

MODELANDO DADOS LONGITUDINAIS DE PROPORÇÕES DE PEGAMENTO DE CAMU-CAMU

Silvano Cesar da COSTA¹
Clarice Garcia Borges DEMÉTRIO²
Eduardo SUGUINO³
Suely Ruiz GIOLO⁴

- **RESUMO:** Dados de proporção são muito comuns nas diversas áreas do conhecimento, principalmente na área agrícola. A distribuição padrão usada para a análise desse tipo de dados é a distribuição binomial, um caso particular de modelos lineares generalizados (McCullagh e Nelder, 1989; Demétrio, 2001). No presente trabalho, dados de proporção observados ao longo do tempo são analisados, ajustando-se quatro modelos. O primeiro modelo é o logístico em parcelas subdivididas no tempo, o segundo incorpora um fator de heterogeneidade na função de variância para levar em conta a (sub) superdispersão. O terceiro é um modelo logístico com uma estrutura de correlação AR(1), que é incorporada na matriz de variância e co-variância, usando-se equações de estimação generalizadas. No último modelo, a correlação existente ao longo do tempo é incorporada por meio de uma variável latente, para a qual se assume uma distribuição a priori, e a variabilidade individual é modelada com um fator de heterogeneidade constante na função de variância estimada do modelo logístico. Essa metodologia é aplicada a um conjunto de dados obtidos de um experimento com Camu-Camu, cujo fruto é a maior fonte natural, conhecida, de vitamina C, sendo uma espécie frutífera nativa das várzeas da Região Norte do Brasil. O objetivo do experimento foi avaliar métodos de enxertia, por garfagem, e tipos de porta-enxertos para serem utilizados na propagação dessa planta em terrenos não inundáveis.
- **PALAVRAS-CHAVE:** Modelos lineares generalizados; equações de estimação generalizadas; análise longitudinal; efeitos aleatórios; modelos lineares generalizados mistos.

¹Departamento de Estatística e Matemática Aplicada, Universidade Estadual de Londrina - UEL, CEP: 86051-990, Londrina, PR, Brasil.

²Departamento de Ciências Exatas, Escola Superior de Agricultura Luiz de Queiroz - ESALQ, Universidade de São Paulo - USP, CEP 13418-900, Piracicaba, SP, Brasil.

³Departamento de Produção Vegetal, Escola Superior de Agricultura Luiz de Queiroz - ESALQ, Universidade de São Paulo - USP, CEP 13418-900, Piracicaba, SP, Brasil.

⁴Departamento de Estatística, Universidade Federal do Paraná - UFPR, CEP 81531-990, Curitiba, PR, Brasil.

1 Introdução

O uso de modelos lineares clássicos não é, em geral, apropriado para analisar dados de proporções, que são muito freqüentes nas mais diversas áreas, pois as pressuposições do modelo não são atendidas. Na análise desse tipo de dados, pode-se utilizar a teoria de modelos lineares generalizados que envolve uma transformação (não dos dados, mas da esperança condicional; nesse contexto a transformação é conhecida como função de ligação), mas o objetivo principal é encontrar uma escala sobre a qual um modelo linear aditivo ocorra. A distribuição padrão utilizada para a análise de dados de proporções é a binomial, sendo um caso particular de modelos lineares generalizados (Nelder e Wedderburn, 1972).

Quando a mesma unidade experimental é observada ao longo do tempo (longitudinal), espera-se que haja uma correlação entre essas unidades, levando à violação da suposição de independência. Uma forma de se levar isso em consideração é modelar explicitamente a estrutura de correlação, utilizando-se a abordagem de equação de estimação generalizada (EEG), dada por Liang e Zeger (1986). Este método permite modelar a variabilidade entre as observações incluindo na análise uma matriz de correlação de “trabalho”.

Uma forma alternativa para se modelarem dados longitudinais é com a inclusão de variáveis latentes no preditor linear considerando-se a correlação entre as mesmas unidades experimentais, assumindo-se uma distribuição de probabilidade, em geral, a distribuição normal, para a variável latente. Uma discussão sobre o assunto é apresentada por McCulloch e Searle (2001).

Neste trabalho, modelos para esse tipo de dados são apresentados e discutidos. Tais modelos devem levar em consideração: a natureza discreta da variável resposta (proporções), a variabilidade entre repetições em cada tratamento e a dependência entre as observações ao longo do tempo.

São estudados para essa finalidade quatro modelos. O primeiro modelo não considera a possível correlação entre as observações ao longo do tempo, tratando-as, portanto, como independentes. O segundo modelo incorpora um fator de dispersão, para considerar a variabilidade existente entre indivíduos, não captada pelo modelo padrão. O terceiro modelo inclui uma matriz de correlação de “trabalho” como uma forma de incorporar a dependência entre observações ao longo do tempo. O último modelo apresentado considera, conjuntamente, o parâmetro de dispersão e uma variável latente (efeito aleatório), o primeiro representando a variabilidade inerente à forma como o experimento foi conduzido e o segundo representando a possível correlação ao longo do tempo.

