

---

CURSO

**Modelos de regressão não linear**

---



58° RBRAS e 15° SEAGRO

---

CURSO

**Modelos de regressão não linear**

---

22 à 26 de julho de 2013

Walmes Marques Zeviani  
Paulo Justiniano Ribeiro Júnior  
Wagner Hugo Bonat

*Laboratório de Estatística e Geoinformação  
Departamento de Estatística  
Universidade Federal do Paraná*

## **Modelos de regressão não linear**

Walmes Marques Zeviani  
[www.leg.ufpr.br/~walmes](http://www.leg.ufpr.br/~walmes)

Paulo Justiniano Ribeiro Júnior  
[www.leg.ufpr.br/~paulojus](http://www.leg.ufpr.br/~paulojus)

Wagner Hugo Bonat  
[www.leg.ufpr.br/~wbonat](http://www.leg.ufpr.br/~wbonat)

Laboratório de Estatística e Geoinformação  
<http://www.leg.ufpr.br>  
Departamento de Estatística  
Universidade Federal do Paraná

Complementos online: <http://www.leg.ufpr.br/mrnl2013>  
Contato: [mrnl2013 @ leg.ufpr.br](mailto:mrnl2013@leg.ufpr.br)

58° RBRAS e 15° SEAGRO  
Campina Grande - PB  
22 à 26 de julho de 2013

# Prefácio

Em modelos regressão não-linear dados observados de uma variável resposta são descritos por uma função de uma ou mais variáveis explicativas que é não linear seus parâmetros. Assim como nos modelos lineares o objetivo é identificar e estabelecer a relação entre variáveis explicativas e resposta. Entretanto, enquanto os modelos lineares definem, em geral, relações empíricas, os modelos não-lineares são, em grande parte das vezes, motivados pelo conhecimento do tipo de relação entre as variáveis. Desta forma, as aplicações surgem nas diversas áreas onde relações físicas, biológicas, cinéticas, químicas, fisiológicas, dentre outras, são estabelecidas por funções não lineares que devem ter coeficientes (parâmetros) identificados (estimados) a partir de dados observados ou experimentais.

O texto apresenta uma seleção de tópicos e exemplos de análises de dados utilizando modelos não-lineares. No primeiro capítulo contextualizamos modelos não lineares dentro de métodos mais usuais de regressão. No capítulo seguinte são apresentados conceitos, características e propriedades básicas de modelos não-linear. É dada ênfase à parametrização dos modelos. Métodos de inferência são revisados no Capítulo 3. Embora o método de mínimos quadrados seja mais usual nesta classe de modelos, destacamos a abordagem mais genérica de inferência baseada na verossimilhança. A prática de ajuste de modelos, explorando os procedimentos e resultados de inferência a predição é apresentada no Capítulo 4, com exemplos de códigos para análises em ambiente computacional. Nos demais capítulos são discutidas algumas extensões dos modelos como o ajuste e comparações de curvas em experimentos planejados (Capítulo 5), a adição de efeitos aleatórios (Capítulo 6) e a modelagem da variância (Capítulo 8). Incluímos ainda uma discussão sobre efeitos da parametrização do ponto de

vista computacional e inferencial.

O texto procura destacar aspectos práticos da análise de modelos não lineares, sem a intenção de cobrir extensivamente o tópico nem de apresentar fundamentações teóricas. A apresentação é compatível com um curso de graduação ou como parte introdutória e aplicada de um curso de pós-graduação. O texto pode ainda atender a profissionais de outras áreas que tenham a necessidade do uso de modelos não-lineares em análises de dados.

O leitor deve ter familiaridade com conceitos básicos de inferência estatística e fundamentos de regressão linear. Ao longo do texto são apresentadas ilustrações computacionais em ambiente R. O conhecimento da linguagem não é essencial para a compreensão e acompanhamento do texto. Entretanto procuramos oferecer exemplos de códigos e procedimentos para manipulação de dados e análises que esperamos ser úteis para usuários deste ambiente computacional. Complementos computacionais, bem como erratas e versões online deste material são disponibilizadas na página de complementos on-line do texto.

