

Uso de modelos bisegmentados para seleção de genótipos

Luiz Roberto Martins Pinto ¹

Renata Alcarde Sermarini ²

Fabyano Fonseca e Silva ³

Charles Leonel Galvão Sanches ⁴

Carlos Tadeu dos Santos Dias ¹

1 Introdução

Assim como em diversas áreas, nos programas de melhoramento genético as avaliações de interesse podem ser realizadas periodicamente em um mesmo indivíduo, de modo a se obterem informações da eficiência de um determinado tratamento ao longo do tempo, ou mesmo conhecer a resistência/tolerância de genótipos a agentes estressantes, tais como doenças, pragas, déficit hídrico, entre outros.

Quando o interesse consiste em avaliar a resistência de genótipos a agentes estressantes, como por exemplo, porcentagem de plantas mortas, tem-se que o comportamento de tal variável respostas é não linear, e dessa forma, tem-se como opção a utilização dos modelos não lineares para a análise de tais dados. Estes apresentam como principal vantagem a informação direta contida nos parâmetros, uma vez que são, geralmente, interpretáveis biologicamente e a comparação dos genótipos é feita por meio da comparação dos parâmetros.

Uma classe de modelos que pode ser muito útil para a análise de dados dessa natureza é aquela composta pelos modelos de regressão *broken line* bisegmentados, propostos neste trabalho para o ajuste de curvas de dano, tratando-se de porcentagem de plantas mortas. Os modelos bisegmentados têm sido utilizados em estudos de dose-resposta em nutrição animal, para se estimar doses ótimas de determinados nutrientes e economicamente viáveis. Pinto e Dias (2012) sugeriram valores biológicos aos parâmetros destes modelos para o melhoramento genético de plantas, ampliando assim sua utilização. Este trabalho tem como objetivo incluir a classe dos modelos bisegmentos como proposta para seleção de genótipos resistentes a agentes estressantes.

¹UESC/Departamento de ciências exatas e tecnológicas.

²ESALQ/USP - Departamento de Ciências Exatas. e-mail: ralcarde@usp.br

³UFV/Departamento de Estatística.

⁴INSTITUTO FEDERAL BAIANO, Guanambi,BA.

2 Material e métodos

2.1 Material

Sanches et al (2008) avaliaram 10 genótipos de cacau *Theobroma cacao*, denotados por: 2, 10, 16, 25, 51, 1188, 2002, 2007, 2008, 2009, desafiados pelo agente etiológico da Ceratocysti wilt (*Ceratocystis cacaofunesta*), quanto a porcentagem de plantas mortas. O experimento foi instalado seguindo um delineamento casualizado em blocos, sendo 3 o número de blocos e cada parcela foi composta por 5 plantas. As avaliações foram realizadas a cada 4 dias a partir da primeira ocorrência de planta morta no experimento, o que ocorreu 15 dias após a inoculação. Neste trabalho foram utilizadas as observações realizadas entre a primeira e oitava avaliação, que podem ser observados na Figura 1, bem como os respectivo perfil médio para proporção de plantas mortas, por parcela.

Os dados aqui apresentados também foram analisados em Pinto et al (2013), porém fazendo uso de diferentes métodos analíticos.

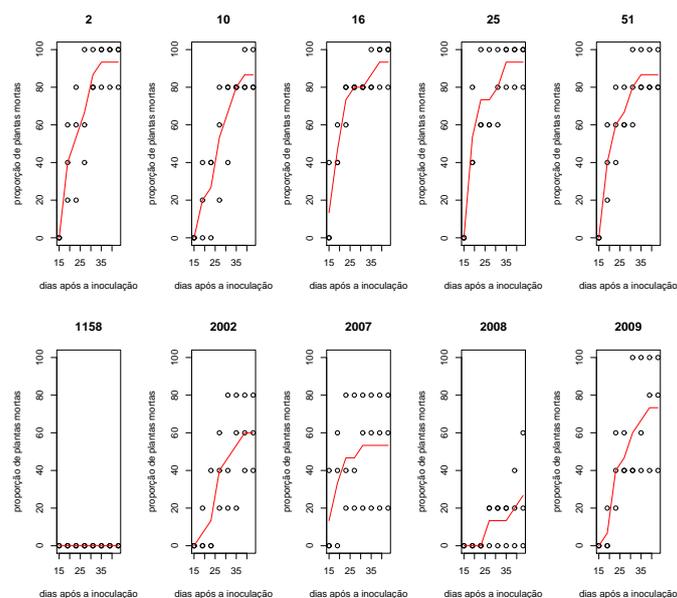


Figura 1: Perfis observados com relação à mortalidade de plantas quando submetidas ao agente etiológico *Ceratocystis cacaofunesta*, considerando-se dez diferentes genótipos de cacau