Os modelos citados são aplicados a um conjunto de dados de Camu-Camu (*Myrciaria dubia* (H.B.K.) McVaugh), resultados de um experimento realizado para se avaliar métodos de enxertia e tipos de porta-enxerto e, também, uma possível associação entre eles, com observações feitas durante oito meses.

2 Material e métodos

Os dados utilizados neste trabalho são provenientes de um experimento realizado para a comparação de métodos de enxertia e tipos de porta-enxertos para Camu-Camu. Este experimento foi instalado no delineamento inteiramente casualizado com 5 repetições e 12 tratamentos no esquema fatorial 3×4 (3 tipos de porta-enxertos: Camu-Camu, goiabeira (*Psidium guajava* L.) e pitangueira (*Eugenia uniflora* L.) e 4 métodos de enxertia: fenda cheia, fenda lateral, inglês simples e fenda de colo). As parcelas foram constituídas de 12 plantas. A notação utilizada para representar os diferentes tratamentos está na Tabela 1.

Tabela 1 - Combinação dos níveis de cada fator, tipos de porta-enxertos e métodos de enxertia. A variável resposta de interesse foi a proporção

Métodos de Enxertia	Porta-enxertos		
	Camu-camu	Goiabeira	Pitangueira
Fenda cheia	T1	T2	T3
Fenda lateral	T4	T5	T6
Inglês simples	T7	T8	T9
Colo	T10	T11	T12

de pegamento, medida mensalmente, durante oito meses (dezembro de 2000 a julho de 2001), com o objetivo de avaliar quais os melhores métodos de enxertia, por garfagem, e tipos de porta-enxertos que podem ser utilizados para propagar essa planta em terrenos não inundáveis.

Para a análise dos dados do experimento descrito, quatro modelos são discutidos e descritos a seguir.

3 Modelo 1: Modelo logístico em parcelas subdivididas

Seja Y a variável aleatória número de pegamentos de enxertos em parcelas de $m = 12$ plantas. A distribuição padrão a ser assumida é a binomial, de índice m e parâmetro π (probabilidade de sucesso), isto é,

$$Y \sim Bin(m, \pi).$$

Assume-se a função de ligação logística e, inicialmente, independência das observações. Tem-se, então, como preditor linear o esquema em parcelas subdivididas, em que as parcelas estão no delineamento inteiramente casualizado com 5 repetições e os tratamentos no esquema fatorial 3×4 (3 tipos de porta-enxertos e 4 métodos de enxertia) e as subparcelas são as observações no tempo (8 tempos), isto é,

$$\eta_s = \text{logit}(\pi_s) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \delta_{ijk} + \tau_\ell + (\alpha\tau)_{i\ell} + (\beta\tau)_{j\ell} + (\alpha\beta\tau)_{ij\ell} \quad (1)$$

sendo $s = 1, \dots, n$, μ o efeito associado à média geral, α_i o efeito associado ao i -ésimo método de enxertia, $i = 1, \dots, 4$, β_j o efeito associado ao j -ésimo porta-enxerto, $j = 1, \dots, 3$, γ_k o efeito associado às repetições, $k = 1, \dots, 5$, δ_{ijk} o efeito associado ao erro de parcelas, τ_ℓ o efeito associado ao ℓ -ésimo tempo em meses, $\ell = 1, \dots, 8$ e $(\alpha\beta)_{ij}$, $(\alpha\tau)_{i\ell}$, $(\beta\tau)_{j\ell}$ e $(\alpha\beta\tau)_{ij\ell}$, os efeitos associados às interações.

Para a verificação do ajuste do modelo é utilizado o gráfico meio-normal (*half-normal plot*) com envelope simulado (Collett, 1991). A comparação entre modelos é feita usando-se o Critério Bayesiano de Schwarz (BIC) e o Critério de Informação de Akaike (AIC), dados por:

$$AIC = -2\ell + 2p \quad \text{e} \quad BIC = -2\ell + p \log n,$$

sendo ℓ o máximo do logaritmo da função de verossimilhança, p o número de parâmetros do modelo e n o número de observações. O modelo com menor valor do critério é o escolhido. A verificação da significância dos efeitos é feita pela comparação da estatística *deviance* com o percentil da distribuição χ^2 com número associado de graus de liberdade, a um nível 100 α % de probabilidade.

4 Modelo 2: Modelo logístico incorporando fator de heterogeneidade constante

Como a coleta e a manipulação do material experimental foram realizadas por diferentes pessoas e a coleta do material também não foi feita na mesma árvore-fonte do Camu-Camu, espera-se que a variância observada seja diferente do que a esperada pelo modelo binomial e, desse modo, um fator de dispersão constante, ϕ , é incorporado no modelo uma vez que o modelo 1 pode não se ajustar satisfatoriamente aos dados. Esse parâmetro é estimado por:

$$\hat{\phi} = \frac{1}{(n-p)} \sum_{s=1}^n \frac{(y_s - m\hat{\pi}_s)^2}{m\hat{\pi}_s(1-\hat{\pi}_s)},$$

em que $\hat{\pi}$ é estimado a partir do modelo 1, pelo método da quase-verossimilhança (Hinde e Demétrio, 1998).

