

Universidade Estadual de Maringá

Departamento de Estatística

Disciplina: 7210 – Modelos Não Lineares

Professor: Dr. Josmar Mazucheli

Acadêmico: André Felipe Berdusco Menezes

Modelos não lineares com erros heterocedásticos

1 Introdução

Tabela 1: Estimativas do viés (raiz quadrada do erro quadrático médio), $\delta = 0.5$.

γ	n	Estimator of δ			Estimator of γ		
		MLE	BCE	PBE	MLE	BCE	PBE
-1.0	10	-0.1586 (0.5244)	-0.0137 (0.4376)	0.0192 (0.4241)	0.0776 (0.1751)	0.0054 (0.1375)	-0.0109 (0.1334)
	20	-0.0765 (0.3162)	-0.0093 (0.2877)	-0.0024 (0.2859)	0.0347 (0.0984)	0.0013 (0.0863)	-0.0022 (0.0858)
	30	-0.0482 (0.2421)	-0.0045 (0.2274)	-0.0017 (0.2267)	0.0221 (0.0745)	0.0003 (0.0682)	-0.0011 (0.0680)
	40	-0.0356 (0.2046)	-0.0033 (0.1952)	-0.0016 (0.1950)	0.0164 (0.0627)	0.0002 (0.0586)	-0.0005 (0.0586)
	50	-0.0291 (0.1809)	-0.0034 (0.1741)	-0.0023 (0.1740)	0.0131 (0.0548)	0.0003 (0.0519)	-0.0002 (0.0519)
-0.3	10	-0.0461 (0.3943)	-0.0029 (0.3426)	0.0069 (0.3320)	0.0754 (0.1727)	0.0035 (0.1360)	-0.0128 (0.1325)
	20	-0.0192 (0.2507)	0.0008 (0.2343)	0.0027 (0.2329)	0.0338 (0.0985)	0.0004 (0.0868)	-0.0030 (0.0863)
	30	-0.0119 (0.1979)	0.0011 (0.1894)	0.0020 (0.1889)	0.0208 (0.0741)	-0.0009 (0.0681)	-0.0023 (0.0680)
	40	-0.0088 (0.1688)	0.0008 (0.1633)	0.0013 (0.1631)	0.0152 (0.0620)	-0.0009 (0.0582)	-0.0017 (0.0582)
	50	-0.0082 (0.1491)	-0.0005 (0.1452)	-0.0001 (0.1451)	0.0121 (0.0546)	-0.0007 (0.0519)	-0.0012 (0.0519)
0.0	10	0.0020 (0.3834)	0.0018 (0.3355)	0.0019 (0.3251)	0.0778 (0.1736)	0.0056 (0.1359)	-0.0108 (0.1319)
	20	0.0023 (0.2442)	0.0021 (0.2290)	0.0022 (0.2275)	0.0349 (0.0987)	0.0015 (0.0866)	-0.0019 (0.0861)
	30	0.0029 (0.1945)	0.0027 (0.1864)	0.0027 (0.1860)	0.0231 (0.0764)	0.0013 (0.0698)	-0.0001 (0.0696)
	40	0.0038 (0.1647)	0.0037 (0.1595)	0.0037 (0.1593)	0.0171 (0.0635)	0.0010 (0.0592)	0.0002 (0.0592)
	50	0.0034 (0.1463)	0.0033 (0.1426)	0.0033 (0.1426)	0.0136 (0.0558)	0.0008 (0.0528)	0.0003 (0.0527)
0.3	10	0.0425 (0.3919)	-0.0003 (0.3408)	-0.0099 (0.3301)	0.0754 (0.1729)	0.0035 (0.1362)	-0.0129 (0.1325)
	20	0.0168 (0.2461)	-0.0030 (0.2302)	-0.0049 (0.2286)	0.0330 (0.0979)	-0.0003 (0.0864)	-0.0037 (0.0860)
	30	0.0117 (0.1969)	-0.0013 (0.1884)	-0.0021 (0.1880)	0.0217 (0.0743)	-0.0001 (0.0681)	-0.0015 (0.0680)
	40	0.0090 (0.1685)	-0.0006 (0.1630)	-0.0011 (0.1628)	0.0157 (0.0619)	-0.0004 (0.0580)	-0.0012 (0.0579)
	50	0.0071 (0.1503)	-0.0006 (0.1463)	-0.0008 (0.1462)	0.0126 (0.0540)	-0.0003 (0.0512)	-0.0007 (0.0512)
1.0	10	0.1553 (0.5115)	0.0109 (0.4266)	-0.0219 (0.4134)	0.0758 (0.1717)	0.0038 (0.1349)	-0.0126 (0.1311)
	20	0.0728 (0.3118)	0.0058 (0.2843)	-0.0010 (0.2825)	0.0351 (0.0999)	0.0016 (0.0877)	-0.0017 (0.0871)
	30	0.0473 (0.2436)	0.0036 (0.2290)	0.0008 (0.2284)	0.0227 (0.0755)	0.0010 (0.0691)	-0.0005 (0.0689)
	40	0.0341 (0.2078)	0.0018 (0.1986)	0.0003 (0.1985)	0.0166 (0.0635)	0.0005 (0.0594)	-0.0003 (0.0593)
	50	0.0279 (0.1825)	0.0022 (0.1759)	0.0012 (0.1759)	0.0132 (0.0553)	0.0004 (0.0524)	-0.0001 (0.0524)

Tabela 2: EMV e suas versões com correção de viés..

Estimador	γ	δ
MLE	-1.4908 (0.3718)	1.4424 (0.2710)
BCE	-1.3976 (0.3613)	1.3522 (0.2533)
PBE	-1.3932 (0.3568)	1.3443 (0.2492)

O presente trabalho tem por objetivo apresentar um modelo alternativo ao proposto por Zeviani et al. (2013) para modelar o crescimento em massa (g) de frutos de goiaba por meio de modelos não lineares com modelagem da variância. É importante mencionar que as amostras são destrutivas, logo não caracteriza um estudo longitudinal.

A Figura 1 exibe o comportamento da variável resposta (massa fresca da goiaba) independente da do dia, é possível notar uma assimetria à direita.

