

Modelos Não Lineares com Erros Heterocedásticos

André F. B. Menezes

Universidade Estadual de Maringá

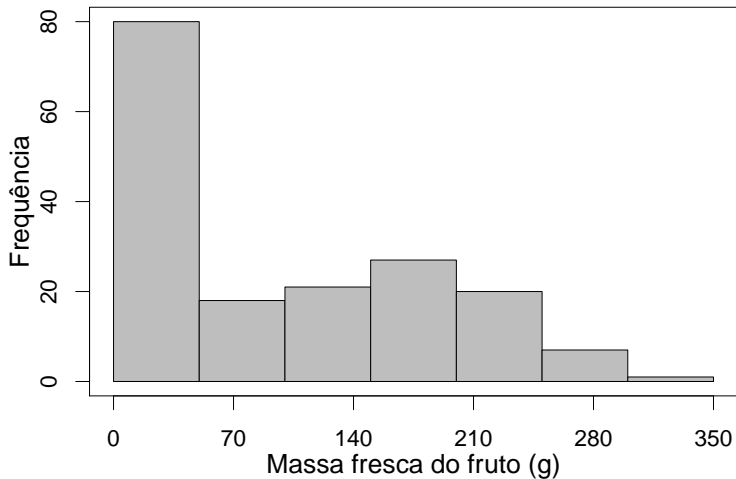
Departamento de Estatística

4 de Agosto de 2017

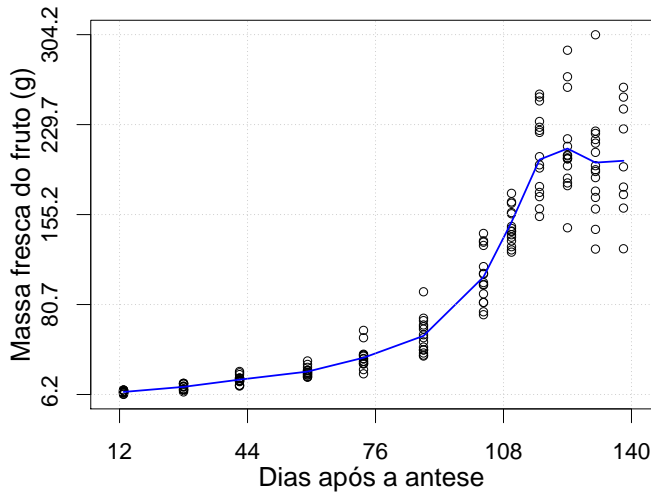
Tópicos

- ▶ Dados retirados em Zeviani et al. (2013).
- ▶ Peso de frutos de goiaba em função dos dias após antese;
- ▶ Autores: diferentes funções de variância;
- ▶ Proposta: distribuição alternativa para os erros aleatórios;
- ▶ Discriminação entre os modelos;
- ▶ Avaliação dos pressupostos.

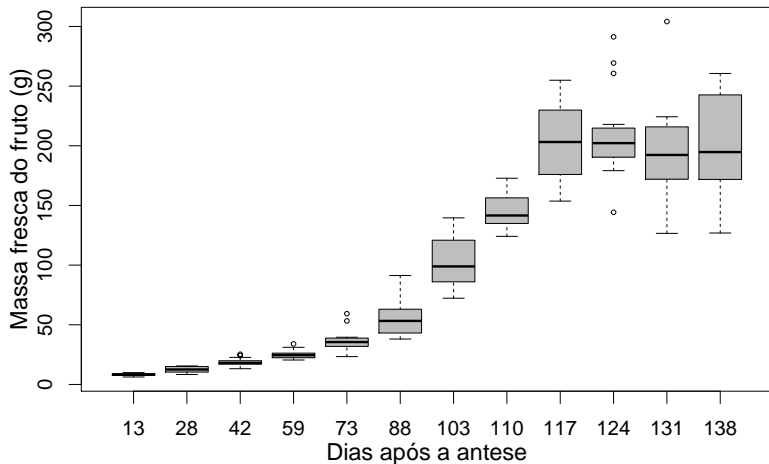
Conjunto de Dados



Conjunto de Dados



Conjunto de Dados



Conjunto de Dados

Medidas descritivas da massa fresca ao longo do tempo.

Dias	Média	Mediana	DP
13	8.33	8.09	1.11
28	12.45	12.55	2.44
42	18.61	17.98	3.58
59	25.04	24.52	3.84
73	36.81	35.54	9.15
88	54.83	53.24	14.63
103	103.60	98.90	22.03
110	144.97	141.59	15.09
117	206.26	203.16	34.29
124	210.01	202.22	37.74
131	195.16	192.31	41.69
138	201.55	194.75	46.25

► Rejeita-se a hipótese nula dos testes de Bartlett e Levene.

