

Para o subitem (ii), tem-se que o modelo Williams tipo II é caracterizado como segue. Seja Y_i variável aleatória com distribuição $\text{Bin}(n_i, R_i)$ e R_i uma variável aleatória com $E(R_i) = \pi_i$ e $V(R_i) = \phi\pi_i(1-\pi_i)$, sendo $\phi \geq 0$ um parâmetro de escala desconhecido. Assim, $Y_i | R_i$ tem distribuição $\text{Bin}(n_i, R_i)$. Sendo R_i uma variável aleatória não observável, incondicionalmente tem-se que

$$E(Y_i) = n_i \pi_i$$

$$V(Y_i) = n_i \pi_i (1 - \pi_i) [1 + \phi(n_i - 1)]$$

Na modelagem da superdispersão com este tipo de modelo, pode-se utilizar o método da quase-verossimilhança para estimação dos parâmetros. Um caso especial do modelo Williams tipo II é a distribuição beta-binomial em que R_i tem distribuição $\text{Beta}(a, b)$, a qual é utilizada algumas vezes como um modelo alternativo para superdispersão (Williams, 1975; Williams, 1982), sendo definida por

$$P(Y = y) = \binom{n}{y} \frac{B(a+y, n+b-y)}{B(a, b)} I_{\{0,1,\dots,n\}}(y),$$

sendo $n > 0$, $a > 0$, e $b > 0$. Mostra-se que

$$E(Y) = \frac{na}{a+b} \quad e \quad V(Y) = \frac{nab(n+a+b)}{(a+b)^2(a+b+1)}.$$

inferência sobre a população no preditor linear de um MLG, permite, em muitos casos, modelar a superdispersão presente e acomodar dados correlacionados (McCulloch, 1997).

Os modelos lineares generalizados mistos (MLGM) surgem em experimentos aleatorizados em que se observam dados discretos da mesma forma como surgem os modelos lineares Gauss-Markov normal (MLGMN) para uma variável contínua. Assumindo aditividade entre efeitos de unidade experimental (UE) e de tratamentos, o modelo hierárquico produzido é aproximadamente normal-normal (nos níveis de parcelas e dados) e a convolução de duas normais gera uma nova distribuição normal. No caso dos dados discretos, a convolução da aproximação normal para o efeito subjacente de parcelas com a distribuição dos dados observados (contagens ou proporções) gera um MLGM na forma de modelo hierárquico (Normal-Poisson ou Normal-Binomial, respectivamente).¹

Um modelo linear generalizado com efeitos aleatórios é definido aplicando-se as mesmas estratégias básicas para ajuste de um MLG aos valores esperados no modelo de efeitos fixos, conforme item 2.1 (McCulloch & Searle, 2001). Quando incorporados efeitos aleatórios no preditor linear, o modelo é alargado para

$$g(\mu) = \eta = X\beta + Zu, \quad (1.13)$$

em que $X_{m \times p}$ e $Z_{m \times q}$ são matrizes do modelo conhecidas e referentes, respectivamente, aos efeitos fixos e aleatórios presentes; $\beta_{p \times 1}$ é o vetor de efeitos fixos; $u_{q \times 1}$ é o vetor de efeitos aleatórios ditos não correlacionados com esperança zero, assumindo uma particular distribuição, como, por exemplo,

¹ GILMOUR, S. G. 2002, Lavras - MG, informação pessoal.

normal multivariada $NM(\mathbf{0},G)$; e $\mu=E(Y|u)$ é o vetor de valores esperados condicionais. G é a matriz de covariâncias, sendo, em geral, uma matriz diagonal, cujos elementos na diagonal são os componentes de variâncias desconhecidos associados aos efeitos aleatórios presentes no modelo.

2.4.2 Estimação

O uso da abordagem dos MLGM's apresenta dois problemas principais (Shall, 1991). Inicialmente, a dificuldade em especificar uma particular distribuição para os efeitos aleatórios do modelo. Em segundo lugar, embora o método da máxima verossimilhança (MV) e variantes sejam padrões para modelos lineares mistos e MLG's, seu uso em MLGM's tem sido limitado a modelos mais simples por necessitar de resoluções de integrais de alta dimensionalidade (McCulloch, 1997).

Algumas propostas têm sido sugeridas para contornar os problemas enunciados no uso de MLGM's. Lee & Nelder (2000) usam uma h-verossimilhança a partir de modelos lineares generalizados hierárquicos como forma de abordar MLGM's, sem necessitar de pressuposições distribucionais para os efeitos aleatórios. Vários outros autores, como Gilmour et al. (1985), Shall (1991) e Breslow & Clayton(1993), sugerem a utilização de algoritmos de maximização conjunta para obtenção das estimativas dos parâmetros por meio da maximização da função quase-logverossimilhança conjunta das observações condicionais e dos efeitos aleatórios.

A partir da segunda proposta, tem-se que os MLGM's podem utilizar a quase-logverossimilhança conjunta, uma extensão dos métodos de verossimilhança, da mesma forma como em modelos de efeitos fixos. A quase-logverossimilhança conjunta é formada pela soma da quase-logverossimilhança

condicional de uma observação $y_i|u_i$, neste caso pertencente à família exponencial, representada por

$$Q(\mu_i; y_i | u_i) = \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)}, \quad (1.14)$$

acrescida pelo logaritmo da forma da função densidade de probabilidade para os efeitos aleatórios desconhecidos; em geral assume-se a distribuição normal. Assim, a quase-logverossimilhança conjunta é expressa por

$$Q(\mu, \mathbf{u}; \mathbf{y}) \propto \sum_{i=1}^n \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi_i)} \right] - \frac{1}{2} (\log |G| + \mathbf{u}' G^{-1} \mathbf{u}). \quad (1.15)$$

Para a obtenção de estimativas dos parâmetros referentes aos efeitos fixos e aleatórios presentes no MLGM, tornam-se necessárias diferenciações de (1.15) em relação aos efeitos e, por conseguinte, a obtenção das equações de máxima quase-verossimilhança. Para a resolução destas equações, vários autores têm proposto alternativas. Gilmour et al. (1985) e Shall (1991) apresentam algoritmos de maximização conjunta, os quais constituem versões aproximadas das equações do modelo misto de Henderson et al. (1959), que surgem a partir da maximização da distribuição conjunta dos dados observados e efeitos aleatórios. Assim, soluções de β e \mathbf{u} podem ser obtidas a partir da resolução das equações do modelo linear generalizado misto dadas por

$$\begin{bmatrix} X'WX & X'WZ \\ Z'WX & Z'WZ + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} X'W\mathbf{y}^* \\ Z'W\mathbf{y}^* \end{bmatrix}, \quad (1.16)$$

em que \mathbf{y}^* é o vetor de variáveis dependentes ajustadas como no MLG.

O procedimento proposto por Breslow & Clayton (1993), baseado numa Quase-verossimilhança Penalizada conjunta(QVP), a qual assume que $\phi = 1$, constitui-se essencialmente no mesmo algoritmo com justificações diferentes. A função de QVP é definida pela introdução de uma função pênalti da forma $\frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}$ na quase-logverossimilhança condicional (1.14) a fim de remediar o fato de que somente a especificação da relação entre a média e a variância não forma uma base suficiente sobre a qual se pode estimar a estrutura de covariâncias (McCulloch & Searle, 2001) . Assim a QVP é dada por

$$QVP(\boldsymbol{\mu}, \mathbf{u}; \mathbf{y}) = Q(\boldsymbol{\mu}; \mathbf{y} | \mathbf{u}) - \frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u} \quad . \quad (1.17)$$

Assim, os estimadores de máxima QVP para os parâmetros do modelo podem ser obtidos, analogamente, por (1.16), a partir da maximização de (1.17).

3 MATERIAL E MÉTODOS

3.1 Material

No presente trabalho foram analisados dois conjuntos de dados com respostas binomiais, mostrados no Anexo A, com fins metodológicos de aplicação. A descrição desses conjuntos de dados é feita a seguir.

3.1.1 Exemplo 1: Ensaio de germinação

Crowder (1978), citado por Breslow & Clayton (1993), apresenta dados sobre a proporção de sementes germinadas (r_i) em 21 bandejas, provenientes de um experimento conduzido num delineamento inteiramente casualizado com os tratamentos repetidos, dispostos num esquema fatorial cruzado 2×2 , sendo duas espécies de semente (*Orobanche aegyptiaco* 75 (= 0), *Orobanche aegyptiaco* 73 (= 1)) e dois diferentes meios para germinação provenientes de extratos de raiz (feijão (= 0), pepino (= 1)). As sementes de *Orobanche* foram colocadas em diluições 1/125 dos referidos extratos.

Os dados deste ensaio, apresentados no Anexo A1, encontram-se bastante discutidos na literatura como exemplo clássico de presença de variação extra-binomial ou superdispersão. O uso destes motiva-se como exemplo-teste para aplicação de algoritmos implementados em ambiente R 1.6.1 (IRAKA & GENTLEMAN, 1996) para MLG e MLGM via QVP, apresentados no Anexo B, para modelagem de dados binomiais, e para fins de validação dos resultados obtidos pelos algoritmos desenvolvidos comparativamente com resultados presentes na literatura, sobretudo com os resultados obtidos por Breslow & Clayton (1993).

