

COMUNICAÇÃO

MODELOS DE PREDIÇÃO PARA SOBREVIVÊNCIA DE PLANTAS DE *Eucalyptus grandis*

Prediction models of *Eucalyptus grandis* plant survival

Telde Natel Custódio¹, Décio Barbin²

RESUMO

Objetivou-se com este trabalho comparar modelos de predição de plantas sobreviventes de *Eucalyptus grandis*. Utilizaram-se os seguintes modelos: modelo linear misto com os dados transformados, utilizando-se as transformações angular e BOX-COX; modelo linear generalizado misto com distribuição binomial e funções de ligação logística, probit e complemento log-log; modelo linear generalizado misto com distribuição Poisson e função de ligação logarítmica. Os dados são provenientes de um experimento em blocos ao acaso, para avaliação de progênies maternas de *Eucalyptus grandis*, aos 5 anos de idade, em que a variável resposta são plantas sobreviventes. Para comparação dos efeitos entre os modelos foram estimadas as correlações de *Spearman* e aplicado o teste de permutação de Fisher. Foi possível concluir que, o modelo linear generalizado misto com distribuição Poisson e função de ligação logarítmica se ajustou mal aos dados e que as estimativas para os efeitos fixos e predição para os efeitos aleatórios, não se diferenciaram entre os demais modelos estudados.

Termos para indexação: Estatística aplicada, modelo linear misto, transformação de dados, modelo linear generalizado misto.

ABSTRACT

The objective of this work was to compare models for prediction of the survival of plants of *Eucalyptus grandis*. The following models were used: linear mixed model with the transformed data, by utilizing the angular transformations and BOX-COX; generalized linear mixed model with binomial distribution and logistic functions, probit and complement log-log links; generalized linear mixed model with Poisson distribution and logarithmic link function. The data came from a randomized block experiment for evaluation of *Eucalyptus grandis* maternal progenies at five years old, in which the variable response are surviving plants. For comparison of the effects among the models the correlations of *Spearman* were estimated and the test of permutation of Fisher was applied. It was possible to conclude that: the generalized linear mixed model with Poisson distribution and logarithmic link function misadjusted to the data; the estimates for the fixed effects and prediction for the random effects did not differ among the to other studied models.

Index terms: Applied statistics, linear mixed model, transformation of data, generalized linear mixed model.

(Recebido em 4 de agosto de 2006 e aprovado em 28 de setembro de 2007)

O gênero *Eucalyptus* é amplamente cultivado no Brasil pela sua importância como espécie botânica de grande diversidade e boa adaptabilidade a vários tipos de ambiente. Assim, são crescentes as pesquisas no sentido de estabelecer florestas de usos múltiplos, voltados principalmente para atender a demanda de madeira para os mercados nacional e internacional, utilizados para diversos fins. Dentre as espécies do gênero *Eucalyptus* destaca-se a espécie *E. grandis*, de ampla diversidade de uso e potencialidade.

Uma vez reconhecida a sua importância, torna-se necessário considerar que o aumento da produção das

florestas não depende somente de técnicas de manejo. Aliado a essas técnicas torna-se necessária a aplicação do melhoramento genético, com o intuito de tornar os povoamentos florestais mais produtivos, aumentando sua adaptabilidade não só ao tipo de utilização, mas também ao tipo de uso final. No melhoramento, uma característica avaliada de grande importância é o número de plantas sobreviventes ao final de um experimento.

A variável sobrevivência de plantas geralmente é estudada na forma de proporção, dividindo-se o número de plantas sobreviventes pelo total de plantas na parcela. Quando não se realiza um estudo adequado das

¹Engenheiro Agrícola, Doutor em Agronomia, Pesquisador – Departamento de Agricultura/DAG – Universidade Federal de Lavras/UFLA – Cx. P. 3037 – 37200-000 – Lavras, MG – telde@ufla.br

²Engenheiro Agrônomo, Doutor em Agronomia, Professor Titular – Departamento de Ciências Exatas/LCE – Escola Superior de Agricultura “Luiz de Queiroz”/ESALQ – Universidade de São Paulo/USP – Avenida Pádua Dias, 11 – Cx. P. 9 – 13418-900 – Piracicaba, SP – debarbin@esalq.usp.br

características da distribuição de frequências, esse tipo de dados são analisados, mediante a técnica da análise de variância ou de regressão, sem se ter a certeza de que todas as exigências necessárias para a aplicação dessas técnicas foram satisfeitas.

