Inferência em Recursos Florestais e Ecologia: A Abordagem da Verossimilhança [1]

JOÃO LUÍS F. BATISTA[2]

16 de outubro de 2008

Resumo

Além dos dois paradigmas mais populares na Estatística, o Freqüentista e o Bayesiano, existe uma terceira abordagem que vem ganhando força na área de Ecologia e de Recursos Naturais. Essa terceira abordagem se fundamenta na interpretação da verossimilhança como evidência relativa para se comparar hipóteses e modelos. Ao contrário dos dois paradigmas tradicionais, o paradigma da verossimilhança não procura usar o conceito de probabilidade como medida de evidência, mas busca na verossimilhança uma medida relativa de evidência para construir a inferência estatística, acentada apenas nos dados.

[1]Palestra no ciclo de seminários do Programa de Pós-Graduação em Estatística e Experimentação Agronômica, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, SP.

[2]Laboratório de Biometria Ecológica — Laboratório de Métodos Quantitativos, Departamento de Ciências Florestais, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Campus Piracicaba.

Nessa palestra, a abordagem da verossimilhança é apresentada, mostrando-se suas diferenças das abordagens tradicionais, sendo apresentados dois exemplos de aplicação. O primeiro exemplo mostra como a modelagem da estrutura de tamanho das árvores de florestas tropicais pode ser modelada de uma maneira quase que "natural", sendo simples a comparação de vários modelos alternativos. A construção de modelos empíricos para predição, que é um procedimento comum em Mensuração Florestal é discutida como um segundo exemplo, mostrando que a abordagem tradicional do "Teste Estatístico de Hipótese" é pouco informativa comparada com os métodos de validação que a a abordagem de verossimilhança permite.

Introdução

Dois paradigmas principais dominam o cenário da Estatística atualmente. O *paradigma Freqüentista* se iniciou com o "*teste estatístico de hipótese*" proposto por Neyman e Pearson na década de 30, onde a abordagem de tomada de decisão é aplicada ao problema de teste de hipóteses. Posteriormente, devido a sugestões de Fisher, sugiram os "*testes de significância*" com o conceito de valor-p. O outro paradigma, chamado *Bayesiano*, difere do paradigma Freqüentista ao assumir que os parâmetros dos modelos deve ser considerados como variáveis aleatórias e uma distribuição, *a distribuição a priori*, deve ser assumida para os parâmetros independemente dos dados observados. Apesar da grande diferença entre esses paradigmas, ambos constroem a inferência estatística tomando o conceito de probabilidade como a medida de evidência estatística para se testar hipóteses.

Existe um terceiro paradigma, bem menos popular, mas que vem gradativamente tomando espaço na área de Ecologia e Recursos Naturais. No paradigma da verossimilhança, como foi chamado por Royall (1997), a inferência estatística é construída somente com base no conceito de verossimilhança. Edwards (1992) desenvolveu detalhadamente essa abordagem de inferência, enquanto que Royal

(1997) apresenta a abordagem da verossimilhança e a compara com os outros dois paradigmas, tecendo sérias críticas a esses.

Nessa apresentação, apresentaremos brevemente o paradigma da Verossimilhança e o aplicaremos a dois exemplos da área de Ecologia e Recursos Florestais.

O Paradigma da Verossimilhança

O paradigma da verossimilhança se acenta em dois conceitos fundamentais. O primeiro deles é a **Lei da Verossimilhança** que é de aceitação geral na Estatística, independenemente do paradigma. Já o segundo conceito é o **Princípio da Verossimilhança** que não é aceito pelos outros paradigmas, mas existe forte discussão sobre sua validade. Hacking (1965), no entatno, argumenta que tanto a Lei da Verossimilhança quanto o Princípio são decorrentes de conceitos fundamentais em que qualquer inferência estatística deve se basear.

A Lei da Verossimilhança

Suponha que tenhamos uma variável aleatória X. Estamos interessados em comparar duas hipóteses sobre o comportamento de X: hipóteses A e B. Foi realizado um estudo e foi obtida uma observação de X, digamos que esta observação tem valor x.

O que as hipótese dizem a respeito dessa observação?

- A hipótese A implica que $X = x$ seria observado com probabilidade $p_A(x)$, enquanto

- A hipótese B implica que $X = x$ seria observado com probabilidade $p_B(x)$.

A **Lei da Verossimilhança** afirma que a observação $X = x$ é uma evidência que favorece a hipótese A sobre a hipótese B se e somente se

$$p_A(x) > p_B(x).$$

Mais ainda, a Lei da Verossimilhança implica que e a **Razão de Verossimilhança**

$$\frac{p_A(x)}{p_B(x)}$$

mede a *força* desta evidência em favor da hipótese A sobre a hipótese B.

A Lei de Verossimilhança também pode ser apresentada em termos de *função de verossimilhança*. Assim, para testar a hipótese A de que $\theta = \theta_A$ contra a hipótese B de que $\theta = \theta_B$, podemos dizer que a observação $X = x$ favores a hipótese A se e somente se

$$\mathcal{L}(\theta_A|x) > \mathcal{L}(\theta_B|x),$$

onde $\mathcal{L}(\theta|x)$ é o valor da função de verossimilhança para o parâmetro θ dada a observação x.

O Princípio da Verossimilhança

O princípio da verossimilhança é mais diretamente estabelecido em termos da *função de verossimilhança*. Suponha que se deseja testar a hipótese A de que $\theta = \theta_A$ contra a hipótese B de que $\theta = \theta_B$ e de que para isso se dispõe de uma observação $X = x$. Igualmente se deseja testar a hipótese de C de que $\beta = \beta_C$ contra a hipótese D de que $\beta = \beta_D$ e que para isso se dispõe de uma observação $Y = y$. O princípio da verossimilhança estabelece que se

$$\frac{\mathcal{L}(\theta_A)}{\mathcal{L}(\theta_B)} = \frac{\mathcal{L}(\beta_C)}{\mathcal{L}(\beta_D)},$$

então a observação $X = x$ referente ao confronto da hipótese A *vis-à-vis* a hipótese B é equivalente à observação $Y = y$ referente ao confronto da hipótese C *vis-à-vis* a hipótese D.

O princípio da verossimilhança afirma que duas observações que geram funções de verossimilhança idênticas são equivalentes em termos de evidência. Ou seja, a *evidência contida nos dados* a respeito de uma hipótese é *totalmente caracterizada* pela função de verossimilhança.