Na Figura 1 está sumariada a estrutura envolvida no experimento, compreendendo o número e tamanho das parcelas experimentais, os fatores, seus níveis, e os dados experimentais resultantes.

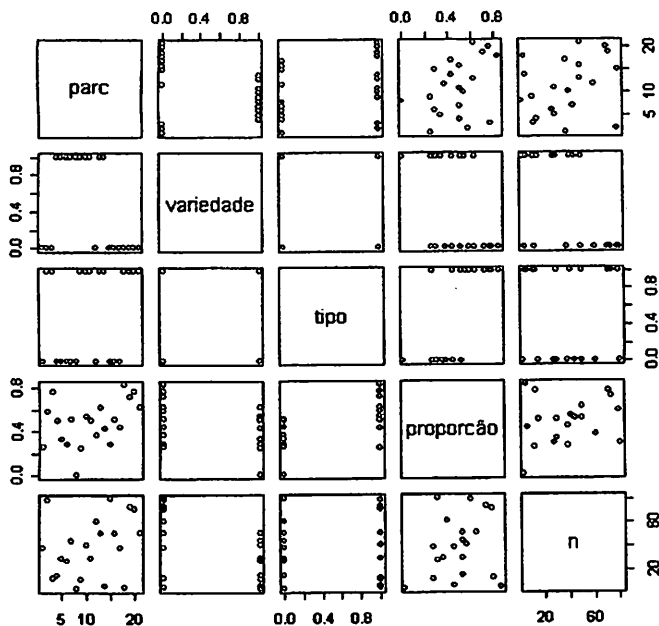


FIGURA 1 Representação gráfica do experimento sobre germinação de semente.

3.1.2 Exemplo 2: Ensaio de sobrevivência

Manly (1978), citado por Hinde & Demétrio (1998), retrata um experimento sobre a sobrevivência de ovos de truta. O experimento constitui-se de caixas contendo quantidades variáveis de ovos, que foram colocadas aleatoriamente em cinco diferentes locais num fluxo e em quatro diferentes

períodos (4, 7, 8 e 11 semanas). Uma caixa de cada local foi amostrada e os números de ovos sobreviventes foram contados.

A Figura 2 apresenta de forma sumariada a estrutura envolvida no experimento, mostrando o número e o tamanho das parcelas experimentais, os fatores, seus níveis e o número de ovos de truta sobreviventes. Os dados estão apresentados no Anexo A2.

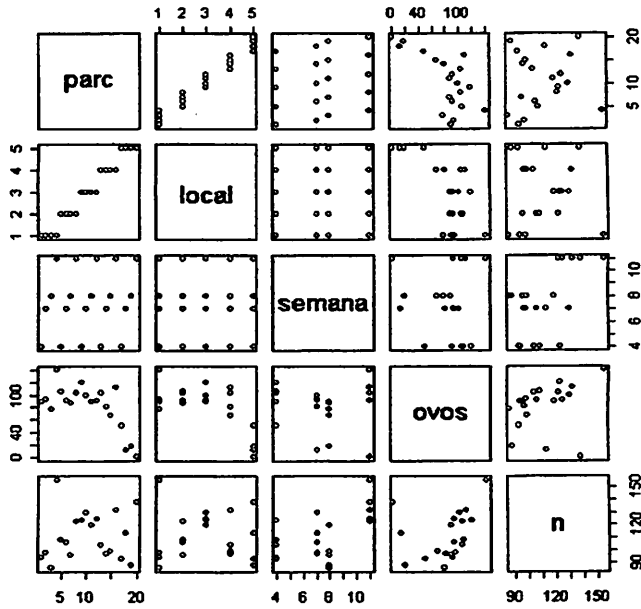


FIGURA 2 Representação gráfica do experimento sobre sobrevivência de ovos de truta.

3.2 Métodos

3.2.1 Modelo linear generalizado binomial

Sendo os dados descritos em 3.1 proporções de sucessos (r_i) em amostras de tamanho (m_i), admitiu-se, primeiramente, o modelo binomial (1.9) para explicar os dados apresentados. Assim, conforme teoria de modelos lineares generalizados considerando efeitos fixos exposta no item 2.1, optou-se pelo uso da função de ligação logística (canônica) como uma transformação do valor esperado que se deseja modelar como uma combinação linear nos parâmetros (1.4), conforme estrutura dos experimentos, devido à interpretação mais simples como o logaritmo da razão de chances. Assim, o MLG ajustado é

$$\begin{aligned} E(R) &= \pi \\ \eta &= \log\left(\frac{\pi}{1-\pi}\right) = X\beta \end{aligned} \quad (1.18)$$

em que

R é o vetor de proporções, de dimensões $n \times 1$, tal que se admite

$$R_i \sim \frac{Bin(m_i, \pi_i)}{m_i}, \text{ com } i = 1, \dots, n;$$

η é o vetor dos preditores lineares, de dimensões $n \times 1$;

X é a matriz do delineamento, de dimensões $n \times p$;

β é o vetor de p parâmetros desconhecidos do preditor linear do modelo, de dimensões $p \times 1$.

Para efetuar o ajuste dos modelos generalizados propostos para os conjuntos de dados descritos no item 3.1, foram utilizados o procedimento

GENMOD do programa SAS 8.10, o comando GLM do programa R 1.6.1 e um algoritmo MLG implementado no programa R 1.6.1, conforme apresentados no Anexo B1 para os dados do experimento de germinação de sementes, para o modelo contendo efeitos principais sem interação. O algoritmo implementado no R utiliza o procedimento de quadrados mínimos ponderados iterativamente para obtenção das estimativas dos parâmetros com uso do algoritmo de Escore de Fisher.

3.2.1.1 Exemplo 1: Ensaio de germinação

As proporções de sementes germinadas (r_{ijk}) em bandejas com m_{ijk} sementes foram modeladas admitindo-se dois possíveis modelos, “M1” com efeitos principais e “M2” incluindo o efeito da interação:

$$M1: \eta_{ijk} = \log\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \alpha_0 + \alpha_1 x_{1(ijk)} + \alpha_2 x_{2(ijk)} \quad (1.19)$$

$$M2: \eta_{ijk} = \log\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \alpha_0 + \alpha_1 x_{1(ijk)} + \alpha_2 x_{2(ijk)} + \alpha_{12} x_{1(ijk)} x_{2(ijk)}$$

em que

η_{ijk} = o preditor linear correspondente à observação r_{ijk} , $k = 1, \dots, n_{ij}$;

$x_{1(ijk)} = 1$ se a variedade de semente i foi *O. aegyptiaco* 73, 0 se for *O. aegyptiaco* 75;

$x_{2(ijk)} = 1$ se o extrato de raiz j foi pepino, 0 se for feijão;

α = representa os efeitos fixos associados com variedade de semente, extrato de raiz e interação.

3.2.1.2 Exemplo 2: Ensaio de sobrevivência

As proporções de ovos sobreviventes (r_{ij}), obtidas a partir da contagem em caixas com m_{ij} ovos, foram analisados conforme MLG descrito por

$$\eta_{ij} = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mu + \alpha_i + \beta_j, \quad (1.20)$$

em que

η_{ij} = o preditor linear correspondente à observação r_{ij} , $i = 1, \dots, 5$ e $j = 1, \dots, 4$;

μ = constante associada a todas as observações;

α_i = o efeito do local i ;

β_j = o efeito do período j .

3.2.2 Verificação de ajuste

Para verificação do ajuste dos modelos descritos no item 3.1.1, utilizou-se a “deviance” como medida de ajustamento conforme descrito no item 2.1.4. A “deviance” para proporções é demonstrada a partir de (1.7).

Seja r_1, r_2, \dots, r_n um conjunto de realizações independentes de variáveis aleatórias $R_i \sim \frac{Bin(m_i, \pi_i)}{m_i}$, com função de densidade discreta dada por (1.9).