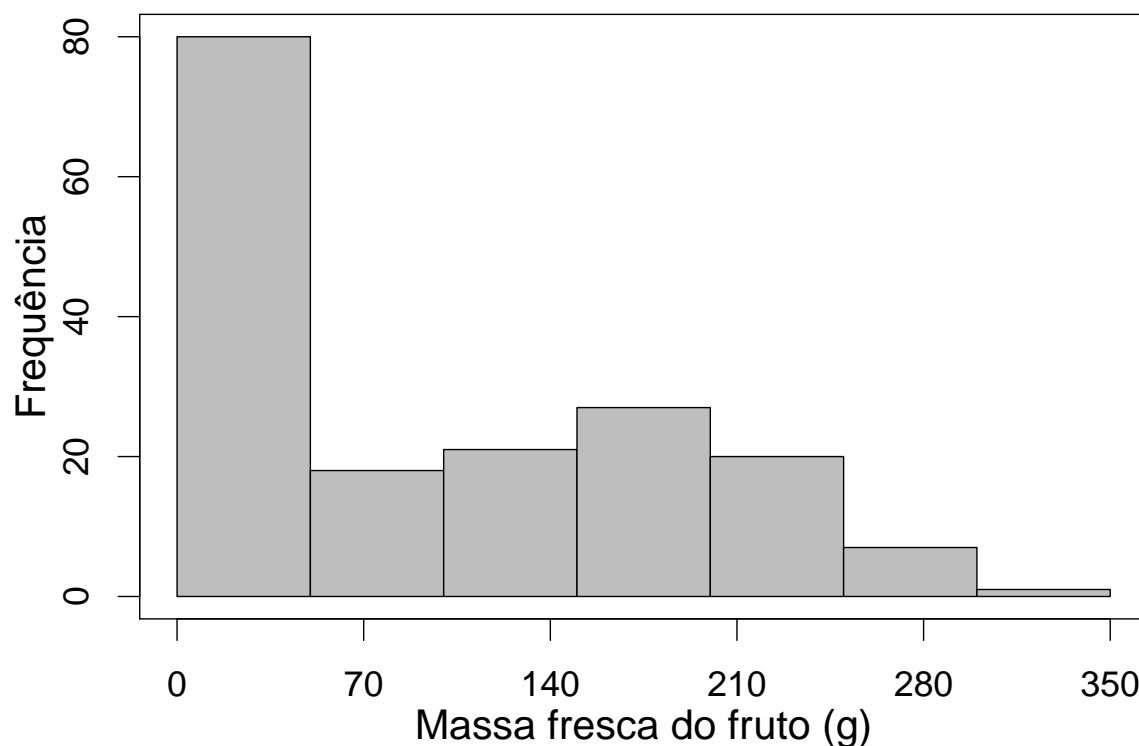


Figura 1: Histograma da massa fresca da goiaba (g).

Uma representação gráfica da variável resposta em função dos dias após a antese é apresentada na Figura 2, a curva representa a função spline. Nota-se um comportamento não linear para o peso assim como variâncias não constantes, este fato evidencia a quebra da usual suposição de homoscedasticidade.

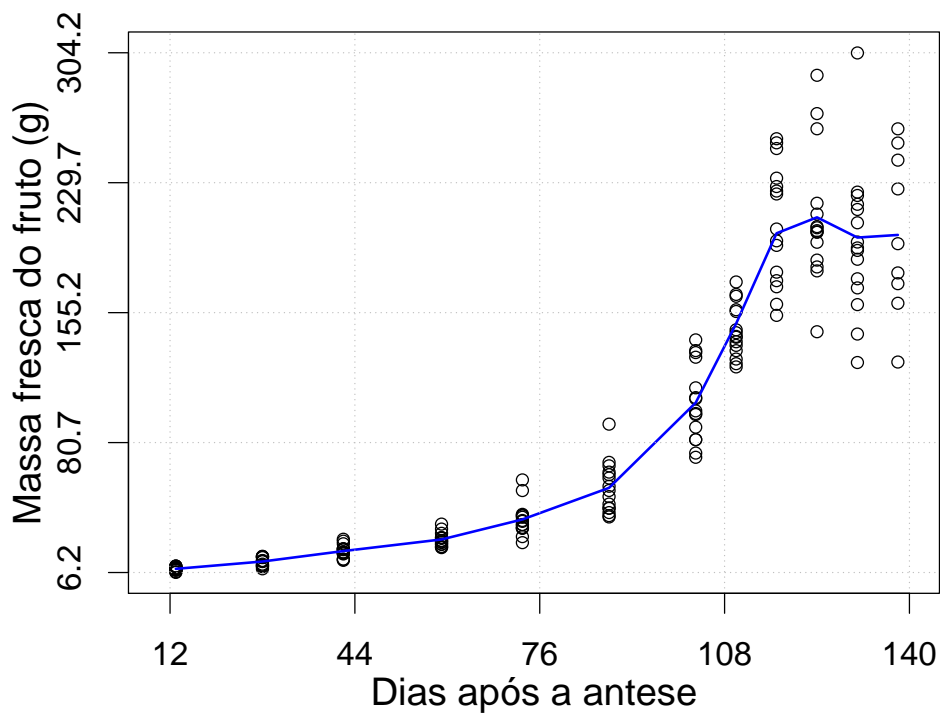


Figura 2: Massa fresca de frutos de goiaba em função dos dias após antese

Na Figura 3 e Tabela 3 é possível notar que as médias bem como as variâncias (desvios-padrões) aumentam com o decorrer dos dias após a antese.

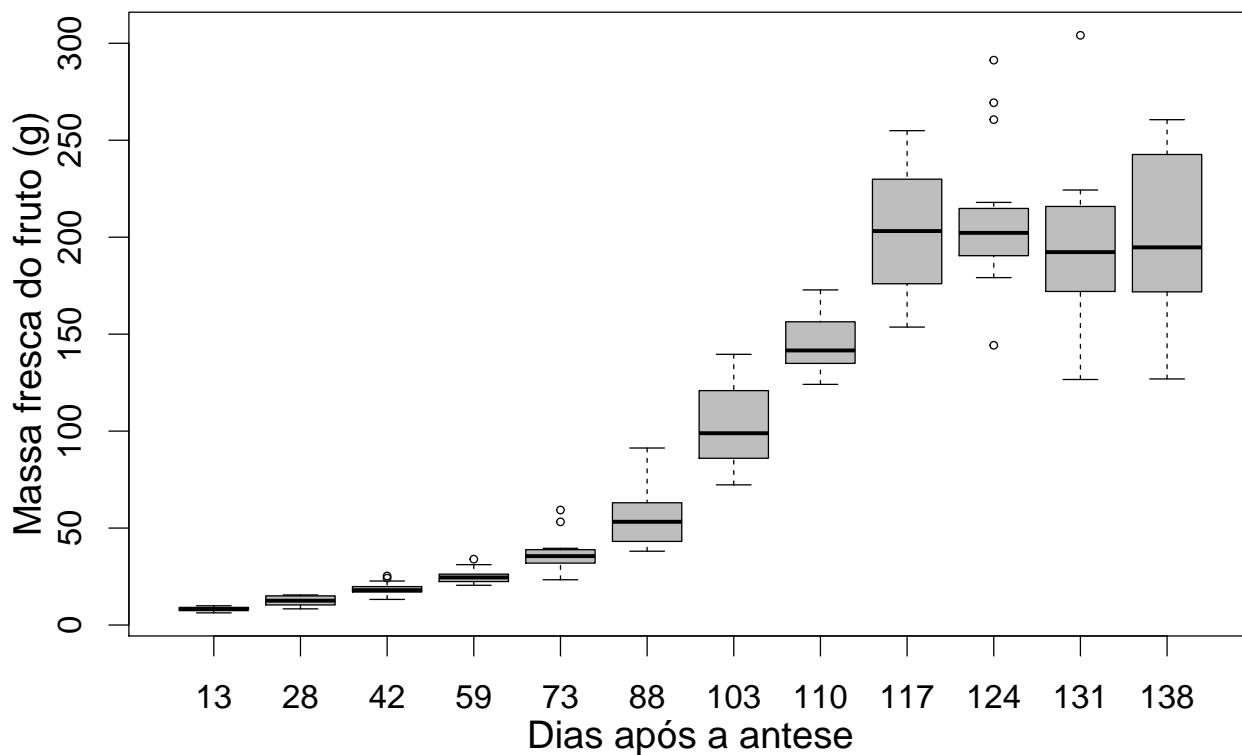


Figura 3: Massa fresca de frutos de goiaba em função dos dias após antese

Este indicativo de heterocedasticidade obtidos pela análise descritiva apresentada nas Figuras 2 e 3 e Tabela 3 também foram avaliados formalmente pelos testes de Bartlett e Levene, em ambos a hipótese nula de homoscedasticidade foi rejeitada (valor- $p < 0.0001$).

