



*Modelos Gaussianos Univariados*¹

JOÃO LUÍS F. BATISTA²
Setembro de 2015

Introdução

Os modelos de predição utilizados nas Ciências Florestais, em geral, e na Mensuração Florestal, em particular, são aplicações da família Gaussiana de distribuições estocásticas. No mais das vezes, os modelos de predição tem uma única variável preditora tendo uma estrutura univariada, por isso, eles podem ser agrupados numa ampla categoria de modelos estocásticos que pode ser designada por “*Modelos Gaussianos Univariados*”. Essa categoria inclui os modelos de regressão linear e não-linear clássicos, mas ela não se resume a esses modelos. O enfoque dos modelos preditivos a partir da família Gaussiana como modelo estocástico básico permite uma visão mais ampla, que reúne as diferentes abordagens

¹Anotações para a disciplina LCF5877 Biometria Florestal Avançada, Departamento de Ciências Florestais, ESALQ, Universidade de São Paulo.

²Departamento de Ciências Florestais, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Campus Piracicaba.

da Teoria Estatística Clássica numa mesma categoria de modelos aplicados: os modelos preditivos.

A Família Gaussiana

A família Gaussiana é definida pela seguinte função de densidade

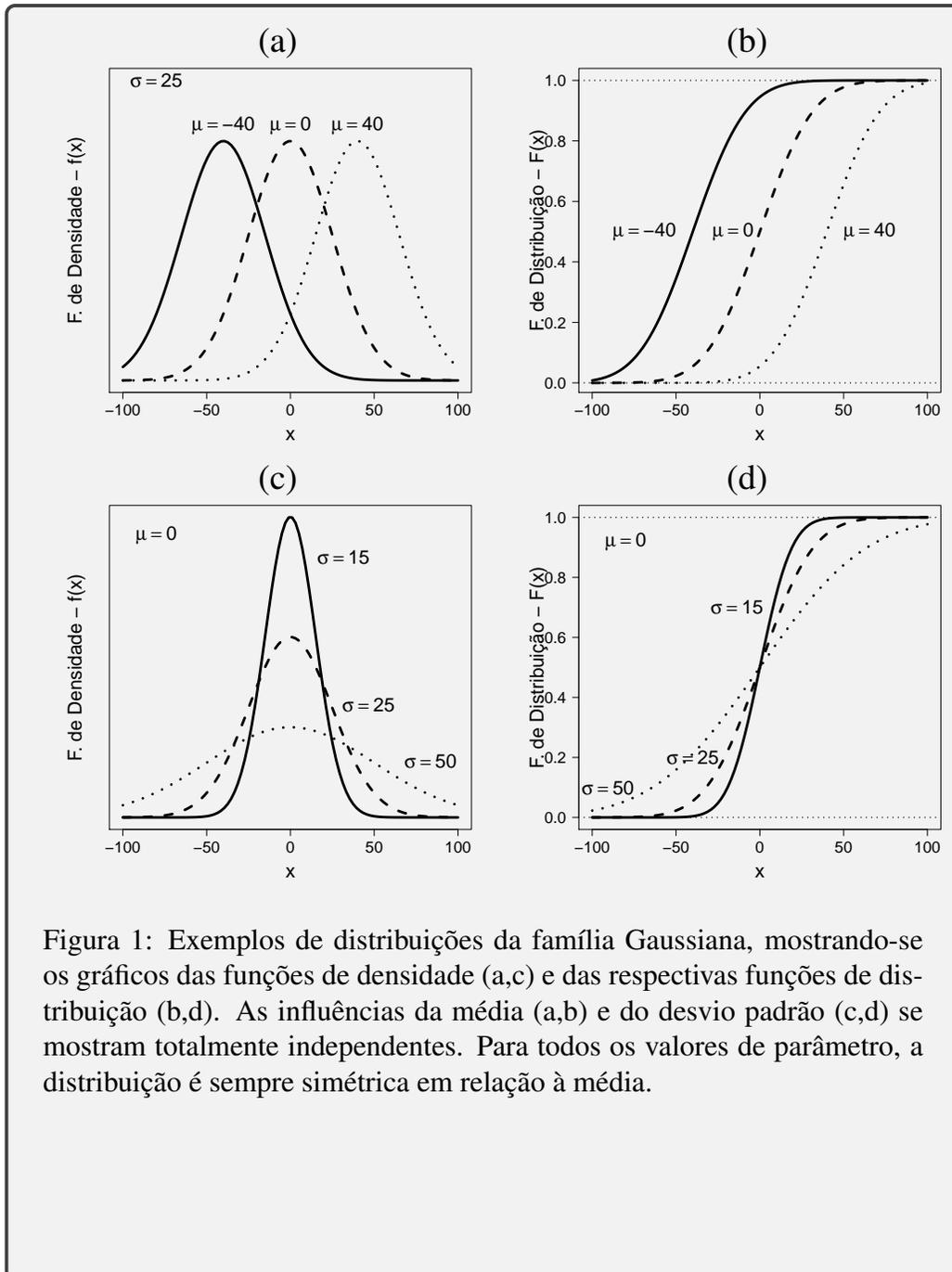
$$f(y) = f(y; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y - \mu)^2}{2\sigma^2}\right] \quad (1)$$
$$-\infty \leq \mu \leq \infty, \sigma > 0; \quad -\infty \leq y \leq \infty.$$

O parâmetro μ é chamado de *média*, sendo um parâmetro de localização, no caso a localização central da distribuição. Já o parâmetro σ é chamado de *desvio padrão*, sendo o parâmetro de escala, isto é, controla a dispersão dos valores ao redor da média μ . Também é comum designar o parâmetro de escala por σ^2 , chamado de *variância*.

A família Gaussiana não tem um parâmetro que controle a forma, logo, as distribuições gaussianas têm sempre a mesma forma fixa. Sua forma é sempre simétrica em relação à posição do parâmetro de localização (*média*), mas é uma simetria com um formato particularmente, tradicionalmente designada como “*forma de sino*” (*bell shape*). Um outro aspecto particular da família Gaussiana é que os seus parâmetros, média e desvio padrão, atuam de modo independente sobre o comportamento das distribuições (figura 1).

Gaussiana Padronizada

A família Gaussiana não possui uma forma fechada para a função de distribuição. Isso implica que, no cálculo de probabilidades, a integração da função de densidade nunca resulta numa expressão matemática explícita, sendo necessário



métodos numéricos³ para obter as probabilidades.

Para facilitar esse processo, costumava-se tabelar os valores de probabilidade para uma única distribuição da família Gaussiana chamada de *Distribuição Gaussiana Padronizada*. Com o desenvolvimento dos computadores eletrônicos, esse procedimento se tornou desnecessário, mas o procedimento para se obter a distribuição padronizada continuou a ser designado por *padronização* de uma variável.

O procedimento consiste em subtrair dos valores da variável (X) a sua média (μ_X), dividindo o resultado pelo desvio padrão (σ_X):

$$Z = \frac{Y - \mu_Y}{\sigma_Y}. \quad (2)$$

Se a variável pertencer à família Gaussiana, a variável resultante (Z) terá distribuição Gaussiana padronizada com média nula ($\mu_Z = 0$) e desvio padrão unitário ($\sigma_Z = 1$).

Exercício

O arquivo `esaligna.csv`^a apresenta dados de biomassa de árvores individuais de *Eucalyptus saligna*.

1. Construa o gráfico de densidade, segundo a Família Gaussiana, que seriam esperados para as medidas de biomassa: tronco, sobra, folha e total.
2. Considerando essas mesmas medidas, compare o gráfico da densidade esperada (item acima) contra a densidade empírica das observações.
3. Utilize o gráfico quantil-quantil para verificar se a Família Gaussiana é uma boa aproximação para as observações dessas medidas.

^aLocalização: <http://cmq.esalq.usp.br/Philodendros/doku.php?id=lcf5877:dados:lista-dados>

³Métodos numéricos são formas de se calcular de modo aproximado o valor de funções. Para efeitos práticos, o cálculo numérico permite obter o valor de uma função com erro negligenciável.

Família Gaussiana e Modelos Preditivos

A Família Gaussiana é um modelo estocástico pertinente para os modelos preditivos pois ela é um modelo padrão (normal) para distribuição de erros de medição. Se o parâmetro da média for fixado em zero (valor nulo), a Família Gaussiana se reduz a distribuições com único parâmetro (desvio padrão) que indica a *imprecisão* das medidas. Nesse caso, a densidade da Gaussiana se torna:

$$f(\varepsilon) = f(\varepsilon; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{\varepsilon^2}{2\sigma^2}\right] \quad (3)$$
$$\sigma > 0; \quad -\infty \leq \varepsilon \leq \infty .$$

Nessa perspectiva, um modelo de predição pode ser visto como uma função matemática determinística, à qual se soma um erro de predição estocástico da Família Gaussiana:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad ; \quad i = 1, 2, \dots, n \quad (4)$$

onde:

i é o índice que indica cada observação a ser predita em particular;

y_i é a medida a ser predita (variável resposta);

x_i é a medida preditora (variável preditora);

β_0, β_1 são coeficientes da função matemática de predição;

ε_i é o erro de predição, que segue a Família Gaussiana, com desvio padrão igual a σ_ε .

Esse modelo implica que a medida a ser predita (variável resposta) seguirá a família Gaussiana, com média igual a função de predição:

$$\mu = \mathbf{E}\{y_i|x_i\} = \beta_0 + \beta_1 x_i ,$$

ou seja, o valor da medida média seguirá a medida preditora (x_i), e cada observação pode ser representado pela expressão:

$$y_i = \mathbf{E}\{y_i|x_i\} + \varepsilon_i.$$

O desvio padrão, contudo, permanecerá constante e igual ao desvio padrão do erro de predição (σ_ε), que é o componente estocástico. Esse desvio padrão representa, portanto, a *imprecisão* da predição gerada pelo modelo.

Note que a função de predição nesse exemplo é a função mais simples possível. A complexidade dessa função, contudo, não muda muito a situação de modelagem em que se trabalha, embora tenha grande impacto nos procedimentos de estimação e inferência. Ao generalizar a função de predição, temos um modelo de predição geral:

$$y_i = f_P(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (5)$$

onde $f_P(\cdot)$ é uma função de predição geral, que depende de p variáveis predictoras ($x_{1i}, x_{2i}, \dots, x_{pi}$). Mas a função de predição representará sempre a média ou valor esperado para medida a ser predita, dado o conhecimento da medidas predictoras:

$$\mathbf{E}\{y_i|x_{1i}, x_{2i}, \dots, x_{pi}\} = f_P(x_{1i}, x_{2i}, \dots, x_{pi}) \quad (6)$$

Essa função de predição geral pode ter qualquer forma matemática desejada, alguns exemplos comuns são:

- Modelo Linear Simples:

$$\mathbf{E}\{y|x\} = \beta_0 + \beta_1 x$$

- Modelo Linear Múltiplo:

$$\mathbf{E}\{y|x_1, \dots, x_p\} = \beta_0 + \sum_{k=1}^p \beta_k x_k$$

- Exemplo de um modelo não-linear de potência bastante simples:

$$\mathbf{E}\{y|x\} = \beta_0 x^{\beta_1}$$

- Exemplo de um modelo não-linear de potência mais complexo:

$$\mathbf{E}\{y|x\} = \beta_0 x^{\beta_1 + \beta_2 \ln(x)}$$

- Exemplo de um modelo não-linear de potência mais complexo com duas medidas preditoras:

$$\mathbf{E}\{y|x_1, x_2\} = \beta_0 x_1^{\beta_1 + \beta_2 \ln(x_1)} x_2^{\beta_3 + \beta_4 \ln(x_2)}$$

- Exemplo de um modelo não-linear exponencial bastante simples:

$$\mathbf{E}\{y|x\} = \beta_0 \exp[\beta_1 x]$$

- Exemplo de um modelo não-linear exponencial mais complexo com duas medidas preditoras:

$$\mathbf{E}\{y|x_1, x_2\} = \beta_0 \exp[\beta_1 x_1^{\beta_2}] x_2^{\beta_3/x_1}$$

Note que no modelo de predição geral acima (5), o erro de predição (ε) permanece se comportando segundo a Família Gaussiana com dispersão constante, isto é, com desvio padrão igual a σ_ε .