Então, a logverossimilhança conjunta de R_i calculada para os parâmetros estimados para o qual θ_i é expresso como uma função de $\hat{\pi}_i$ (modelo corrente) é definida como

$$l(\hat{\pi}, \mathbf{r}) = \sum_{i=1}^n m_i \left[r_i \log \left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i} \right) + \log(1-\hat{\pi}_i) \right] + \log \left(\frac{m_i}{m_i r_i} \right) \quad (1.21)$$

Assim, a “deviance” é derivada por

$$D(\mathbf{r}, \hat{\pi}) = 2[l(\mathbf{r}, \mathbf{r}) - l(\hat{\pi}, \mathbf{r})]$$

$$D(\mathbf{r}, \hat{\pi}) = \sum_{i=1}^n m_i \left[r_i \log \left(\frac{r_i}{1-r_i} \right) + \log(1-r_i) \right] + \log \left(\frac{m_i}{m_i r_i} \right) - \left\{ \sum_{i=1}^n m_i \left[r_i \log \left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i} \right) + \log(1-\hat{\pi}_i) \right] + \log \left(\frac{m_i}{m_i r_i} \right) \right\}$$

ou ainda

$$D(\mathbf{r}, \hat{\pi}) = \sum_{i=1}^n 2m_i \left\{ r_i \left[\log \left(\frac{r_i}{1-r_i} \right) - \log \left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i} \right) \right] + \log(1-r_i) - \log(1-\hat{\pi}_i) \right\},$$

o que resulta na forma mais simplificada

$$D(\mathbf{r}, \hat{\pi}) = \sum_{i=1}^n 2m_i \left[r_i \log \left(\frac{r_i}{\hat{\pi}_i} \right) + (1-r_i) \log \left(\frac{1-r_i}{1-\hat{\pi}_i} \right) \right]. \quad (1.22)$$

Visou-se, com isto, encontrar o modelo mais simples e que explicasse de maneira satisfatória os dados, facilitando as interpretações (princípio da parcimônia). Assim, realizou-se o teste da adequação do modelo corrente com p parâmetros linearmente independentes por meio da estatística $\chi^2_{(n-p; \alpha)}$; se a “deviance” binomial residual (1.22), equivalente a Sp (1.8) ($\phi = 1$), fosse inferior a este quantil superior, não haveria evidências para supor que o acréscimo de parâmetros no modelo forneceria ganhos no ajuste; caso contrário, o modelo

proposto não comportaria a variação presente nos dados, a qual em muitos casos se deve à variação extra-binomial presente, como descrita no item 2.3.

A adequação do modelo linear generalizado binomial foi complementarmente verificada pela estimação do parâmetro de escala (ϕ); dentre vários estimadores encontrados na literatura, utilizou-se o estimador dado pela razão da estatística X^2 generalizada de Pearson, dada por

$$X^2 = \sum_{i=1}^n \frac{(r_i - \hat{\pi}_i)^2 w_i}{V(\hat{\pi}_i)} = (\mathbf{z} - \hat{\boldsymbol{\eta}})' \mathbf{W} (\mathbf{z} - \hat{\boldsymbol{\eta}}), \quad (1.23)$$

com graus de liberdade ($n - p$), resultando em

$$\hat{\phi} = \frac{X^2}{n - p}, \quad (1.24)$$

sendo \mathbf{z} a variável dependente ajustada, \mathbf{W} a matriz de pesos, $\boldsymbol{\eta}_i$ o preditor linear e p o posto da matriz do modelo corrente como definidos no item 2.1.4.

A estimativa (1.24) foi utilizada para checagem de superdispersão ou subdispersão, respectivamente indicadas por valores calculados significativamente maiores ou menores que 1.

3.2.3 Modelo linear generalizado binomial misto

Evidenciada a presença de variação extrabinomial, de acordo com 3.2.2, efetuou-se a modelagem dos dados correspondentes pelo uso da abordagem que utiliza a incorporação de efeitos aleatórios no preditor linear, como mostrado no item 2.4, ou seja, por um modelo linear generalizado misto.

Seja r_1, r_2, \dots, r_n um conjunto de realizações condicionalmente independentes de variáveis aleatórias $R_i | u_i \sim \frac{Bin(m_i, \pi_i)}{m_i}$, com função de densidade discreta dada por

$$R_i | u_i \sim iid. f_{R_i|u_i}(r_i | u_i)$$

$$f_{R_i|u_i}(r_i | u_i) = \exp \left\{ m_i \left[r_i \log \left(\frac{\pi_i}{1-\pi_i} \right) + \log(1-\pi_i) \right] + \log \left(\frac{m_i}{m_i r_i} \right) \right\}. \quad (1.25)$$

O MLGM para análise destas proporções condicionais foi construído utilizando uma função de ligação logística, resultando no seguinte modelo proposto:

$$\pi = E(\mathbf{R} | \mathbf{u})$$

$$\log \left(\frac{\pi}{1-\pi} \right) = \eta = X\beta + Z\mathbf{u}, \quad (1.26)$$

em que

$\mathbf{R}|\mathbf{u}$ é o vetor de proporções condicionalmente independentes, de dimensões $n \times 1$, tal que se admite $R_i | u_i \sim \frac{Bin(m_i, \pi_i)}{m_i}$ (1.25);

η é o vetor dos preditores lineares, de dimensões $n \times 1$;

X é a matriz do delineamento referente aos efeitos fixos, de dimensões $n \times p$;

Z é a matriz do delineamento referente aos efeitos aleatórios, de dimensões $n \times q$;

β é o vetor de efeitos fixos desconhecidos do preditor linear do modelo, de dimensões $p \times 1$;

\mathbf{u} é o vetor de efeitos aleatórios desconhecidos do preditor linear do modelo, de dimensões $q \times 1$, assumindo-se $\mathbf{u} \sim N_q(\mathbf{0}, G)$;

Assim, com base no item 2.4 e tendo-se admitido o modelo (1.26), submetem-se os dados binomiais a uma análise pela teoria dos MLGM, conforme descrito nos trabalhos de Gilmour et al. (1985), Shall (1991) e Breslow & Clayton (1993), assumindo-se os efeitos aleatórios normalmente distribuídos.

A quase-logverossimilhança condicional para dados binomiais é dada por

$$Q(\pi_i; r_i | u_i) = m_i \left[r_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + \log(1 - \pi_i) \right]. \quad (1.27)$$

Gilmour et al. (1985) e Shall (1991) utilizaram algoritmos de maximização conjunta para obter os estimadores dos vetores de parâmetros β e \mathbf{u} . Estimadores de máxima quase-logverossimilhança, tendo estimativas bem próximas às de máxima verossimilhança, podem ser obtidos com base nestes algoritmos citados, por meio da maximização da quase-logverossimilhança conjunta expressa por

$$Q(\pi, \mathbf{u}; \mathbf{r}) \propto \sum_{i=1}^n m_i \left[r_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + \log(1 - \pi_i) \right] - \frac{1}{2} (\log |G| + \mathbf{u}' G^{-1} \mathbf{u}) \quad (1.28)$$

O algoritmo sugerido por Breslow & Clayton (1993) obtém estimadores para os vetores de parâmetros β e \mathbf{u} a partir de diferenciações de uma quase-logverossimilhança penalizada (QVP), de (1.17), dada pela soma da quase-logverossimilhança condicional (1.27) com uma função pênalti, a qual, para o modelo (1.26), é

$$QVP(\pi, \mathbf{u}; \mathbf{r}) = \sum_i m_i \left[r_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + \log(1 - \pi_i) \right] - \frac{1}{2} \mathbf{u}' \mathbf{G}^{-1} \mathbf{u}. \quad (1.29)$$

Em ambos os algoritmos, as diferenciações realizadas resultam num sistema de equações pouco tratável; porém, com base na variável dependente ajustada dada por

$$\mathbf{y}^* = \boldsymbol{\eta} + \Delta(\mathbf{r} - \boldsymbol{\pi}),$$

tem-se que esta pode ser aproximada por um modelo linear misto (Henderson et al., 1959; Harville, 1977) da forma

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}^*,$$

em que

$$\begin{aligned} \mathbf{e}^* &= \Delta(\mathbf{r} - \boldsymbol{\pi}); \text{Cov}(\mathbf{e}^*) = \mathbf{W}^{-1}; \text{Cov}(\mathbf{u}) = \mathbf{G}; \\ \mathbf{E}(\mathbf{Y}^*) &= \mathbf{X}\boldsymbol{\beta}; \text{Cov}(\mathbf{Y}^*) = \mathbf{W}^{-1} + \mathbf{Z}\mathbf{G}\mathbf{Z}'. \end{aligned}$$

Logo, aqueles autores sugerem o uso de algoritmo iterativo similar, em que um modelo generalizado misto é ajustado para encontrar soluções para estimativas dos vetores de parâmetros β e \mathbf{u} por

Os MLGM's foram ajustados, com estimativas dos parâmetros obtidas via maximização da função de quase-verossimilhança penalizada conjunta (1.29), por meio da macro GLIMMIX do programa SAS 8.10, por algoritmo implementado no programa R 1.6.1 com base no algoritmo descrito em Shall (1991) e Breslow & Clayton (1993) e pelo comando glimmPQL do programa R 1.6.1, presente na biblioteca MASS, conforme apresentados no Anexo B2 para o ensaio de germinação de sementes para o modelo de efeitos principais.