Tabela 3: Medidas descritivas da massa fresca ao longo do tempo.

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

Assim como foi realizado por Zeviani et al. (2013) considerarei no ajuste do modelo não linear funções para a variância com intuito de contemplar as evidências amostrais expostas nesta seção. Nos modelos de regressão não lineares é usual considerar que os erros aleatórios seguem distribuição Normal e consequentemente a variável resposta condicionada a covariável também. Alternativamente, proponho considerar que os erros aleatórios seguem distribuição Gumbel (Valor Extremo), assim como a normal essa distribuição pertence a família locação-escala o que facilita na estrutura de regressão. Além disso, ela acomoda casos de assimetria a direita.

2 Metodologia

Os modelos de regressão, em sua forma univariada, são utilizados para descrever o comportamento de uma variável aleatória Y_i , como uma função condicional a(s) variável(is) explicativa(s) \mathbf{x} . Em particular, nos modelos não lineares a relação entre Y_i e \mathbf{x} é descrita por uma função esperança não linear nos parâmetros. O modelo é então expresso por:

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon} \quad (1)$$

em que $\mathbf{y} = (y_1, \dots, y_n)^\top$ e $\mathbf{x} = (x_1, \dots, x_n)^\top$ são os vetores de variáveis resposta e explicativa, respectivamente, $\boldsymbol{\theta}$ é o vetor de parâmetros desconhecidos, $f(\mathbf{x}, \boldsymbol{\theta})$ é uma função das variáveis regressoras e dos parâmetros chamada de função esperança ou função de regressão e $\boldsymbol{\varepsilon}$ é o vetor de erros aleatórios. Em geral, assume-se que os erros são variáveis aleatórias independentes e identicamente distribuídas normais com média $\mathbf{0}$ e variância constante $\sigma_i^2 = \sigma^2 \mathbf{I}$. Uma vez que a distribuição Normal pertence a família locação-escala temos que:

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

Sob o ponto de vista prático a suposição de homoscedasticidade é muito instável, uma alternativa é supor que a variância é função multiplicativa dos níveis da covariável. Assim como os autores Zeviani et al. (2013), vamos considerar as funções potência e exponen-

cial da covariável como alternativa ao modelo homocedástico. Formalmente, define-se as seguintes funções para a variância:

$$\sigma_i^2 = \begin{cases} \sigma^2 & \text{constante} \\ \sigma^2 \cdot \mathbf{x}_j^{2\lambda} & \text{potência} \\ \sigma^2 \cdot \exp\{2\lambda \mathbf{x}_j\} & \text{exponencial} \end{cases} \quad (3)$$

em que j denota os níveis da covariável \mathbf{x} . Note que para $\lambda = 0$, nos casos potência e exponencial, têm-se o modelo homocedástico como caso particular.

Outra suposição que pode não ser atendida é a normalidade dos erros aleatórios e consequentemente a normalidade da variável resposta condicionada aos níveis da covariável (2). Neste sentido proponho uma alternativa, supondo que os erros aleatórios seguem distribuição Gumbel com parâmetro de locação $\mu = 0$ e escala σ . É importante lembrar que a média e a variância da distribuição Gumbel são $\mathbb{E}(X) = \mu + \gamma \sigma$ e $\mathbb{V}(X) = \sigma^2 \pi^2/6$, respectivamente, em que $\gamma \simeq 0.5572$ é a constante de Euler.

Assim como a Normal a distribuição Gumbel também pertence a família locação-escala, dessa forma temos que

$$Y_i \mid \mathbf{x} \sim \text{Gumbel}(f(\mathbf{x}, \boldsymbol{\theta}), \sigma_i), \quad (4)$$

ou seja, a variável aleatória Y_i condicionada ao vetor de covariáveis \mathbf{x} segue distribuição Gumbel com parâmetro de locação $f(\mathbf{x}, \boldsymbol{\theta})$ e escala σ . Logo, a média e variância do vetor de variáveis resposta observado dado a(s) covariável(is) são definidos, respectivamente, por:

$$\mathbb{E}(Y_i \mid \mathbf{x}) = f(\mathbf{x}, \boldsymbol{\theta}) + \gamma \sigma_i \quad \text{e} \quad \mathbb{V}(Y_i \mid \mathbf{x}) = \frac{\sigma_i^2 \pi^2}{6} \quad (5)$$

em que σ_i^2 pode ser uma das três funções definidas em (3).

Inferências para o vetor de parâmetros $\boldsymbol{\theta}$ podem ser realizada sob o paradigma da verossimilhança (ver, por exemplo, Lehmann and Casella, 1998; Pawitan, 2001). Para o caso estabelecido em (2) a função de verossimilhança para $\boldsymbol{\theta}$ e σ_i pode ser escrita na forma:

$$L(\boldsymbol{\theta}, \sigma_i \mid \mathbf{y}, \mathbf{x}) = (2\pi)^{-n/2} (\sigma_i^2)^{-n/2} \exp \left[-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} \right)^2 \right] \quad (6)$$

Assim, a função log-verossimilhança é dada por:

$$\ell(\boldsymbol{\theta}, \sigma_i \mid \mathbf{y}, \mathbf{x}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_i^2) - \frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} \right)^2 \quad (7)$$

Em contrapartida, supondo que os erros são variáveis aleatórias independentes e identicamente distribuídas segundo a distribuição Gumbel (4), então a função de verossimilhança para $\boldsymbol{\theta}$ e σ_i pode ser expressa na forma:

$$L(\boldsymbol{\theta}, \sigma_i \mid \mathbf{y}, \mathbf{x}) = \frac{1}{\sigma_i^n} \exp \left[\sum_{i=1}^n -\frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} - \exp \left(-\frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} \right) \right] \quad (8)$$

Logo, a log-verossimilhança é dada por:

$$\ell(\boldsymbol{\theta}, \sigma_i \mid \mathbf{y}, \mathbf{x}) = -n \log(\sigma_i) - \sum_{i=1}^n \frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} - \sum_{i=1}^n \exp \left(-\frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} \right) \quad (9)$$

Em ambos os casos, as estimativas de máxima verossimilhança, $\hat{\boldsymbol{\theta}}$ e $\hat{\sigma}_i$ de $\boldsymbol{\theta}$ e σ_i são obtidas maximizando (7) ou (9) em relação a $\boldsymbol{\theta}$ e σ_i . Testes de hipóteses e intervalos de confiança para os parâmetros são realizados a partir das propriedades assintóticas dos estimadores de máxima verossimilhança (Lehmann and Casella, 1998).