Exercício

Utilizando os mesmos dados de *E. saligna*, simule conjuntos de dados seguindo o modelo de predição geral, variando a função matemática de predição segundo os modelos apresentados acima. Tome a biomassa total como medida resposta e o DAP e altura como medidas preditoras.

Construa gráficos de dispersão dos dados simulados, grafando a linha de tendência e o modelo matemático que gerou os dados.

Modelos com Parâmetros Constantes

O modelo gaussiano mais simples é aquele que estabelece os parâmetros média (μ) e desvio padrão (σ) como parâmetros constantes, ou seja, o comportamento da

medida resposta não é adequadamente explicado por nenhuma medida preditora. Esse modelo é igual a própria definição da Família Gaussiana, mas pode ser definido na forma de um modelo preditivo:

$$y_i = \mu + \varepsilon_i \quad ; \quad i = 1, 2, 3, \dots, n \quad (7)$$

onde

- i é o índice das observações;
- y_i é a medida resposta;
- μ é a média geral;
- ε_i é o erro de predição, que possui distribuição Gaussiana com média nula e desvio padrão σ_ε .

A função de densidade desse modelo é a própria função de densidade da Família Gaussiana (equação 1), mas com desvio padrão do erro de predição:

$$f(y) = f(y; \mu, \sigma_\varepsilon) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp \left[-\frac{(y - \mu)^2}{2\sigma_\varepsilon^2} \right] \quad (8)$$
$$-\infty \leq \mu \leq \infty, \sigma_\varepsilon > 0; \quad -\infty \leq y \leq \infty.$$

Estimativas de Máxima Verossimilhança

Nesse modelo, as funções de verossimilhança e log-verossimilhança são as funções da Família Gaussiana:

$$\mathcal{L}\{\mu, \sigma_\varepsilon\} = (2\pi)^{-n/2} \sigma_\varepsilon^{-n} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (y_i - \mu)^2 \right] \quad (9)$$

$$\mathbf{L}\{\mu, \sigma_\varepsilon\} = \frac{n}{2} \ln 2\pi + n \ln \sigma_\varepsilon + \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (y_i - \mu)^2 \quad (10)$$

Para se obter as estimativas de máxima verossimilhança (MLE), iguala-se a zero as derivadas parciais e soluciona-se para os parâmetros desejados, obtêm-se:

$$\frac{\partial \mathbf{L}\{\mu, \sigma_\varepsilon\}}{\partial \mu} = \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^n (y_i - \mu) = 0 \quad \implies \quad \hat{\mu} = \frac{\sum_{i=1}^n y_i}{n}$$

$$\frac{\partial \mathbf{L}\{\mu, \sigma_\varepsilon\}}{\partial \sigma_\varepsilon} = \frac{n}{\sigma_\varepsilon} - \frac{1}{\sigma_\varepsilon^3} \sum_{i=1}^n (y_i - \mu)^2 = 0 \quad \implies \quad \hat{\sigma}_\varepsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2}$$

Modelo Nulo

Esse modelo pode ser considerado um modelo nulo para predição da medida resposta de interesse, pois ele é o modelo mais simples que ignora a relevância de qualquer medida preditora para a predição da medida resposta. Assim, a comparação com esse modelo permite verificar a relevância das medidas predictoras individualmente ou em grupo, no processo de predição.

Exemplo de Modelo com Parâmetros Constantes

O arquivo `sitio-florin.csv`⁴ apresenta dados de parcela de floresta de eucalipto com dados da altura média das árvores dominantes (`mhdom`).

```
> head(florin)
  fazenda parcela mhdom idade   vol    g
1      1      101   8.5   2.5 18.47  4.05
2      1      102  19.0   2.5 110.63 16.53
3      1      103  13.2   2.5  63.42 11.06
4      1      104  12.5   2.5  56.82 10.25
5      1      105  10.4   2.5  18.82  4.39
6      1      106   6.4   2.5   8.00  2.10
>
```

Ajustar o modelo com parâmetros constantes a esses dados consiste simplesmente em encontrar a média e o desvio padrão amostrais, mas ao considerar esse modelo como *modelo nulo* é importante que se ajuste esse modelo numa estrutura que permita a comparação desse modelo com outros modelos. Uma forma de fazer isso é utilizando a função “`gls`” do pacote “`nlme`”, que gera um objeto da classe “`gls`”:

⁴Localização: <http://cmq.esalq.usp.br/Philodendros/doku.php?id=lcf5877:dados:lista-dados>

```
> M.cte <- gls( mhdom ~ 1, data=florin , method="ML")
> class(M.cte)
[1] "gls"
> summary(M.cte)
Generalized least squares fit by maximum likelihood
Model: mhdom ~ 1
Data: florin
      AIC      BIC    logLik
2710.566 2718.651 -1353.283

Coefficients:
              Value Std.Error  t-value p-value
(Intercept) 19.48789 0.2938701 66.31463      0

Standardized residuals:
      Min      Q1      Med      Q3      Max
-2.17314895 -0.87801528 0.05182428 0.79901678 2.19377611

Residual standard error: 6.022544
Degrees of freedom: 421 total; 420 residual
>
```

Pode se extrair facilmente as MLEs do modelo ajustado, bem como explorá-lo graficamente:

```
(Intercept)
 19.48789
> M.cte$sigma
[1] 6.022544
> plot(M.cte)
> qqnorm(residuals(M.cte))
> qqline(residuals(M.cte))
>
```

Exercício

Utilizando os dados de *E. saligna*, ajuste o modelo Gaussiano de parâmetros constantes para as medidas de biomassa (tronco, sobra, folha e total).

Modelos Categóricos

Um modelo Gaussiano um pouco mais complexo, mas ainda sem a presença de medidas predictoras, é o modelo Gaussiano ajustado a níveis de uma medida cate-

górica. No lugar do parâmetro da média geral, temos uma média para cada nível da medida categórica:

$$y_{ij} = \mu_j + \varepsilon_{ij} \quad (11)$$

onde

- i é o índice das observações dentro de cada nível da medida categórica ($i = 1, 2, \dots, n_j$);
- j é o índice dos níveis da medida categórica ($j = 1, 2, \dots, k$);
- y_{ij} é a medida resposta;
- μ_j é a média do nível j ;

O comportamento do erro de predição distingue dois tipos de modelos categóricos:

Modelo Categórico Homocedástico: o erro de predição ε_{ij} possui distribuição Gaussiana com média nula e desvio padrão constante (σ_ε), independentemente do nível da medida categórica. Esse modelo tem k médias e um desvio padrão σ_ε , portanto, $k + 1$ parâmetros.

Modelo Categórico Heteroscedástico: o erro de predição ε_{ij} possui distribuição Gaussiana com média nula, mas com um desvio padrão para cada nível da medida categórica: σ_j ($j = 1, 2, \dots, k$). Esse modelo tem k médias e k desvios padrão σ_j , portanto, $2k$ parâmetros.

Estimativas de Máxima Verossimilhança

As estimativas de máxima verossimilhança (MLE) dependem da função de log-verossimilhança que é ligeiramente diferente no dois tipos de modelos categóricos. Um aspecto importante para obter as MLEs é que independência das observações implica a independência das estimativas dos parâmetros para cada nível da medida categórica.

Modelo Categórico Homocedástico. A função de log-verossimilhança possui uma média para cada nível da medida categórica (k médias) e o desvio padrão do erro de predição:

$$\begin{aligned} \mathbf{L}\{\mu_1, \dots, \mu_k, \sigma_\varepsilon\} &= \frac{\sum_{j=1}^k n_j}{2} \ln(2\pi) + \\ &+ \ln(\sigma_\varepsilon) \sum_{j=1}^k n_j + \frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2 \quad (12) \\ & \quad i = 1, 2, \dots, n_j; \quad j = 1, 2, \dots, k \end{aligned}$$

As primeiras derivadas em relação aos parâmetros igualadas a zero resultam nas MLEs:

$$\frac{\partial \mathbf{L}\{\mu_1, \dots, \mu_j, \sigma_\varepsilon\}}{\partial \mu_j} = \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{n_j} (y_{ij} - \mu_j) = 0 \quad \implies \quad \hat{\mu}_j = \frac{\sum_{i=1}^{n_j} y_{ij}}{n_j}$$

$j = 1, 2, \dots, k$

$$\begin{aligned} \frac{\partial \mathbf{L}\{\mu_1, \dots, \mu_j, \sigma_\varepsilon\}}{\partial \sigma_\varepsilon} &= \frac{\sum_{j=1}^k n_j}{\sigma_\varepsilon} - \frac{1}{\sigma_\varepsilon^3} \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2 = 0 \\ \implies \quad \hat{\sigma}_\varepsilon &= \sqrt{\frac{1}{\sum_{j=1}^k n_j} \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2} \end{aligned}$$

Note que as médias são estimadas para cada nível da medida categórica, mas a estimativa do desvio padrão é realizada com os dados agrupados, tomando-se os desvios de cada observação em relação à média do seu respectivo nível.