Para a verificação do ajuste do modelo, é utilizado o gráfico meio-normal (*half-normal plot*) com envelope simulado (Collett, 1991). A comparação entre modelos é feita usando-se o Critério Bayesiano de Schwarz (BIC) e o Critério de Informação de Akaike (AIC), sendo a verificação da significância dos efeitos dos fatores feita pela estatística

$$F = \frac{D_1 - D_2}{r\hat{\phi}},$$

em que D_1 e D_2 são valores da estatística *deviance* para modelos encaixados com $p_1 < p_2$ parâmetros e $r = p_2 - p_1$. Sob H_0 a estatística F tem distribuição F , com r e $n - p$ graus de liberdade.

5 Modelo 3: Modelo logístico incorporando matriz de correlação

Esse modelo considera a correlação entre as observações ao longo do tempo, pela inclusão de uma matriz de correlação de “trabalho”, com o uso de equações de estimação generalizadas. Assume-se o modelo logístico com preditor linear:

$$\eta_s = \text{logit}(\pi_s) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad (2)$$

sendo μ o efeito associado à média geral, α_i o efeito associado ao i -ésimo método de enxertia, $i = 1, \dots, 4$, β_j o efeito associado ao j -ésimo porta-enxerto, $j = 1, \dots, 3$, $(\alpha\beta)_{ij}$ o efeito da interação entre o i -ésimo método de enxertia e o j -ésimo porta-enxerto.

A estimação dos parâmetros é semelhante ao caso em que se supõem observações independentes, porém, na matriz de variâncias e co-variâncias é introduzida uma matriz de correlação da seguinte forma:

$$\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2},$$

sendo \mathbf{A}_i uma matriz diagonal formada pelos elementos da função de variância $V(\mu_i)$.

Essa matriz é utilizada na obtenção das equações de estimação generalizadas, dadas por:

$$\sum_{i=1}^n \frac{\partial \mu'_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0},$$

como definido em Zeger e Liang (1986). O vetor $\boldsymbol{\beta}$ de parâmetros é estimado a partir de um algoritmo cujas etapas são:

1. calcula-se uma estimativa inicial de $\boldsymbol{\beta}$ considerando-se um modelo linear generalizado padrão, assumindo-se independência;
2. obtém-se a estimativa da matriz de correlação de trabalho \mathbf{R} baseada nos resíduos e estrutura assumida;
3. calcula-se uma estimativa da matriz de co-variância $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$;
4. atualiza-se o valor de $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}_{r+1} = \boldsymbol{\beta}_r + \left[\sum_{i=1}^n \frac{\partial \mu'_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right]_r^{-1} \left[\sum_{i=1}^n \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \right]_r^{-1}, \text{ e}$$

5. repete-se o processo, a partir de (2), até a convergência.

Algumas especificações para a estrutura da matriz de correlação de trabalho têm sido sugeridas por Liang e Zeger (1986), entre a independente, esférica, não-estruturada, m-dependente e auto-regressiva AR(1). O estimador do parâmetro α para a estrutura AR(1) é dado por:

$$\hat{\alpha} = \frac{1}{n-p} \sum_{i=1}^n r_{ijk} r_{ij,k+1},$$

em que r_{ij} é o resíduo de Pearson, obedecendo-se a lei de formação da matriz que é:

$$\text{Corr}(y_{ijk}, y_{ij,k+1}) = \alpha^{|k-k'|}.$$

A seleção da matriz de correlação pode ser feita ajustando-se modelos com estruturas de co-variâncias alternativas e usando-se o Critério de Informação de Akaike (AIC) para sua escolha, como apresentado em Dobson (2001). A verificação de ajuste do modelo pode ser completada com gráficos de resíduos.

A matriz de variâncias e co-variâncias das estimativas dos parâmetros β , é dada por:

$$\mathbf{V}_r(\hat{\beta}) = \mathfrak{S}^{-1} \mathbf{C} \mathfrak{S}^{-1},$$

sendo $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}_i}$,

$$\mathfrak{S} = \sum_{s=1}^n \mathbf{D}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{D}_i,$$

a matriz de informação e

$$\mathbf{C} = \sum_{s=1}^n \mathbf{D}_i' \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)' \hat{\mathbf{V}}_i^{-1} \mathbf{D}_i.$$

Assintoticamente, $\hat{\beta}$ tem distribuição $N(\beta, \mathbf{V}_r(\hat{\beta}))$ e, inferências a respeito dos parâmetros podem ser feitas usando-se a estatística de Wald, dada por:

$$W = [\hat{\beta} - \beta^0]' \hat{\mathbf{V}}^{-1}(\hat{\beta}) [\hat{\beta} - \beta^0],$$

em que $\hat{\mathbf{V}}(\hat{\beta})$ denota a matriz de variância e co-variância assintótica de $\hat{\beta}$.