Outros Elementos do Paradigma

O princípio da verossimilhança é o ponto central de discussão do paradigma da verossimilhança. A aceitação desse princípio gera as seguintes implicações:

- A evidência dos dados é sempre relativa, isto é, só é possível comparar uma hipótese contra outra.

  Não é possível fazer inferência a respeito de uma única hipótese.

- A completa especificação do modelo estocástico é necessária para se obter a função de verossimilhança, sem ela não é possível fazer inferência.

- **Nada mais** é necessário além da especificação completa do modelo estocástico e dos dados observados para se fazer inferência, logo não são necessários:

  - a definição de um *espaço amostral* dos dados (fundamento freqüentista), pois é estatisticamente irrelevante a forma *como* os dados foram coletados ou quais dados emphpoderiam ter sido observados;
  - a definição de uma *distribuição a priori* para os parâmetros do modelo (fundamento bayesiano), pois é estatisticamente irrelevante o que o pesquisador pensa ou acredita sobre os dados.

Nessa breve apresentação do paradigma muitos outros elementos estão ausentes, mas para um aprofundamento no assunto acreditamos ser mais pertinente a consulta à literatura.

Literatura sobre o Paradigma da Verossimilhança

As referências mais recentes são os livros de Edwardas e Royall. O livro de Edwards (1992), cuja primeira edição data de 1972, desenvolve uma teoria da inferência baseada na verossimilhança de modo detalhado e completo. Já Royall (1997), apresenta o paradigma nos seus elementos principais e o contrasta com os outros dois paradigmas, tecendo duras críticas à abordagem freqüentista. Hacking (1965) discute a fundamentação lógica da Estatística mostrando como a lei e o princípio da verossimilhança são a base para uma boa inferência estatística, justificando assim a abordagem da verossimilhança. Já Berger e Wolpert (1988) apresentam em profundidade o princípio da verossimilhança, discutindo suas diferenças da abordagem freqüentista, novamente com vantagens para o princípio da verossimilhança.

Exemplos da abordagem da verossimilhança na área de Ecologia e Recursos Naturais podem ser encontradas em Hilborn e Mangel (1997) e em Burnham e Anderson (2002).

Exemplo 1: Estrutura de Tamanho de Florestas

A estrutura de tamanho das árvores numa florestal, ou simplesmente *a estrutura da floresta*, é representada pela distribuição das medidas dos DAP (diâmetros à altura do peito) dos troncos das árvores. A distribuição de DAP releva importantes aspectos de uma florestas e de seu funcionamento. Por exemplo, florestas plantadas tendem a ter sempre uma distribuição unimodal, às vêzes bimodal, enquanto que as florestas nativas tendem sempre a uma distribuição amodal semelhante à distribuição exponencial.

O problema

Nesse exemplo, a questão central é como modelar a estrutura de florestas nativas, o que envolve dois aspectos principais:

Qual a distribuição mas apropriada? Qual distribuição probabilística é apresentada como melhor modelo dos dados.

Um mesmo modelo é o melhor em qualquer grau de agregação? O modelo melhor modelo da estrutura da floresta é o melhor para todas as áreas, ou a variação da estrutura da floresta implica em modelos distintos para locais distintos.

Os Dados

Os dados se referem a 87 parcelas de floresta ombrófila do Maranhão, totalizando 11992 árvores, sendo que o número de árvores por parcela foi sempre superior a 47.

Os Modelos

O modelo mais tradicional para florestas nativas é a distribuição exponencial que estabelece que a razão das freqüências de classes diamétricas adjacentes é constante:

$$f_E(x) = \lambda\, \mathrm{e}^{-\lambda x}, \qquad 0 \leq x \leq \infty, \quad \lambda > 0$$

Um modelo mais flexível para a estrtura da floresta é a distribuição Weibull que implica em uma estrutura mais variável na freqüência das árvores por classes de tamanho:

$$f_W(x) = \frac{\gamma}{\beta^\gamma} (x-\alpha)^{\gamma-1} \exp\left(-\frac{(x-\alpha)^\gamma}{\beta^\gamma}\right), \qquad 0 \leq x \leq \infty, \quad \alpha > 0, \beta > 0, \gamma > 0$$

onde:

α — é o parâmetro de localização. Nesse caso é conhecido: $\alpha = 15cm$, pois é o de DAP mínimo para medição das árvores.

β — é o parâmetro de escala.

γ — é o parâmetro de forma. A distribuição exponencial pode ser considerada um caso particular da Weibull, onde $\gamma = 1$.

Comparação de Modelos

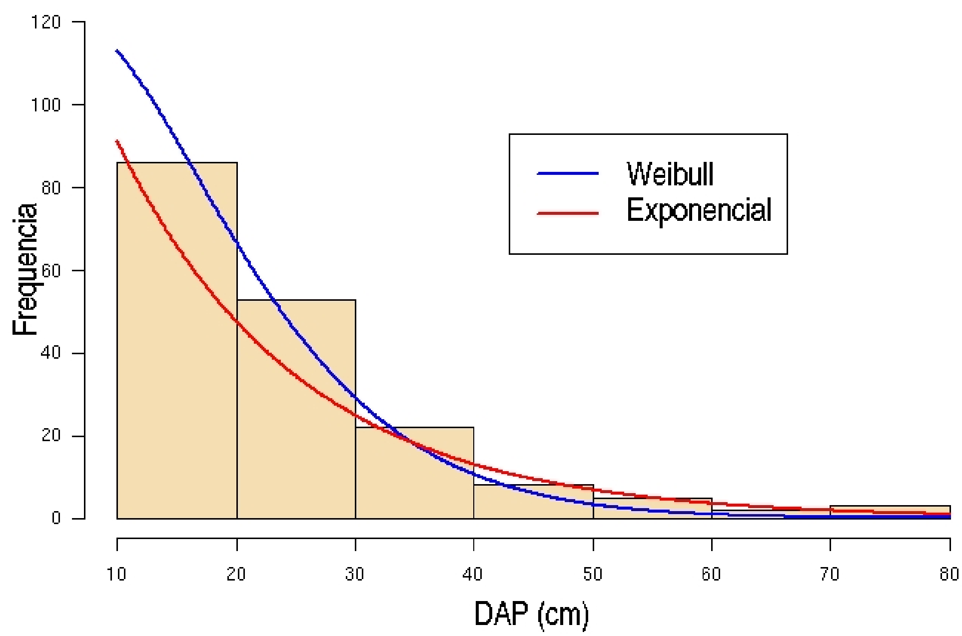
O critério de comparação de modelos é o **Critério de Informação de Akaike**, que penaliza na log-verosimilhança negativa do modelo o número de parâmetros.