Para a modelar a massa fresca de frutos de goiaba em função dos dias após a antese considerou-se o modelo Gompertz com função esperança expressa por:

$$f(\mathbf{x}, \boldsymbol{\theta}) = \theta_1 - (\theta_1 - \theta_2) \exp \{ -\exp [\theta_3 (x - \theta_4)] \} \quad (10)$$

em que θ_1 representa a assíntota (massa final do fruto), θ_2 é o intercepto, θ_3 é proporcional a taxa de crescimento e θ_4 é o tempo (dias após a antese) em que ocorre a taxa máxima de crescimento (inflexão).

Resíduos

A análise de resíduos é uma etapa crucial em qualquer análise de regressão, seja linear ou não-linear, pois permite avaliar as suposições empregadas pelo modelo adotado. Nesse sentido, apresento nesta seção os resíduos utilizados para os modelos com erros normais e Gumbel

Erros Normais

Usualmente, assume-se que os erros tem distribuição Normal, ou seja,

$$\varepsilon_i \sim \mathcal{N}(0, \sigma_i) \quad (11)$$

em que σ_i pode ser uma das função definidas em (3). Assim, os erros “preditos”, pois são variáveis aleatórias, comumente denominado de resíduos são definidos por:

$$\hat{\varepsilon}_i = y_i - \hat{y}_i \quad (12)$$

e os correspondente resíduos padronizados são:

$$\hat{\varepsilon}_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}_i} \quad (13)$$

sendo que $\hat{y}_i = \mathbb{E}[Y_i \mid \mathbf{x}]$, que no caso normal resulta em $f(\mathbf{x}, \hat{\boldsymbol{\theta}})$, pois a distribuição do vetor de variáveis aleatórias Y_i condicional ao vetor de covariável \mathbf{x} é tem média definida em (2) tem média $f(\mathbf{x}, \hat{\boldsymbol{\theta}})$.

Assim, sob a suposição de erros normais, os resíduos padronizados, definidos em (13), tem distribuição Normal com $\mu = 0$ e $\sigma = 1$, isto é,

$$\hat{\varepsilon}_i \sim \mathcal{N}(0, 1). \quad (14)$$

Erros Gumbel

Alternativamente, foi proposto neste trabalho outra distribuição para os erros aleatórios, isto é,

$$\varepsilon_i \sim \text{Gumbel}(0, \sigma_i) \quad (15)$$

em que σ_i foi definido em (3). Os resíduos, bem como a versão padronizada deles são definidas como no caso Normal (ver Equações (12) e (13)). No entanto, em virtude da distribuição dos erros ser Gumbel, a distribuição dos resíduos padronizados também será Gumbel, pela propriedade da família de distribuição locação-escala.

Note que, $\hat{y}_i = \mathbb{E}[Y_i \mid \mathbf{x}] = f(\mathbf{x}, \boldsymbol{\theta}) + \gamma \sigma_i$, uma vez que a distribuição Gumbel pertence a família locação-escala, pode-se facilmente mostrar que

$$\hat{\varepsilon}_i \sim \text{Gumbel}(-\gamma, 1) \quad (16)$$

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

Para avaliar a distribuição dos resíduos padronizados, (14) e (16), os testes de aderência de Kolmogorov-Smirnov (KS), Anderson-Darling (AD) e Cramér-von Mises (CvM) foram utilizados. Além disso, o gráfico QQ-Plot para os resíduos padronizados versus os percentis teóricos também foi construído

3 Resultados e Discussões

Tendo em vista os dados sobre o crescimento de frutos de goiaba, em gramas, e a função esperança definida em (10), têm-se nas Tabelas 4 e 5 as estimativas de máxima verossimilhança e os erros padrão assintóticos do vetor de parâmetros $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \sigma, \lambda)$ considerando as diferentes funções de variância e erros aleatórios.

No caso dos modelos com função de variância constante, percebe-se que o modelo com erros Gumbel apresentou estimativas e erros padrão menores para todos os parâmetros, exceto θ_3 . Ressalta-se que o intercepto (θ_2) do modelo Gumbel homocedástico é menor do que o intercepto do modelo com erros normais. Estes resultados estão expostos separadamente na Tabela 4.

Tabela 4: 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

Em relação aos modelos com função de variância potência e exponencial pode-se verificar pela Tabela 5 que os modelos com erros normais apresentaram maiores estimativas e erros padrão para todos os parâmetros com exceção de θ_3 ao ser comparado com o modelo de erros aleatórios Gumbel.

Em particular, o modelo com função de variância exponencial possui maior estimativa do intercepto (θ_1). Este parâmetro é importante pois fornece informações sobre a massa da goiaba antes da maturação da flor. Observa-se que os demais parâmetros são bastante similares dentro da mesma função de variância, ou seja, mesmo supondo diferente distribuições para os erros aleatórios, mas com a mesma função de variância os resultados das estimativas e erros padrão são bem próximos.

As principais diferenças ocorrem sob o parâmetro θ_1 e quando altera a função de variância. Também é importante ressaltarmos que o parâmetro λ , que indica a presença de heterocedasticidade foi significativo para todas as configurações, isto é, conforme o teste assintótico de Wald ele é diferente de zero. As mesmas conclusões foram verificadas por meio do teste da razão de verossimilhanças.

Tabela 5: 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

Uma vez que a inferência foi realizada sob o paradigma da verossimilhança e a mesma é uma medida de compatibilidade da amostra observada com o modelo ajustado, assim foi utilizado neste trabalho os critérios de informação AIC e BIC como forma de discriminar e comparar os modelos. Estes critérios utilizam-se da log-verossimilhança avaliada em suas estimativas e penalizam os modelos pelo seus número de parâmetros e o tamanho da amostra. Assim, o melhor modelo é aquele com os menores valores de AIC e BIC. Os valores destas medidas estão apresentados na Tabela 6. Observa-se que o modelo com função exponencial independentemente da distribuição dos erros apresentou os menores valores de AIC e BIC, indicando um melhor ajuste do que os demais. Curiosamente, o modelo com erros Gumbel apresentou menores valores para o caso homocedástico e função exponencial. No entanto as diferenças entre os modelos com erros normais e Gumbel é mínima, conforme estes critérios.

Tabela 6: Discriminação entre modelos.

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

Nas Figuras 4, 5 e 6 são apresentada as curvas ajustadas dos modelos com diferentes distribuições para os erros e funções de variância. Verifica-se as curvas ajustadas para os modelos homocedásticos diferem significativamente (ver Figura 4) Em contrapartida, os modelos heterocedásticos apresentam curvas similares como mostram as Figuras 5 e 6.