6 Modelo 4: Modelo considerando fator de heterogeneidade constante e um efeito aleatório

Uma outra forma de se modelar a correlação entre as observações ao longo do tempo é adicionando-se um efeito aleatório no preditor linear. Este último modelo, leva em conta o parâmetro de heterogeneidade constante e uma variável latente (efeito aleatório), conjuntamente. Nesse caso, considera-se, inicialmente, o modelo

de superdispersão com heterogeneidade constante, obtendo-se $\hat{\phi}$ a partir do preditor dado por (1), que é o modelo maximal. Em seguida, fixa-se $\phi = \hat{\phi}$ que é incorporado na função de quase-verossimilhança (Wedderburn, 1974), com preditor linear dado por:

$$\eta_s = \text{logit}(\pi_s) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \delta_{ijk} + \tau_\ell + (\alpha\tau)_{i\ell} + (\beta\tau)_{j\ell} + (\alpha\beta\tau)_{ij\ell} + \sigma Z_{ij\ell k} \quad (3)$$

sendo μ o efeito associado à média geral, α_i o efeito associado ao i -ésimo método de enxertia, $i = 1, \dots, 4$, β_j o efeito associado ao j -ésimo porta-enxerto, $j = 1, \dots, 3$, τ_ℓ o efeito associado ao ℓ -ésimo tempo em meses, $\ell = 1, \dots, 8$, σ o efeito aleatório associado ao i -ésimo indivíduo, δ_{ijk} o efeito associado ao erro de parcelas, $s = 1, \dots, n$ e $(\alpha\beta)_{ij}$, $(\alpha\tau)_{i\ell}$, $(\beta\tau)_{j\ell}$ e $(\alpha\beta\tau)_{ij\ell}$ interações entre os efeitos principais. Assume-se que $Z_{ij\ell k} \sim N(0, 1)$.

Nesse caso, não é possível a obtenção de equações de quase-verossimilhança de forma explícita e, assim, usa-se o método de quadratura Gaussiana (Abramowitz e Stegun, 1970), considerando-se que há apenas um efeito aleatório no preditor linear a ser estimado.

Nos dois últimos modelos apresentados, estudos adicionais precisam ser realizados para a verificação de ajustes, pois ainda não existem técnicas disponíveis.

Após a escolha do modelo, podem-se utilizar contrastes para hipóteses específicas relativas aos parâmetros. As estatísticas calculadas são baseadas na distribuição qui-quadrado assintótica da razão da verossimilhança ou na estatística escore, com graus de liberdade determinado pelo número de linhas linearmente independentes. No presente caso, para o teste dos contrastes de médias foi utilizada a estatística escore dada por:

$$E = \mathbf{U}' (\boldsymbol{\beta}^0) \hat{\mathbf{V}}_0(\hat{\boldsymbol{\beta}}) \mathbf{U} (\boldsymbol{\beta}^0)$$

que tem distribuição assintótica com q graus de liberdade sob H_0 , em que q é o número de linhas linearmente independentes.

7 Resultados e discussão

Para a realização das análises estatísticas foi utilizado o programa SAS - *Statistical Analysis System*, versão 8.2, cujos programas utilizados podem ser obtidos em Costa (2003). Pela Figura 1 pode-se notar, para cada tratamento, os números observados de pagamentos a partir de 12 garfos vivos de Camu-Camu enxertados em Camu-Camu, goiabeira e pitangueira (colunas 1, 2 e 3, respectivamente) por 4 tipos de enxertia, fenda cheia, fenda lateral, inglês simples e colo (linhas 1, 2, 3 e 4, respectivamente).