Modelo Categórico Heteroscedástico. Nesse caso, cada nível da medida categórica tem a sua própria média e desvio padrão, totalizando $2k$ parâmetros no

modelo. Assim, a função de log-verossimilhança tem a seguinte forma:

$$\begin{aligned} \mathbf{L}\{\mu_1, \dots, \mu_k, \sigma_1, \dots, \sigma_k\} &= \frac{\sum_{j=1}^k n_j}{2} \ln(2\pi) + \\ &+ \sum_{j=1}^k n_j \ln(\sigma_j) + \\ &+ \sum_{j=1}^k \frac{1}{2\sigma_j^2} \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2 \end{aligned} \quad (13)$$

$i = 1, 2, \dots, n_j; \quad j = 1, 2, \dots, k$

Como as estimativas nos nível da medida categórica são independentes, as primeiras derivadas em relação à média e o desvio padrão são obtidas em cada nível ($j = 1, 2, \dots, k$), resultando nas respectivas MLEs:

$$\begin{aligned} \frac{\partial \mathbf{L}\{\mu_1, \dots, \mu_j, \sigma_1, \dots, \sigma_k\}}{\partial \mu_j} &= \frac{1}{\sigma_j^2} \sum_{i=1}^{n_j} (y_{ij} - \mu_j) = 0 \implies \hat{\mu}_j = \frac{\sum_{i=1}^{n_j} y_{ij}}{n_j} \\ \frac{\partial \mathbf{L}\{\mu_1, \dots, \mu_j, \sigma_1, \dots, \sigma_k\}}{\partial \sigma_j} &= \frac{n_j}{\sigma_j} - \frac{1}{\sigma_j^3} \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2 = 0 \\ &\implies \hat{\sigma}_j = \sqrt{\frac{1}{n_j} \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2} \end{aligned}$$

Note que as MLEs das médias são mesmas nos dois tipos de modelo categórico, somente a MLE do desvio padrão é diferente.

ExEMPLO DE Modelo Categórico

Utilizando novamente os dados do arquivo `sitio-florin.csv`⁵, podemos tomar a idade dos povoamentos (`idade`) como uma variável categórica que indica alturas médias diferentes para as árvores dominantes (`mhdom`).

O modelo categórico homocedástico também pode ser ajustado com a função “`gls`”, que permite que o modelo seja graficamente avaliado:

⁵Localização: <http://cmq.esalq.usp.br/Philodendros/doku.php?id=lcf5877:dados:lista-dados>

```
> M.homo <- gls( mhd0m ~ factor(idade), data=florin, method="ML" )
> summary(M.homo)
Generalized least squares fit by maximum likelihood
Model: mhd0m ~ factor(idade)
Data: florin
      AIC      BIC    logLik
2080.453 2104.709 -1034.226

Coefficients:
              Value Std.Error  t-value p-value
(Intercept)  13.57059 0.2177802  62.31324      0
factor(idade)3.5  7.12616 0.3361241  21.20098      0
factor(idade)4   11.07386 0.9712334  11.40185      0
factor(idade)4.5 12.45868 0.3817787  32.63325      0
factor(idade)5.5 13.33482 0.5151132  25.88716      0

Correlation:
      (Intr) f()3.5 fct()4 f()4.5
factor(idade)3.5 -0.648
factor(idade)4   -0.224 0.145
factor(idade)4.5 -0.570 0.370 0.128
factor(idade)5.5 -0.423 0.274 0.095 0.241

Standardized residuals:
      Min      Q1      Med      Q3      Max
-3.65949488 -0.63962632 -0.03734345  0.68355979  2.66755040

Residual standard error: 2.822594
Degrees of freedom: 421 total; 416 residual
>
> plot(M.homo)
> qqnorm(residuals(M.homo))
> qqline(residuals(M.homo))
>
```

A obtenção das MLEs, na forma como foram apresentadas na exposição teórica acima, envolve um pouco mais de manipulação do objeto, pois a “gls” sendo uma generalização dos modelos lineares clássicos, ajusta a medida categórica através de *variáveis indicadoras*:

```
> tmp1 <- coef(M.homo)
> tmp1 <- tmp1 + c(0, rep(tmp1[1], 4))
> M.homo.MLE <- data.frame(idade=c(2.5,3.5,4.,4.5,5.5), m=tmp1, row.names=1:5 )
> M.homo.MLE$s <- M.homo$sigma
> M.homo.MLE
  idade      m      s
1  2.5 13.57059 2.822594
2  3.5 20.69675 2.822594
3  4.0 24.64444 2.822594
4  4.5 26.02927 2.822594
5  5.5 26.90541 2.822594
>
```

A melhor maneira de explorar os dados e o modelo categórico é através de um gráfico de dispersão dos dados que inclua as médias por idade (figura 2):

```
> plot( mhdom ~ idade, data=florin, xlab="Idade_(anos) ",
+       ylab="Altura_Média_das_Árvores_Dominantes")
> points( m ~ idade, data=M.homo.MLE, pch=15, col="red", cex=2.5)
>
```

O modelo categórico heteroscedástico também é ajustado com a função “`gls`”, mas é necessário definir uma estrutura particular de variância, para fazer o desvio padrão seja estimado para cada nível da “categoria idade”. A estrutura de variância é definida pelo argument “`weights`”:

```
weights = varIdent(form= ~ 1 | factor(idade))
```

Nessa definição, a função “`varIdent`” indica que cada nível da medida categórica (`factor(idade)`) possui um desvio padrão próprio.

Assim, o modelo heteroscedástico pode ser ajustado e avaliado graficamente:

```
M.hetero <- gls( mhdom ~ factor(idade), data=florin, method="ML",
+               weights = varIdent(form= ~ 1 | factor(idade)) )
> summary(M.hetero)
Generalized least squares fit by maximum likelihood
Model: mhdom ~ factor(idade)
Data: florin
      AIC      BIC   logLik
2080.48 2120.907 -1030.24

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | factor(idade)
Parameter estimates:
      2.5      3.5      4.5      4      5.5
1.0000000 0.8200773 0.9512615 0.9161780 0.7639242

Coefficients:
              Value Std.Error  t-value p-value
(Intercept)  13.57059 0.2368057  57.30685     0
factor(idade)3.5  7.12616 0.3289391  21.66407     0
factor(idade)4   11.07386 0.9722019  11.39049     0
factor(idade)4.5 12.45868 0.4015940  31.02307     0
factor(idade)5.5 13.33482 0.4543532  29.34901     0

Correlation:
              (Intr) f()3.5 fct()4 f()4.5
factor(idade)3.5 -0.720
factor(idade)4   -0.244 0.175
```

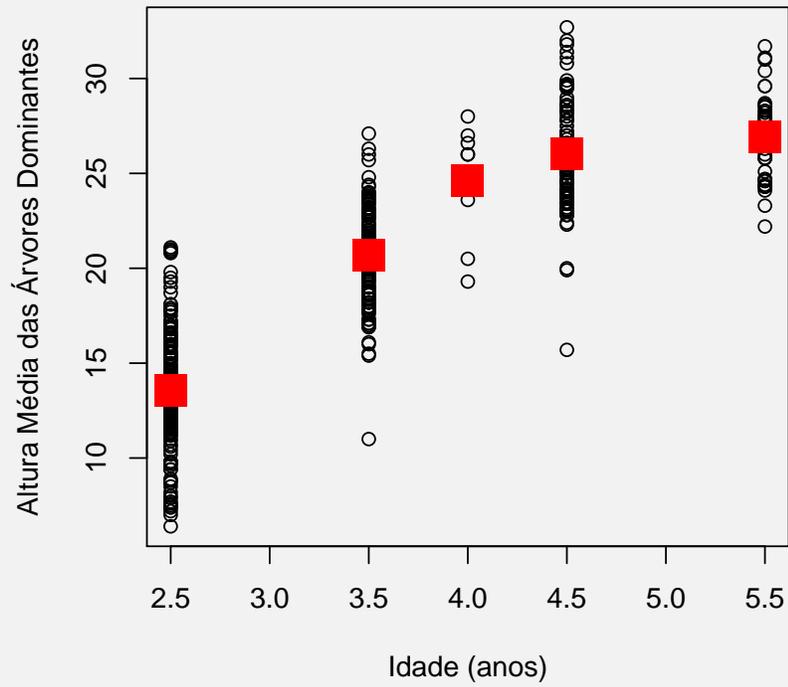


Figura 2: Gráfico de dispersão da altura média das árvores dominantes em função da idade para florestas de eucalipto. Os quadrados em vermelho são a altura média em cada idade.

```
factor(idade)4.5 -0.590 0.425 0.144
factor(idade)5.5 -0.521 0.375 0.127 0.307

Standardized residuals:
      Min      Q1      Med      Q3      Max
-3.85255717 -0.62654933 -0.04495628  0.68010786  2.54403792

Residual standard error: 3.069179
Degrees of freedom: 421 total; 416 residual
>
> plot(M.hetero)
> qqnorm(residuals(M.hetero))
> qqline(residuals(M.hetero))
>
```

Como no moodelo anterior, a função “`glm`” utiliza variáveis indicadoras para ajustar o modelo tanto no caso das médias quanto no caso dos desvios padrão. Logo, um pouco de manipulação do objeto se faz necessário para se obter as MLEs na forma com que foram apresentadas na exposição teórica acima.

```
> tmp1 = coef(M.hetero)
> tmp1 = tmp1 + c(0, rep(tmp1[1],4))
> s = M.hetero$sigma
> tmp2 = coef(M.hetero$modelStruct)
> tmp2 <- exp(c(0,tmp2)) * s
> M.hetero.MLE <- data.frame(idade=c(2.5,3.5,4.,4.5,5.5), m=tmp1, s=tmp2,
+ row.names=1:5 )
> M.hetero.MLE
  idade      m      s
1  2.5 13.57059 3.069179
2  3.5 20.69675 2.516964
3  4.0 24.64444 2.919592
4  4.5 26.02927 2.811914
5  5.5 26.90541 2.344620
>
```

Para a comparação dos modelos categóricos as MLEs de cada um podem ser reunidas numa mesma tabela:

```
> M.cat <- merge(M.homo.MLE, M.hetero.MLE, by="idade",
+ suffixes=c(".homo", ".hetero") )
> print(M.cat)
  idade  m.homo  s.homo m.hetero s.hetero
1  2.5 13.57059 2.822594 13.57059 3.069179
2  3.5 20.69675 2.822594 20.69675 2.516964
3  4.0 24.64444 2.822594 24.64444 2.919592
4  4.5 26.02927 2.822594 26.02927 2.811914
5  5.5 26.90541 2.822594 26.90541 2.344620
>
```

Note que as MLEs das médias são as mesmas em ambos os modelos, o que muda é apenas as MLES dos desvios padrão.

A comparação desses dois modelos categóricos com o modelo constante, em termos do critério de Akaike (AIC), mostra que ambos são melhores que o modelo constante, mas praticamente não possuem diferença entre si:

```
> anova(M.cte, M.homo, M.hetero, test=F)
      Model df      AIC      BIC    logLik
M.cte      1  2 2710.566 2718.651 -1353.283
M.homo      2  6 2080.453 2104.709 -1034.226
M.hetero    3 10 2080.480 2120.907 -1030.240
>
```

A melhor comparação entre os modelos categóricos, no entanto, é a apresentação gráfica (figura 3).

```
> attach(M.cat)
> plot( mhdom ~ idade, data=florin, xlab="Idade_(anos)",
+       ylab="Altura_Média_das_Árvores_Dominantes", xlim=c(2,6) )
> points( m.homo ~ idade, M.cat, pch=12, col="red", cex=1.8, lwd=2)
> arrows( idade, m.homo-s.homo, idade, m.homo+s.homo,
+         code=3, angle=90, length=0.1, col="red", lwd=2)
> points( m.hetero ~ I(idade+0.2), M.cat, pch=12, col="blue", cex=1.8, lwd=2)
> arrows( idade+0.2, m.hetero-s.hetero, idade+0.2, m.hetero+s.hetero,
+         code=3, angle=90, length=0.1, col="blue", lwd=2)
> legend(3.7, 12, legend=c("Homocedástico","Heterocedástico"), lty=1, lwd=3,
+        col=c("red","blue"))
> detach(M.cat)
```

Exercício

Utilizando os mesmos dados do exemplo acima, ajuste os modelos Gaussianos categóricos tendo como medida resposta o volume de madeira (`vol`) e como medida categórica a idade (`idade`).