$$\text{AIC} = -2\,\ln\left[\mathcal{L}(\text{modelo})\right] + 2\,p$$

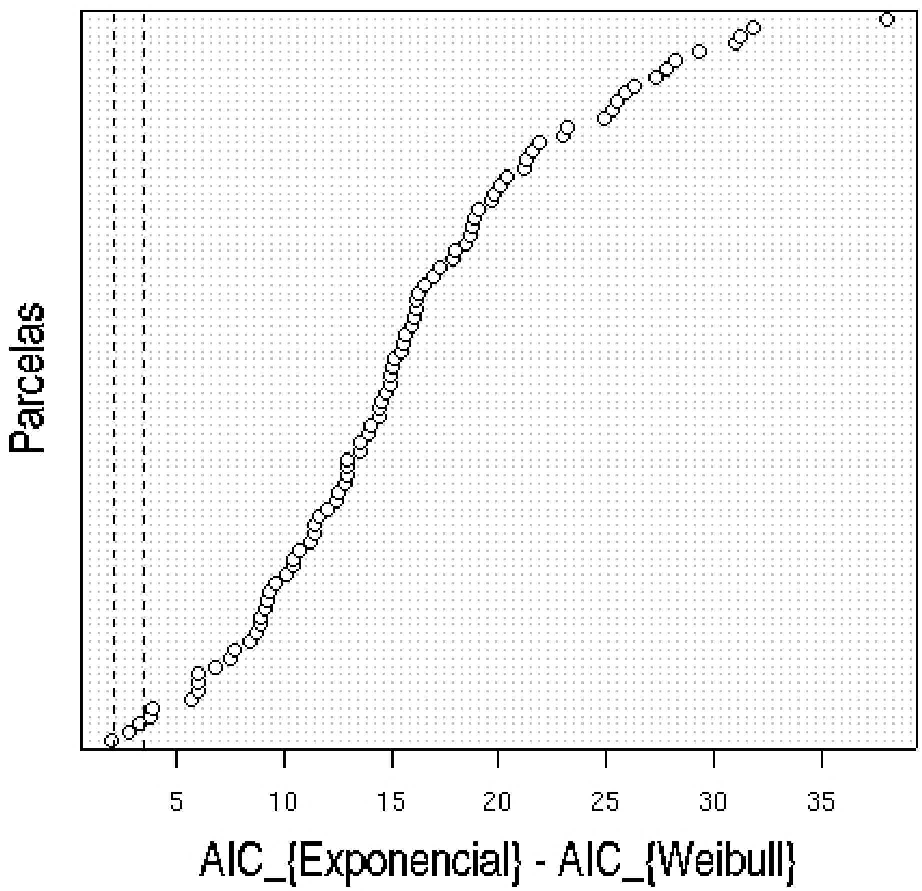
onde $\mathcal{L}(\text{modelo})$ é a função de verossimilhança do modelo e p é o número de parâmetros do modelo.

Resultados

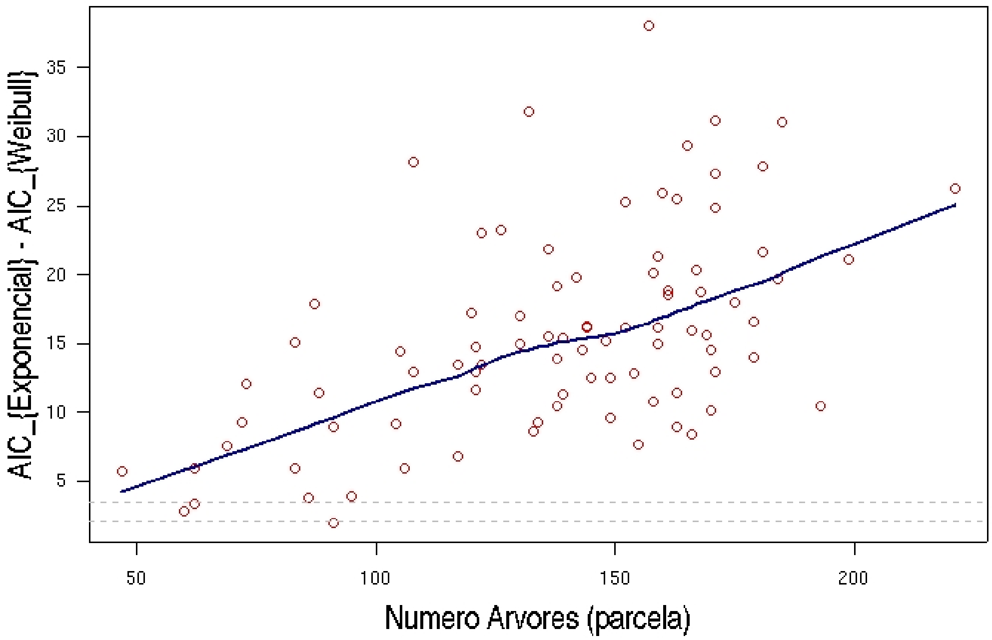
Quando os modelos são ajsutados a todas as árvores, ignorando-se a informação de parcela, a comparação mostra um comportamento bastante distinto dos modelos, com vantagem da distribuição Weibull (menor AIC).



Quando os modelos são ajsutados para cada parcela, constata-se que de fato a distribuição Weibull é superior à distribuição exponencial. Somente em três parcelas a distribuição exponencial teve um desempenho próximo à distribuição Weibull.



Verifica-se que quanto maior o número de árvores na parcela, maior a superioridade da distribuição Weibull, o que é uma forte evidência de que ela é um modelo superior para esse tipo de dado.



Exemplo 1: Conclusão

Ao utilizar a abordagem da verossimilhança a comparação de modelos pode ser facilmente utilizada pelo critério de Akaice, e essa comparação pode ser realizada em diferentes níveis dos dados ou com diferentes graus de agregação. A comparação global é sempre possível, seja diretamente, seja pela agregação dos resultados parciais a cada nível de agregação.

Exemplo 2: Modelos Empíricos em Mensuração Florestal

Tipo de Modelo: Equação de Volume

Equação de volume é um modelo empírico utilizado para a predição do volume de madeira de árvores em pé. O modelo é desenvolvido com base em árvores abatidas onde o volume pode ser medido e relacionado a medidas que podem ser obtidas da árvore em pé, geralmente o diâmetro à altura do peito (DAP) e a altura total da árvore.