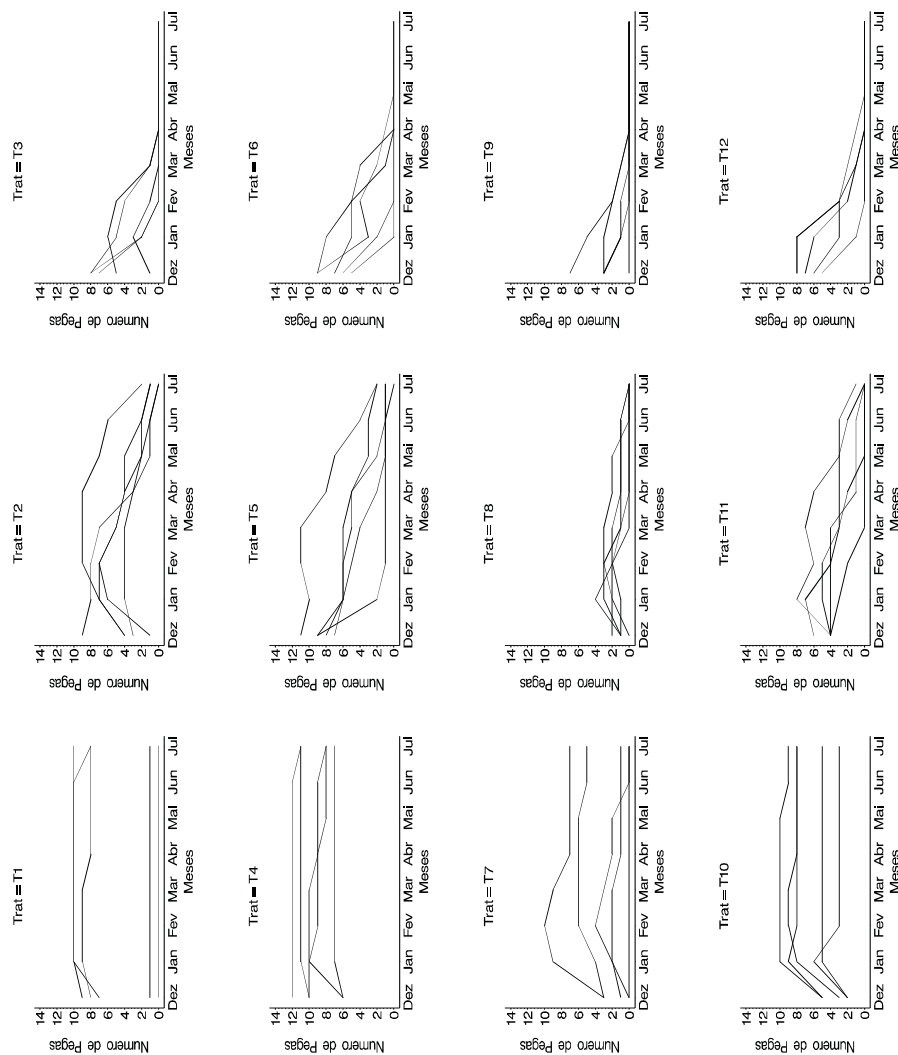


FIGURA 1 - Gráficos de dispersão dos números observados de pagamentos a partir de 12 garfos vivos de Camu-Camu enxertados em Camu-Camu, goiabeira e pitangueira (colunas 1, 2 e 3, respectivamente) por 4 tipos de enxertia, fenda cheia, fenda lateral, inglês simples e colo (linhas 1, 2, 3 e 4, respectivamente).

Verifica-se que apesar da variabilidade entre repetições, em geral, o Camu-Camu como porta-enxerto apresenta uma maior estabilidade ao longo dos meses e uma porcentagem maior de pegamento, sendo T1 o tratamento com maior variabilidade; a goiabeira como porta-enxerto apresenta uma ligeira estabilidade até o mês de março, a partir desse mês, o número de pegamentos tende a zero e a pitangueira, como porta-enxerto, foi a que menos se adaptou, pois o número de pegamentos tende a zero de forma mais acentuada desde o início.

Utilizando-se os quatro modelos considerados, tem-se que o logístico padrão em parcelas subdivididas não é adequado, uma vez que não leva em consideração a correlação entre os tempos e foi notado pelos resultados que a *deviance* residual (174,5) é muito menor que o número de graus de liberdade (336) do modelo. O gráfico meio-normal com envelope simulado (*half-normal plot*), apresentado na Figura 2, mostra os pontos residuais fora do envelope simulado, sugerindo a falta de ajuste causada por uma subdispersão.

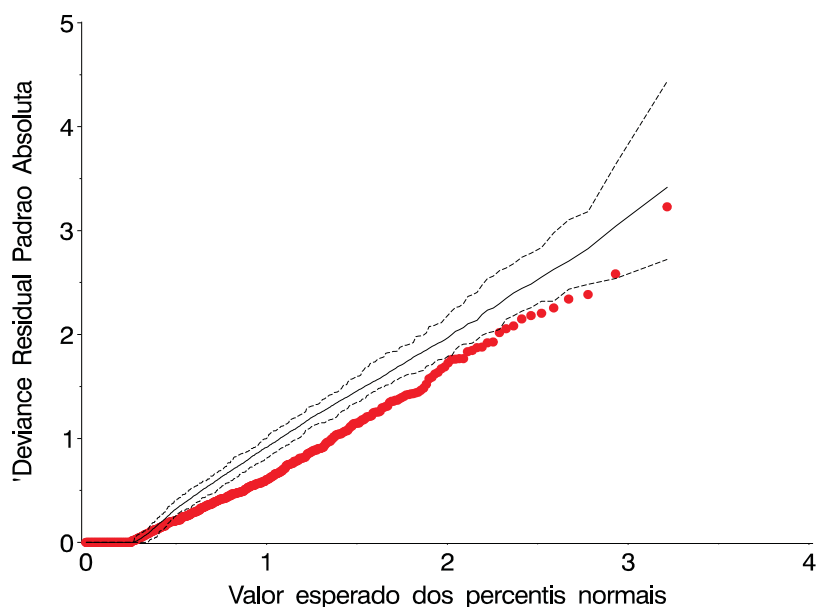


FIGURA 2 - Gráfico meio normal considerando-se o modelo logístico em parcelas subdivididas.