Definição

Usualmente a relação entre y e x é descrita por:

$$y = f(x, \theta) + \varepsilon \quad (1)$$

em que

- ▶ $y = (y_1, \dots, y_n)^\top$ e $x = (x_1, \dots, x_n)^\top$;
- ▶ $f(x, \theta)$ é a função esperança;
- ▶ θ é o vetor de parâmetros desconhecidos;
- ▶ ε é o vetor de erros aleatórios.

Erros Normal

- ▶ Distribuição dos erros

$$\varepsilon_i \sim \mathcal{N}(0, \sigma_i) \quad (2)$$

- ▶ Distribuição da resposta condicionada a covariável

$$Y_i | \mathbf{x} \sim \mathcal{N}(f(\mathbf{x}, \boldsymbol{\theta}), \sigma_i^2) \quad (3)$$

- ▶ Função log-verossimilhança

$$\ell(\boldsymbol{\theta}, \sigma_i | \mathbf{y}, \mathbf{x}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_i^2) - \frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} \right)^2 \quad (4)$$

Erros Gumbel

- ▶ Distribuição dos erros

$$\varepsilon_i \sim \text{Gumbel}(0, \sigma_i) \quad (5)$$

- ▶ Distribuição da resposta condicionada a covariável

$$Y_i \mid \mathbf{x} \sim \text{Gumbel}(f(\mathbf{x}, \boldsymbol{\theta}), \sigma_i) \quad (6)$$

- ▶ Função log-verossimilhança

$$\begin{aligned} \ell(\boldsymbol{\theta}, \sigma_i \mid \mathbf{y}, \mathbf{x}) = & -n \log(\sigma_i) - \sum_{i=1}^n \frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} \\ & - \sum_{i=1}^n \exp \left(-\frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} \right) \end{aligned}$$

Metodologia

Modelo para a média (locação)

$$f(\mathbf{x}, \boldsymbol{\theta}) = \theta_1 - (\theta_1 - \theta_2) \exp \{ - \exp [\theta_3 (x - \theta_4)] \} \quad (7)$$

- ▶ θ_1 : assíntota (massa final do fruto);
- ▶ θ_2 : intercepto;
- ▶ θ_3 : proporcional a taxa de crescimento;
- ▶ θ_4 : tempo (dias após a antese) em que ocorre a taxa máxima de crescimento (inflexão).

Modelo para a variância (escala)

$$\sigma_i^2 = \begin{cases} \sigma^2 & \text{(constante)} \\ \sigma^2 \cdot \mathbf{x}_j^{2\lambda} & \text{(potência)} \\ \sigma^2 \cdot \exp \{ 2 \lambda \mathbf{x}_j \} & \text{(exponencial)} \end{cases} \quad (8)$$

Resíduos

- ▶ Resíduo padronizados

$$\hat{\varepsilon}_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}_i} \quad (9)$$

- ▶ Erros Normais

$$\hat{y}_i = f(\mathbf{x}, \boldsymbol{\theta}) \Rightarrow \hat{\varepsilon}_i \sim \mathcal{N}(0, 1) \quad (10)$$

- ▶ Erros Gumbel

$$\hat{y}_i = f(\mathbf{x}, \boldsymbol{\theta}) + \gamma \sigma_i \Rightarrow \hat{\varepsilon}_i \sim \text{Gumbel}(-\gamma, 1) \quad (11)$$

em que $\gamma \simeq 0.5572$ é a constante de Euler.

Comparação e Diagnóstico

- ▶ Critérios de informação AIC e BIC;
- ▶ Gráficos com as curvas ajustadas;
- ▶ Gráficos QQ-Plot para os resíduos;
- ▶ Gráficos dos resíduos versus tempo (covariável);
- ▶ Teste de aderência sobre o resíduos (KS, CvM, AD).

Recursos computacionais

- ▶ Software SAS® 9.4:
 - ▶ PROC NLMIXED: ajuste dos modelos.
- ▶ Software R, versão 3.3.2:
 - ▶ Recursos gráficos.

Resultados

Estimativas e erros padrão para os modelos homocedásticos.