O ajuste de equações de volume, como os modelos empíricos em geral, segue tradicionalmente a abordagem de regressão linear ou não-linear.

Conjunto de Dados

Nesse exemplo utiliziaremos dados de 1514 árvores de florestas plantadas de *Eucalyptus grandis* da região central do Estado de São Paulo.

Estatísticas descritivas das **1514** árvores.

VARIÁVEL	MÍNIMO	MEDIANA	MÉDIA	MÁXIMO
Idade (anos)	1.90	3.30	4.20	14.40
DAP (cm)	3.82	10.82	11.43	38.20
Altura (m)	5.40	16.72	17.31	44.91
Volume (dm^3)	4.59	69.95	115.87	2012.11

Modelo Spurr

Um dos modelos trandicionais de equação de volume tem o nome de *modelo Spurr* em homenagem ao florestal norte-americano Stephen H. Spurr, sendo um modelo linear simples:

$$v_i = \beta_0 + \beta_1 \left(d_i^2 \, h_i \right) + \varepsilon_i$$

onde:

v_i — volume das árvores,

d_i — DAP das árvores,

h_i — altura total das árvores,

$\beta_0,\ \beta_1$ — coeficientes de regressão,

$\varepsilon_i \sim N(0, \sigma^2)$ — erro aleatório.

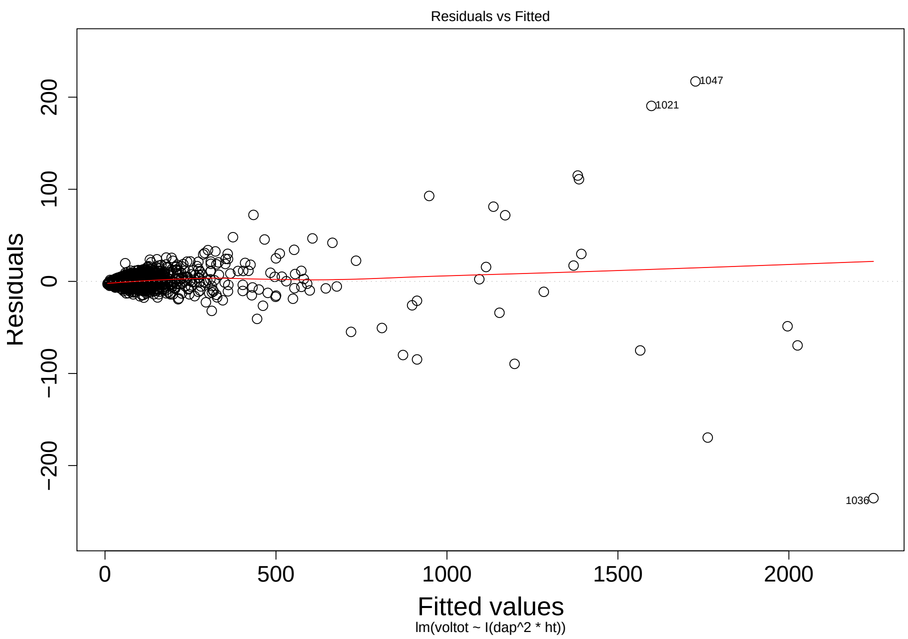
Nessa forma de apresentação de modelo, a maior ênfase é colocada na *relação funcional* entre variável resposta e variável ou variáveis preditoras. Nesse caso, a relação é uma relação linear simples.

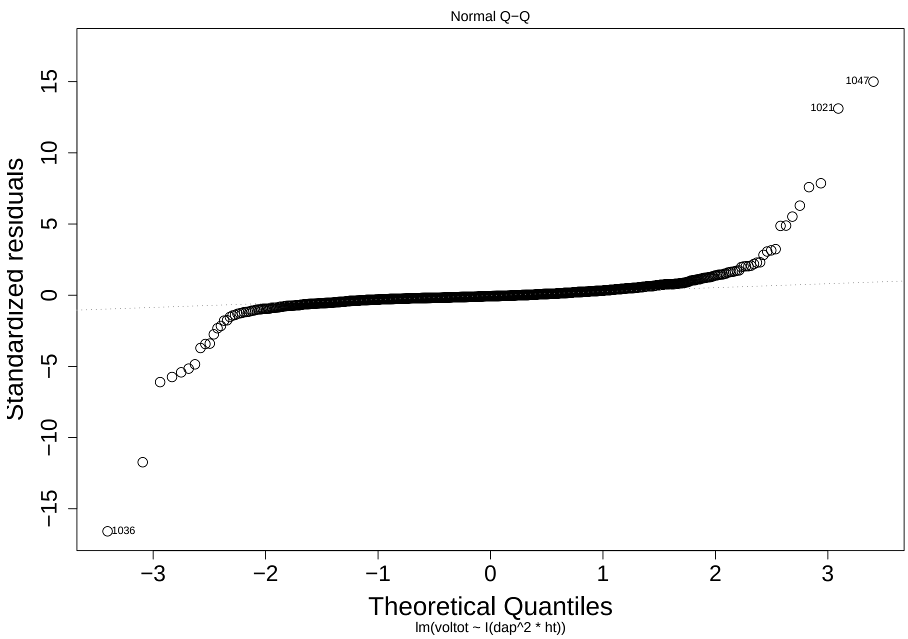
Na Biometria Florestal, muito esforço de pesquisa foi colocado na busca de relações funcionais apropriadas, mas, ainda hoje, o procedimento padrão é essencialmente empírico: ajusta-se uma série de modelos diferentes e seleciona-se "*o melhor*".