Usando-se o modelo de subdispersão com fator de heterogeneidade constante tem-se $\hat{\phi} = 0,4640$. Os resultados, porém, mostram que esse modelo também não se ajusta bem aos dados, conforme se observa pelo gráfico meio-normal com envelope simulado (Figura 3). Esse comportamento era, de certa forma, esperado, uma vez que a dependência entre as observações ao longo do tempo não foi considerada.

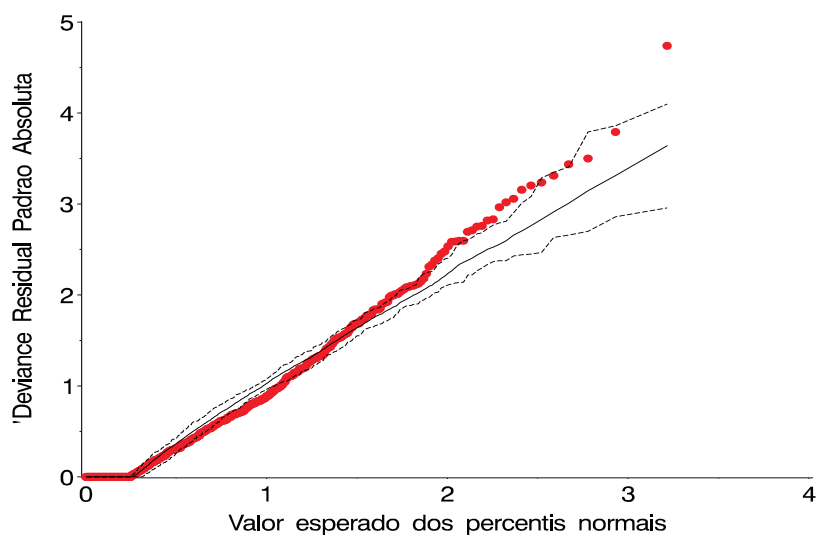


FIGURA 3 - Gráfico meio normal, considerando-se o modelo logístico com fator de heterogeneidade constante.

A matriz de correlação observada, eliminados os efeitos de Métodos de Enxertia, Tipos de Porta-enxerto e interação Método \times Porta-enxerto, apresentada na Tabela 2, mostra que a correlação decresce ao longo do tempo indicando, desse modo, que a estrutura de correlação de trabalho AR(1) pode ser adequada para o modelo 3.

Tabela 2 - Matriz de correlação observada

	Dez.	Jan.	Fev.	Mar.	Abr.	Maio	Jun.	Jul.
Dez.	1							
Jan.	0,6835	1						
Fev.	0,6256	0,8951	1					
Mar.	0,6287	0,8429	0,9318	1				
Abr.	0,5369	0,7382	0,8450	0,9307	1			
Maio	0,5414	0,7117	0,8217	0,8960	0,9636	1		
Jun.	0,5689	0,7030	0,7934	0,8668	0,9380	0,9635	1	
Jul.	0,5809	0,7018	0,7885	0,8293	0,8812	0,9069	0,9551	1

Pela análise de *deviance*, baseada nas estatísticas escores para a estrutura AR(1), apresentada na Tabela 3 observa-se que a interação Método de Enxertia \times Tipo de Porta-enxerto não é significativa, havendo significância dos efeitos principais.

Tabela 3 - Análise de *deviance*, considerando-se modelo logístico com matriz de correlação na estrutura AR(1)

Causa de variação	g.l.	<i>deviance</i>	Valor p
Método	3	14,41	0,0024
Porta-enxerto	2	18,25	0,0001
Método \times Porta-enxerto	6	7,67	0,2633

Os valores das estatísticas para os contrastes entre os efeitos principais são apresentados na Tabela 4, e mostram que o método da Fenda Cheia e Fenda Lateral não diferem entre si mas diferem dos métodos de Inglês Simples e Fenda de Colo, ao nível de significância de 5% e, ainda, que Fenda de Colo difere de Inglês Simples. Tem-se, ainda, que o porta-enxerto Camu-Camu difere dos porta-enxertos Goiabeira e Pitangueira, sendo que esses também diferem entre si.