O desenvolvimento do material foi motivado pelos estudos de mestrado e doutorado do primeiro autor na UFLA (Universidade Federal de Lavras) e posteriormente pelas parcerias e trabalhos em conjunto estabelecidas com a equipe do LEG (Laboratório de Estatística e Geoinformação) da UFPR (Universidade Federal do Paraná)

Os autores são gratos a Comissão Organizadora da 58<sup>a</sup> RBras e 15<sup>o</sup> SEAGRO, e à RBras - Região Brasileira da Sociedade Internacional de Biometria pela confiança, oportunidade, incentivo e paciência para desenvolvimento e produção do presente texto. A produção de um material desta natureza só se completa com sua utilização e revisão crítica por parte de leitores e interessados. Desta forma esperamos receber comentários e sugestões pela página ou email associados ao texto.

# Conteúdo

<b>Prefácio</b>	<b>5</b>
<b>1 Modelos de regressão</b>	<b>1</b>
<b>2 Modelos não lineares</b>	<b>13</b>
2.1 Motivações . . . . .	14
2.2 Propriedades . . . . .	15
2.3 Interpretação . . . . .	22
2.4 Reparametrização . . . . .	25
2.4.1 Reparametrização 1:1 - funções MC com ênfase no $PF_q$	27
2.4.2 Reparametrização 2:2 - Parábolas com ênfase no PC	29
<b>3 Verossimilhança</b>	<b>33</b>
<b>4 Uma única curva</b>	<b>39</b>
4.1 Ajuste e diagnóstico . . . . .	39
4.2 Extendendo o modelo não linear . . . . .	45
4.3 Comparando parametrizações . . . . .	46
4.3.1 Reparametrização . . . . .	47
4.3.2 Medidas de curvatura . . . . .	47
4.4 Intervalos de confiança . . . . .	49
4.5 Bandas de confiança . . . . .	52
4.6 Comparando modelos . . . . .	53
<b>5 Curvas em experimentos planejados</b>	<b>57</b>
<b>6 Efeito aleatório</b>	<b>69</b>
6.1 Liberação de potássio . . . . .	69
6.2 Produção de soja . . . . .	78
<b>7 Efeito da parametrização</b>	<b>83</b>
<b>8 Modelagem da variância</b>	<b>89</b>
<b>Referências bibliográficas</b>	<b>98</b>





# Modelos de regressão

Modelos de regressão são usados para descrever o comportamento de uma variável aleatória (v.a.)  $Y$  como uma função de condicionantes  $x$  chamadas de variáveis explicativas, covariáveis ou estímulos. Com tais modelos busca-se encontrar e descrever padrões de homogeneidade dentre a heterogeneidade observada em um conjunto de observações de  $Y$ . Tal procedimento permite (i) explicar, ao menos parcialmente o comportamento de  $Y$ ; (ii) quantificar a influência das condicionantes em  $Y$ ; (iii) selecionar, mediante algum critério, as condicionantes relevantes; (iv) prever o comportamento de  $Y$  para estados observados e também não observados das condicionantes  $x$ ; (v) avaliar a incerteza associada a esse processo. O estudo de uma série de problemas práticos, nas mais diversas áreas, podem ser formatado sob esta abordagem, o que faz com que os modelos de regressão sejam largamente adotados e, provavelmente, a mais utilizada dentre as metodologias estatísticas.

Nessa introdução, vamos considerar que as condicionantes podem ser resumidas em medidas de uma única variável explicativa  $x$ . O objetivo na adoção de modelos de regressão é associar de alguma forma a distribuição (de probabilidades ou densidade de probabilidades) da v.a.  $Y$  a valores de  $x$  que denotamos por  $[Y|x]$ . A Figura 1.1 ilustra a ideia de regressão de forma genérica mostrando diferentes distribuições para  $Y$  para cada um dos valores de  $x$ . Entretanto, tal descrição genérica é de pouca valia na prática por ser de difícil interpretação e estimação a partir de dados reais. As diversas especificações de modelos de regressão definem restrições e condições sobre esta especificação genérica.

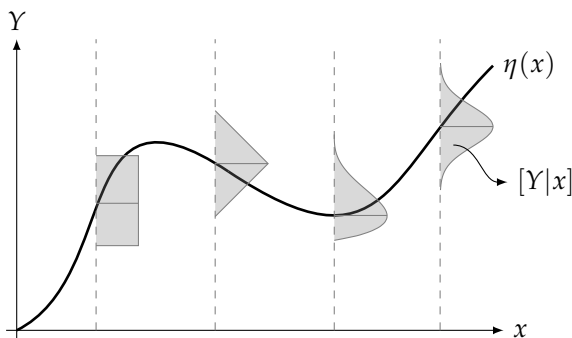


Figura 1.1: Representação esquemática genérica de um modelo de regressão.