Os dois primeiros algoritmos citados obtêm as estimativas dos parâmetros por meio da resolução do sistema de equações dos MLGM's (1.30),

$$W = \text{diag} \left\{ \frac{\partial \eta_i}{\partial \pi_i} \right\} V(\gamma_i | \eta_i) \left[\frac{\partial \eta_i}{\partial \pi_i} \right]^{-1} = \text{diag} \left\{ m_i \left[\frac{x_i'(1-x_i)}{x_i} \right]_2 \right\} = \text{diag} \{ m_i x_i'(1-x_i) \} \cdot$$

$$\Delta = \text{diag} \left\{ \frac{\partial \eta_i}{\partial \pi_i} \right\} = \text{diag} \left\{ \frac{\partial}{\partial \pi_i} \left[\log \left(\frac{x_i}{1-x_i} \right) \right] \right\} = \text{diag} \left\{ \frac{x_i'(1-x_i)}{1} \right\}$$

tem-se que

$$\eta_i = \log \left(\frac{x_i}{1-x_i} \right) = \beta' x_i + \alpha' z_i + \gamma_i$$

em que, para um MLGM para dados binomiais,

$$\begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} X'W X & X'W Z \\ Z'W X & Z'W Z + G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'W y^* \\ Z'W y^* \end{bmatrix} \quad (1.30)$$

sendo que o algoritmo implementado no R 1.6.1 utiliza o algoritmo de Escore de Fisher no modelo generalizado e algoritmo EM para os componentes de dispersão referentes aos efeitos aleatórios, enquanto as estimativas obtidas pelo comando `glmmPQL` do programa R 1.6.1 são derivadas de (1.29) usando o procedimento de quadrados mínimos reponderados iterativamente.

3.2.3.1 Exemplo 1: Ensaio de germinação

As proporções condicionalmente independentes de sementes germinadas ($r_{ijk}|b_k$) em bandejas com m_{ijk} sementes foram modeladas admitindo-se dois possíveis modelos, “M1” com efeitos principais e “M2” incluindo o efeito da interação:

$$\begin{aligned} M1: \quad \text{logit}(\text{Pr} = 1 | x_{ijk}, b_k) &= \alpha_0 + \alpha_1 x_{1(ijk)} + \alpha_2 x_{2(ijk)} + b_k \\ M2: \quad \text{logit}(\text{Pr} = 1 | x_{ijk}, b_k) &= \alpha_0 + \alpha_1 x_{1(ijk)} + \alpha_2 x_{2(ijk)} + \alpha_{12} x_{1(ijk)} x_{2(ijk)} + b_k \end{aligned} \quad (1.31)$$

em que

b_k = o efeito aleatório associado à parcela k , considerando $b_k \sim N(0, \sigma_p^2)$.

3.2.3.2 Exemplo 2: Ensaio de sobrevivência

As proporções condicionalmente independentes de ovos sobreviventes no local i e período j ($r_{ij}|b_{ij}$) em caixas com m_{ij} ovos foram modeladas admitindo-se um MLGM dado por

$$\eta_{ij} = \log(\text{Pr} = 1 | \alpha_i, \beta_j, b_{ij}) = \mu + \alpha_i + \beta_j + b_{ij}, \quad (1.32)$$

em que

b_{ij} = o efeito aleatório da parcela, considerando $b_{ij} \sim N(0, \sigma_p^2)$.

3.2.4 Verificação de ajuste

Verificou-se o ajuste do MLGM proposto por meio do cálculo da estimativa do componente extra de dispersão ($\hat{\sigma}^2$) conforme descrito por Shall (1991). O teste para a superdispersão ou subdispersão presente foi realizado verificando se os valores de $\hat{\sigma}^2$ eram significativamente maiores ou menores que 1, respectivamente (Shall, 1991) O estimador sugerido foi calculado de forma análoga a (1.24) e aproximado pelo algoritmo, conforme apresentado no Anexo B2 por

$$\hat{\sigma}^2 = \frac{\left(\mathbf{z} - X\hat{\beta} - Z\hat{\mathbf{u}}\right)'W\left(\mathbf{z} - X\hat{\beta} - Z\hat{\mathbf{u}}\right)}{n - r(X)}, \quad (1.33)$$

em que $r(X)$ é o posto da matriz de efeitos fixos.

4 RESULTADOS E DISCUSSÃO

A metodologia descrita no item 3.2 foi aplicada aos conjuntos de dados apresentados em 3.1 conforme estruturas de delineamento e de tratamento anteriormente descritas. Os resultados e discussões para os exemplos utilizados estão apresentados a seguir, tomando cada exemplo individualmente.

4.1 Exemplo 1: Ensaio de germinação

Primeiramente foram ajustados os modelos de regressão logística ordinários para estes dados. A Tabela 4 apresenta as estimativas de máxima verossimilhança dos coeficientes de regressão logística para os modelos ajustados, de (1.19), para as 21 proporções de sementes germinadas por meio de programas executados nos pacotes estatísticos SAS 8.10 e R 1.6.1. As estimativas dos parâmetros obtidas foram idênticas para as três análises do MLG proposto.

Verifica-se que para o modelo de efeitos principais “M1”, de (1.19), a estimativa de ϕ , de (1.24), foi de 2,1284, excedendo o valor assumido para o modelo ($\phi=1$), evidenciando presença de variação extra não comportada pelo modelo sugerido. O valor da “deviance” residual, de (1.22), para o modelo “M1” com 18 graus de liberdade, foi de 39,6859, tendo ($Pr > D = 0,0023$), rejeitando-se a hipótese de que o modelo está bem ajustado e somando forte evidência de variação extra não modelada ou má especificação do modelo.

A partir da evidência do efeito da interação dos fatores presente, conforme mostrado na Figura 3, ajustou-se o modelo “M2”, de (1.19), incluindo o efeito da interação da variedade de semente e tipo de raiz. Para este modelo, a “deviance” residual, de (1.22), com 17 graus de liberdade, foi de 33,2778, a qual

TABELA 4 Estimativas dos parâmetros para os modelos de regressão logística descritos em (1.19) por três algoritmos para MLG.

Variável	ALGORITMOS		
	GENMOD – SAS	GLM – R	ALGORITMO – MLG
	8.10		
	$\hat{\beta}(EP)$	$\hat{\beta}(EP)$	$\hat{\beta}(EP)$
Modelo com efeitos principais			
Intercepto	-0,4300(0,1137)	-0,4300(0,1137)	-0,4300(0,1137)
Variedade	-0,2705(0,1547)	-0,2705(0,1547)	-0,2705(0,1547)
Extrato raiz	1,0647(0,1442)	1,0647(0,1442)	1,0647(0,1442)
$\hat{\phi}$	2,1284	2,1284	2,1284
Modelo com interação			
Intercepto	-0,5582(0,1260)	-0,5582(0,1260)	-0,5582(0,1260)
Variedade	0,1459(0,2232)	0,1459(0,2232)	0,1459(0,2232)
Extrato raiz	1,3182(0,1775)	1,3182(0,1775)	1,3182(0,1775)
Interação	-0,7781(0,3064)	-0,7781(0,3064)	-0,7781(0,3064)
$\hat{\phi}$	1,8618	1,8618	1,8618

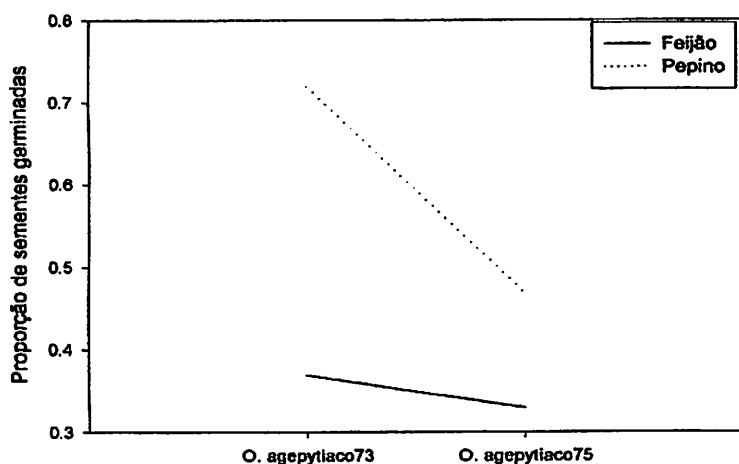


FIGURA 3 Representação gráfica da proporção média de sementes germinadas em função das variedades *O. aegyptiaco* 73 e 75 e dos extratos de raízes de feijão e pepino.

superou o valor do $\chi^2_{(17;0,05)} = 27,5871$, rejeitando-se a hipótese de ajustamento do modelo. A estimativa de ϕ , de (1.24), para o modelo sugerido foi igual a 1,8618, excedendo o valor assumido para o modelo ($\phi=1$), evidenciando-se presença de superdispersão.

A Figura 4 apresenta as proporções observadas e correspondentes valores ajustados pelos modelos “M1” e “M2”, de (1.19), obtidos pela inversa da função de ligação logística correspondentes, respectivamente:

$$M1: \hat{\pi}_{ijk} = \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{1(ijk)} + \hat{\alpha}_2 x_{2(ijk)})}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{1(ijk)} + \hat{\alpha}_2 x_{2(ijk)})}$$

$$M2: \hat{\pi}_{ijk} = \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{1(ijk)} + \hat{\alpha}_2 x_{2(ijk)} + \hat{\alpha}_{12} x_{1(ijk)} x_{2(ijk)})}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{1(ijk)} + \hat{\alpha}_2 x_{2(ijk)} + \hat{\alpha}_{12} x_{1(ijk)} x_{2(ijk)})}$$

As proporções ajustadas pelo modelo “M2”, de (1.19), apresentam valores médios ajustados mais conformáveis com os valores observados no experimento, conforme Figura 4.