Modelos Lineares Homocedásticos

Os modelos lineares homocedásticos são os modelos da “*Regressão Linear Clássica*”, simples ou múltipla. Para uma apresentação mais compacta, trataremos dos

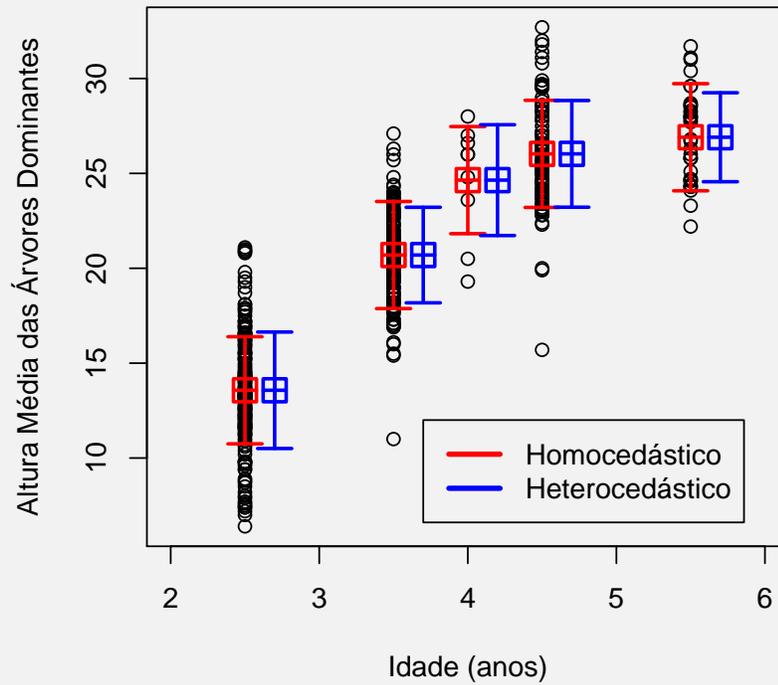


Figura 3: Gráfico de dispersão da altura média das árvores dominantes em função da idade para florestas de eucalipto com os modelos Gaussianos caetóricos.

modelos múltiplos na sua forma matricial:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (14)$$

onde

\mathbf{y} é o vetor das n observações da medida resposta:

$$\mathbf{y}' = [y_1 \quad y_2 \quad \dots \quad y_n]$$

\mathbf{X} é a matrix modelo com as n observações das p medidas predictoras:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

$\boldsymbol{\beta}$ é o vetor com os coeficientes do modelo

$$\boldsymbol{\beta}' = [\beta_1 \quad \beta_2 \quad \dots \quad \beta_p]$$

$\boldsymbol{\varepsilon}$ é o vetor dos erros de predição:

$$\boldsymbol{\varepsilon}' = [\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_p]$$

Estimativas de Máxima Verossimilhança

A função de log-verossimilhança também pode ser apresentada com base na forma matricial:

$$\mathbf{L}\{\boldsymbol{\beta}, \sigma_\varepsilon\} = \frac{n}{2} \ln 2\pi + n \ln \sigma_\varepsilon + \frac{1}{2\sigma_\varepsilon^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (15)$$

Igualando-se a zero as primeiras derivadas em relação aos coeficientes de regressão obtém-se um sistema de equações lineares (“*Sistema de Equações Normais*”):

$$\frac{\partial \mathbf{L}\{\boldsymbol{\beta}, \sigma_\varepsilon\}}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma_\varepsilon^2} (-\mathbf{X})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0 \implies \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} \quad (16)$$

cuja solução são as MLEs dos coeficientes de regressão:

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{y}. \quad (17)$$

A MLE do desvio padrão do erro de predição é obtida em função das MLEs dos coeficientes de regressão:

$$\begin{aligned} \frac{\partial \mathbf{L}\{\boldsymbol{\beta}, \sigma_\varepsilon\}}{\partial \sigma_\varepsilon} &= \frac{n}{\sigma_\varepsilon} - \frac{1}{\sigma_\varepsilon^3} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0 \\ \implies \hat{\sigma}_\varepsilon &= \sqrt{\frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})} \end{aligned} \quad (18)$$

Exemplo de Modelo Linear Homocedástico

Retomando o exemplo da altura média das árvores dominantes em florestas de eucalipto, podemos ajustar um modelo linear da altura das árvores dominantes em função da idade. Muito modelos lineares candidatos poderiam ser testados, contudo ajustaremos apenas um a título de exemplo. A posição das médias das alturas das árvores dominantes em cada idade (figura 2) sugere uma relação côncava, isto é, à medida que a idade cresce, o aumento na altura das árvores dominantes decresce. Um modelo linear possível para esse padrão é o modelo parabólico:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$

onde y_i seria altura média das árvores dominantes e x_i a idade do povoamento.

Como modelo de regressão linear, esse modelo pode ser ajustado pela função “lm”, o que é muito conveniente para avaliação numérica e gráfica do modelo:

```
> M.parabola.lm <- lm( mhd0m ~ idade + I(idade^2), data=florin )
> summary(M.parabola.lm)

Call:
lm(formula = mhd0m ~ idade + I(idade^2), data = florin)

Residuals:
    Min       1Q   Median       3Q      Max
-10.0441  -1.7441  -0.1242   1.8559   7.6070

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -18.3829     2.0960  -8.771  <2e-16 ***
idade       16.4641     1.1714  14.055  <2e-16 ***
```

Modelos Gaussianos Univariados

```
I(idade^2)  -1.4855    0.1542   -9.636   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.85 on 418 degrees of freedom
Multiple R-squared:  0.7776,    Adjusted R-squared:  0.7766 
F-statistic: 730.9 on 2 and 418 DF,  p-value: < 2.2e-16

> plot(M.parabola.lm)
```

Mas, esse modelo também pode ser ajustado pela função “gls” do pacote “nlme”, que também permite a análise do modelo ajustado, mas com um pouco mais de trabalho de programação:

```
> M.parabola <- gls(mhdom ~ idade + I(idade^2), data=florin, method="ML")
> summary(M.parabola)
Generalized least squares fit by maximum likelihood
Model: mhdom ~ idade + I(idade^2)
Data: florin
      AIC      BIC    logLik
2081.596 2097.767 -1036.798

Coefficients:
              Value Std. Error  t-value p-value
(Intercept) -18.382853  2.0959637  -8.770597    0
idade        16.464120  1.1714359  14.054648    0
I(idade^2)   -1.485507  0.1541588  -9.636213    0

Correlation:
      (Intr) idade
idade  -0.991
I(idade^2)  0.970 -0.993

Standardized residuals:
      Min      Q1      Med      Q3      Max
-3.53679440 -0.61414535 -0.04372286  0.65350967  2.67861580

Residual standard error: 2.839889
Degrees of freedom: 421 total; 418 residual
>
> plot(M.parabola)
> abline(h=0, lty=2, col="red")
> qqnorm(residuals(M.parabola))
> qqline(residuals(M.parabola), col="red", lwd=2)
```

Note que o modelo obtido pelos dois métodos é exatamente o mesmo, tanto nas MLEs dos coeficientes de regressão, quanto na MLEs do desvio padrão do erro de predição.

Os modelos podem ser comparados utilizando o critério de Akaike e o “*Teste da Razão de Verossimilhança*” (*L.Ratio*):

```
> anova(M.cte, M.homo, M.parabola.lm)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
M.cte      1  2 2710.566 2718.651 -1353.283
M.homo     2  6 2080.453 2104.709 -1034.226 1 vs 2 638.1127 <.0001
M.parabola.lm 3  4 2081.597 2097.767 -1036.798 2 vs 3  5.1436 0.0764
> AIC(M.cte, M.homo, M.parabola.lm)
      df      AIC
M.cte  2 2710.566
M.homo  6 2080.453
M.parabola.lm 4 2081.596
```

Nessa comparação faz sentido comparar o modelo categórico e o modelo linear através de um teste hierárquico como o teste da razão de verossimilhança (L.Ratio)? Sim! Esse teste é conhecido em regressão linear como “*teste da falta de ajuste*” (“*lack of fitness test*”), sendo geralmente apresentado na forma de um “teste F”. Note que o modelo linear (parabólico) possui menos parâmetro que o modelo categórico (3 contra 6) e apresenta uma “*perda*” da qualidade de ajuste. Essa “*perda*” não é “*significativa*” (p-value > 0.05), nem é “*relevante*” (diferença de AIC < ln(8)).

Esse modelo linear também pode ser comparado aos dois modelos anteriores utilizando um gráfico comparativo (figura 4).

Exercício

Utilizando os dados das florestas de eucalipto, ajuste um ou mais modelos Gaussianos lineares homocedásticos do volume de madeira (vol) em função da idade (idade). Compara os modelos obtidos com os modelos constante e categórico.

Modelos Não-Lineares Homocedásticos

O ponto característicos dos modelos Gaussianos não-lineares é que cada modelo não-linear é um caso absolutamente particular. Dizer que a função de predição ter um formato não-linear é o mesmo que dizer que ela pode ter qualquer forma, exceto a linear. Logo, cada forma não-linear particular representa um modelo distinto de todos os demais e que requer um estudo matemático próprio para a sua

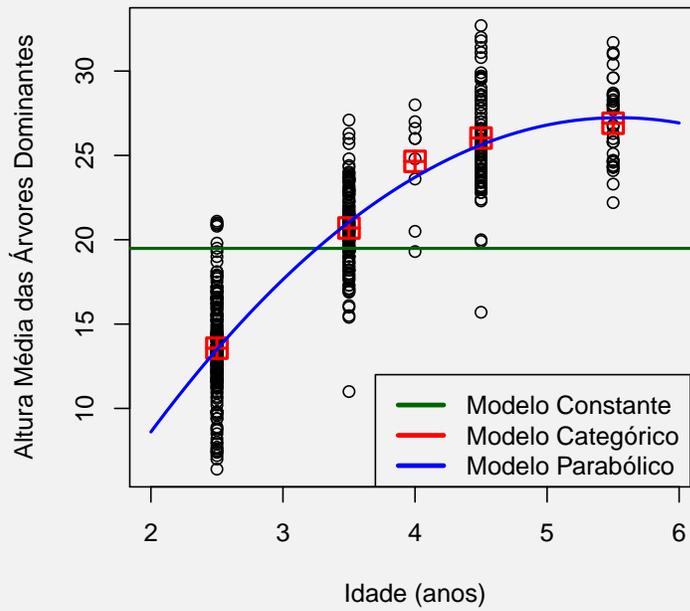


Figura 4: Comparação dos modelos constantes, categóricos e linear (parabólico) através do gráfico de dispersão da altura média das árvores dominantes em função da idade em florestas de eucalipto.

compreensão.