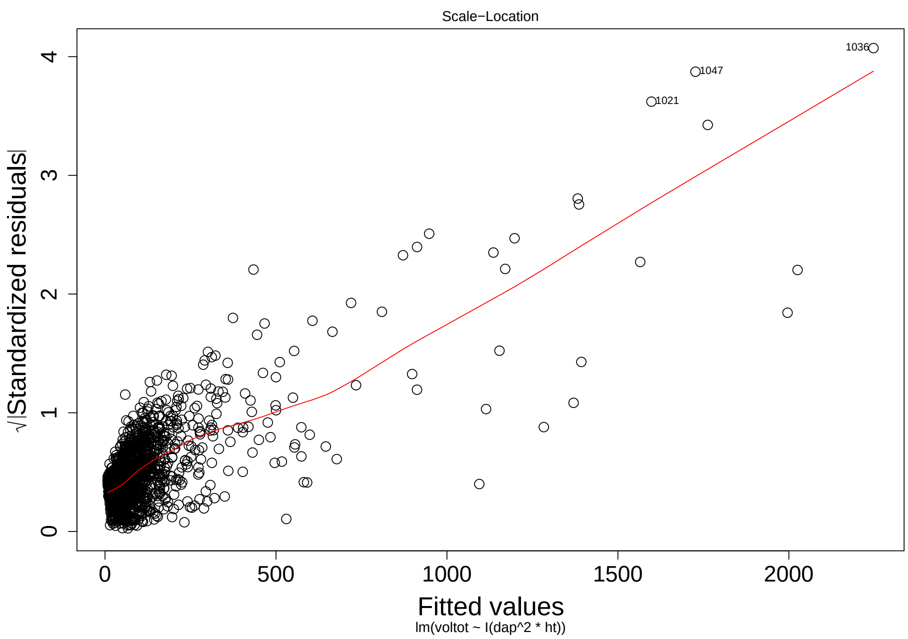
As estatísticas tradicionais de ajuste revelam um bom ajuste do modelo aos dados:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.7107	0.4445	8.35	0.0000
I(dap^2 * ht)	0.0342	0.0001	491.89	0.0000
Erro Padrão da Estimativa	14.84			
R^2	0.9938			
R^2 ajustado	0.9938			

Vejamos os gráficos dos resíduos:


Residuals vs Fitted
Residuals
Fitted values
lm(voltot ~ I(dap^2 * ht))
1047
1021
1036
200
100
0
−100
−200
0
500
1000
1500
2000


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(voltot ~ I(dap^2 * ht))
1047
1021
1036
15
10
5
0
−5
−10
−15
−3
−2
−1
0
1
2
3



Nota-se que em termos de relação linear e normalidade o modelo está bem, mas a pressuposição de homoscedasticidade é claramente inapropriada. Quais são as conseqüências negativas da heteroscedasticidade?

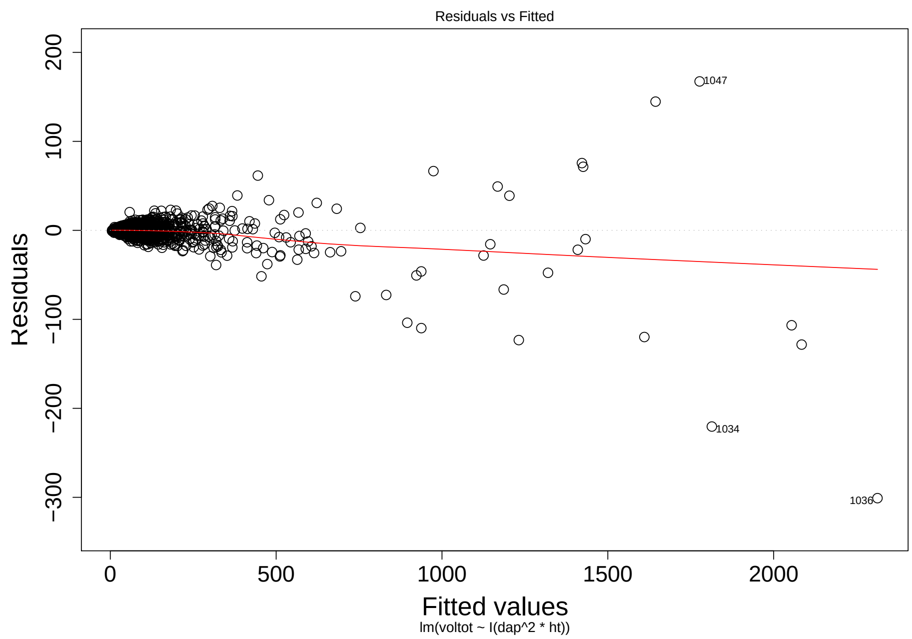
A solução tradicional, em Biometria Florestal, é a emphregressão ponderada ou a *transformação* da variável resposta.