Crowder (1978), citado por Breslow & Calyton (1993), observou que existe uma variação dentro de parcelas que excede à predita pelos modelos acima ajustados, caracterizando o fenômeno da superdispersão. A Figura 1 mostra a plotagem das proporções com as parcelas demonstrando uma variabilidade bem pronunciada entre as parcelas dentro de cada tratamento.

Vários autores, analisando os dados de Crowder (1978), têm sugerido formas de acomodar a superdispersão presente. Crowder (1978), citado por Willians (1982), assume o modelo beta-binomial e estimação de máxima verossimilhança. Breslow & Clayton (1993) buscaram modelar a variação extra-binomial por um modelo linear generalizado misto (MLGM) via quase-verossimilhança penalizada, ou seja, por meio da incorporação de efeito

aleatório no preditor linear (efeito de parcela) considerando o componente de extra dispersão igual a 1. Hinde & Demétrio (1998) utilizaram vários modelos com funções de ligação logística e complemento log-log baseados na inclusão de parâmetro de superdispersão e métodos de quase-verossimilhança (modelos com função de variância mais geral).

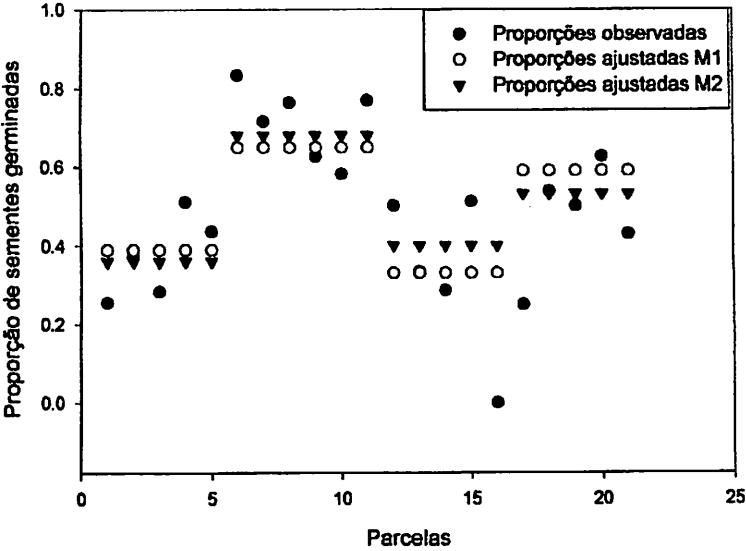


FIGURA 4 Representação gráfica das proporções de sementes germinadas observadas e ajustadas pelos modelos binomiais padrão de efeitos principais “M1” e com interação “M2”.

Com base na evidência anunciada, partiu-se para o ajuste de modelos com inclusão do efeito aleatório de parcela no preditor linear (1.31) como forma de modelar a variação extra-binomial por um MLGM. Devido à dificuldade na

convergência do algoritmo, optou-se por re-expressar os ensaios binomiais na forma de ensaios de Bernoulli (1's e 0's) (Littell et al., 1996).

A Tabela 5 apresenta as estimativas de máxima quase-verossimilhança penalizada dos coeficientes de regressão logística obtidas pela macro Glimmix-SAS 8.10, pelo algoritmo MLGM implementado na linguagem R 1.6.1 e pelo comando glmmPQL do R 1.6.1 para os efeitos fixos dos modelos generalizados mistos com ausência e presença da interação dos fatores (1.31) e as estimativas dos componentes de dispersão referentes ao parâmetro de extra dispersão e ao efeito aleatório da parcela.

TABELA 5 Estimativas dos parâmetros para os MLGM's descritos em (1.31) por três algoritmos para MLGM.

Variável	ALGORITMOS		
	Glimmix – SAS 8.10	glmmPQL – R	ALGORITMO - MLGM
	$\hat{\beta}(EP)$	$\hat{\beta}(EP)$	$\hat{\beta}(EP)$
Modelo com efeitos principais			
Intercepto	-0,3754(0,1821)	-0,3841(0,1643)	-0,3754(0,1821)
Variedade	-0,3634(0,2284)	-0,3427(0,2085)	-0,3634(0,2284)
Extrato raiz	1,0116(0,2236)	1,0175(0,2030)	1,0116(0,2236)
$\hat{\sigma}_p^2$	0,1242	0,0857	0,1242
$\hat{\sigma}^2$	0,9805	0,9871	0,9805
Modelo com interação			
Intercepto	-0,5422(0,1904)	-0,5441(0,1656)	-0,5422(0,1904)
Variedade	0,07679(0,3081)	0,0970(0,2746)	0,07679(0,3081)
Extrato raiz	1,3385(0,2699)	1,3268(0,2347)	1,3385(0,2699)
Interação	-0,8251(0,4296)	-0,8053(0,3817)	-0,8251(0,4296)
$\hat{\sigma}_p^2$	0,09780	0,0551	0,09780
$\hat{\sigma}^2$	0,9855	0,9899	0,9855

Observa-se que as estimativas obtidas pela macro Glimmix do SAS 8.10 e algoritmo MLGM apresentam os mesmos resultados para ambos os modelos, com e sem interação, denotando uma boa caracterização do algoritmo para fins

de ajuste de um MLGM via quase-verossimilhança penalizada conjunta, tendo assumido pressuposição distribucional de normalidade para o efeito aleatório presente.

Os resultados obtidos pelo comando `glmmPQL` do software R 1.6.1 apresentaram resultados similares com relação aos dois outros procedimentos de análise, por utilizar uma implementação diferente a partir da QVP (1.29). Porém, as estimativas obtidas pelo `glmmPQL` foram bem próximas das estimativas obtidas por Breslow & Clayton (1993) e Hinde & Demétrio (1998) via máxima verossimilhança, demonstrando uma robustez dos estimadores de QVP. Salienta-se que a especificação do modelo de estrutura fatorial dos tratamentos na linguagem adotada no R 1.6.1 segue a notação de Wilkinson & Rogers (1973), o que complica a interpretação. Observa-se, ainda, que os erros padrão das estimativas obtidas pelo `glmmPQL` do R 1.6.1 foram menores do que os obtidos pelo `glimmix` – SAS 8.10 e pelo algoritmo MLGM implementado, denotando maior precisão destes estimadores.

Para todos os algoritmos de análise foram calculadas as estimativas do parâmetro referente ao componente de extra dispersão (σ^2), de (1.33), para fins de verificação da acomodação da superdispersão (Shall, 1991), e a estimativa do parâmetro relacionado com a variância do efeito aleatório de parcela (σ_p^2).

Para o modelo “M1”, de (1.31), as estimativas alcançadas para o algoritmo MLGM e SAS 8.10 foram iguais a 0,9805 e 0,1242, enquanto as obtidas no R foram de 0,9871 e 0,0857, respectivamente, para os componentes de extra dispersão e componentes de variância da parcela. Como os valores dos componentes de extra dispersão estão bem próximos a 1, valor admitido e fixo para este no modelo binomial, indicam que a variância está consistente com a distribuição assumida, isto é, os dados não fornecem evidência de superdispersão para o modelo assumido.

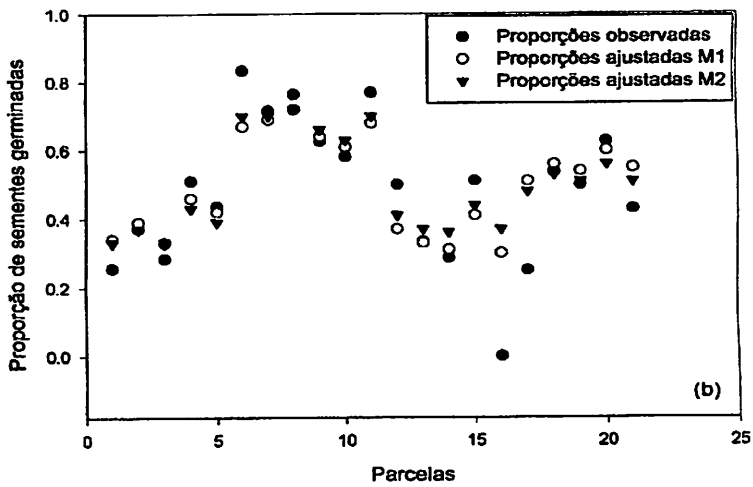
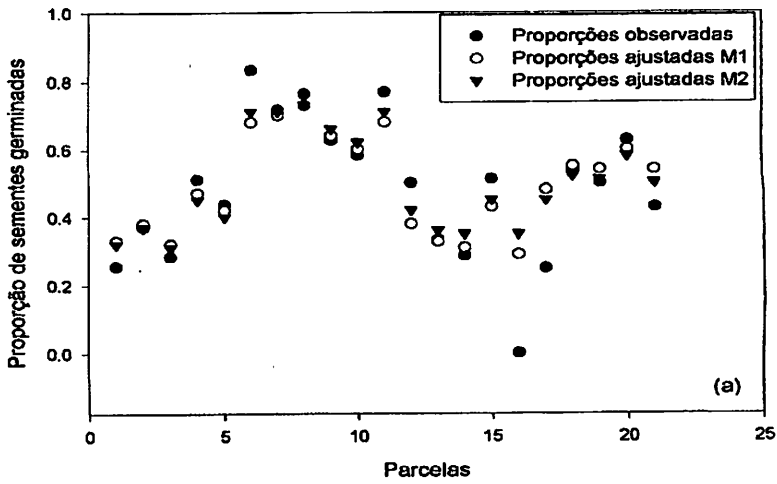


FIGURA 5 Proporções de germinação ajustadas pelos modelos mistos binomiais de efeitos principais “M1” e com interação “M2” por meio do glimmix SAS 8.10 e algoritmo MLGM (a) e pelo glmmPQL R (b).