Parâmetro	Normal		Gumbel	
	Estimativa	EP	Estimativa	EP
θ_1	205.7624	3.8869	174.9926	3.2893
θ_2	18.7061	3.3240	17.0740	2.6626
θ_3	0.0937	0.0106	0.0970	0.0121
θ_4	107.3518	0.8229	105.6639	0.7540
σ	24.6528	1.3215	20.9402	1.2269

Resultados

Estimativas e erros padrão para os modelos heterocedásticos

Parâmetro	Normal (\mathbf{x}_j^λ)		Gumbel (\mathbf{x}_j^λ)		Normal $(\exp \{ \lambda \mathbf{x}_j \})$		Gumbel $(\exp \{ \lambda \mathbf{x}_j \})$	
	Estimativa	EP	Estimativa	EP	Estimativa	EP	Estimativa	EP
θ_1	225.6572	11.3297	195.1575	8.3010	262.5170	30.7963	234.6978	26.5578
θ_2	6.7222	0.6569	6.5025	0.6363	6.6483	0.9684	6.3273	0.8400
θ_3	0.0493	0.0035	0.0515	0.0045	0.0445	0.0035	0.0450	0.0039
θ_4	109.9760	2.3351	108.6075	2.0379	117.3240	5.0846	117.4804	5.0985
σ	0.0132	0.0042	0.0146	0.0050	1.0583	0.1667	0.9000	0.1491
λ	1.6371	0.0744	1.5878	0.0797	0.0301	0.0018	0.0305	0.0018

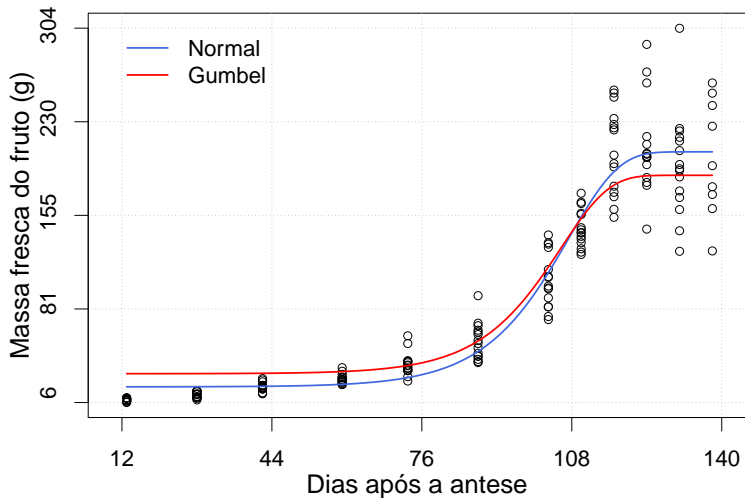
Resultados

Discriminação entre modelos

Critério	σ		$\sigma \cdot x_j^\lambda$		$\sigma \cdot \exp \{ \lambda x_j \}$	
	Normal	Gumbel	Normal	Gumbel	Normal	Gumbel
AIC	1619	1610	1418	1422	1402	1401
BIC	1634	1626	1437	1441	1421	1420