Outra implicação é que a forma não-linear implica na ausência de uma solução analítica explícita para as MLEs. Portanto, a única forma de ajuste das MLEs é através de métodos numéricos, de forma, que o estudo matemático de um modelo não-linear ajuda a compreensão do modelo, mas não influencia em nada a teoria estatística de estimação por máxima verossimilhança.

Estimativas de Máxima Verossimilhança

Nos modelos não-lineares, a forma geral das funções de verossimilhança e log-verossimilhança são:

$$\mathcal{L}\{\mu, \sigma_\varepsilon\} = (2\pi)^{-n/2} \sigma_\varepsilon^{-n} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (y_i - f(x_i; \beta_0, \dots, \beta_p))^2 \right] \quad (19)$$

$$\mathbf{L}\{\mu, \sigma_\varepsilon\} = \frac{n}{2} \ln 2\pi + n \ln \sigma_\varepsilon + \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (y_i - f(x_i; \beta_0, \dots, \beta_p))^2 \quad (20)$$

onde $f(x_i; \beta_0, \dots, \beta_p, \sigma_\varepsilon)$ é a função de predição não-linear, que é apresentada como tendo apenas uma medida preditora (x_i), mas que poderia ter duas ou mais.

Igualando-se a zero as primeiras derivadas parciais em relação aos parâmetros obtem-se:

$$\begin{aligned} \frac{\partial \mathbf{L}\{\beta_0, \dots, \beta_p, \sigma_\varepsilon\}}{\partial \beta_j} &= \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^n [y_i - f(x_i; \beta_0, \dots, \beta_p)] \frac{\partial f(x_i; \beta_0, \dots, \beta_p)}{\partial \beta_j} = 0 \\ \implies \hat{\beta}_j &= \sum_{i=1}^n [y_i - f(x_i; \hat{\beta}_0, \dots, \hat{\beta}_p)] \frac{\partial f(x_i; \hat{\beta}_0, \dots, \hat{\beta}_p)}{\partial \beta_j} = 0 \end{aligned}$$

O que mostra que a MLE de um dado coeficiente do modelo ($\hat{\beta}_j$) pode depender das MLEs de todos os demais coeficientes ($\hat{\beta}_0, \dots, \hat{\beta}_p$). Esse é justamente o carácter *não-linear* do modelo! Portanto, as MLEs dos coeficientes só pode ser obtida por métodos numéricos iterativos.

A MLE do desvio padrão do erro de predição é obtida em função das MLEs dos

coeficientes de regressão:

$$\frac{\partial \mathbf{L}\{\beta_0, \dots, \beta_p, \sigma_\varepsilon\}}{\partial \sigma_\varepsilon} = \frac{n}{\sigma_\varepsilon} - \frac{1}{\sigma_\varepsilon^3} \sum_{i=1}^n [y_i - f(x_i; \hat{\beta}_0, \dots, \hat{\beta}_p)]^2 = 0$$

$$\Rightarrow \hat{\sigma}_\varepsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n [y_i - f(x_i; \hat{\beta}_0, \dots, \hat{\beta}_p)]^2}$$

Um modelo não-linear de uso frequente na Biometria Florestal como função de predição é o modelo da Chapman-Richards, que graficamente gera uma função sigmoideal que é típica da relação entre o tamanho de um organismo e sua idade. Ele ilustra bem o problema da não-linearidade, que resulta nas primeiras derivadas se tornarem funções complexas dos coeficientes de regressão:

$$f(x_i, \beta_0, \beta_1, \beta_2) = \beta_0 [1 - \exp(-\beta_1 x_i)]^{\beta_2} \quad (21)$$

$$\frac{\partial f(x_i, \beta_0, \beta_1, \beta_2)}{\partial \beta_0} = [1 - \exp(-\beta_1 x_i)]^{\beta_2}$$

$$\frac{\partial f(x_i, \beta_0, \beta_1, \beta_2)}{\partial \beta_1} = \beta_0 \beta_2 [1 - \exp(-\beta_1 x_i)]^{\beta_2 - 1} [-\exp(-\beta_1 x_i)] (-\beta_1)$$

$$\frac{\partial f(x_i, \beta_0, \beta_1, \beta_2)}{\partial \beta_2} = \beta_0 [1 - \exp(-\beta_1 x_i)]^{\beta_2} \ln [1 - \exp(-\beta_1 x_i)]$$

Exemplo de Modelo Não-Linear Homocedástico

Podemos utilizar o modelo de Chapman-Richards como modelo não linear para relação entre a altura média das árvores dominantes e a idade de povoamentos de eucalipto. As MLEs podem ser obtidas utilizando a função “nls” que faz parte do pacote estatístico básico do R e o objetivo gerado também pode ser avaliado graficamente:

```
> M.richards <- nls( mhd0m ~ b0*(1 - exp(-b1*idade))^b2, data=florin,
+ start=list(b0=27, b1=0.7, b2=2.5))
> summary(M.richards)

Formula: mhd0m ~ b0 * (1 - exp(-b1 * idade))^b2

Parameters:
  Estimate Std. Error t value Pr(>|t|)
```

```

b0 29.5186 0.9328 31.644 < 2e-16 ***
b1 0.7984 0.0889 8.980 < 2e-16 ***
b2 5.3549 1.1294 4.742 2.91e-06 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.864 on 418 degrees of freedom

Number of iterations to convergence: 4
Achieved convergence tolerance: 4.795e-06

> plot(fitted(M.richards), residuals(M.richards))
> lines(lowess(fitted(M.richards), residuals(M.richards)), col="red")
> abline(h=0, lty=2)
> qqnorm(residuals(M.richards))
> qqline(residuals(M.richards), col="red", lwd=2)
>

```

Como se trata de um modelo complexo (não-linear) é importante verificar se as MLEs obtidas apresentam estabilidade. Isso é realizado visualmente através do gráfico da verossimilhança perfilhada, utilizando-se a função “plot.profmle”⁶. A figura 5 com as verossimilhança perfilhada para cada MLE é obtida pelo seguinte código do R:

```

> source("tese-plot-profmle.R")
> M.richards.prof <- profile(M.richards)
> plot.profmle(M.richards.prof, which=1)
> plot.profmle(M.richards.prof, which=2)
> plot.profmle(M.richards.prof, which=3)

```

Note que apenas a MLE do coeficiente ($\hat{\beta}_1$) apresenta certa simetria na curva perfilhada, mas o intervalo de log-verossimilhança de $\ln(8)$ não inclui o valor nulo (zero) em nenhuma das MLEs dos coeficientes.

É possível comparar o modelo não-linear com os modelos ajustados anteriormente:

```

> anova(M.cte, M.homo, M.parabola.lm, M.richards)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
M.cte      1  2 2710.566 2718.651 -1353.283
M.homo      2  6 2080.453 2104.709 -1034.226 1 vs 2  638.1127 <.0001
M.parabola.lm  3  4 2081.597 2097.767 -1036.798 2 vs 3   5.1436  0.0764
M.richards  4  4 2085.813 2101.983 -1038.906
> AIC(M.cte, M.homo, M.parabola.lm, M.richards)
      df      AIC
M.cte  2 2710.566
M.homo  6 2080.453

```

⁶Localização do arquivo com código: <http://cmq.esalq.usp.br/Philodendros/doku.php?id=lcf5877:rscripts:rscripts!>

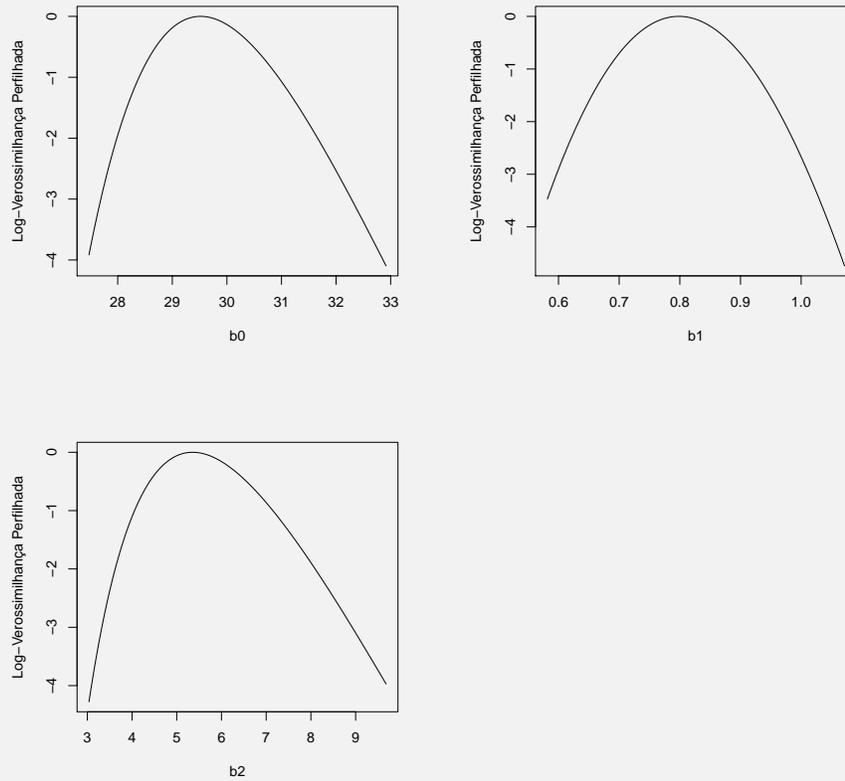


Figura 5: Verossimilhança perfilhada para as MLEs dos coeficientes de regressão do modelo Chapman-Richards ajustado pela função “nls”, aos dados de altura média de árvores dominantes em função da idade em florestas de eucalipto.

```
M.parabola.lm 4 2081.596
M.richards    4 2085.813
> anova(M.cte, M.homo, M.richards, M.parabola.lm )
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
M.cte      1  2 2710.566 2718.651 -1353.283
M.homo     2  6 2080.453 2104.709 -1034.226 1 vs 2  638.1127 <.0001
M.richards 3  4 2085.813 2101.983 -1038.906 2 vs 3   9.3600  0.0093
M.parabola.lm 4  4 2081.597 2097.767 -1036.798
```

A comparação com os modelos anteriores mostra que esse modelo não-linear teve desempenho inferior ao modelo linear (parabólico), pois sua performance foi marcadamente inferior ao modelo categórico. A comparação gráfica (figura 6), no entanto, sugere uma pequena diferença entre modelo linear e não-linear.

Exercício

Utilizando os dados das florestas de eucalipto, ajuste modelos Chapman-Richards como modelo não-linear homocedásticos do volume de madeira (`vol`) em função da idade (`idade`). Compara os modelos ajustados anteriormente.

Modelos Lineares Heteroscedásticos

Os modelos lineares heteroscedásticos diferem dos modelos clássicos de regressão linear pelo fato da variabilidade dos erros de predição (ε_i) não ser constante. Isso implica que a variabilidade não pode ser representada por uma constante (σ_ε), mas deve ser modelada por uma função das medidas predictoras $\mathbf{x}_i = [x_{1i}, \dots, x_{pi}]$ e com parâmetros $\alpha = [\alpha_1, \dots, \alpha_p]$: $f_\sigma(\mathbf{x}_i; \alpha)$.