Para o modelo “M2”, de (1.31), considerando o efeito de interação, as estimativas alcançadas para os componentes de extra dispersão e de variância da parcela pelo algoritmo MLGM e Glimmix-SAS 8.10 foram similares e iguais a 0,9855 e 0,0978, enquanto as obtidas no R foram de 0,9899 e 0,0551, respectivamente. A inclusão do efeito da interação, a qual foi significativa no modelo, aproximou a estimativa de σ^2 a 1, evidenciando melhor acomodação da variação extrabinomial.

A Figura 5 apresenta as proporções observadas e correspondentes aos valores ajustados pelos modelos “M1” e “M2”, de (1.31), calculados a partir da inversa da função de ligação logística por meio das estimativas obtidas pelo glimmix SAS 8.10 e algoritmo MLGM (a) e pelo comando glmpQL do programa R 1.6.1, conforme apresentado abaixo:

$$M1: \hat{\pi}_{ijk} = \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{1(ijk)} + \hat{\alpha}_2 x_{2(ijk)} + \hat{b}_k)}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{1(ijk)} + \hat{\alpha}_2 x_{2(ijk)} + \hat{b}_k)}$$

$$M2: \hat{\pi}_{ijk} = \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{1(ijk)} + \hat{\alpha}_2 x_{2(ijk)} + \hat{\alpha}_{12} x_{1(ijk)} x_{2(ijk)} + \hat{b}_k)}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{1(ijk)} + \hat{\alpha}_2 x_{2(ijk)} + \hat{\alpha}_{12} x_{1(ijk)} x_{2(ijk)} + \hat{b}_k)}$$

A Figura 5(a) mostra que os valores das proporções ajustadas pelo modelo com efeito da interação ajusta-se suavemente de forma mais adequada aos dados observados, o mesmo ocorrendo na Figura 5(b), devido ao fato de o efeito da interação no experimento mostrar-se apropriado no modelo.

4.2 Exemplo 2: Ensaio de Sobrevivência

A Tabela 6 apresenta algumas de muitas soluções possíveis para as estimativas dos parâmetros para o modelo binomial via máxima

verossimilhança, de (1.20), para as 20 proporções de ovos sobreviventes em função das variáveis classificatórias local e período, por meio de programas executados nos pacotes estatísticos SAS 8.10 e R 1.6.1.

Verifica-se que a estimativa de ϕ , obtida de (1.24), para as análises realizadas, foi de 5,3303, excedendo o valor assumido para o modelo ($\phi=1$), evidenciando presença de superdispersão. O valor da “deviance” residual, de (1.22), para o modelo com 12 graus de liberdade, foi de 64,495, sendo superior ao quantil do qui-quadrado com 12 graus de liberdade para o nível de significância adotado ($\alpha = 0,05$), rejeitado-se a hipótese de ajustamento do modelo.

TABELA 6 Estimativas dos parâmetros para o modelo binomial padrão (1.20) por três algoritmos para MLG.

Variável	ALGORITMOS		
	GENMOD – SAS 8.10	GLM – R	ALGORITMO – MLG
	$\hat{\beta}(EP)$	$\hat{\beta}(EP)$	$\hat{\beta}(EP)$
Intercepto	-2,4279(0,1919)	4,6358(0,2810)	1,0032(0,0502)
Local 01	4,6138(0,2502)	0	1,6454(0,1630)
Local 02	4,1970(0,2317)	-0,4168(0,2461)	1,2286(1,1445)
Local 03	3,3717(0,2014)	-1,2421(0,2194)	0,4033(0,1119)
Local 04	3,6629(0,2131)	-0,9509(0,2287)	0,6945(0,1244)
Local 05	0	-4,6138(0,2500)	-2,9685(0,1439)
Período (4 sem)	2,4500(0,2341)	0	1,9872(0,1684)
Período (7 sem)	0,2798(0,1640)	-2,1702(0,2381)	-0,1829(0,1118)
Período (8 sem)	0,1243(0,1648)	-2,3256(0,2456)	-0,3384(0,1144)
Período (11 sem)	0	-2,4499(0,2338)	-0,4627(0,1029)
$\hat{\phi}$	5,3303	5,3303	5,3303

A Figura 6 mostra as proporções de ovos de truta observados e ajustados pelo modelo (1.20), obtidos pela inversa da função ligadora denotada por

$$\hat{\pi}_{ij} = \frac{\exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)}{1 + \exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)},$$

para as 20 unidades experimentais. Denota-se que o modelo ajustado não descreve de forma satisfatória os dados deste experimento.

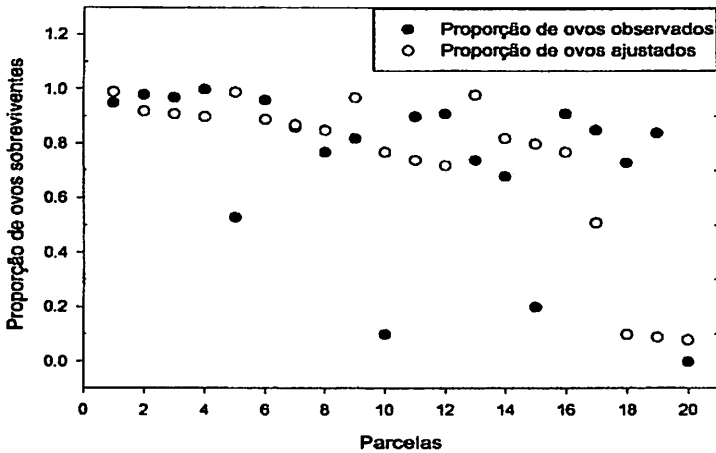


FIGURA 6 Proporções de ovos de truta sobreviventes observados e ajustados pelo modelo (1.20) para as 20 parcelas.

Em virtude da inadequação do modelo proposto, de (1.20), em explicar os dados, propôs-se, analogamente, modelar esta variabilidade por meio de um MLG com inclusão do efeito aleatório da parcela experimental, conforme (1.32).

Com base neste modelo (1.32), as estimativas de máxima quase-verossimilhança dos parâmetros para os efeitos fixos de local e período resultaram em iguais proporções ajustadas para as análises realizadas pela macro glimmix do SAS 8.10 e algoritmo MLGM e estas foram próximas às obtidas por

meio das estimativas provenientes pelo comando glmmPQL do programa R 1.6.1, conforme mostrado na Tabela 7.

Estimaram-se, ainda, os componentes de dispersão ou variância relativos à parcela e à variação extra, de (1.33), iguais a 0,655 e 0,9097, respectivamente, para as análises realizadas no SAS 8.10 e algoritmo GLMM. A análise oriunda do glmmPQL do programa R 1.6.1 resultou em estimativas equivalentes a 0,2330 e 0,9392 para estas componentes. Nota-se que a estimativa da componente de variância para o efeito aleatório de parcela para as análises diferem substancialmente devido à diferente implementação utilizada a partir de (1.29), utilizada pelo comando glmmPQL.

TABELA 7 Estimativas dos parâmetros para os MLGM's descritos em (1.32) por três algoritmos para MLGM.

Variável	ALGORITMOS		
	Glimmix – SAS 8.10	GLM – R	ALGORITMO – MLG
	$\hat{\beta}(EP)$	$\hat{\beta}(EP)$	$\hat{\beta}(EP)$
Intercepto	-2,8666(0.5877)	4,3625(0,4367)	0,9593(0,1392)
Local 01	4,6104(0.6401)	0	1,4586(0,4065)
Local 02	4,4150(0.6404)	-0,2536(0,4375)	1,2632(0,4058)
Local 03	3,6052(0.6230)	-1,0710(0,4190)	0,4534(0,3888)
Local 04	4,0877(0.6433)	-0,6846(0,4303)	0,9360(0,4081)
Local 05	0	-4,5426(0,4321)	-3,1518(0,4043)
Período (4 sem)	2,5860(0.5996)	0	1,9112(0,3827)
Período (7 sem)	0,5889(0.5534)	-2,0049(0,4109)	-0,0851(0,3410)
Período (8 sem)	0,4805(0.5502)	-2,1060(0,4080)	-0,1936(0,3378)
Período (11 sem)	0	-2,4710(0,4131)	-0,6740(0,3460)
$\hat{\sigma}_p^2$	0,6550	0,2330	0,6550
$\hat{\sigma}^2$	0,9097	0,9392	0,9097

Com base no critério sugerido por Shall (1991), aproximado por (1.33), tem-se que o valor estimado para a componente de extra dispersão foi,

relativamente, próximo a 1, o que nos permite aferir que o modelo misto comporta de forma satisfatória a variação extra-binomial presente.