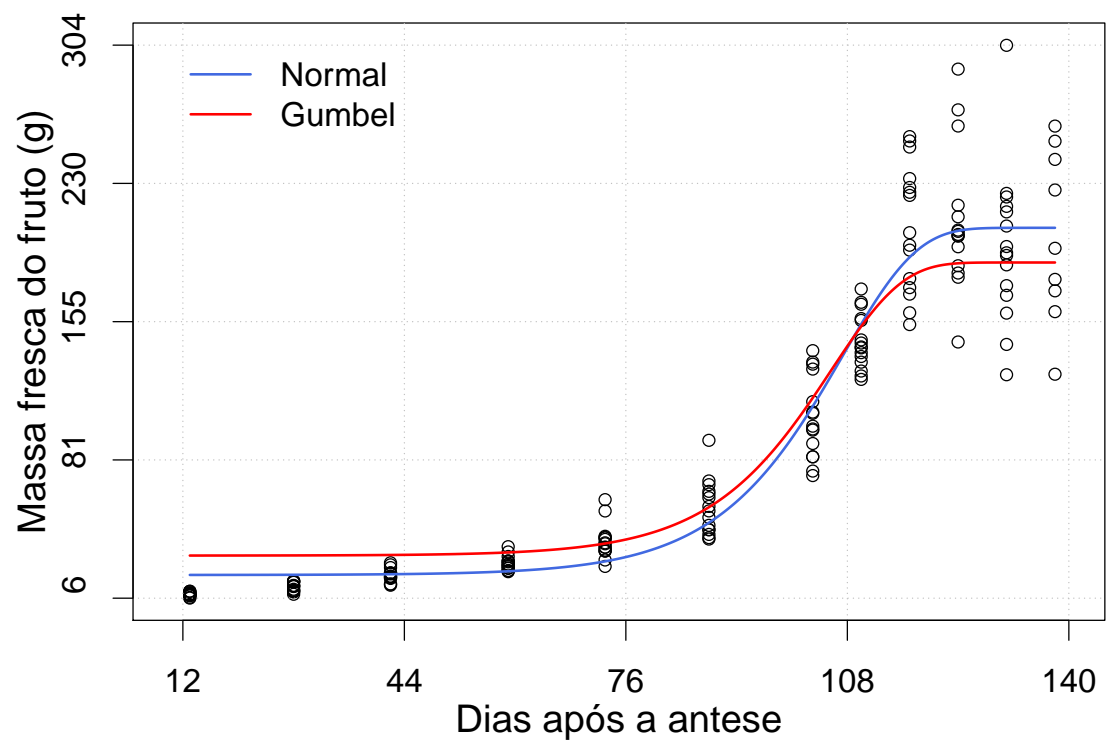


Figura 4: Curvas ajustadas conforme distribuição dos erros e função variância constante.

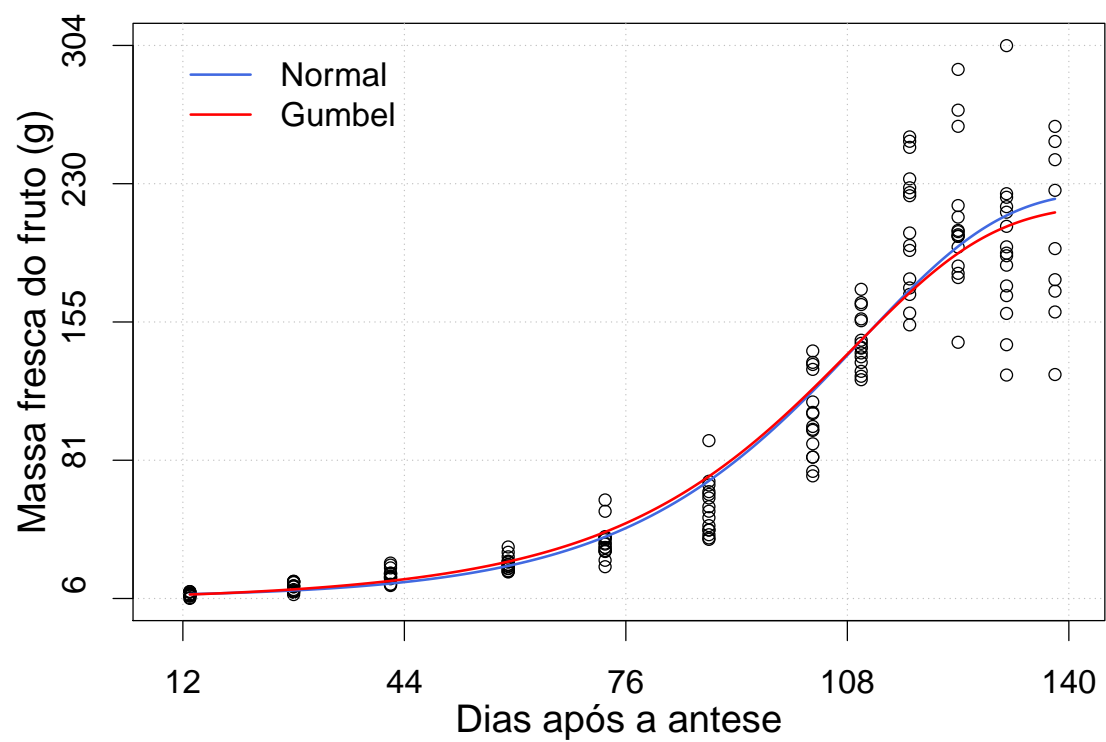


Figura 5: Curvas ajustadas conforme distribuição dos erros e função variância potência.

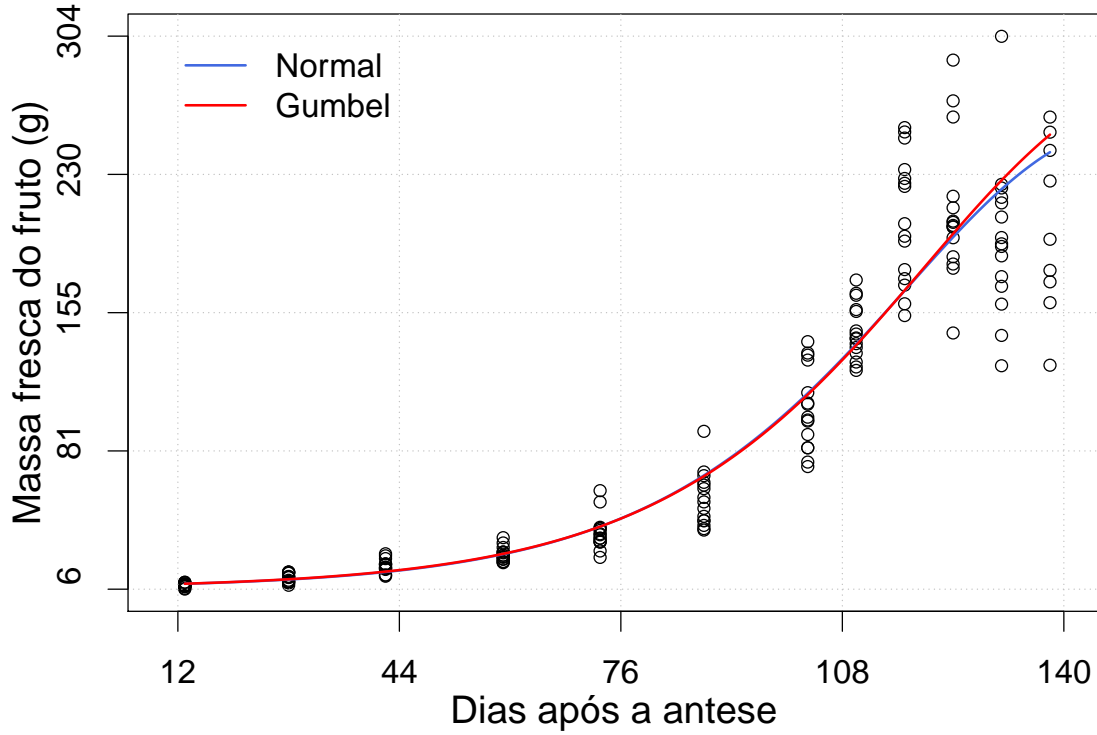


Figura 6: Curvas ajustadas conforme distribuição dos erros e função variância exponencial.