Existem diversas funções apropriadas para modelar como a variabilidade dos erros de predição muda em função das medidas predictoras. Mas, a situação mais frequente nas medidas trabalhadas na Biometria Florestal é que essa variabilidade se mostre crescente com a medida preditora, seguindo a “Lei de Taylor”, isto é, seguindo uma função de potência:

$$f_\sigma(x_i; \sigma, \delta) = \sigma x_i^\delta \quad (22)$$

onde x_i é a medida preditora, σ é o parâmetro de proporcionalidade e δ é o expo-

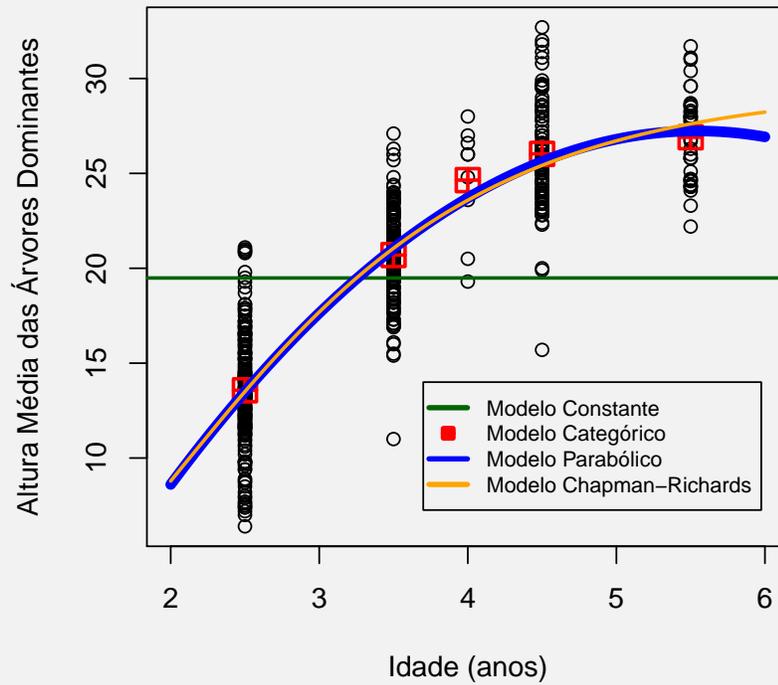


Figura 6: Comparação dos modelos constantes, categóricos, linear (parabólico) e não-linear (Chapman-Richards) através do gráfico de dispersão da altura média das árvores dominantes em função da idade em florestas de eucalipto.

ente da função de potência.

Estimativas de Máxima Verossimilhança

Para facilitar o desenvolvimento da estimativa por máxima verossimilhança, é conveniente expressar a função de predição, ainda que tenha uma forma linear, não na forma matricial, mas na forma de uma função:

$$f_{\mu}(x_{i0}, x_{i1}, \dots, x_{ip}; \beta_0, \dots, \beta_p) = \mathbf{X}'\boldsymbol{\beta} = \sum_{j=0}^p \beta_j x_{ij} \quad (23)$$

onde o índice i ($i = 1, 2, \dots, n$) indica as observações, o índice j ($j = 0, 1, 2, \dots, p$) indica as medidas preditoras (x_{ij}), sendo $x_{i0} = 1$ e β_j os coeficientes de regressão.

A função de log-verossimilhança nesse modelos é semelhante àquela dos modelos lineares clássicos, mas o desvio padrão constante (σ_{ε}) é substituído pela função de variabilidade:

$$\mathbf{L}\{\boldsymbol{\beta}, \boldsymbol{\alpha}\} = \frac{n}{2} \ln(2\pi) + \sum_{i=1}^n \ln(f_{\sigma}(\mathbf{x}_i; \boldsymbol{\alpha})) + \sum_{i=1}^n \frac{\left(y_i - \sum_{j=0}^p \beta_j x_{ij}\right)^2}{2 [f_{\sigma}(\mathbf{x}_i; \boldsymbol{\alpha})]^2} \quad (24)$$

As primeiras derivadas em relação aos coeficientes de regressão gera um sistema não-linear de equações:

$$\frac{\partial \mathbf{L}\{\boldsymbol{\beta}, \boldsymbol{\alpha}\}}{\partial \beta_k} = - \sum_{i=1}^n \frac{\left(y_i - \sum_{j=0}^p \beta_j x_{ij}\right) x_{ik}}{[f_{\sigma}(\mathbf{x}_i; \boldsymbol{\alpha})]^2} = 0$$

Para obter $\hat{\beta}_k \implies \sum_{i=1}^n \frac{x_{ik} \sum_{j=0}^p \beta_j x_{ij}}{[f_{\sigma}(\mathbf{x}_i; \boldsymbol{\alpha})]^2} - \sum_{i=1}^n \frac{y_i x_{ik}}{[f_{\sigma}(\mathbf{x}_i; \boldsymbol{\alpha})]^2} = 0 \quad (25)$

Note que esses sistema de equação depende simultaneamente de todos os coeficientes de regressão e dos parâmetros da função de variabilidade e, portanto, não possui solução algébrica explícita.

As primeiras derivadas em relação aos parâmetros da função de variabilidade gera um segundo sistema de equações que necessita ser solucionado para se obter as

MLEs:

$$\text{Para obter } \hat{\alpha}_h \implies \frac{\partial L\{\boldsymbol{\beta}, \boldsymbol{\alpha}\}}{\partial \alpha_h} = \sum_{i=1}^n \frac{1}{f_\sigma(\mathbf{x}_i; \boldsymbol{\alpha})} \left[\frac{\partial f_\sigma(\mathbf{x}_i; \boldsymbol{\alpha})}{\partial \alpha_h} \right] - \sum_{i=1}^n \frac{\left(y_i - \sum_{j=0}^p \beta_j x_{ij} \right)^2}{[f_\sigma(\mathbf{x}_i; \boldsymbol{\alpha})]^3} \left[\frac{\partial f_\sigma(\mathbf{x}_i; \boldsymbol{\alpha})}{\partial \alpha_h} \right] = 0 \quad (26)$$

As MLEs são obtidas pela solução simultânea dos dois sistemas de equações (25) e (29), que resultam das primeiras derivadas da função de log-verossimilhança em relação dos coeficientes de regressão e em relação aos parâmetros da função de variabilidade. Essa solução só pode ser obtidas por métodos numéricos iterativos.

Exemplo de Modelo Linear Heteroscedástico

Os modelos lineares heteroscedásticos podem ser ajustados utilizando a função “gls”, definindo o argumento “weights” em termos de uma função para variância (desvio padrão) dos erros de predição. A heteroscedasticidade das principais medidas trabalhadas na Biometria Florestal segue em geral a “Lei de Taylor”, de modo que a função apropriada para a variância é a “varPower”.

```
> M.parabola.hetero <- gls( mhdom ~ idade + I(idade^2), data=florin, method="ML",
+   weights = varPower(form= ~ idade) )
> summary(M.parabola.hetero)
Generalized least squares fit by maximum likelihood
  Model: mhdom ~ idade + I(idade^2)
  Data: florin
      AIC      BIC    logLik
2080.872 2101.085 -1035.436

Variance function:
Structure: Power of variance covariate
Formula: ~idade
Parameter estimates:
  power
-0.2075533

Coefficients:
              Value Std.Error   t-value p-value
(Intercept) -18.696029  2.0624308  -9.065045    0
idade        16.639371  1.1365247  14.640571    0
I(idade^2)   -1.508271  0.1477817 -10.206076    0

Correlation:
(Intr) idade
```

```
idade      -0.990
I(idade^2)  0.968 -0.993

Standardized residuals:
      Min      Q1      Med      Q3      Max
-3.73319928 -0.62364170 -0.05207721  0.65408184  2.65242063

Residual standard error: 3.637661
Degrees of freedom: 421 total; 418 residual
```

O modelo gerado pode ser avaliado graficamente para se avaliar as duas premissas básicas dos modelos de predição: (1) simetria dos erros em relação ao valor nulo e (2) distribuição Gaussiana dos erros.

```
> plot(fitted(M.parabola.hetero), residuals(M.parabola.hetero))
> lines(lowess(fitted(M.parabola.hetero), residuals(M.parabola.hetero)), col="red"
)
> abline(h=0, lty=2)
> qqnorm(residuals(M.parabola.hetero))
> qqline(residuals(M.parabola.hetero), col="red")
```

Infelizmente, objetos da classe “gls” não podem ser perfilhados com a função “profile”. Para avaliar a curva de verossimilhança perfilhada, é necessário ajustar o modelo utilizando a função “mle” do pacote “stats4” (distribuição padrão do **R**).

A função “mle” tem como primeiro argumento uma *função de logverossimilhança negativa (NLL)* que ao ser minimizada produz as MLEs. No modelo em questão a função NLL é

```
> nll.parabola.hetero <- function(b0, b1, b2, s, d)
+ {
+   x <- florin$idade
+   y <- florin$mhdom
+   mu <- b0 + b1*x + b2*x^2
+   sigma <- s*x^d
+   - sum(dnorm(y, mean=mu, sd=sigma, log=TRUE))
+ }
> formals(nll.parabola.hetero) <- alist(b0= -18.6960, b1= 16.6394, b2= -1.5083,
+   s = 3.6377, d= -0.2075)
>
```

A função “formals” associa os valores de inicialização à função NLL a ser minimizada. Note que, para maior eficiência, os valores iniciais são definidos como iguais às MLEs geradas pela função “gls”.

A função “mle” gera um objeto que pode ser perfilhado e, conseqüentemente, a curva de verossimilhança perfilhada pode ser obtida:

```
> M.parabola.hetero.mle <- mle(nll.parabola.hetero)
> M.parabola.hetero.mle.prof <- profile(M.parabola.hetero.mle)
There were 50 or more warnings (use warnings() to see the first 50)
```

A figura 12 apresenta as curvas obtidas, que mostram o bom comportamento das MLEs.

```
> plot.profmle(M.parabola.hetero.mle.prof)
```

A comparação do modelo linear heteroscedástico contra o modelo homocedástico através do AIC, mostra que a modelagem gerou uma pequena melhora, mas que não chega a ser marcante.

```
> AIC(M.parabola, M.parabola.hetero)
          df      AIC
M.parabola      4 2081.596
M.parabola.hetero 5 2080.872
```

A comparação dos coeficientes de regressão mostra pouca diferença entre os dois modelos

```
> coef(M.parabola)
(Intercept)      idade      I(idade^2)
-18.382853    16.464120    -1.485507
> coef(M.parabola.hetero)
(Intercept)      idade      I(idade^2)
-18.696029    16.639371    -1.508271
> coef(M.parabola.hetero.mle)
          b0      b1      b2      s      d
-18.6959991 16.6394038 -1.5082827 3.6376989 -0.2075051
```

e a comparação gráfica (figura 13) sugere absolutamente nenhuma mudança.