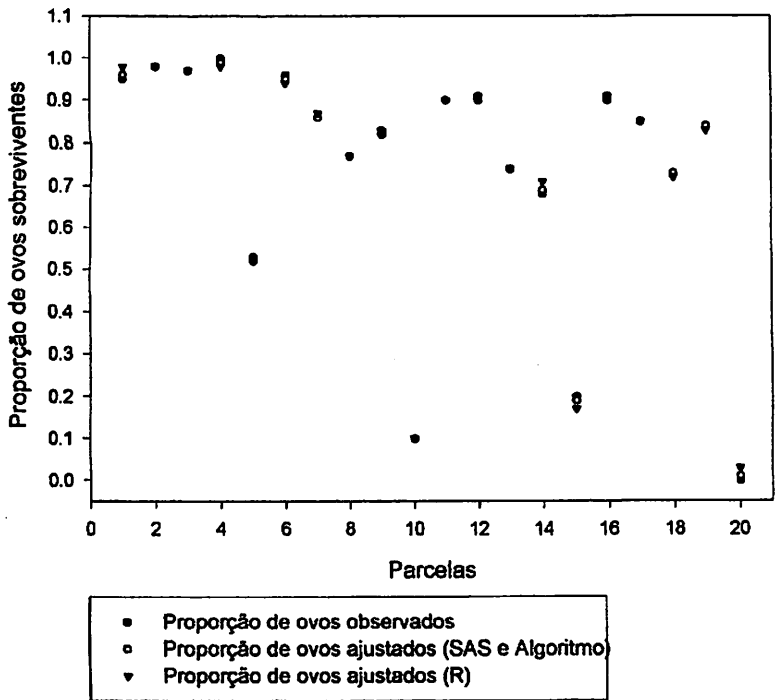


FIGURA 7 Proporções de ovos de truta sobreviventes observados e ajustados pelo modelo (1.32) para as 20 parcelas.

A Figura 7 apresenta a plotagem das proporções de ovos de truta sobreviventes observados e ajustados pelo modelo (1.32), obtidos analogamente pela inversa da função de ligação

$$\hat{\pi}_{ij} = \frac{\exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{b}_{ij})}{1 + \exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{b}_{ij})},$$

com substituição pelas estimativas dos parâmetros correspondentes ao modelo. Nota-se que o modelo misto proposto explica de forma bastante satisfatória a variabilidade inerente aos dados experimentais.

5 CONCLUSÕES

- A aplicação dos modelos lineares generalizados mistos para fins de acomodação da variação extrabinomial presente nos dados referentes aos ensaios de germinação e sobrevivência constitui uma técnica bastante adequada;
- O algoritmo implementado em ambiente R 1.6.1 mostrou-se bastante consistente com softwares de renomada precisão, nos dois exemplos utilizados para o ajuste de modelos lineares generalizados mistos com ligação logística e pressuposição de normalidade para o efeito aleatório presente.

REFERÊNCIAS BIBLIOGRÁFICAS

BARTLETT, M. S. The use of transformations. **Biometrics**, Washington, v. 3, n. 1, p. 39-53, 1947.

BRESLOW, N. E.; CLAYTON, D. G. Approximate inference in generalized linear mixed models. **Journal of the American Statistical Association**, Alexandria, v. 88, n. 421, p. 9-25, Mar. 1993.

BROOKS, S. P.; MORGAN, B. J. T.; RIDOUT, M. S.; PACK, S. E. Finite mixture models for proportions. **Biometrics**, Washington, v. 53, n. 3, p.1097-1115, Sept. 1997.

CORDEIRO, G. M. Modelos lineares generalizados. VII SINAPE. Campinas – SP, 1986. 286 p.

CROWDER, M. J. Beta binomial ANOVA for proportions. **Journal of the Royal Statistical society Applied Statistics, Series C**, London, v.27, n. 1, p.34-37, 1978.

DEAN, C. B. Testing overdispersion in poisson and binomial regression models. **Journal of the American Statistical Association**, Alexandria, v. 87, n. 418, p. 451-457, June 1992.

DEMÉTRIO, C. G. B. Modelos Lineares Generalizados em Experimentação Agronômica. **Reunião Anual da RBRAS**, 46.; **SEAGRO**, 9., 2001, Piracicaba. Piracicaba: ESALQ/USP, 2001. 113 p.

DOBSON, A. J. **An introduction to generalized linear models**. London: Chapman and Hall. 1990. 173 p.

FITZMAURICE, G. M. Model selection with data overdispersed data. **The Statistician**, London, v. 46, n. 1, p.81-91, 1997.

GANIO, L. M.; SCHAFER, D. W. Diagnostics for overdispersion. **Journal of the American Statistical Association**, Alexandria, v. 87, n. 419, p. 795-804, Sept. 1992.

GILMOUR, A. R.; ANDERSON, R. D.; RAE, A. L. The analysis of binomial data by a generalized linear mixed model. **Biometrika**, London, v. 72, n. 3, p. 593-599, Dec. 1985.

GRAYBILL, F. A. **Theory and application of the linear model**. Belmont: Duxbury Press, 1976. 703 p.

HALL, D. B. Zero-inflated poisson and binomial regression with random effects: A case study. **Biometrics**, Washington, v.56, n. 4, p.1030-1039, Dec. 2000.

HARVILLE, D. A. Maximum likelihood approaches to variance component estimation and to related problems. **Journal of the American Statistical Association**, Alexandria, v. 72, n. 358, p. 320-339, June 1977.

HENDERSON, C. R.; KEMPTHORNE, O.; SEARLE, S. R.; KROSIGK, C. M. von. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, Washington, n. 1, p. 192-218, Mar. 1959.

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

HINES, R. J. O'HARA; LAWLESS, J. F. Modeling overdispersion in toxicological mortality data grouped over time. **Biometrics**, Washington, v. 49, n. 1, p. 107-121, Mar. 1993.

IM, S.; GIANOLA, D. Mixed models for binomial data with an application to lamb mortality. **Journal of the Royal Statistical society Applied Statistics, Series C**, London, v. 37, n. 2, p. 196-204, 1988.

IRAKA, R.; GENTLEMAN, R. R. A language for data analysis and graphics. **Journal of computation and graphical statistics**, Alexandria, v. 5, n. 3, p. 299-314, Sept. 1996.

LEE, Y.; NELDER, J. A. Two ways of modeling overdispersion in non-normal data. **Journal of the Royal Statistical society Applied Statistics, Series C**, London, v. 49, n. 4, p. 591-598, 2000.

LITTLE, R. C.; MILLEKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. **SAS[®]System for mixed models**. Cary: SAS Institute, 1996. 633 p.

MANLY, B. Regression models for proportions with extraneous variance. **Biometric-Praximetrie**, Genbloux, v. 18, n. 1, p. 1-18, 1978.

McCULLAGH, P.; NELDER, J. A. **Generalized linear model**. 2. ed. London: Chapman and Hall, 1989. 511 p.

McCULLOCH, C. E. Maximum likelihood algorithms for generalized linear mixed models. **Journal of the American Statistical Association**, Alexandria, v.92, n.437, p.162-170, Mar. 1997.

McCULLOCH, C. E.; SEARLE, S. R. **Generalized, linear, and mixed models**. New York: E. Willey-Interscience, 2001. 324 p.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3. ed. McGraw-Hill, 1974. 564 p.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear model. **Journal of the Royal Statistical Society A**, London, v. 135, n. 3, p. 370-384, 1972.

SHALL, R. Estimation in generalized linear models with random effects. **Biometrika**, London, v. 78, n. 4, p. 719-727, Dec. 1991.

VIEIRA, A. M. C. **Modelos para dados de proporções com superdispersão aplicados ao controle biológico**. 1998. 61 p. Dissertação (Mestrado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, São Paulo, SP.

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

WILKINSON, G. N.; ROGERS, C. E. Symbolic description of factorial models for the analysis of variance. **Journal of the Royal Statistical society Applied Statistics, Series C**, London, v. 22, n. 3, p. 392-399, 1973.

WILLIAMS, D. A. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. **Biometrics**, Washington, v. 31, n. 4, p. 949-952, Dec. 1975.

WILLIAMS, D. A. Extra-binomial variation in logistic linear models. **Journal of the Royal Statistical society Applied Statistics, Series C**, London, v. 31, n. 2, p.144-148, 1982.

ANEXOS

ANEXOS	PÁGINA
ANEXO A.....	58
A1 Exemplo 1: Ensaio de Germinação.....	58
A2 Exemplo 2: Ensaio de sobrevivência	59
 ANEXO B.....	 60
B1 Algoritmos para modelos lineares generalizados (MLG).....	60
B2 Algoritmos para modelos lineares generalizados mistos (MLGM) ..	62

ANEXO A Conjuntos de dados referentes aos ensaios de germinação e sobrevivência utilizados neste trabalho, com fins metodológicos de análise.