2.2 Métodos

Inicialmente ajustou-se um modelo para eliminar o efeito de blocos e posteriormente foram utilizados modelos de regressão *broken line* bisegmentados (ROBBINS et al., 1979; PORTZ et al., 2000; ROBBINS et al., 2006; PESTI et al., 2009; PESTI & VEDENOV, 2011) para descrever o comportamento da mortalidade das plantas ao longo do tempo. O ponto de quebra do modelo, ou o *broken line*, é o ponto a partir do qual há uma mudança significativa no comportamento da variável resposta, nesse caso, corresponde ao ponto para o qual houve uma mudança

na mortalidade das plantas. Para o experimento em questão, o modelo é dado por:

$$Y_{ijk} = \alpha_i + \beta_{1i}(\rho_i - x_k) + \beta_{2i}(x_k - \rho_i) + \varepsilon_{ijk},$$

em que $i = 1, \dots, 10$, $j = 1, 2, 3$, $k = 1, \dots, 8$, Y_{ijk} corresponde a porcentagem de plantas mortas observada no i -ésimo genótipo, na j -ésima repetição no k -ésimo tempo eliminado o efeito do bloco, α_i corresponde a ordenada do ponto de quebra para o i -ésimo genótipo, ρ_i corresponde a abscissa do ponto de quebra para o i -ésimo genótipo, β_{1i} é o coeficiente angular da reta para valores de $x_k < \rho_i$ para o i -ésimo genótipo, β_{2i} é o coeficiente angular da reta para valores de $x_k \geq \rho_i$ para o i -ésimo genótipo e ε_{ijk} corresponde ao efeito do acaso.

De modo a captar a grande variabilidade existente entre os pontos de quebra e coeficientes angulares para os diferentes genótipos, foram incluídos efeitos aleatórios no modelo.

Os parâmetros foram estimados pelo método da máxima verossimilhança por meio da função `n.lme` da biblioteca `nlme` (PINHEIRO et al, 2012) para o software estatístico R e a comparação dos modelos foi feita pelo teste da razão de verossimilhanças. Em modelos com as características aqui apresentadas, os valores iniciais apresentados para a estimação dos parâmetros é muito importante e pode influenciar na convergência do algoritmo, para tanto, os valores iniciais foram obtidos pela simples observação dos gráficos na Figura 1.

Tendo ajustado o modelo de regressão *broken line* bisegmentado, procedeu-se ao agrupamento dos genótipos, de tal modo a construir classes de equivalência para avaliar o comportamento dos mesmos. Logo, foi realizada uma análise de cluster fazendo uso dos valores preditos, obtidos pelo ajuste dos modelos de regressão *broken line* bisegmentada por genótipo, por meio da função `cutree` da biblioteca `car`, também do *software* estatístico R. O método utilizado para a análise de cluster foi o método de Ward, avaliado por meio dos critérios de informação AIC e BIC.

A verificação da qualidade do ajuste se deu pela construção de gráficos, entre eles valores observados *versus* valores preditos e gráfico de resíduos.

3 Resultados e discussão

Apenas um termo aleatório foi considerado não significativo no modelo, o termo referente a taxa de crescimento da proporção de plantas mortas no segundo intervalo observado (β_{2i}), ou seja, para $x_k \geq \rho_i$. Portanto, a taxa de crescimento posterior ao ponto de quebra é semelhante com relação a todos os genótipos observados, embora o ponto de quebra varie de acordo com cada genótipo. Na Tabela 1 é apresentado o teste da razão de verossimilhanças para a hipótese ($H_0 : \sigma_{b_2}^2 = 0$), a qual não é rejeitada e o segundo modelo apresentado na tabela, foi o modelo selecionado.

Na Figura 2 são apresentados gráficos para verificação da qualidade de ajuste do modelo selecionado, perfil médio observado *versus* curva ajustada, para cada genótipo avaliado. Sendo

este um indicativo de bom ajuste do modelo de regressão *broken line* bisegmentado.

Tabela 1: Teste da razão de verossimilhanças para o coeficiente angular para $x_k \geq \rho_i$

Modelo	gl	AIC	BIC	logVer	Razão	valor- p
Completo	15	2035,394	2087,603	-1002,697		
$\sigma_{b_2}^2 = 0$	11	2034,210	2072,497	-1006,105	6,816603	0,1459

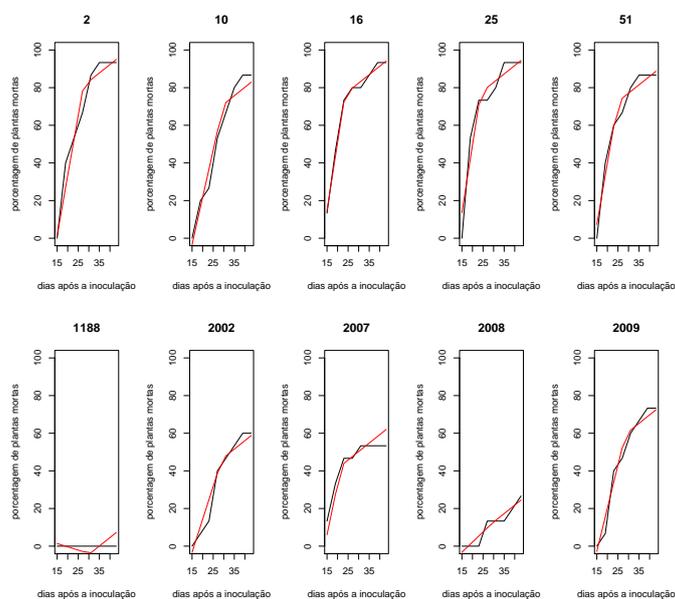


Figura 2: Perfil médio e modelo de regressão *broken line* bisegmentado ajustado para cada genótipo