Mas, para que esses dados sejam analisados mediante essas técnicas é necessário que eles atendam às pressuposições nas quais as análises se baseiam. Em certos casos quando não é possível satisfazer as pressuposições para análise de variância ou de regressão na própria escala observada, transformações têm sido feitas para uma escala mais apropriada com o objetivo de se conseguir homogeneidade de variâncias e uma distribuição aproximadamente normal.

No caso de proporções de plantas sobreviventes, o arco seno da raiz quadrada dessas proporções, seria então utilizado em uma análise de variância ou regressão. Essa operação é conhecida como transformação angular, sendo uma das mais utilizadas (SNEDECOR & COCHRAN, 1989; STEEL & TORRIE, 1980). Portanto, essa transformação é feita admitindo-se que a variável resposta segue uma distribuição binomial. Uma outra opção de transformação seria a transformação de Box & Cox (1964). Assim, espera-se que a transformação homogeneizará a variância, o que é uma das exigências para a validade dos testes estatísticos. Porém, em muitas situações práticas, a transformação necessária para estabilizar a variância pode não ser a mesma para se obter uma resposta linear.

Outra abordagem seria a utilização das proporções sem o uso de transformação, assumindo-a como sendo proveniente de uma distribuição binomial, ou ainda, a utilização somente do número de plantas sobreviventes, admitindo-o como sendo de uma distribuição de Poisson, e proceder à análise pela metodologia dos modelos lineares generalizados.

Os modelos lineares generalizados foram desenvolvidos por Nelder & Wedderburn (1972), visando à análise de dados associados a distribuições pertencentes à família exponencial com um parâmetro. Segundo McCullagh & Nelder (1989), um modelo linear generalizado é composto por um componente aleatório associado à distribuição da variável resposta; um componente sistemático linear nos parâmetros, denominado preditor linear; uma função de ligação, que liga o componente aleatório ao componente sistemático, ou seja, relaciona a

média ao preditor linear. Incorporando efeitos aleatórios no preditor linear, têm-se os modelos lineares generalizados mistos. Nesse sentido, a função de ligação deve ser incorporada aos estimadores dos efeitos fixos, preditores dos efeitos aleatórios e estimadores dos componentes de variância dos efeitos aleatórios, conforme Schall (1991).

Devido às abordagens citadas, objetivou-se com este trabalho comparar modelos de predição de plantas sobreviventes de *Eucalyptus grandis*, considerando-se as diferentes metodologias estatísticas, através dos seguintes modelos: modelo linear misto utilizando-se as transformações angular e BOX-COX; modelo linear generalizado misto com distribuição binomial e funções de ligação logística, probit e complemento log-log; modelo linear generalizado misto com distribuição Poisson e função de ligação logarítmica.

Foram utilizados dados provenientes de um experimento para avaliação de progênies de *Eucalyptus grandis*, aos 5 anos de idade, em que a variável resposta são plantas sobreviventes. O experimento foi instalado no delineamento em blocos casualizados, com 6 plantas por parcela, 10 repetições e 25 progênies maternas. Na avaliação da sobrevivência foram atribuídos os valores 1 para as plantas vivas e 0 para as mortas.

Para verificação da normalidade dos dados aplicou-se o teste de *Shapiro-Wilk*, através do PROC UNIVARIATE do SAS (SAS INSTITUTE, 2001), considerando os dados como proporções e contagem de plantas sobreviventes. Foram constatados que os mesmos não apresentaram normalidade, justificando assim a transformação dos dados ou a análise pela teoria de modelos lineares generalizados.