A1 Exemplo 1: Ensaio de Germinação

Dados de germinação de sementes de *O. aegyptiaco* (Crowder, 1978) em função de variedade de semente e tipo de raiz.

Variedade de sementes (x1)	Extrato de raiz (x2)	Sementes germinadas (y _i)	Sementes na bandeja (m _i)	Proporção de germinação (r _i)
0	0	10	39	0.26
0	0	23	62	0.37
0	0	23	81	0.28
0	0	26	51	0.51
0	0	17	39	0.44
0	1	5	6	0.83
0	1	53	74	0.72
0	1	55	72	0.76
0	1	32	51	0.63
0	1	46	79	0.58
0	1	10	13	0.77
1	0	8	16	0.50
1	0	10	30	0.33
1	0	8	28	0.29
1	0	23	45	0.51
1	0	0	4	0.00
1	1	3	12	0.25
1	1	22	41	0.54
1	1	15	30	0.50
1	1	32	51	0.63
1	1	3	7	0.43

A2 Exemplo 2: Ensaio de sobrevivência

Dados de sobrevivência de ovos de truta em função de local e período.

Parcela	Local	Período	Ovos sobreviventes (y_i)	Total de ovos (m_i)	Proporção de ovos sobreviventes (r_i)
1	1	4	89	94	0.95
2	1	7	94	98	0.96
3	1	8	77	86	0.90
4	1	11	141	155	0.91
5	2	4	106	108	0.98
6	2	7	91	106	0.86
7	2	8	87	96	0.91
8	2	11	104	122	0.85
9	3	4	119	123	0.97
10	3	7	100	130	0.77
11	3	8	88	119	0.74
12	3	11	91	125	0.73
13	4	4	104	104	1.00
14	4	7	80	97	0.82
15	4	8	67	99	0.68
16	4	11	111	132	0.84
17	5	4	49	93	0.53
18	5	7	11	113	0.10
19	5	8	18	88	0.20
20	5	11	0	138	0.00

ANEXO B Modelos de regressão logística para os dados do ensaio de germinação, considerando: modelo de efeitos principais M1, de (1.19), e o modelo de efeitos principais, reescritos na forma de 0's e 1's, incorporando efeito aleatório de parcela M1, de (1.31).

B1 Algoritmos para modelos lineares generalizados (MLG)

B1.1 Algoritmo para ajuste de MLG para dados binomiais usando ligação logit implementado para uso no software R 1.6.1.

```
library(MASS)
seeds <- read.table("Arquivo de dados ", header=T)

# Matriz de delineamento ( X )
xm <- matrix(1,21,1)
x1 <- seeds$x1
x2 <- seeds$x2
X <- cbind(xm, x1, x2)
n <- nrow(X)
p <- ncol(X)

# Vetor de observações ( y )
m <- seeds$n
r <- seeds$r
y <- r/m

b <- ginv(t(X)%*%X) %*% t(X) %*% y
b1 <- matrix(0,3,1)

k <- 0
while ( t(b1-b) %*% (b1-b) > 0.00000001 )
{
  k <- k + 1
}

## Vetor de estimativas dos valores ajustados ( mi )
b1 <- b
eta <- X %*% b
```

```

mi <- exp(eta)/(1+exp(eta))

## Vetor de variáveis dependentes ajustadas ( z )
z <- eta + (diag(c(1/(mi*(1-mi))),n,n)%*% (y-mi))
w <- diag(c(m*mi*(matrix(1,n,1)-mi)),n,n)

xwx <- t(X) %*% w %*% X
Mv <- ginv(xwx)
xwz <- t(X) %*% w %*% z

b <- Mv %*% xwz
}

## Vetor de estimativas dos parâmetros e de erros padrão ( sol )
se <- c(sqrt(diag(Mv)))
sol <- cbind(b, se)

## Estimativa do parâmetro de dispersão ( phi )
phi <- ( t(z-eta) %*% w %*% (z-eta))/(n-p)

```

B1.2 Comando glm do software R 1.6.1

```

library(MASS)
library(GLMMGibbs)
data(seeds)
a <- seeds$x1
b <- seeds$x2
r <- seeds$r
n <- seeds$n
p <- cbind(r, n - r)
seeds <- data.frame( a, b, p)
ajuste.mlg <- glm( p ~ a + b, family = binomial, data = seeds)
summary(ajuste.mlg)

```

B1.3 Procedimento GENMOD do SAS 8.10

```

proc genmod data = seeds;
model r/n = x1 x2 / dist = binomial link = logit;
run;

```

B2 Algoritmos para modelos lineares generalizados mistos (MLGM)

B2.1 Algoritmo para ajuste de MLGM para dados binomiais, via quase-verossimilhança penalizada (QVP), implementado para uso no software R 1.6.1.

```
library(MASS)
data <- read.table("Arquivo de dados", header=T)
attach(data)

count <- matrix(0,1,2)
counter <- count

# Matriz de delineamento (X)
xm <- matrix(1,831,1)
xa <- x1
xb <- x2
Ap <- diag(21)
xf <- cbind(xm,xa,xb)
xr <- cbind(p1,p2,p3,p4,p5,p6,p7,p8,p9
           p10,p11,p12,p13,p14,p15,p16,p17,p18,p19,p20,p21)
X <- cbind(xf, xr)
n <- nrow(X)
p <- ncol(X)

# Vetor de observações (y)
m <- data$n
y <- data$y

xwx <- t(X) %*% x
xwy <- t(X) %*% y
Mi <- ginv(xwx)

b <- Mi %*% xwy
br <- matrix(0,p,1)
ba <- matrix(0,p,1)

j <- 0
while ( t(b-br) %*% (b-br) > 0.000000000000000001 )
{
```

```
j <- j + 1
```

```
# Vetor de estimativas dos valores ajustados ( mi )
```

```
br <- b  
eta <- x %>% b  
mi <- exp(eta)/(1+exp(eta))
```

```
# Vetor de variáveis dependentes ajustadas ( z )
```

```
z <- eta + (diag(c(1/(mi*(1-mi))),n,n)%*(y-mi))  
w <- diag(c(m*mi*(matrix(1,n,1)-mi)),n,n)
```

```
# Vetor de estimativas de parâmetros ( b )
```

```
xwx <- t(X) %>% w %>% X  
Mxi <- ginv(xwx)  
Cp <- Mxi[4:p,4:p]
```

```
xwz <- t(X) %>% w %>% z  
b <- Mxi %>% xwz
```

```
k <- 1
```

```
while ( t(b-ba) %>% (b-ba) > 0.000000000000000001 )
```

```
{
```

```
k <- k + 1
```

```
count[1,1] <- j
```

```
count[1,2] <- k
```

```
counter <- rbind(counter,count)
```

```
ba <- b
```

```
int <- b[1:1]
```

```
trat <- b[2:3]
```

```
par <- b[4:p]
```

```
eps <- z - x %>% b
```

```
## Estimativa do componente de extra-dispersão ( Ve )
```

```
Ve <- (t(eps) %>% w %>% eps)/(n-3)
```

```
## Estimativa do componente de variância de parcela ( Vp )
```

```
Vp <- (t(par) %>% ginv(Ap) %>% par + sum(diag(Cp))) / 21
```

```
# Efeitos aleatórios de Parcelas:
```

```
P <- matrix(0,p,p)
```

```

for (i in 4:p)
{
  P[i,i] <- 1/Vp
}

```

Atualiza para efeitos aleatórios de Parcelas

```

eta <- X %*% b
mi <- exp(eta)/(1+exp(eta))
w <- diag(c(mi*(matrix(1,n,1)-mi)),n,n)

Mx <- t(X) %*% w %*% X + P
Mxi <- ginv(Mx)
Cp <- Mxi[4:p,4:p]

xwz <- t(X) %*% w %*% z
b <- Mxi %*% xwz
}
}

```

Vetor de estimativas dos parâmetros e erros padrão (sol)

```

se <- c(sqrt(diag(Mxi)))
sol <- cbind(b, se)

```

B2.2 Comando glmmPQL do software R 1.6.1

```

library(nlme)
library(mass)
data <- read.table("Arquivo de dados", header=T)
a <- data$x1
b <- data$x2
parc <- factor(data$parc)
y <- data$y
ajuste.mlglmm <- glmmPQL(y ~ a + b, random = ~ 1 | parc, family = binomial)
summary(ajuste.mlglmm)

```

B2.3 Macro **glimmix** associada ao procedimento **MIXED** do software SAS 8.10

```
data new;
set seeds;
do i=1 to r; y=1;
output;
end;
do i=1 to ur; y=0;
output;
end;

%include 'C:\ Arquivos de programas\ SASInstitute\ SAS\ V8\ STAT\
SAMPLE\glimmix.sas';
%glimmix (data=new, procopt=method=reml,
          stmts=%str(
            class parc;
            model y = x1 x2 x1*x2/solution;
            random parc/solution;
            parms (0.12) (1) / eqcons=2;
                 /* parâmetro de escala é fixo
                 (phi=1) */
            ),
          error=binomial,
          link=logit);

run; quit;
```