Tabela 4 - Teste escore para os contrastes entre os efeitos principais, considerando-se a estrutura de correlação AR(1)

Contrastes	g.l.	X^2	Valor p
Fendas cheia e lateral <i>vs</i> Inglês simples e Fenda de colo	1	12,39	0,0004
Fendas cheia <i>vs</i> Fenda lateral	1	3,24	0,0720
Inglês simples <i>vs</i> Fenda de colo	1	9,84	0,0017
Camu-camu <i>vs</i> Goiabeira e Pitangueira	1	10,58	0,0011
Goiabeira <i>vs</i> Pitangueira	1	13,56	0,0002

Em razão da forma como foi conduzido o experimento, acredita-se que é possível haver mais de uma fonte de variabilidade, além daquela atribuída à correlação relativa ao tempo. Assim, foi considerado o modelo incluindo o fator de heterogeneidade constante e o efeito aleatório.

Pela Tabela 5, observa-se que a interação Tipo de porta-enxerto \times Tempo é significativa, o que, de certa forma, confirma os resultados gráficos dos valores observados, como apresentados nas Figuras 4 e 5.

Tabela 5 - Análise de *deviances* para o modelo considerando a correlação ao longo do tempo e o efeito aleatório conjuntamente

Causa de variação	g.l.	<i>Deviance</i>	F	Valor de p	σ (erro)
σ	1				1,5084 (0,1601)
Método de enxertia	3	14,7	4,9	0,0047	1,3339 (0,1443)
Tipo de porta-enxerto	2	33,3	16,65	< 0,0001	0,9303 (0,1120)
Método \times Tipo	6	9,0	1,5	0,1984	0,8610 (0,1047)
Tempo	7	142,8	20,40	< 0,0001	0,9418 (0,1135)
Método \times Tempo	21	16,0	0,76	0,7683	0,9394 (0,1130)
Tipo \times Tempo	14	191,3	13,66	< 0,0001	0,9653 (0,1162)
Mét. \times Tipo \times Tempo	42	10,7	0,25	0,999	0,9549 (1,0211)

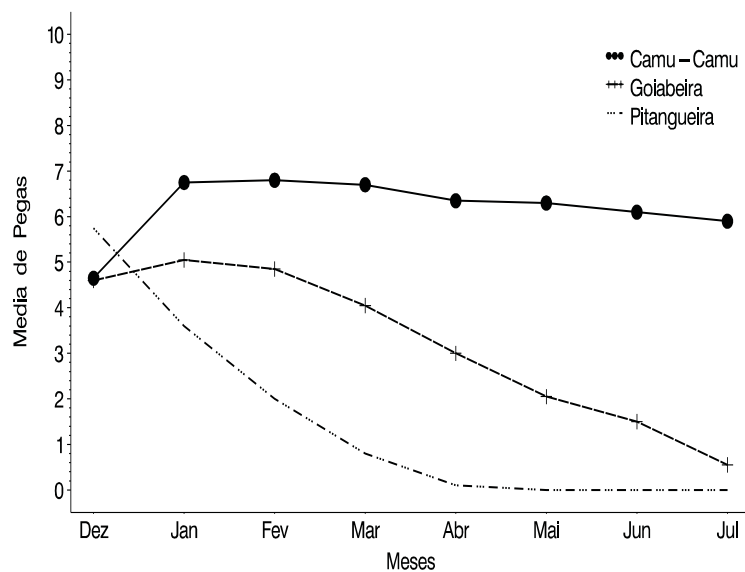


FIGURA 4 - Totais observados para espécies.

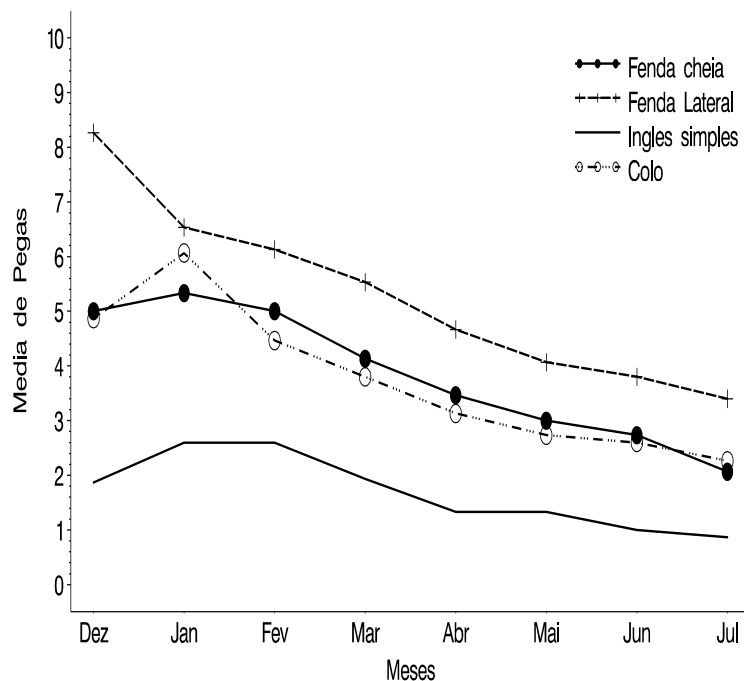


FIGURA 5 - Totais observados para métodos de enxertias.