Exercício

Utilizando os dados das florestas de eucalipto, ajuste o modelo linear heteroscedástico do volume de madeira (vol) em função da idade (idade). Compara os modelos ajustados anteriormente.

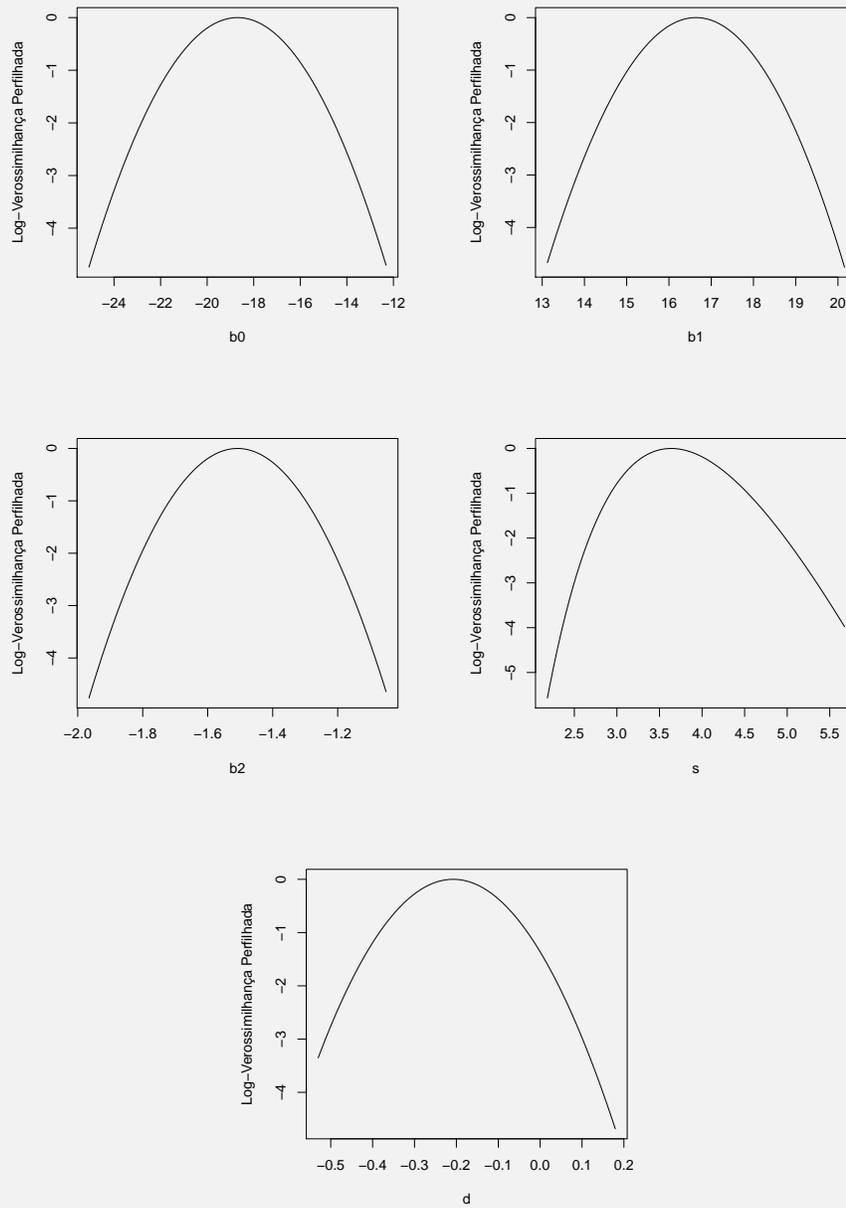


Figura 12: Curvas de logverossimilhança perfilhada para as MLEs dos parâmetros do modelo linear heteroscedástico.

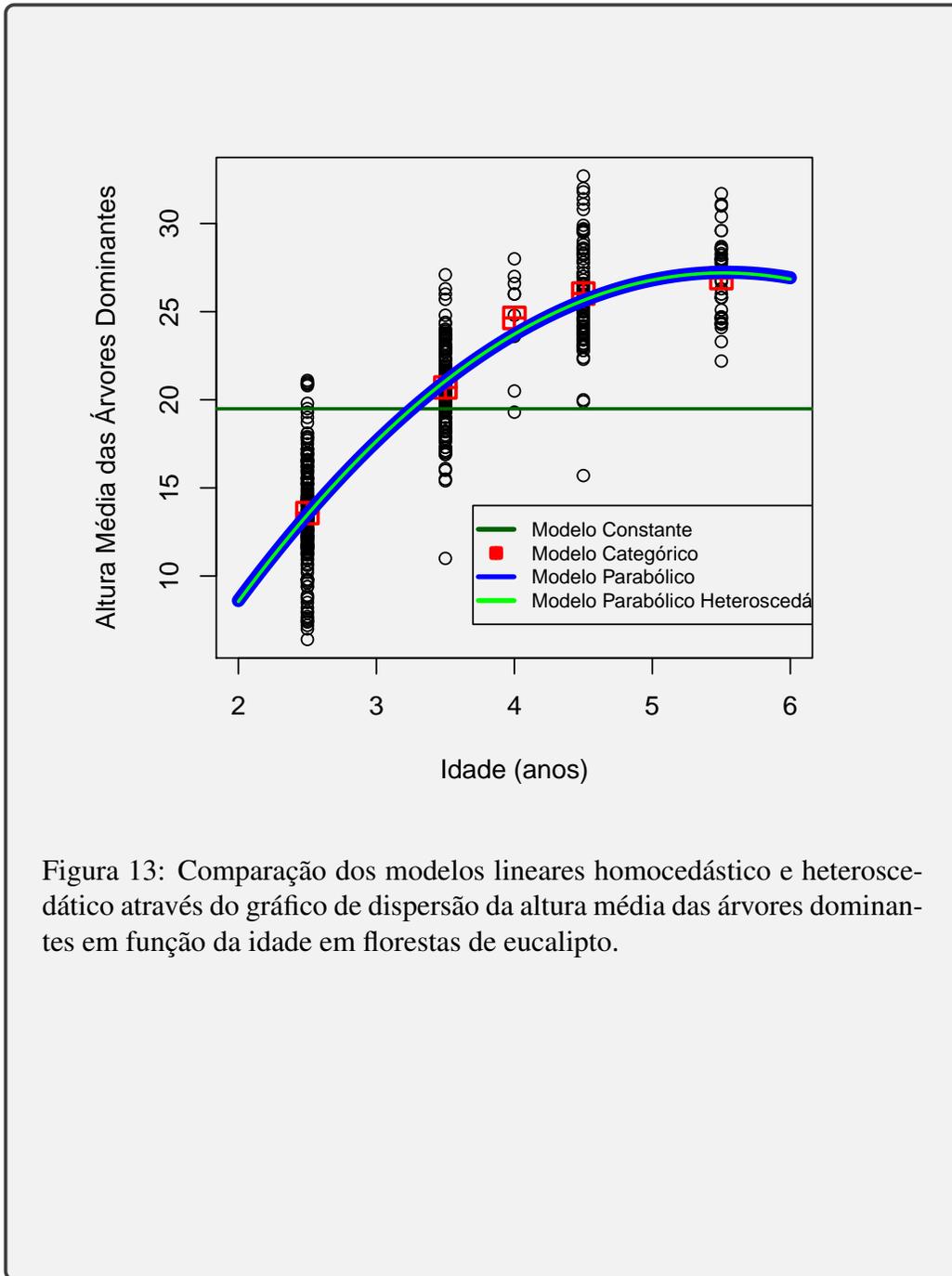


Figura 13: Comparação dos modelos lineares homocedástico e heteroscedástico através do gráfico de dispersão da altura média das árvores dominantes em função da idade em florestas de eucalipto.

Modelos Não-Lineares Heteroscedásticos

Como no modelo linear, a heteroscedasticidade nos modelos não-lineares implica que o desvio padrão do erro de predição (σ_ε) deixa de ser constante e passa a ser definido por uma função. Esse aspecto implica em mais um aspecto dos dados que deve ser modelado, mas a *Lei de Taylor* também é apropriada nesse caso, uma vez que o fenômeno da heteroscedasticidade está sempre associado à variável resposta a ser modelada. Assim, a função de potência permanece, a princípio, como a função mais apropriada para modelar a heteroscedasticidade.

Estimativas de Máxima Verossimilhança

A função de log-verossimilhança do modelo não-linear heteroscedástico tem uma forma muito complexa:

$$L\{\boldsymbol{\beta}, \boldsymbol{\alpha}\} = \frac{n}{2} \ln(2\pi) + \sum_{i=1}^n \ln(f_\sigma(\mathbf{x}_i; \boldsymbol{\alpha})) + \sum_{i=1}^n \frac{[y_i - f_\mu(\mathbf{x}_i; \boldsymbol{\beta})]^2}{2 [f_\sigma(\mathbf{x}_i; \boldsymbol{\alpha})]^2} \quad (27)$$

de forma que o sistema de equações que deveria ser solucionado para se obter as MLEs quase nunca terá uma solução analítica:

$$\frac{\partial L\{\boldsymbol{\beta}, \boldsymbol{\alpha}\}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{[y_i - f_\mu(\mathbf{x}; \boldsymbol{\beta})]}{[f_\sigma(\mathbf{x}; \boldsymbol{\alpha})]^2} \left[\frac{\partial f_\mu(\mathbf{x}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right] = 0 \quad (28)$$

$$\begin{aligned} \frac{\partial L\{\boldsymbol{\beta}, \boldsymbol{\alpha}\}}{\partial \boldsymbol{\alpha}} &= \sum_{i=1}^n \frac{1}{f_\sigma(\mathbf{x}; \boldsymbol{\alpha})} \left[\frac{\partial f_\sigma(\mathbf{x}; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \right] - \\ &\quad - \sum_{i=1}^n \frac{[y_i - f_\mu(\mathbf{x}; \boldsymbol{\beta})]^2}{[f_\sigma(\mathbf{x}; \boldsymbol{\alpha})]^3} \left[\frac{\partial f_\sigma(\mathbf{x}; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \right] = 0 \end{aligned} \quad (29)$$

Assim, as MLEs são sempre obtidas pela maximização da função de log-verossimilhança ou, na prática, pela minimização da função de log-verossimilhança *negativa*.

Exemplo de Modelo Não-Linear Heteroscedástico

Tomando-se novamente o modelo de Chapman-Richards, introduzimos a função de potência como função para o desvio padrão do modelo, de modo que o modelo

passa a ter 5 parâmetros: 3 coeficientes de regressão e 2 parâmetros para função de variância.