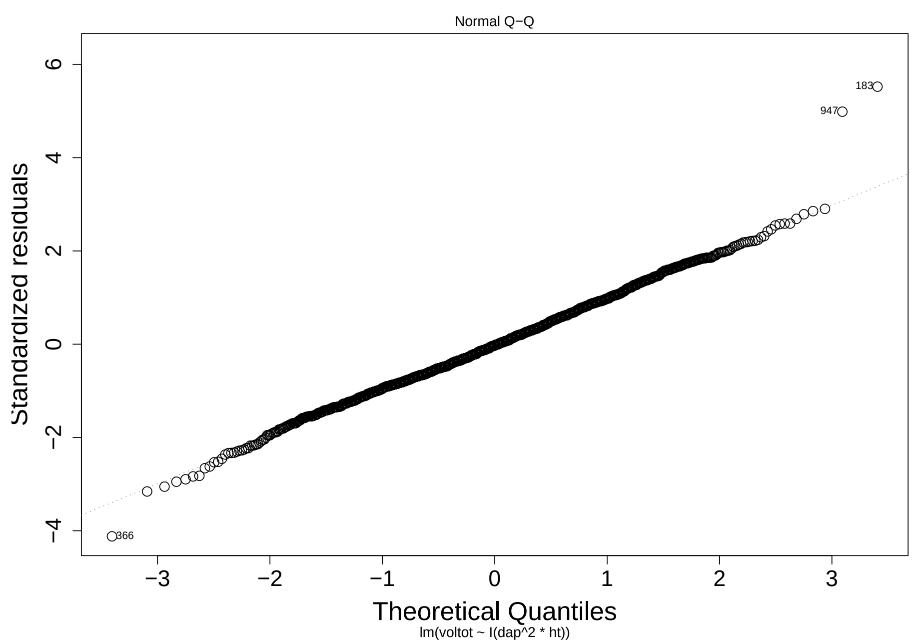
Modelo Spurr Ponderado

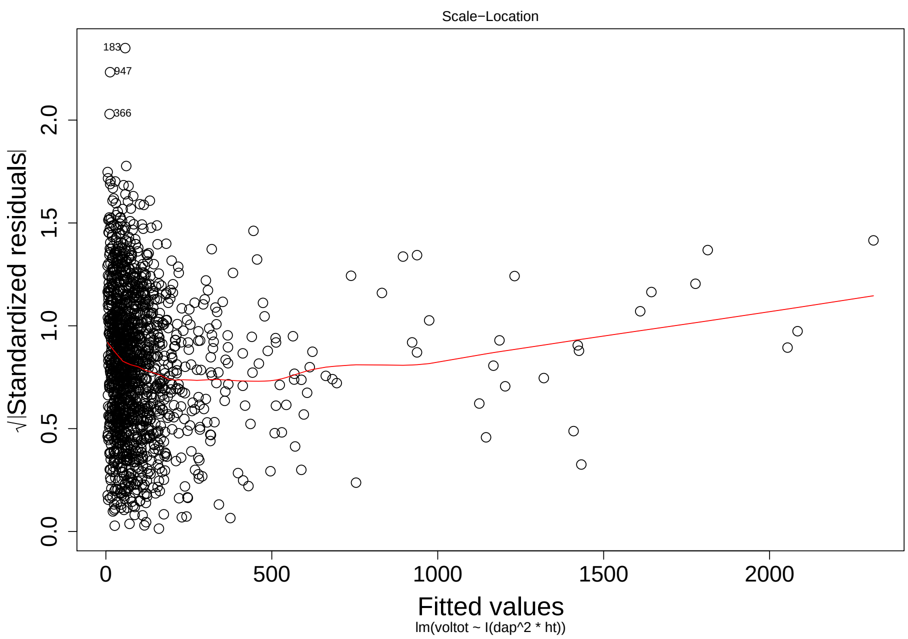
A ponderação geralmente é realizada com a potência do inverso da variável peditora. Nesse caso, a ponderação que tornou a variância homogênena foi

$$\left[\frac{1}{d_i^2\, h_i}\right]$$

resultando nos seguintes gráficos de resíduo:


Residuals vs Fitted
Residuals
Fitted values
lm(voltot ~ I(dap^2 * ht))
1047
1034
1036
200
100
0
−100
−200
−300
0
500
1000
1500
2000


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(voltot ~ I(dap^2 * ht))
183
947
366
6
4
2
0
−2
−4
−3
−2
−1
0
1
2
3



A ponderação manteve a relação linear e a normalidade, solucionando o problema de heteroscedasticidade. Mas o que acontece com as estimativas dos parâmetros e com a utilização do modelo para predição?

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.2779	0.0543	23.53	0.0000
I(dap^2 * ht)	0.0353	0.0001	468.58	0.0000
Erro Padrão da Estimativa	0.002294			
R^2	0.9932			
R^2 ajustado	0.9932			

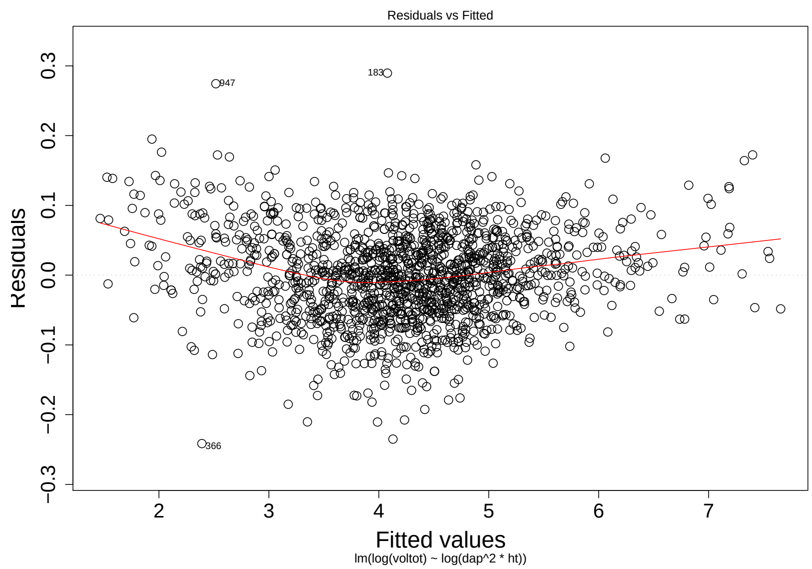
Modelo Spurr Transformado

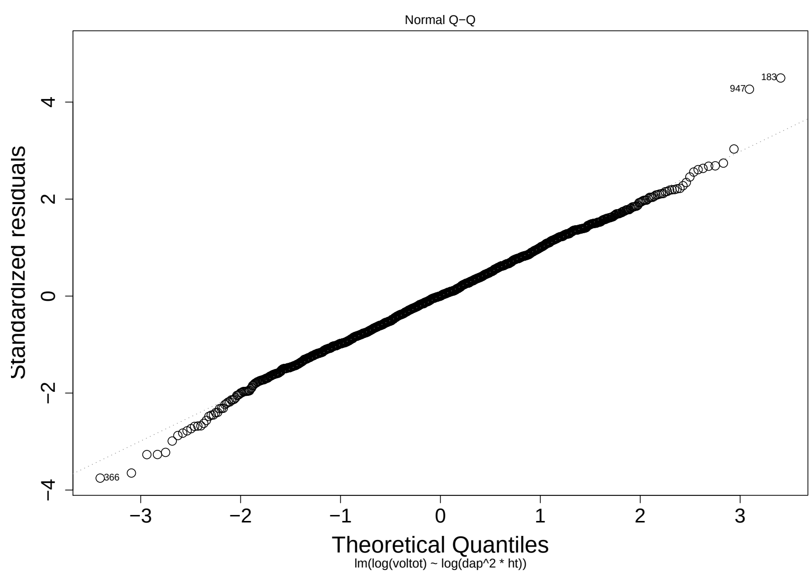
As variáveis de produção florestal, como volume ou biomassa de madeira, são heteroscedásticas "*por natureza*". Há muito que florestais utilizam a escala logarítmica para tratar essas variáveis, assim o modelo Spurr é freqüentemente utilizado na nessa escala:

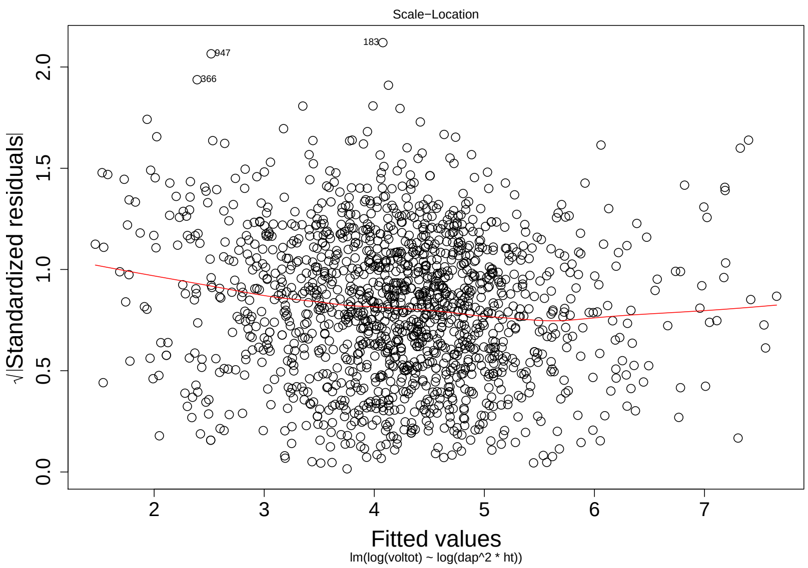
$$\ln(v_i) = \beta_0 + \beta_1 \ln\left(d_i^2\, h_i\right) + \varepsilon_i$$

onde os termos são os mesmos da apresentação do modelo acima.