Assim, foram considerados os seguintes modelos para fins de comparação:

A – Modelo linear misto com transformação angular:

$$\text{arco seno } \sqrt{\pi_{ij}} = \mu + p_i + b_j + e_{ij};$$

B – Modelo linear misto com transformação BOX-COX:

$$\pi_{ij}^{1,3} = \mu + p_i + b_j + e_{ij};$$

C – Modelo linear generalizado misto com distribuição binomial e função de ligação logística:

$$\ln\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \mu + p_i + b_j ;$$

D – Modelo linear generalizado misto com distribuição binomial e função de ligação probit:

$$\Phi^{-1}(\pi_{ij}) = \mu + p_i + b_j ;$$

E – Modelo linear generalizado misto com distribuição binomial e função de ligação complemento log-log:

$$\ln[-\ln(1 - \pi_{ij})] = \mu + p_i + b_j ;$$

F – Modelo linear generalizado misto com distribuição Poisson e função de ligação logarítmica:

$$\ln(y_{ij}) = \mu + p_i + b_j ,$$

em que π_{ij} é a proporção de plantas sobreviventes para a progênie i no bloco j ; y_{ij} é o número de plantas sobreviventes para a progênie i no bloco j ; m é uma constante inerente a todas as observações (média geral); p_i é o efeito da i -ésima progênie ($i = 1, 2, \dots, 25$), considerado aleatório; b_j é o efeito do j -ésimo bloco ($j = 1, 2, \dots, 10$), considerado fixo; e_{ij} é o erro aleatório.

Os modelos foram ajustados utilizando-se o programa SAS (SAS INSTITUTE, 2001), através da PROC MIXED para os modelos A e B, e por meio da macro GLIMMIX para os modelos C, D, E e F.

Para comparação das estimativas dos efeitos fixos e efeitos aleatórios preditos, foram estimadas as correlações de *Spearman* entre os efeitos dos modelos, com o intuito de se verificar o grau de associação entre eles, e também aplicado o teste de permutação de Fisher, para se verificar se os modelos diferem estatisticamente entre si.

Para a verificação da qualidade do ajuste dos modelos C, D, E e F (presença de superdispersão ($\phi > 1$) ou subdispersão ($\phi < 1$), estimou-se o parâmetro de dispersão ($\hat{\phi}$), baseado na estatística de Pearson X^2 generalizada (JFRGENSEN, 1987), dado por:

$$\hat{\phi} = \frac{X^2}{n - p} = \frac{1}{n - p} \sum_{i=1}^n \frac{(\pi_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} ,$$

em que: $V(\hat{\mu}_i)$ é a função de variância estimada sob o modelo que está sendo ajustado, sendo: n igual ao número de observações; p igual ao número de parâmetros envolvidos no modelo; π_i são as proporções observadas de plantas sobreviventes; $\hat{\mu}_i$ são as proporções de plantas sobreviventes estimadas pelo modelo.

As estimativas para o parâmetro de dispersão ($\hat{\phi}$), considerando os modelos C, D, E e F estão apresentadas na Tabela 1.

Tabela 1 – Estimativas do parâmetro de dispersão.

Modelo	Estimativa
C	0,9772
D	0,9794
E	0,9838
F	0,3988

Verifica-se que os valores das estimativas do parâmetro de dispersão, para os modelos C, D e E estão bem próximos de um, valor admitido no modelo binomial, indicando que a variância está consistente com a distribuição assumida, isto é, os dados não fornecem evidências da presença de superdispersão ou subdispersão para os modelos assumidos. Para o modelo F, observa-se que este apresentou uma estimativa muito inferior a um, valor admitido no modelo Poisson, caracterizando a presença de subdispersão, ou seja, a variância não está consistente com a distribuição assumida para este modelo, indicando que o modelo está mal ajustado aos dados. Assim os modelos C, D e E, apresentaram um melhor ajuste em comparação ao modelo F.

Na Tabela 2 estão apresentadas as estimativas dos efeitos fixos considerando-se os diferentes modelos de ajuste.

Pode-se constatar pelas estimativas dos efeitos fixos, que estas apresentaram valores bem próximos entre os diferentes modelos, indicando assim que estes apresentaram ajustes semelhantes, o que também pode ser verificado pela Figura 1, onde estão plotados os valores para as estimativas de efeitos fixos para os modelos ajustados.

A Tabela 3 apresenta os valores das correlações de *Spearman* dos efeitos fixos entre os modelos, onde pode-se verificar altas correlações, confirmando então que esses modelos apresentaram ajustes bastante coincidentes.

Tabela 2 – Estimativas dos efeitos fixos ajustados para os diferentes modelos.