Nos exemplos a seguir ilustramos alguns modelos frequentemente utilizados na literatura que são definidos impondo diferentes restrições e condições sobre a especificação genérica já mencionada. A primeira e mais comum restrição é assumir que todos os  $Y_i$  pertencem à uma mesma família de distribuições. Assumiremos aqui alguma família paramétrica conhecida definida para a função de probabilidade ou densidade de probabilidade  $[Y|x, \theta]$  que depende do valor de  $x$  e é indexada por  $\theta$ , um escalar ou vetor de parâmetros, usualmente desconhecido, a ser estimado a partir de dados disponíveis. A especificação do modelo se completa com uma segunda restrição sobre o relacionamento de alguma característica da distribuição de  $Y$  com a covariável  $x$ . Tipicamente relaciona-se o parâmetro que indexa a distribuição, ou alguma função desse parâmetro a  $x$ . Vamos impor também a condição de que tal relação é descrita por uma função com expressão fechada, que é a mesma para qualquer valor de  $x$ .

A Figura 1.2 mostra um exemplo que impõe: (i) que qualquer  $Y_i$  possui distribuição gaussiana; (ii) os  $Y_i$  possuem a mesma variância e média que depende do respectivo  $x_i$  pela relação  $E[Y_i|x_i] = \mu_i = \beta_0 + \beta_1 x_i$ . Este modelo pode ser escrito na forma a seguir.

$$\begin{aligned} Y_i &\sim N(\mu_i, \sigma^2) \\ \mu_i &= g(\beta, x_i) = \beta_0 + \beta_1 x_i. \end{aligned} \tag{1.1}$$

O modelo é chamado de *modelo de regressão linear simples*, sendo *linear* por impor que  $g(\mu_i)$  é uma função linear dos *coeficientes*  $(\beta_0, \beta_1)$  que são os elementos do vetor de parâmetros  $\beta$  e *simples* por tal função envolver apenas

uma variável explicativa  $x$ . O modelo pode ser escrito na forma alternativa,

$$\begin{aligned} Y_i | x_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2). \end{aligned} \tag{1.2}$$

Entretanto, a possibilidade de escrever na forma (1.2) é uma particularidade do modelo gaussiano e a especificação (1.1) é preferida e adotada nos exemplos a seguir por ter uma forma mais geral que permite explicitar mais claramente as restrições impostas por cada um dos modelos. Definimos ainda a função  $\nu(\cdot) = g^{-1}(\cdot)$  que nesse caso é dada por  $\nu(\mu_i) = \nu(\beta, x_i) = \mu_i = \beta_0 + \beta_1 x_i$ . Esta é chamada *função de ligação* por definir a relação do parâmetro da distribuição de  $Y$  com a covariável  $x$ . No exemplo a função de ligação é simplesmente dada pela função identidade, mas veremos outras formas nos exemplos a seguir.

O modelo de regressão linear é portanto definido pelas restrições de distribuição gaussiana homocedástica, ligação identidade com a  $E[Y|x]$  e função linear dos parâmetros. Cada uma destas restrições pode ser alterada ou relaxada na definição de modelos alternativos. Por exemplo, a forma linear na relação entre  $Y$  e  $x$  pode ser inadequada ou restritiva em um particular contexto. Possíveis alternativas incluem assumir: (i) uma regressão polinomial, como por exemplo uma forma quadrática  $\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$ ; (ii) uma ligação não identidade, como por exemplo  $\mu_i = \exp(\beta_0 + \beta_1 x_i)$  ou (iii) algum modelo intrinsecamente não-linear, por exemplo dado por algum sistema de equações diferenciais. Outras alternativas não consideradas neste texto incluem ainda modelos locais como regressão segmentada, *lowess*, *splines* ou modelos aditivos generalizados (GAM's).