Com o interesse de realizar uma análise de diagnóstico para os modelos com diferentes erros aleatórios e funções para variância, os testes de aderência KS, CvM e AD foram utilizados para comparar os resíduos teóricos com os observados. A Tabela 7 apresenta os resultados dos testes. Verifica-se que no caso de variância constante os resíduos não seguem os pressupostos estabelecidos em ambos os erros. Curiosamente, para o caso em que da função de variância potência, os resíduos do modelo Gumbel não seguem os pressupostos empregado, ou seja, a hipótese nula de todos os testes foram rejeitadas. Por fim, para o caso da variância exponencial nota-se que ambos os resíduos são compatíveis com os percentis teóricos da Normal e Gumbel.

Tabela 7: Estatística (valor- p) dos 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)

Nas Figuras 7, 8 e 9 são apresentados os gráficos QQ-Plots para ambos os resíduos. Ressalta-se que para o caso homocedástico, Figura 7, ambos os resíduos não possuem comportamento adequados, isto é, não formam uma reta com inclinação um e intercepto zero. Já os modelos com as funções de variância potência e exponencial, independente da distribuição dos resíduos, possuem um comportamento aproximadamente igual a reta com inclinação um e intercepto 0.

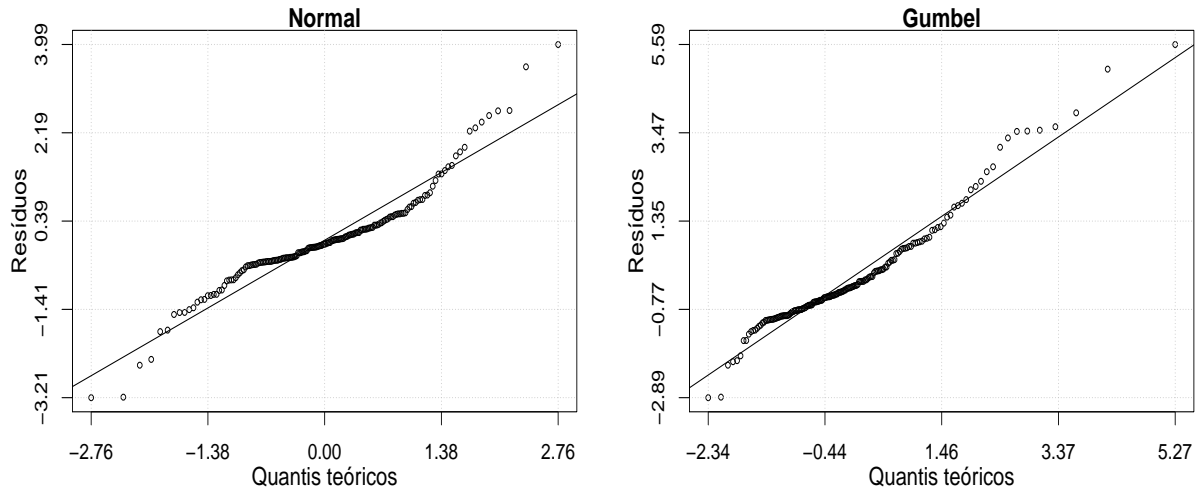


Figura 7: QQ-Plots para os resíduos considerando variância constante.

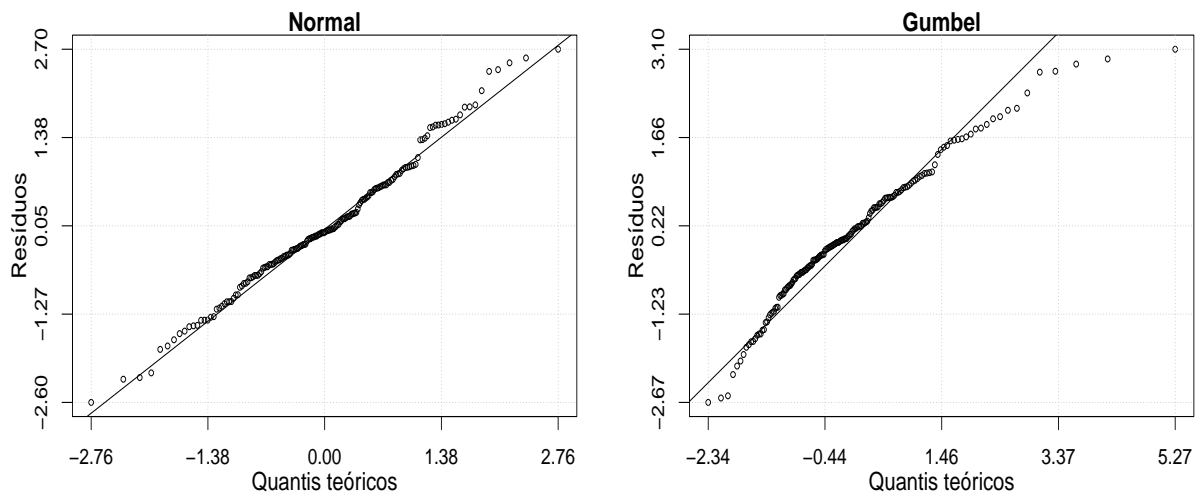


Figura 8: QQ-Plots para os resíduos considerando função de variância potência.

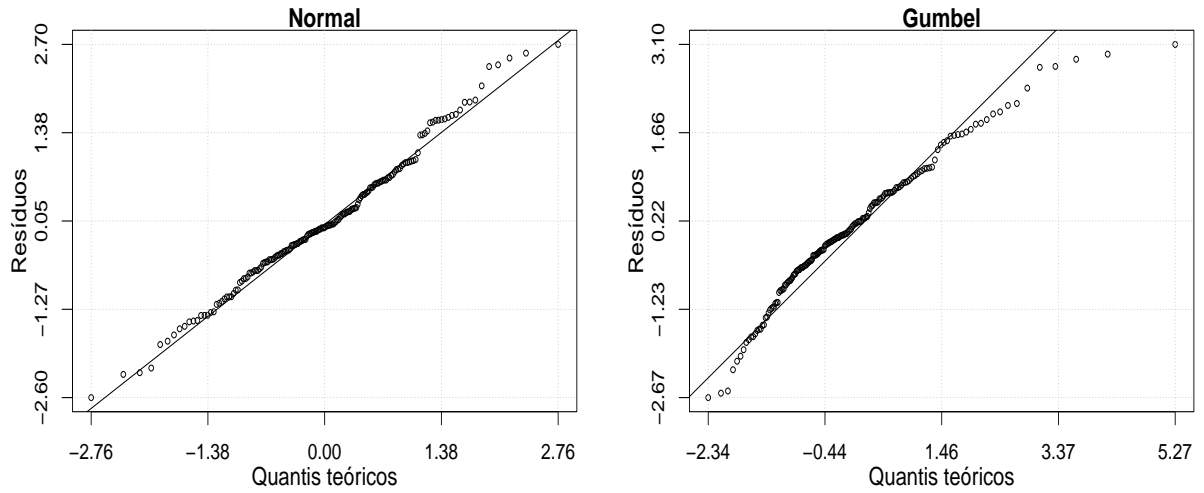


Figura 9: QQ-Plots para os resíduos considerando função de variância exponencial.