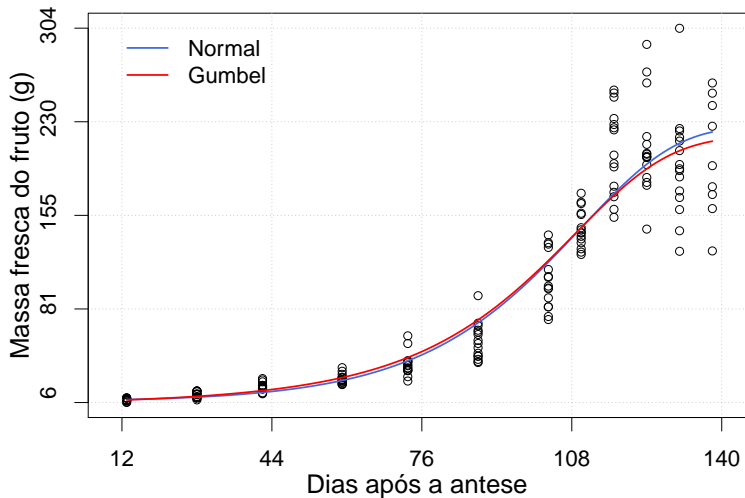
Resultados

Função variância constante



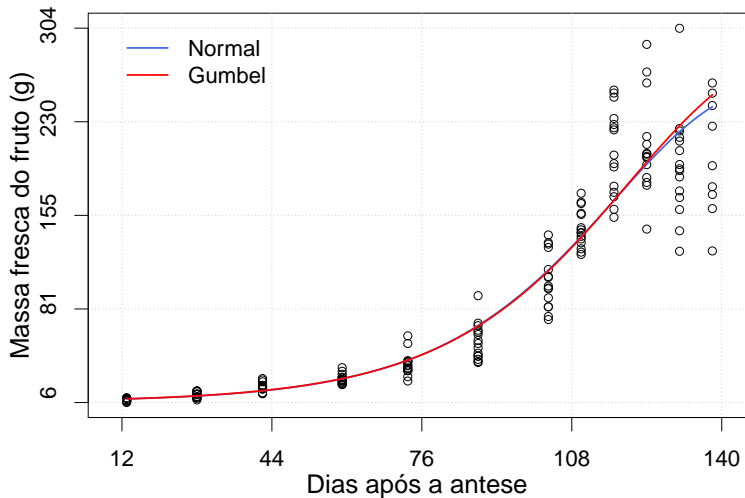
Resultados

Função variância potência



Resultados

Função variância exponencial



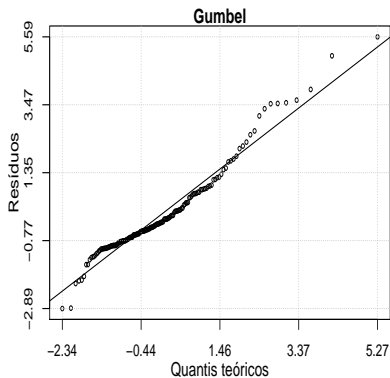
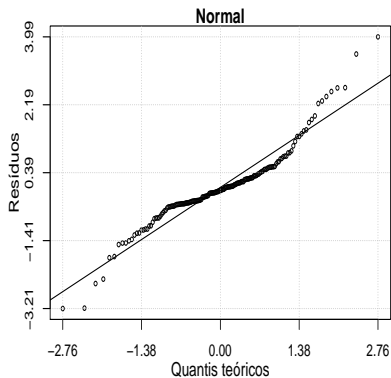
Resultados

Testes de aderência para os resíduos

Teste	σ		$\sigma \cdot x_j^\lambda$		$\sigma \cdot \exp \{ \lambda x_j \}$	
	Normal	Gumbel	Normal	Gumbel	Normal	Gumbel
KS	0.124 (0.010)	0.109 (0.033)	0.050 (0.783)	0.111 (0.027)	0.089 (0.125)	0.073 (0.307)
CvM	0.807 (0.007)	0.515 (0.036)	0.093 (0.621)	0.527 (0.034)	0.188 (0.293)	0.171 (0.333)
AD	4.467 (0.005)	3.037 (0.026)	0.655 (0.597)	2.667 (0.041)	1.108 (0.305)	1.062 (0.326)

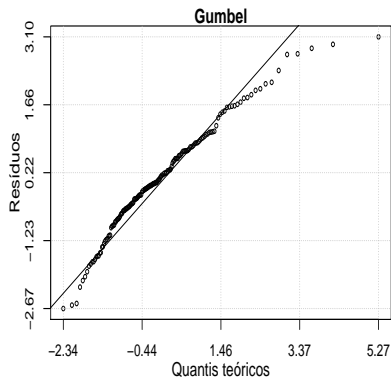
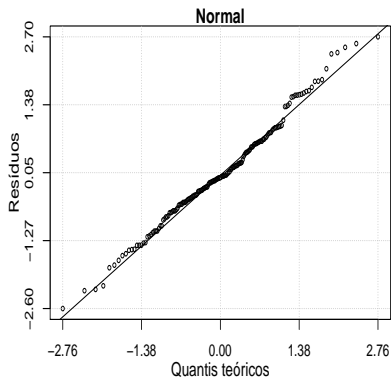
Resultados

Função variância constante



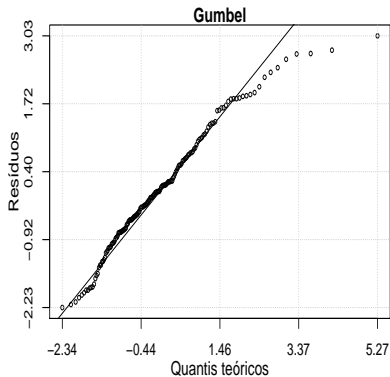
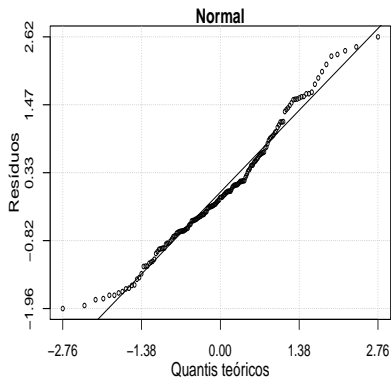
Resultados