Com todas as suas suposições, o modelo linear pode aparentar ser bastante restritivo impondo uma estrutura talvez arbitrariamente simples na relação entre  $Y$  e  $x$  sendo razoável que extensões com formas mais flexíveis sejam consideradas. No que se segue motivamos tais extensões por um argumento que explora a descrição de modelos lineares como aproximações de uma função em um certo intervalo (VENABLES; DICHMONT, 2004). Mostramos que o modelo linear gaussiano em sua forma mais simples pode ser obtido por uma aproximação de primeira ordem de uma função e que aproximações de maior ordem geram termos que podem induzir heterogeneidade de variâncias e assimetrias, o que motiva extensões nas alternativas de modelagem.

Como já mencionado, os modelos de regressão procuram relacionar o comportamento de uma variável resposta  $Y$  aos valores de uma variável explicativa  $x$ . Entretanto, quando verificado em dados reais, tal relaciona-

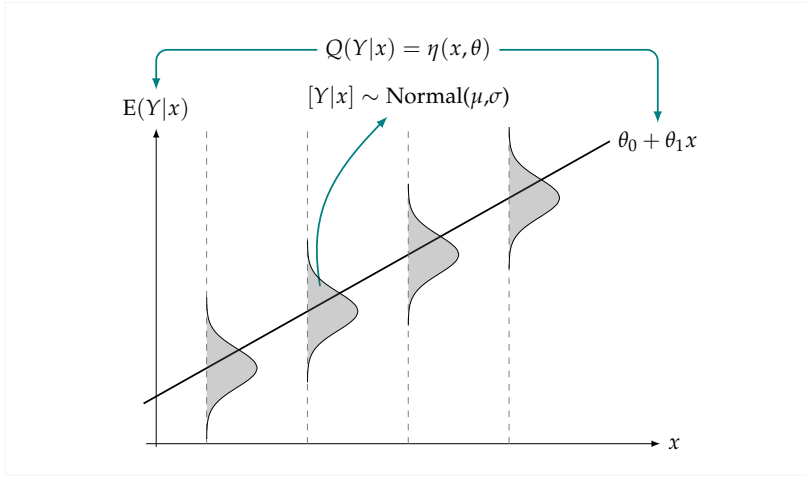


Figura 1.2: Modelo de regressão gaussiana (linear simples).

mento não é em geral perfeito e podemos considerar que  $Y$  é explicado não apenas por  $x$  mas também por desvios da relação entre esses que denotamos por uma variável  $z$  considerada aleatória. Desta forma, podemos considerar que a variável  $Y$  é descrita por alguma função não exatamente determinada

$$y = h(x, z) ,$$

em que é razoável assumir que os desvios  $z$  provém de variável aleatória centrada em zero. É usual que esperamos que o modelo descreva bem os dados na vizinhança de um particular valor  $x_0$ , tipicamente no centro dos valores disponíveis para  $x$ . Sendo a função contínua e suave nessa vizinhança. A aproximação de Taylor de primeira ordem para  $h(x, z)$  ao redor do ponto  $(x_0, 0)$  pode ser escrita na forma

$$\begin{aligned} y = h(x, z) &\approx h(x_0, 0) + \frac{\partial h(x_0, 0)}{\partial x} (x - x_0) + \frac{\partial h(x_0, 0)}{\partial z} (z - 0) \\ &= h(x_0, 0) + h_x(x_0, 0)(x - x_0) + h_z(x_0, 0)(z - 0) \end{aligned}$$

e considerando os valores das funções de  $(x_0, 0)$  como coeficientes obtemos a expressão do modelo de regressão linear simples

$$y = h(x, z) \approx \beta_0^{(c)} + \beta_1(x - x_0) + \sigma z = \beta_0 + \beta_1 x + \sigma z,$$

que pode ser escrito com a covariável centrada com coeficientes  $(\beta_0^{(c)}, \beta_1)$  ou não centradas com coeficientes  $(\beta_0 = \beta_0^{(c)} + \beta_1 x_0, \beta_1)$ .