Função de Predição:

$$\mathbf{E}\{y_i|x_i\} = \beta_0 [1 - \exp(-\beta_1 x_i)]^{\beta_2} \quad (30)$$

Função de Desvio Padrão:

$$\mathbf{S}\{y_i|x_i\} = \sigma x_i^\delta \quad (31)$$

Esse modelo pode ser ajustado pela função “gnls” do pacote “nlme2”, que é uma generalização do ajuste dos Modelos Não-Lineares Clássicos (função “nls”). Como valores iniciais, podemos fornecer as MLEs obtidas para o modelo homocedástico:

```
> M.richards.hetero <- gnls( mhdom ~ b0*(1 - exp(-b1*idade))^b2, data=florin,
+   start=list(b0=29.5, b1=0.798, b2=5.35),
+   weights = varPower(form= ~ idade) )
> summary(M.richards.hetero)
Generalized nonlinear least squares fit
Model: mhdom ~ b0 * (1 - exp(-b1 * idade))^b2
Data: florin
      AIC      BIC    logLik
2085.759 2105.972 -1037.88

Variance function:
Structure: Power of variance covariate
Formula: ~idade
Parameter estimates:
  power
-0.1798878

Coefficients:
      Value Std.Error  t-value p-value
b0 29.38607  0.8529682  34.45154     0
b1  0.81214  0.0863003   9.41063     0
b2  5.53208  1.1469223   4.82341     0

Correlation:
      b0      b1
b1 -0.930
b2 -0.881  0.990

Standardized residuals:
      Min      Q1      Med      Q3      Max
-3.58732619 -0.62738444 -0.02968884  0.65716658  2.68714138

Residual standard error: 3.551215
Degrees of freedom: 421 total; 418 residual
>
```

As premissão do modelo também podem ser verificadas através da análise gráfica:

```
> plot(fitted(M.richards.hetero), residuals(M.richards.hetero))
> abline(h=0, lty=2)
> lines(lowess(fitted(M.richards.hetero), residuals(M.richards.hetero)), col="red")
>
> qqnorm(residuals(M.richards.hetero))
> qqline(residuals(M.richards.hetero), col="red")
```

Os objetos da classe “gnls” também não podem ser perfilhados, assim para se obter as curvas de verossimilhança perfilhada é necessário utilizar o mesmo esquema utilizado no modelo linear heteroscedástico, definindo-se a função NLL

```
> nll.richards.hetero <- function(b0, b1, b2, s, d)
+ {
+   mu = exp(b0)*(1 - exp(-b1*florin$idade))^b2
+   sigma = s*(florin$idade)^d
+   if( b0 > 0 && b1 > 0 && b2 > 0 && sigma > 0 )
+     - sum(dnorm(florin$mhdom, mean=mu, sd=sigma, log=TRUE))
+   else
+     NA
+ }
> formals(nll.richards.hetero) <- alist(b0=log(29.38607), b1=0.81214, b2=5.53208,
+ s=3.551215, d=-0.1798878)
>
```

utilizando a função NLL para se obter objeto da classe “mle”, perfilhando esse objeto e, finalmente, construindo-se os gráficos (figura 19):

```
> M.richards.hetero.mle <- mle(nll.richards.hetero)
> M.richards.hetero.mle.prf <- profile(M.richards.hetero.mle)
> plot.profmle(M.richards.hetero.mle.prf)
```

A diferença enter os modelos não-lineares homocedástico e heteroscedático é praticamente desprezível tanto em termos dos coeficientes de regressão quanto em termos do critério de Akaike, enquanto que em termos gráficos, a diferença é imperceptível (figura 20).

```
> coef(M.richards)
      b0      b1      b2
29.5186020 0.7983768 5.3549389
> coef(M.richards.hetero)
      b0      b1      b2
29.3860685 0.8121396 5.5320797
> AIC(M.richards, M.richards.hetero)
      df      AIC
M.richards      4 2085.813
M.richards.hetero 5 2085.759
>
```

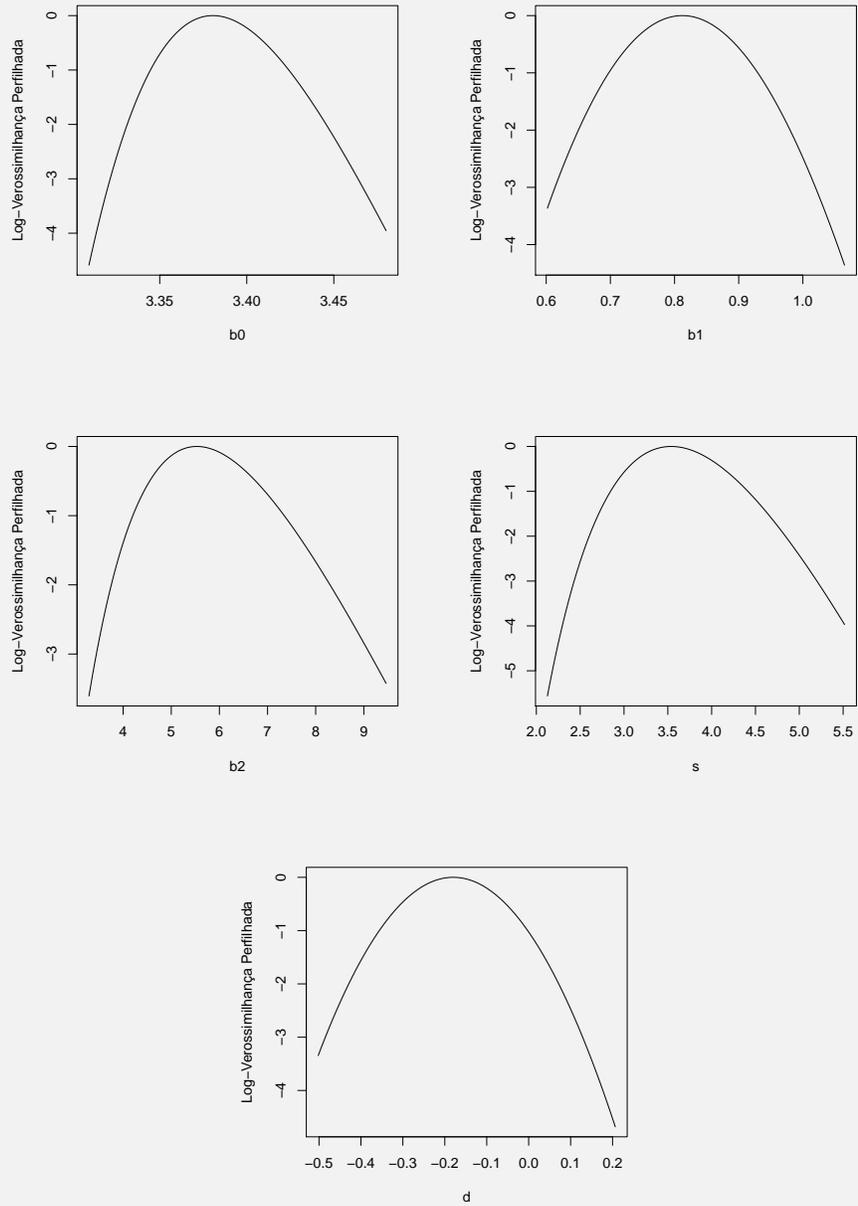


Figura 19: Curvas de logverossimilhança perfilhada para as MLEs dos parâmetros do modelo não-linear heteroscedástico.

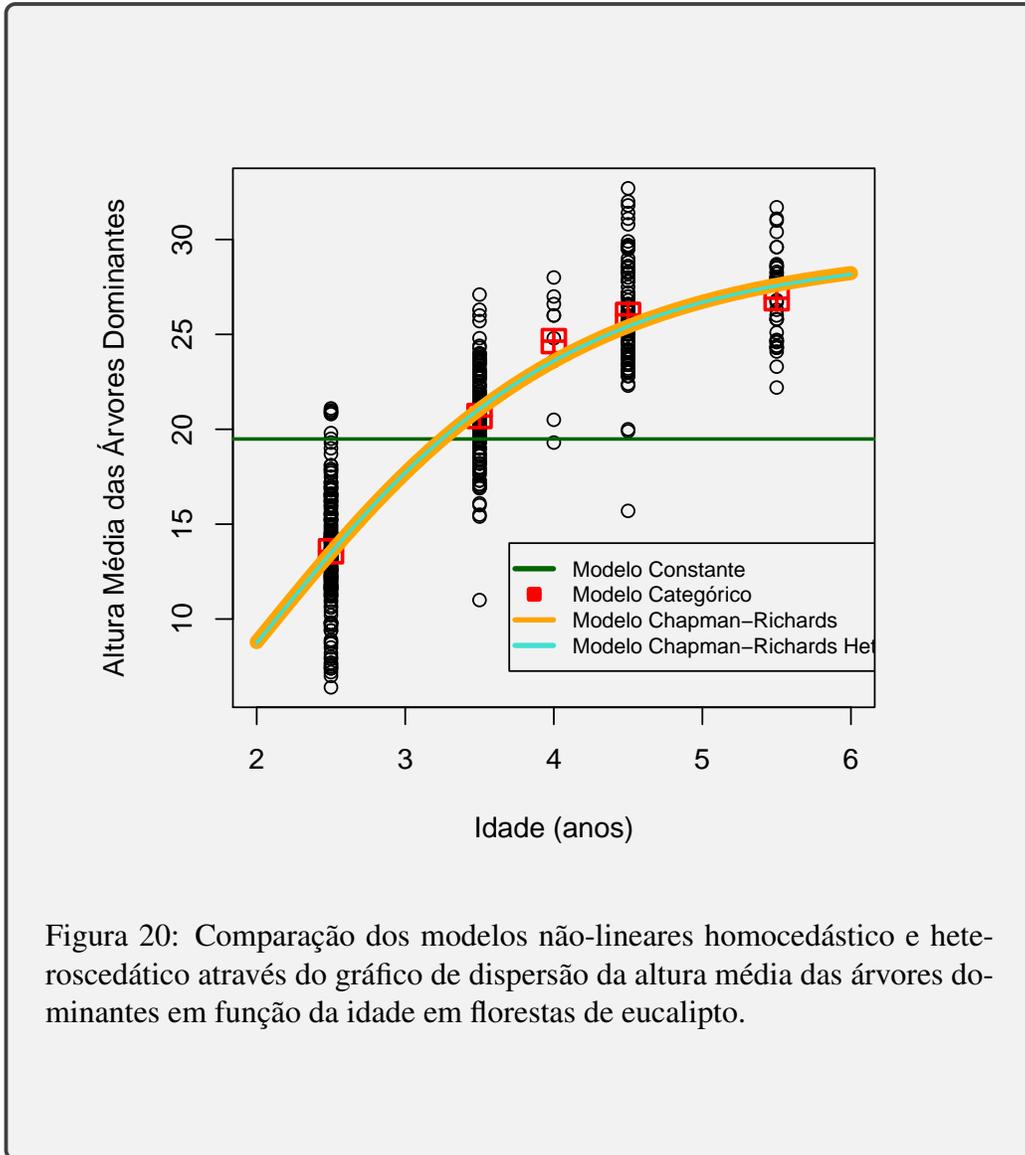


Figura 20: Comparação dos modelos não-lineares homocedástico e heteroscedástico através do gráfico de dispersão da altura média das árvores dominantes em função da idade em florestas de eucalipto.

Exercício

Utilizando os dados das florestas de eucalipto, ajuste modelos Chapman-Richards como modelo não-linear heteroscedásticos do volume de madeira (`vol`) em função da idade (`idade`). Compara os modelos ajustados anteriormente.