As estimativas dos parâmetros do modelo de regressão *broken line* bisegmentado, bem como os respectivos erros-padrão, apresentados entre parênteses, são: $\hat{\alpha} = 53,56 (9,33)$, $\hat{\rho} = 26,92 (1,20)$, $\hat{\beta}_1 = -4,65 (0,83)$ e $\hat{\beta}_2 = 0,90 (0,22)$. Na Tabela 2 são apresentados valores preditos por genótipo e o grupo ao qual pertence cada genótipo, obtido pela análise de cluster, indicando assim: os genótipos 1188 e 2008 como resistentes, os genótipos 2002, 2007 e 2009 como moderadamente resistentes e os genótipos 2, 10, 16, 25 e 51 como susceptíveis.

Tabela 2: Valores preditos por genótipo e grupos

Genótipo	Valores preditos			Grupo
	a	r	b_2	
2	27,45	0,52	-1,75	3
10	17,50	2,91	-0,37	3
16	23,11	-3,33	-2,56	3
25	23,72	-3,04	-2,52	3
51	18,99	-1,91	-1,87	3
1188	-57,59	3,51	5,00	1
2002	-7,24	2,24	1,15	2
2007	-10,58	-5,05	-0,72	2
2008	-41,30	2,48	3,58	1
2009	5,94	1,67	0,07	2

4 Considerações finais

Embora os dados deste trabalho apresentem um comportamento característico não linear muitas vezes explicado por modelos como o Logístico, o Gompertz e o de Richards, o modelo de regressão *broken line* bisegmentado forneceu um bom ajuste aos dados. A análise de agrupamento a partir dos valores preditos por genótipo, permitiu uma classificação dos mesmos, em resistentes, susceptíveis e moderadamente susceptíveis, uma vez que foram três os grupos sugeridos pelo método do Ward. Cabe salientar que o ajuste de tais modelos é relativamente simples, quando comparado ao ajuste de modelos não lineares tradicionais, e a obtenção de valores iniciais para os parâmetros pode ser feita por meio de análise visual dos gráficos dos perfis observados, sendo esta uma característica desejada em estudos de regressão.

Referências

- [1] PESTI G.M.; VEDENOV, D. An economic comparison of several models fitted to nutritional response data. **Journal of Animal Science** 89, p. 3344-3349, 2011.
- [2] PESTI, G.M.; VEDENOV, D.; CASON, J.A.; BILLARD, L. A comparison of methods to estimate nutritional requirements from experimental data. **British Poultry Sc** 50 p.16-32, 2009.
- [3] PINTO, L.R.M.; DIAS, C.T.S. Significado biológico dos parâmetros do modelo de regressão ‘broken-line’ bi-segmentado para o melhoramento genético de plantas. In: 57^a Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria, 2012, Piracicaba.
- [4] PINTO, L.R.M.; SANCHES, C.L.G.; DIAS, C.T.S.; LOGUERCIO, L.L. Advantages of multivariate analysis of profiles for studies with temporal variation of treatments effects in plants. **International Journal of Plant Sciences**. 174(1), p.85-96, 2013.
- [5] PINHEIRO, J.; BATES, D.; DEBROY, S.; SARKAR, D. **Linear and Nonlinear Mixed Effects Models**. V.3.1-105, Set. 2012.
- [6] Portz, L.; Dias, C.T.S.; Cyrino, J.E.P. Regressão segmentada como modelo na determinação de exigências nutricionais de peixes. **Scientia Agricola** 57, p. 601-607, 2000.
- [7] Robbins, K.R., Norton, H.W.; Baker, D.H. Estimation of nutrient requirements from growth data. **J Nutr.** 109: p. 1710-1714. 1979.
- [8] Robbins, K.R.; Saxton, A.M.; Southern, L.L. Estimation of nutrient requirements using broken-line regression analysis. **Journal of Animal Science** 84: p. 155-165, 2006
- [9] R Core Team (2012). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [10] Sanches, C.L.G.; Pinto, L.R.M.; Pomella, A.W.V.; Silva, S.D.V.M.; Loguercio, L.L. Assessment of resistance to ceratocystis cacaofunesta in cacao genotypes. **Eur J Plant Pathol** 122, p. 517-528, 2008.