Similarmente a expansão pode ser feita para um conjunto de covariáveis  $x_j$  levando ao modelo de regressão linear múltiplo. Vizinhanças mais amplas podem ser aproximadas por expansões de maior ordem levam ao modelo de regressão polinomial. Para a aproximação de segunda ordem obtemos:

$$\begin{aligned}
y &= h(x, z) \\
&\approx h(x_0, 0) + \frac{\partial h(x_0, 0)}{\partial x}(x - x_0) + \frac{\partial h(x_0, 0)}{\partial z}(z - 0) + \\
&+ \frac{1}{2!} \left[ \frac{\partial^2 h(x_0, 0)}{\partial x^2}(x - x_0)^2 + 2 \frac{\partial^2 h(x_0, 0)}{\partial x \partial z}(x - x_0)(z - 0) + \frac{\partial^2 h(x_0, 0)}{\partial z^2}(z - 0)^2 \right] \\
&= h(x_0, 0) + h_x(x_0, 0)(x - x_0) + h_z(x_0, 0)(z - 0) + \\
&+ \frac{1}{2!} [h_{xx}(x_0, 0)(x - x_0)^2 + 2h_{xz}(x_0, 0)(x - x_0)(z - 0) + h_{zz}(x_0, 0)(z - 0)^2] \\
&= h(x_0, 0) + h_x(x_0, 0)(x - x_0) + \frac{1}{2} h_{xx}(x_0, 0)(x - x_0)^2 + \\
&+ \left\{ h_z(x_0, 0)z + h_{xz}(x_0, 0)(x - x_0)z + \frac{1}{2} h_{zz}(x_0, 0)z^2 \right\} \\
&= \beta_0^{(c)} + \beta_1^{(c)}(x - x_0) + \beta_2(x - x_0)^2 + \left\{ \sigma_1 z + \sigma_2(x - x_0)z + \sigma_3 z^2 \right\} \\
&= \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon_1 + \epsilon_2 + \epsilon_3
\end{aligned}$$

em que os três primeiros termos definem um polinômio quadrático em  $x$  e os termos entre chaves descrevem desvios desse polinômio. Neste caso o desvio é composto por três componentes, sendo o primeiro ( $\epsilon_1$ ) um desvio "usual", o segundo ( $\epsilon_2$ ) um termo de interação entre  $x$  e  $z$  que pode portanto induzir heterogeneidade de variâncias e um terceiro ( $\epsilon_3$ ) que pode gerar assimetrias e curtoses gerando desvios do padrão gaussiano.

A discussão acima sugere que extensões do modelo linear devem considerar (i) relações não lineares entre  $Y$  e  $x$ , (ii) heterogeneidade de variâncias e (iii) não normalidade, possivelmente nessa ordem de relevância. Nos exemplos a seguir consideramos cada uma destas três generalizações.

Na Figura 1.3 ilustramos o seguinte modelo *não-linear*

$$\begin{aligned}
Y_i | x_i &\sim N(\mu_i, \sigma^2) \\
\mu_i &= g(\beta_0, \beta_1, x_i) = \frac{\theta_a x}{\theta_v + x_i} \quad ,
\end{aligned} \tag{1.3}$$

que generaliza o anterior sem impor que  $E[Y|x]$  seja dada por uma relação linear entre os parâmetros, em que  $g(\cdot)$  é uma função não-linear do vetor

de parâmetros  $\theta = (\theta_a, \theta_v)$ . A distribuição dos  $Y$ 's é ainda suposta gaussiana o que implica em erros aditivos à função média sendo ainda possível escrever o modelo na forma alternativa

$$Y_i|x_i = \frac{\theta_a x_i}{\theta_v + x_i} + \epsilon_i \quad (1.4)$$

$$\epsilon_i \sim N(0, \sigma^2).$$

Nesse exemplo temos que:

$$\mu_i = \frac{\theta_a x_i}{\theta_v + x_i}$$

$$\mu_i^{-1} = \frac{\theta_v + x_i}{\theta_a x_i} = \frac{1}{\theta_a} + \frac{\theta_v}{\theta_a} \frac{1}{x_i}$$

$$\mu_i^{-1} = \beta_0 + \beta_1 x_i^*$$

e portanto, o modelo é definido por uma *função inversa* de ligação  $\nu(\cdot)$ , uma *reparametrização* ( $\beta_0 = 1/\theta_a$  e  $\beta_1 = \theta_v/\theta_a$ ) e utilizando a covariável  $x^* = 1/x$ .