Sempre que o tempo ou distância estão envolvidas como covariável na análise de regressão, é prudente checar a suposição de erros independentes. A correlação dos erros pode ser detectada por meio do gráfico do tempo versus os resíduos. Estes gráficos estão expostos nas Figuras 10, 11 e 12. Pelos gráficos dos resíduos padronizados em função dos valores da covariável verifica-se uma disposição de pontos na forma de corneta para os modelos homocedásticos (ver Figura 10) Em contrapartida, utilizando outras funções de variância os resíduos exibem um comportamento aleatório, como era de se esperar.

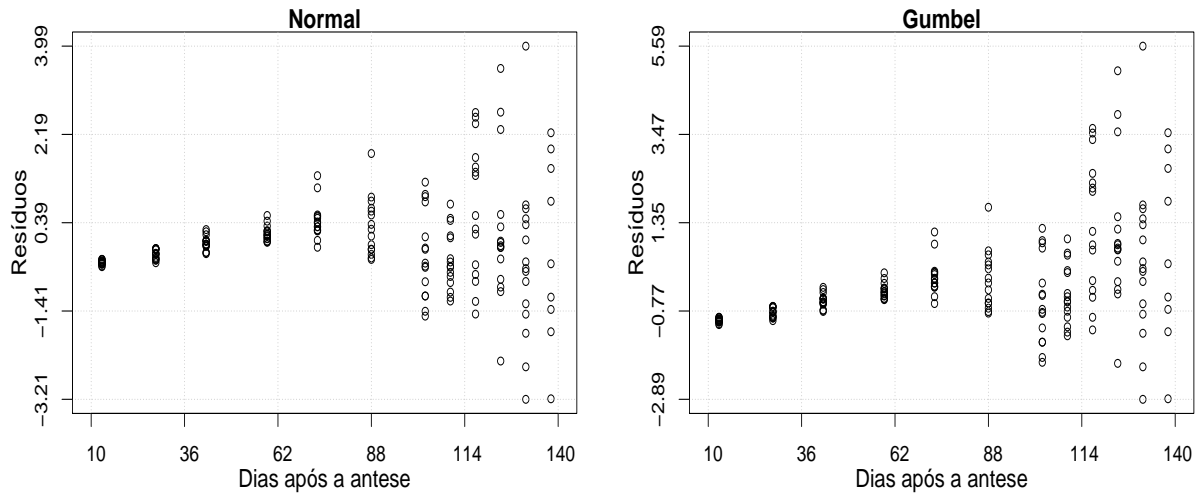


Figura 10: Resíduos versus covariável considerando variância constante.

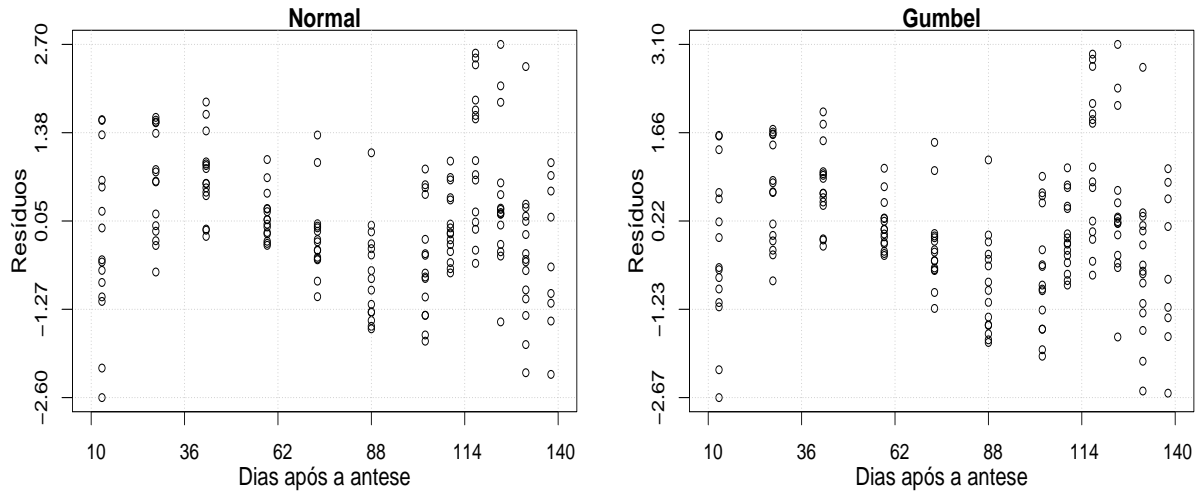


Figura 11: Resíduos versus covariável considerando função de variância potência.

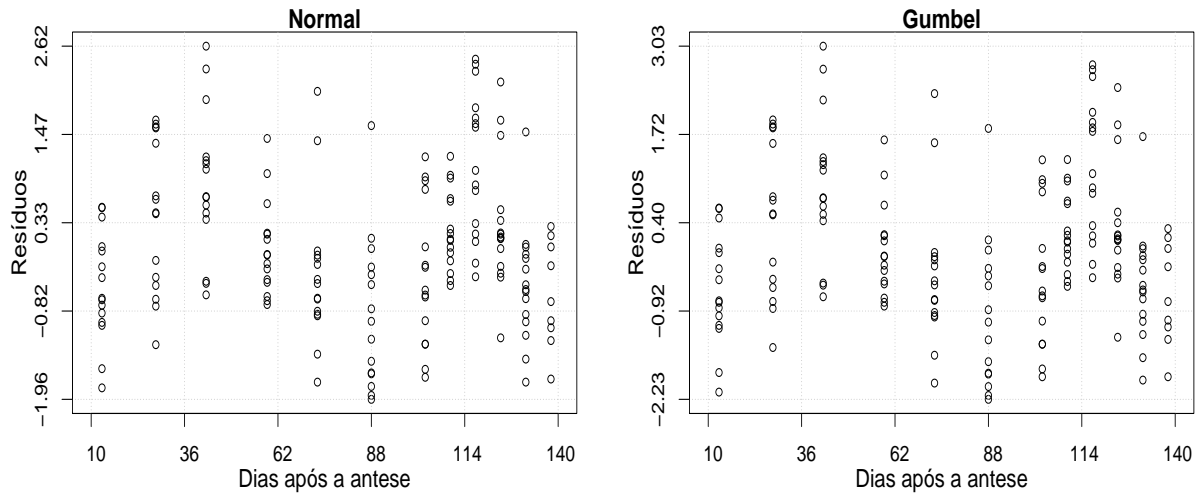


Figura 12: Resíduos versus covariável considerando função de variância exponencial.

Finalmente, considerando os critérios de discriminação bem como a análise de diagnóstico realizada pelos gráficos e testes pode-se concluir que os modelos com função de variância exponencial e erros normais e Gumbel produziram melhores ajustes dentre todos os modelos avaliados, isto significa que este modelo consegue explicar minimamente a variabilidade dos dados. Embora a distribuição dos erros seja diferente ambos os modelos levaram as mesmas conclusões.