Qual é o desempenho do modelo transformado?





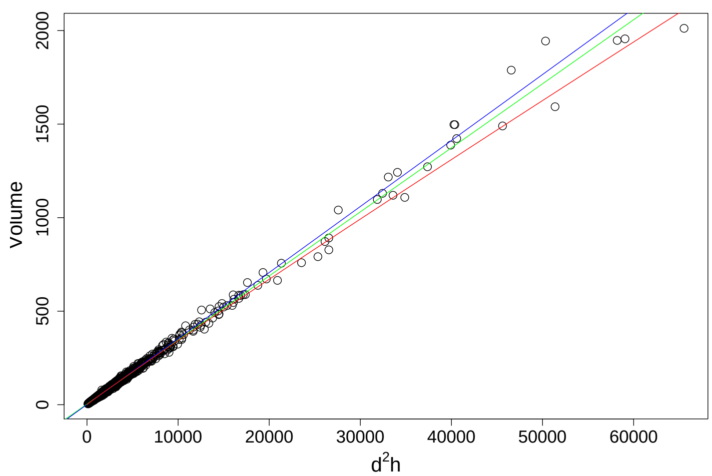


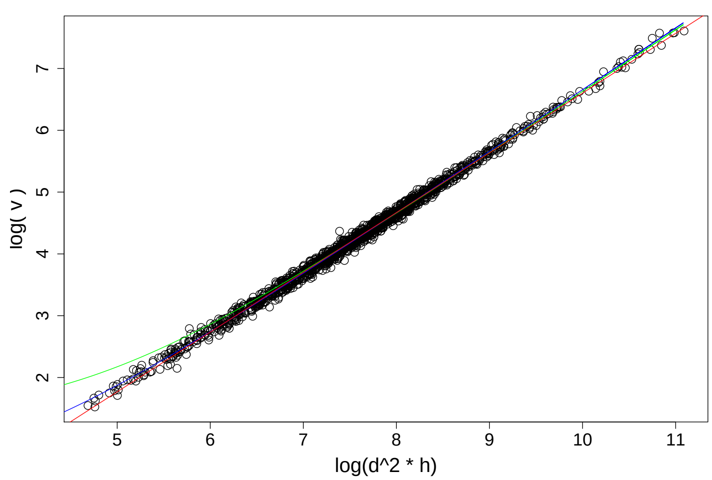
A heteroscedasticidade é corrigida mas ao custo de uma *incômoda* tendência que surge no gráfico de dispersão dos resíduos. Mas qual a influência da transformação nas estimativas dos parâmetros e nas predições do modelo?

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	−3.0637	0.0127	−241.22	0.0000
log(dap^2 * ht)	0.9665	0.0017	579.38	0.0000
Erro Padrão da Estimativa	0.06444			
R^2	0.9955			
R^2 ajustado	0.9955			

Como se Comparam os modelos?

Os modelos são bastane distintos para árvores muito grandes ou muito pequenas, mas quase não diferem para as árvores de tamanho intermediário que são a maioria.





A Abordagem da Verossimilhança no Exemplo 2

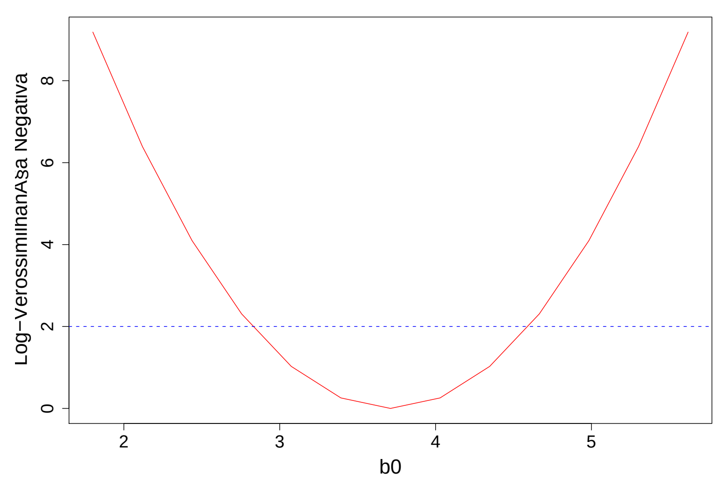
Ao invés de se colocar o modelo de Spurr como um *modelo linear*, podemos colocálo como um *modelo Gaussiano*, onde o volume das árvores (v) é uma variável aleatória Gaussiana com densidade:

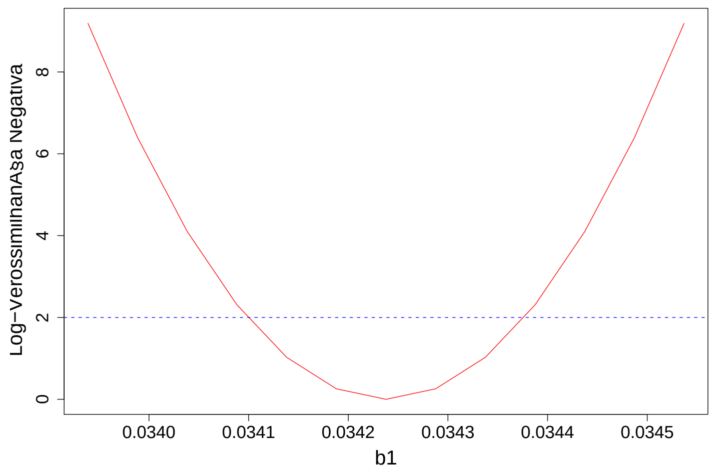
$$f(v) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left[\frac{v - (\beta_0 + \beta_1\, d^2\, h)}{\sigma}\right]^2\right]$$