Bloco	Modelo				
	A	B	C	D	E
1	0,7088	0,6713	0,6602	0,6610	0,6645
2	0,6611	0,6488	0,6392	0,6386	0,6356
3	0,8102	0,7560	0,7570	0,7548	0,7498
4	0,7823	0,7402	0,7433	0,7423	0,7381
5	0,7021	0,6824	0,6741	0,6737	0,6729
6	0,7852	0,7331	0,7296	0,7274	0,7211
7	0,7675	0,7350	0,7365	0,7337	0,7282
8	0,7531	0,7186	0,7089	0,7057	0,7004
9	0,7112	0,6859	0,6741	0,6759	0,6812
10	0,8075	0,7699	0,7707	0,7714	0,7720

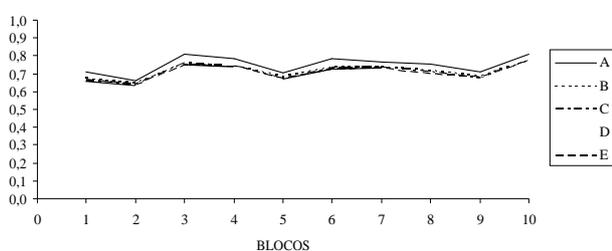


Figura 1 – Estimativas dos efeitos fixos para os diferentes modelos de ajuste.

Observa-se através da Figura 1 uma ligeira diferença entre os valores do modelo A em relação aos demais modelos.

Tabela 3 – Valores da correlação de *Spearman* dos efeitos fixos, considerando os diferentes modelos de ajuste.

Modelo	A	B	C	D	E
A	–	0,9394 (0,0001)	0,9301 (0,0001)	0,9394 (0,0001)	0,9394 (0,0001)
B		–	0,9970 (0,0001)	1,0000 (0,0001)	1,0000 (0,0001)
C			–	0,9970 (0,0001)	0,9970 (0,0001)
D				–	1,0000 (0,0001)
E					–

Os efeitos aleatórios preditos, considerando os diferentes modelos de ajuste, estão apresentados na Tabela 4.

Constata-se pelos resultados dos efeitos aleatórios preditos, apresentados na Tabela 4, que estes, da mesma forma que para as estimativas dos efeitos fixos,

apresentaram valores bem próximos, quando se ajustaram os modelos sob os diferentes enfoques, indicando que esses modelos apresentaram ajustes semelhantes.

Tabela 4 – Efeitos aleatórios preditos ajustados para os diferentes modelos.

Progênie	Modelo				
	A	B	C	D	E
1	0,6892	0,6671	0,6532	0,6531	0,6513
2	0,9404	0,8851	0,8720	0,8787	0,8913
3	0,8705	0,8253	0,8204	0,8232	0,8287
4	0,7661	0,7282	0,7163	0,7172	0,7180
5	0,8880	0,8470	0,8416	0,8457	0,8533
6	0,7991	0,7844	0,7871	0,7887	0,7912
7	0,7414	0,7045	0,6915	0,6920	0,6923
8	0,8133	0,7982	0,7984	0,8004	0,8044
9	0,8619	0,8132	0,8095	0,8117	0,8162
10	0,7560	0,7385	0,7405	0,7404	0,7388
11	0,8490	0,8116	0,8095	0,8124	0,8180
12	0,7575	0,7450	0,7405	0,7414	0,7419
13	0,7579	0,7316	0,7285	0,7292	0,7300
14	0,9693	0,9303	0,9093	0,9180	0,9318
15	0,8665	0,8336	0,8311	0,8344	0,8407
16	0,9029	0,8403	0,8311	0,8372	0,8492
17	0,8151	0,7880	0,7871	0,7892	0,7927
18	0,8800	0,8405	0,8311	0,8348	0,8419
19	0,7953	0,7701	0,7641	0,7663	0,7697
20	0,7118	0,6937	0,6915	0,6911	0,6877
21	0,4361	0,4744	0,4161	0,4187	0,4364
22	0,6602	0,6547	0,6402	0,6408	0,6413
23	0,7611	0,7344	0,7285	0,7292	0,7294
24	0,8019	0,7445	0,7405	0,7416	0,7431
25	0,8578	0,8227	0,8204	0,8224	0,8262

Na Figura 2, estão plotados os valores dos efeitos aleatórios preditos considerando os diferentes modelos de ajuste, onde se pode observar que os modelos apresentaram ajustes bem semelhantes.