4 Considerações Finais

Os modelos não lineares com erros normais homocedásticos são amplamente utilizados na modelagem de crescimento de animais ou frutos. No entanto, na análise destes dados é muito comum a violação da suposição de homoscedasticidade de variâncias, bem como a normalidade dos resíduos. Neste trabalho considerando os dados disponibilizados em Zeviani et al. (2013), relacionado ao crescimento de frutos de goiaba, procedeu-se com o ajuste de diferentes modelos não lineares, sob distintas especificações dos erros aleatórios ε_i e da variância σ_i^2 . Em particular, dado um modelo da forma $y_{ij} = f(x_{ij}, \boldsymbol{\theta}) + \varepsilon_{ij}$, foi adotada a função de regressão Gompertz para $f(x_{ij}, \boldsymbol{\theta})$, assumindo-se que $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2)$ e $\varepsilon_{ij} \sim \text{Gumbel}(0, \sigma_i^2)$, com σ_i^2 sendo constante ($\sigma_i^2 = \sigma^2$), função potência ($\sigma_i^2 = \mathbf{x}_j^\lambda$) e função exponencial ($\sigma_i^2 = \exp\{\lambda \mathbf{x}_j\}$) da covariável dias após a antese. Ressalta-se que a mesma análise foi considerado por Zeviani et al. (2013) com as mesmas funções para variância e considerando os erros aleatórios normais. Nesse sentido, o presente estudo distingue-se dos autores considerando a distribuição Gumbel para os erros aleatórios.

No que tange os resultados, em geral foi verificado que os modelos com erros normais e Gumbel dentro da mesma função de variância forneceram ajustes satisfatório e equivalentes, resultando em mesmas conclusões. A principal diferença entre os modelos com diferentes distribuições dos erros ocorreu nas estimativas e erros padrão do intercepto (θ_1). Para fins de comparação entre os modelos considerou-se os critérios de informação AIC e BIC. Além disso, para avaliar as suposições dos erros, os resíduos dos modelos foram avaliados por meio de métodos gráficos e os testes KS, CvM e AD foram utilizados para comparar os resíduos com os percentis da Normal e Gumbel.

Finalmente, é importante destacar que de maneira oposta a Zeviani et al. (2013) o ajuste dos modelos foram realizada no software SAS via PROC NLIMIXED (SAS, 2010). No apêndice deste texto são apresentadas as linhas de comandos utilizadas.

Referências

- Lehmann, E. J., Casella, G., 1998. Theory of Point Estimation. Springer Verlag.
- Pawitan, Y., 2001. In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press, Oxford.
- SAS, 2010. The NLIMIXED Procedure, SAS/STAT[®] User's Guide, Version 9.4. Cary, NC: SAS Institute Inc.
- Zeviani, W., Júnior, P. J. R., Bonat, W. H., 2013. Modelos de regressão não linear. In: 58° RBRAS e 15° SEAGRO.

Apêndice

Neste apêndice são apresentadas as linhas de comandos do SAS, utilizadas na obtenção das estimativas de máxima verossimilhança.

Código 1: Códigos utilizados no SAS

```
1  /* Importando dados */
2  proc import out = dados replace dbms = tab
3      datafile = '...\goiaba.txt';
4      getnames = yes;
5  run;
6
7  ***** Modelos Homocedasticos *****;
8
9
10 /*Modelo Gompertz com ei ~ Normal(0, sigma) */
11
12 proc nlmixed data = dados tech = tr df = 99999;
13     parms theta1 = 190, theta2 = 18, theta3 = 0.09, theta4 = 107, sigma = 1;
14     bounds sigma > 0;
15     mu = theta1 - (theta1 - theta2) * exp(-exp(theta3 * (daa - theta4)));
16     sigma2i = sigma**2;
17     ll = - 0.5 * log(2*constant("pi")) - 0.5 * log(sigma2i) - 0.5 / (sigma2i) * (peso
18         - mu)**2;
19     model peso ~ general(ll);
20 run;
21
22 /*Modelo Gompertz com ei ~ Gumbel(0, sigma) */
23
24 proc nlmixed data = dados tech = tr df = 99999;
25     parms theta1 = 190, theta2 = 18, theta3 = 0.09, theta4 = 107, sigma = 20;
26     bounds sigma > 0;
27     mu = theta1 - (theta1 - theta2) * exp(-exp(theta3 * (daa - theta4)));
28     Ex = mu + sigma * constant('EULER');
29     zedd = (peso - mu) / sigma;
30     ll = -zedd - exp(-zedd) - log(sigma);
31     model peso ~ general(ll);
32 run;
33
34 ***** Modelos Heterocedasticos *****;
35
36 /*Modelo Gompertz com ei ~ Normal(0, sigma) e funcao variancia potencia*/
37
38 proc nlmixed data = dados tech = quaneu update=bfgs df = 99999;
39     parms theta1 = 190, theta2 = 18, theta3 = 0.09, theta4 = 107, sigma = 1, lambda =
40         1;
41     bounds sigma > 0, lambda > 0;
42     mu = theta1 - (theta1 - theta2) * exp(-exp(theta3 * (daa - theta4)));
43     sigmai = sigma * (daa ** lambda);
44     ll = - 0.5 * log(2*constant("pi")) - log(sigmai) - 0.5 * ((peso - mu) / sigmai )
45         **2;
46     model peso ~ general(ll);
47 run;
48
49
```

```

50 /*Modelo Gompertz com ei ~ Gumbel(0, sigma) e funcao variancia potencia*/
51
52 proc nlmixed data = dados tech = tr df = 99999;
53     parms theta1 = 190, theta2 = 18, theta3 = 0.09, theta4 = 107, sigma = 20, lambda
        = 1;
54     bounds sigma > 0, lambda > 0;
55     sigmai = sigma * (daa ** lambda);
56     mu = theta1 - (theta1 - theta2) * exp(-exp(theta3 * (daa - theta4) ));
57     Ex = mu + sigmai * constant('EULER');
58     zedd = (peso - mu) / sigmai;
59     ll = -zedd - exp(-zedd) - log(sigmai);
60     model peso ~ general(ll);
61 run;
62
63 /*Modelo 4-Gompertz com ei ~ Normal(0, sigma) e funcao variancia exponencial*/
64 proc nlmixed data = dados tech = quanew update=bfgs df = 99999;
65     parms theta1 = 190, theta2 = 18, theta3 = 0.09, theta4 = 107, sigma = 1, lambda
        = 1;
66     bounds sigma > 0, lambda > 0;
67     mu = theta1 - (theta1 - theta2) * exp(-exp(theta3 * (daa - theta4) ));
68     sigmai = sigma * exp(daa * lambda);
69     ll = - 0.5 * log(2*constant("pi")) - log(sigmai) - 0.5 * ( (peso - mu) / sigmai )
        **2;
70     model peso ~ general(ll);
71 run;
72
73 /*Modelo 4-Gompertz com ei ~ Gumbel(0, sigma) e funcao variancia exponencial*/
74 proc nlmixed data = dados tech = tr df = 99999;
75     parms theta1 = 190, theta2 = 18, theta3 = 0.09, theta4 = 107, sigma = 20, lambda
        = 1;
76     bounds sigma > 0, lambda > 0;
77     sigmai = sigma * exp(daa * lambda);
78     mu = theta1 - (theta1 - theta2) * exp(-exp(theta3 * (daa - theta4) ));
79     Ex = mu + sigmai * constant('EULER');
80     zedd = (peso - mu) / sigmai;
81     ll = -zedd - exp(-zedd) - log(sigmai);
82     model peso ~ general(ll);
83 run;

```