$$Y_i|x_i \sim N(\mu_i, \sigma^2) \quad (1.5)$$

$$\mu_i = g(\beta_0, \beta_1, x_i) = (\beta_0 + \beta_1 x_i^*)^{-1}.$$

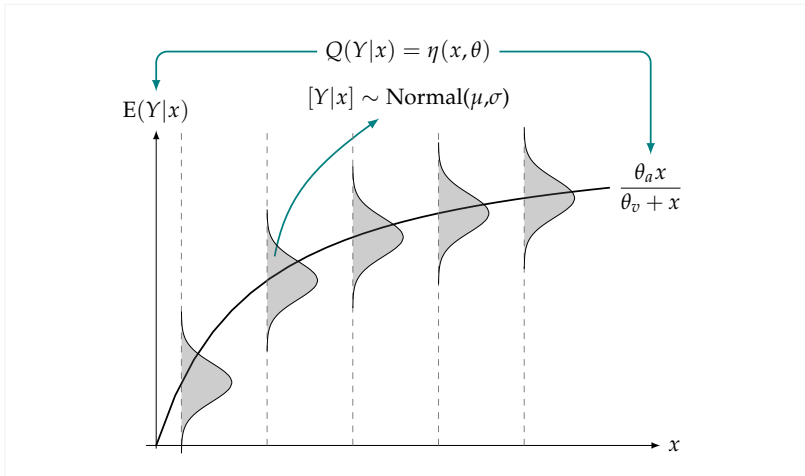


Figura 1.3: Modelo de regressão não linear gaussiano.

O modelo apresentado pode portanto ser dito ser um *modelo não-linear* gaussiano de  $Y$  como função de  $x$  ou, equivalentemente, um *modelo linear*

*generalizado* com distribuição gaussiana, função de ligação inversa e covariável  $1/x$ . Dessa forma, algoritmos de uma ou outra metodologia podem ser usados para ajustar o modelo para um determinado conjunto de dados e devem fornecer resultados equivalentes retornando as mesmas estimativas (após reparametrizações) e valores preditos. Na linguagem R pode-se fazer o ajuste com funções como `nls()` (*non linear least-squares*) ou `glm()` (*generalised linear models*).

Os casos considerados até aqui modelam a  $E[Y|x]$  como função de  $x$  e assumindo variância constante. Na Figura 1.4 pode-se ver uma generalização na qual a variância de  $Y$  se altera para diferentes valores de  $x$  segundo uma função que depende dos valores de  $x$  e é parametrizada por um (escalar ou vetor). O modelo considerado possui a forma

$$Y_i|x_i \sim \text{N}(\mu_i, \sigma^2) \quad (1.6)$$

$$\mu_i = g_\mu(\beta_0, \beta_1, x_i) = \frac{\theta_a x_i}{\theta_v + x_i}$$

$$\sigma_i^2 = g_\sigma(\alpha, x_i) \propto (\mu_i)^\delta$$

com a função de variância parametrizada por  $\alpha = (\beta_0, \beta_1, \delta)$ .

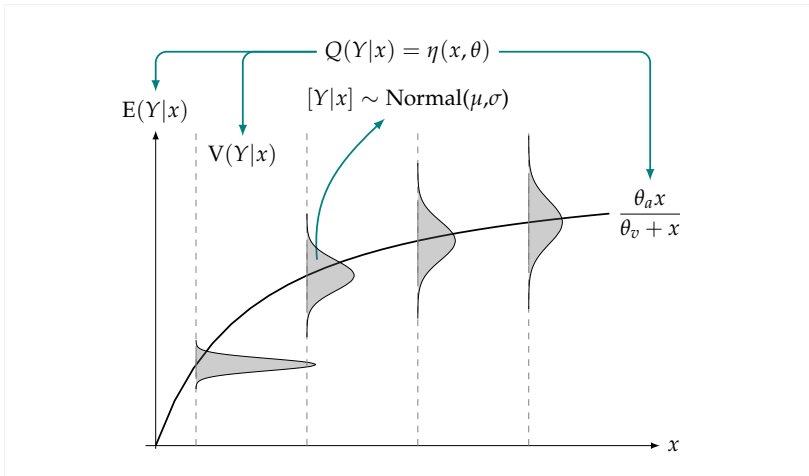


Figura 1.4: Modelo de regressão não linear gaussiano heterocedástico.