Verifica-se também que o modelo A se diferenciou um pouco mais dos demais modelos, apresentando valores ligeiramente maiores para os efeitos aleatórios preditos.

Apesar de o modelo A apresentar valores dos efeitos fixo e aleatório, um pouco maiores em relação aos demais modelos, há recomendações desaconselhando o uso da transformação angular quando as proporções se aproximam de um dos extremos, isto é, menor que 0,20 ou maior que 0,80 (STEEL & TORRIE, 1980). Em casos

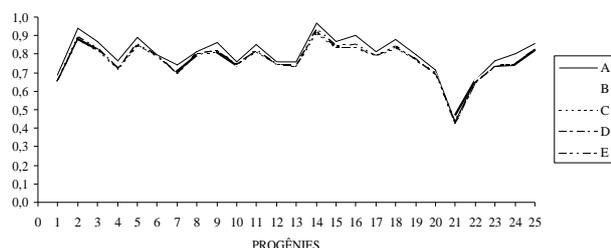


Figura 2 – Efeitos aleatórios preditos para os diferentes modelos de ajuste.

extremos, Bartlett (1947) sugere a substituição de 0 por $1/4$ e n por $n - 1/4$, deixando os outros valores inalterados.

A Tabela 5 mostra os valores das correlações de Spearman dos efeitos aleatórios entre os diferentes modelos de ajuste, quando se observa que os modelos apresentaram altas correlações, indicando assim que os mesmos apresentaram ajustes bastante coincidentes.

Tabela 5 – Valores da correlação de Spearman dos efeitos aleatórios preditos, considerando os diferentes modelos de ajuste.

Modelo	A	B	C	D	E
A	–	0,9746 (0,0001)	0,9709 (0,0001)	0,9790 (0,0001)	0,9785 (0,0001)
B		–	0,9975 (0,0001)	0,9975 (0,0001)	0,9969 (0,0001)
C			–	0,9977 (0,0001)	0,9975 (0,0001)
D				–	0,9998 (0,0001)
E					–

O nível de significância (α) obtidos através do teste de Permutação de Fisher, para comparação das estimativas dos efeitos fixos e aleatórios entre os diferentes modelos de ajuste, estão apresentados nas Tabelas 6 e 7, respectivamente.

Tabela 6 – Nível de significância (α) obtidos na comparação das estimativas dos efeitos fixos, considerando os diferentes modelos de ajuste.

Modelo	A	B	C	D	E
A	–	0,4233	0,2927	0,2699	0,2251
B		–	0,9994	0,9984	0,9953
C			–	1,0000	0,9999
D				–	0,9999
E					–

Tabela 7 – Nível de significância (α) obtidos na comparação das estimativas dos efeitos aleatórios preditos, considerando os diferentes modelos de ajuste.

Modelo	A	B	C	D	E
A	–	0,8315	0,6645	0,7148	0,8003
B		–	0,9990	0,9996	1,0000
C			–	1,0000	0,9993
D				–	1,0000
E					–

Verifica-se pelos resultados apresentados na Tabela 6 e 7, através do nível de significância (α) obtidos, que as estimativas dos efeitos fixos e aleatórios, ajustados para os diferentes modelos, não diferem estatisticamente entre si.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARTLETT, M. S. The use of transformations. *Biometrics*, Washington, v. 3, n. 1, p. 39-52, 1947.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society, B*, [S.l.], v. 26, p. 211-252, 1964.
- JÆRGENSEN, B. Exponential dispersion models: with discussion. *Journal of the Royal Statistical Society, Series B*, [S.l.], v. 49, n. 2, p. 127-162, 1987.
- McCULLAGH, P.; NELDER, J. **Generalized linear models**. 2. ed. London: Chapman and Hall, 1989. 511 p.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society*, [S.l.], v. 135, p. 370-384, 1972.
- SAS INSTITUTE. **Release 8.02**. Cary, 2001.
- SCHALL, R. Estimation in generalized linear models with random effects. *Biometrika*, Washington, v. 78, p. 719-727, 1991.
- SNEDECOR, G. W.; COCHRAN, W. G. **Statistical methods**. 8. ed. Iowa: Iowa State University, 1989. 503 p.
- STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics**. 2. ed. New York: McGraw-Hill, 1980. 633 p.