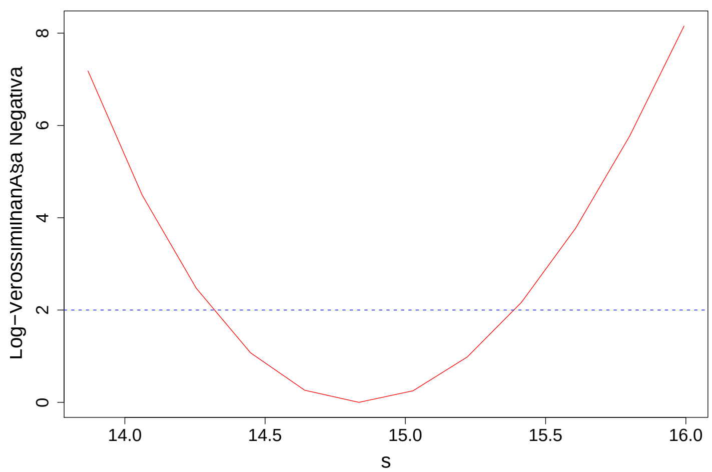
Esse modelo é equivalente ao modelo linear simples apresentado acima, mas pode ser tratado como um modelo com três parâmetros que são estimados por máxima verossimilhança, através de um processo de maximização da função de verossimilhança. Nesse, caso o processo de maximização é o mesmo que o método de quadrados mínimos dos modelos lineares, pois existe solução analítica para o problema de maximização. Entretanto, foi utilizando um algoritmo de optimização que gerou as seguintes estimativas:

	Estimate	Std. Error
b0	3.71074252	4.441572e-01
b1	0.03423787	6.955812e-05
s	14.83449093	2.695844e-01

Como no modelo linear, os resíduos desse modelo deveria ser analisado para verificar se as pressuposições são razoáveis. O comportamento das estimativas é investigado analisando-se diretamente a curva de **log-verossimilhança negativa perfilhada**:




Log−VerossimilhanÃ§a Negativa
0
2
4
6
8
0.0340
0.0341
0.0342
0.0343
0.0344
0.0345
b1


Log−VerossimilhanÃ§a Negativa
0
2
4
6
8
14.0
14.5
15.0
15.5
16.0
s

Novo Modelo

O problema de heteroscedasticidade, entretanto, persiste e deve ser abordado. Os dados sugerem que tanto a média como a variância do volume variam linearmente com a variável preditora, logo o modelo potencial seria:

$$
\begin{aligned}
f(v) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left[\frac{v-\mu}{\sigma}\right]^2\right] \\
\mu &= \beta_0 + \beta_1\,(d^2\,h) \\
\sigma &= \exp(\alpha_1)\,(d^2\,h)^w
\end{aligned}
$$

Os parâmetros desse modelo também podem ser estimados por máxima verossimilhança, mas o processo de estimação não é mais equivalente ao método de quadrados mínimos oridinários, uma vez que não há solução analítica para o problema de maximização. Sendo um problema maximização não linear, o ajuste desse modelo é bem mais complexo.

Mas as estimativas geradas são:

	Estimate	Std. Error
b0	1.33936070	6.504917e-02
b1	0.03519417	7.398218e-05
a1	-5.28485330	1.271703e-01
w	0.89314122	1.667432e-02

Todo o trabalho de ajuste é justificado? O que se ganha a se definir o modelo nessa outra forma?

Modelo Log-Normal

No modelo log-Spurr, o ajuste da relação linear é realizada na escala logarítmica. Nesse caso, para não haver transformação da variável resposta (volume de madeira) é necessário trabalhar com outro modelo estatístico: a distribuição Log-Normal.

$$
\begin{aligned}
f_L(v) &= \frac{1}{v\,\sigma_L\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left[\frac{\ln(v)-\mu_L}{\sigma_L}\right]^2\right] \\
\mu_L &= \beta_0 + \beta_1\,\ln(d^2\,h)
\end{aligned}
$$

Nesse caso, as predição serão realizadas não através da média, mas da mediana:

$$\widehat{v} = \exp\left[\beta_0 + \beta_1 \, \ln(d^2 \, h)\right] = \exp[\mu_L] = \left[\text{mediana}|(d^2 \, h)\right].$$

A predição sem viés requer a correção do viés logarítmico, que também é necessária no modelo log-Spurr:

$$\widehat{v}^* = \widehat{v} \, \exp\left[\frac{1}{2}\, \sigma_L^2\right].$$

Mas as estimativas de máxima verossimilhança são:

	Estimate	Std. Error
b0	-3.0637043	0.012694618
b1	0.9665144	0.001667397
ls	0.0644044	0.001169456

Note que as estimativas são equivalentes aos modelos log-Spurr ajustado por quadrados mínimos.

Comparação dos Modelos

A comparação dos modelos na abordagem da verossimilhança se baseia na razão da verossimilhança, mas o número de parâmetros deve ser considerado. Assim, o Critério de Informação de Akaike (AIC) é a medida mais utilizada para confrontar os modelos.

MODELO	AIC
Linear	12468.93^a
Linear Ponderado	8754.13
Logarítmico	-4002.51^*
Gaussiano: média como função linear	12468.93^a
Gaussiano: média e desvio como função linear	8716.24
Log-Normal	8811.89

a – modelos equivalentes

* – transformação da escala da variável resposta, não comparável.

Todo o trabalho complexo de ajuste é justificado? O que se ganha ao se definir os modelos dessa outra forma?

Referências Bibliográficas

Berger, J. O. e Wolpert, R. L. *The likelihood principle*, volume 6 of *Lecture Notes - Mongraph Series, Gupta, Shanti S. (ed.)*. Hayward: Institute of Mathematical Statistics, 1988.

Burnham, K. P. e Anderson, D. R. *Model selection and multimodel inference: a pratical information-theoretic approach*. New York: Springer-Verlag, 2002.

Edwards, A. *Likelihood: expanded edition*. London/Baltimore: The John Hopkins University Press, 1992.

Hacking, I. *Logic of statistical inference*. Cambridge: Cambridge at the University Press, 1965.

Hilborn, R. e Mangel, M. *The ecological detective: confronting models with data*. Number 28 in Monographs in Population Biology. Princeton: Princeton University Press, 1997.

Royall, R. *Statistical evidence: a likelihood paradigm*. Boca Raton: Chapman & Hall, 1997.