Função variância potência



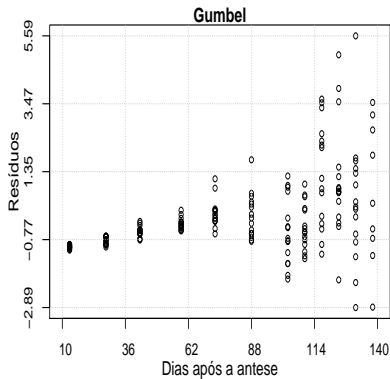
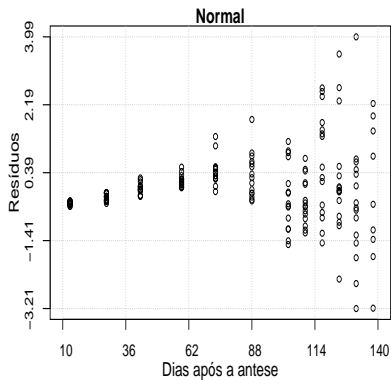
Resultados

Função variância exponencial



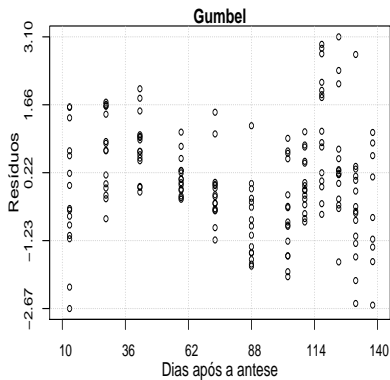
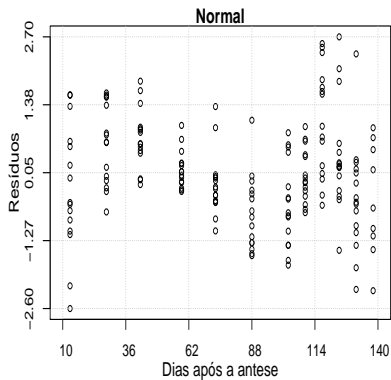
Resultados

Função variância constante



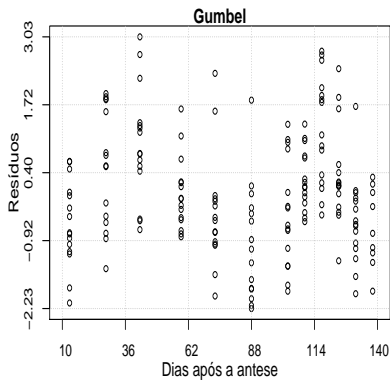
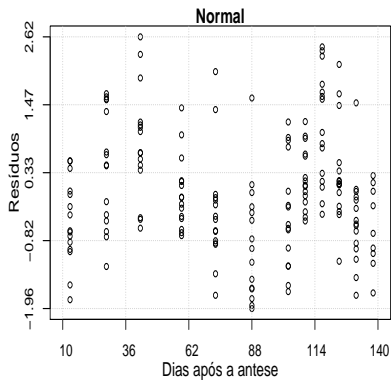
Resultados

Função variância potência



Resultados

Função variância exponencial



Considerações Finais

- ▶ Os modelos homocedásticos diferem bastante conforme a especificação dos erros aleatórios.
- ▶ Os modelos heterocedásticos apresentaram melhor ajuste.
- ▶ As maiores diferenças nas estimativas e erros padrão foram sob o intercepto (θ_1).
- ▶ De acordo com a discriminação e análise de resíduo os modelos com erros normais e Gumbel e função de variância exponencial apresentaram melhores ajustes.

Referências

Referências

- [1] Huet, S.; Bouvier, A.; Poursat, M. A.; Jolivet, E. **Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples**. Springer, 2004
- [2] Pinheiro, J. C.; Bates, D. M. **Mixed-Effects Models in S and S-PLUS**. Springer, 2000
- [3] SAS **The NLMIXED Procedure**, SAS/STAT User's Guide. Version 9.4. Cary, NC: SAS Institute Inc., 2010.
- [4] Zeviani, W.; Júnior, P. J. R.; Bonat, W. H. **Modelos de regressão não linear**. 58° RBRAS e 15° SEAGRO. 2013