Os níveis de significância para os contrastes dos métodos de enxertia são apresentados na Tabela 6. Os resultados apenas confirmam as diferenças observadas na Figura 5, ou seja, fenda cheia não difere da fenda lateral e nem de colo, e todas diferem do inglês.

Tabela 6 - Valores das probabilidades para os teste escores para os contrastes dos efeitos dos métodos de enxertia

	Fenda cheia	Fenda lateral	Inglês
Fenda lateral	0,1026		
Inglês	0,0005	< 0,0001	
Colo	0,0984	0,6182	< 0,0001

Embora não haja métodos para a comparação dos modelos 3 e 4, o modelo 4 parece mais razoável do que o modelo 3, por levar em consideração, além da correlação ao longo do tempo, um parâmetro de dispersão que incorpora a variabilidade inerente à forma como o experimento foi conduzido.

Conclusões

Neste trabalho alguns modelos para dados de proporção observados ao longo do tempo foram discutidos.

As análises realizadas para o experimento com Camu-Camu, mostraram que, quando a variável resposta é observada ao longo do tempo, há a necessidade de se levar em consideração a correlação existente entre as observações.

Entre as várias formas de se modelar essa correlação, há indícios de que a inclusão de um fator aleatório, para o qual se pressupõe uma distribuição *a priori*, no preditor linear, faz que o modelo ajustado seja melhor, captando a variabilidade não considerada quando só da inclusão do fator de dispersão constante.

No caso desse experimento, a inclusão do efeito aleatório e do parâmetro de dispersão constante, conjuntamente, explica melhor o comportamento dos dados, embora ainda haja necessidade de maiores estudos em relação a técnicas de seleção de modelos.

Agradecimentos

Os autores agradecem à CAPES e à FAPESP pelo suporte financeiro.

COSTA, S. C. da; DEMÉTRIO, C. G. B.; SUGUINO, E.; GIOLO, S. R. Modelling longitudinal proportion data from successful grafting on Camu-Camu. *Rev. Mat. Estat.*, São Paulo, v.22, n.3, p.113-127, 2004.

■ **ABSTRACT:** *Proportion data are very common in several research areas, specially in agriculture. The standard distribution used for analysing this kind of data is the binomial distribution, a particular case of generalized linear models (McCullagh e Nelder, 1989; Demétrio, 2001). In this paper repeated proportion data over time are analysed using four models. The first model is a logistic split-plot in time while the second incorporates a heterogeneity factor in the variance function to account for (under) overdispersion. The third one is a logistic model with an AR(1) correlation structure which is incorporated in the variance-covariance matrix using generalized estimating equations. In the last model the correlation existing over time is incorporated by means of a latent variable, assumed to have a priori distribution, and the individual variability is modelled with a heterogeneity factor in the variance function estimated from the logistic split-plot in a time model. The methodology is applied to a data set obtained from an experiment with Camu-Camu, the major source of vitamin C in Brazil. The plants were collected from low and flat lands alongside a watercourse located in the North of Brazil. The aim of the experiment was to evaluate grafting methods and understock types to be used for propagating the plant in dry lands.*

- **KEYWORDS:** *Generalized linear models, generalized estimating equations; longitudinal analysis; random effects; generalized linear mixed models.*

Referências

- ABRAMOWITZ, M.; STEGUN, I. *Handbook of mathematical functions*. New York: Dover, 1970. 1044p.
- COLLETT, D. *Modelling binary data*. London: Chapman & Hall, 1991. 369p.
- COSTA, S. C. *Modelos lineares generalizados mistos para dados longitudinais*, 2003. 110f. Tese (Doutorado em Estatística). Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2003.
- DEMÉTRIO, C. G. B. Modelos lineares generalizados em experimentação agrônômica. In: REUNIÃO ANUAL DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 46., 2001, Piracicaba. *Anais...* Piracicaba: ESALQ/USP, 2001. p.113.
- DOBSON, A. J. *An introduction to generalized linear models*. 2nd. ed. London: Chapman & Hall, 2001. 225p.
- HINDE, J. P.; DEMÉTRIO, C. G. B. *Overdispersion: models and estimation*. In: SINAPE, 13., 1998, Caxambu. *Anais...* Caxambu: SINAPE, 1998. 73p.
- LIANG, K. Y.; ZEGER, S. L. Longitudinal data analysis using generalized linear models. *Biometrika*, Cambridge, v.73, n.1, p.13-22, 1986.
- McCULLAGH, P.; NELDER, J. A. *Generalized linear models*. 2nd. ed. London: Chapman & Hall, 1989. 511p.
- McCULLOCH, C. E.; SEARLE, S. R. *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons, 2001. 325p.
- WEDDERBURN, R. W. M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, Cambridge, v.61, p.439-447, 1974.
- ZEGER, S.; LIANG, K. Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, Washington, v.42, p.121-130, 1986.

Recebido em 23.01.2003.

Aprovado após revisão em 18.12.2004.