Os modelos considerados até aqui assumem erros simétricos (e gaussianos) ao redor da média. Pode-se generalizar tal restrição assumindo uma outra distribuição para  $Y$ . Tal distribuição deve pertencer a uma

outra família (não-gaussiana) de distribuições que podem ou não incluir como casos particulares os modelos de transformação da variável resposta  $Y$ . Esses casos particulares remetem aos modelos anteriores em uma outra escala da variável resposta e portanto não vamos considerá-los aqui. Vamos nos ater aos casos nos quais a escolha da distribuição é feita por adequação à natureza dos dados em questão.

A Figura 1.5 ilustra um modelo em que os valores da variável resposta  $Y$  são restritos ao intervalo unitário e podem ser interpretados como proporções. No exemplo assume-se a distribuição beta para  $Y$  e desta forma o modelo de regressão garante a restrição nos valores da resposta, bem como possui desvios com variâncias e assimetrias que variam com diferentes valores de  $x$ . O modelo é definido por

$$Y_i|x_i \sim \text{Beta}(\mu_i, \phi) \tag{1.7}$$

$$\mu_i = g_\mu(\beta_0, \beta_1, x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}.$$

A expressão da distribuição beta é usualmente dada em função de parâmetros  $(\alpha, \beta)$  tal que  $E[Y] = \alpha/(\alpha + \beta)$ . Nesta especificação adotamos uma forma reparametrizada da distribuição beta (??) com parâmetros  $\mu = \alpha/(\alpha + \beta)$  e  $\phi = \alpha + \beta$  que é conveniente para modelos de regressão relacionando valores da média  $\mu$  com os de  $x$  através da função de ligação  $\nu(\cdot) = g^{-1}(\cdot)$  que garanta valores da média no intervalo unitário. No exemplo assumimos ser dada pela função logística com  $\nu(\mu) = \log(\mu/1 - \mu)$ . Esta especificação é adotada por Ferrari e Cribari-Neto (2004) na proposição do modelo de regressão beta. O modelo tem a estrutura de um GLM mas pode, equivalentemente ser visto como um modelo beta não linear de regressão uma vez que  $E[Y]$  se relaciona com  $x$  segundo uma função não linear dos coeficientes do vetor  $\beta$ .

Bonat, Ribeiro Jr e Zeviani (2012) apresentam uma revisão e comparação baseada em diferentes aplicações de modelos de regressão para respostas no intervalo unitário. Neste trabalho os autores exploram uma especificação genérica para modelos de regressão com respostas no intervalo unitário que inclui diferentes distribuições para resposta, transformações e funções de ligação, de forma que diversos modelos propostos na literatura surgem como casos particulares.

Modelos para variáveis aleatórias discretas são ilustrados nas Figu-



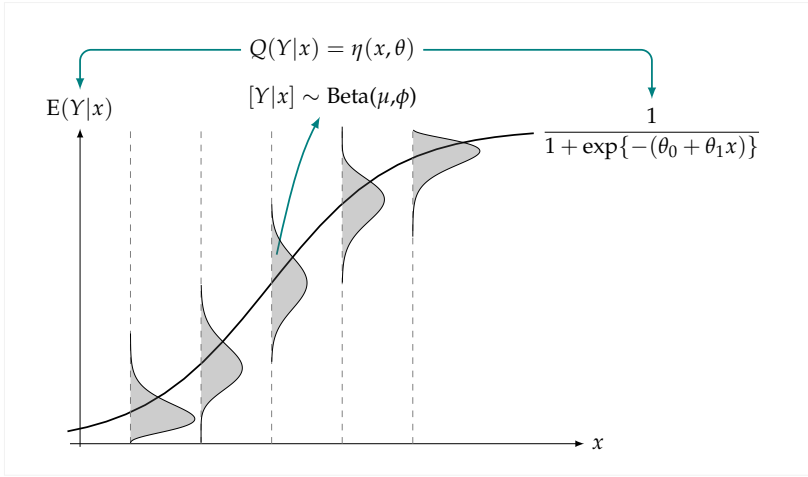


Figura 1.5: Modelo de regressão beta.

ras 1.6 em que

$$Y_i|x_i \sim B(n = 20, p_i) \quad (1.8)$$

$$\mu_i = np_i = g_\mu(\beta_0, \beta_1, x_i) = n \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} = \frac{n}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

e 1.7 em que

$$Y_i|x_i \sim P(\lambda_i) \quad (1.9)$$

$$\lambda_i = g_\lambda(\beta_0, \beta_1, x_i) = e^{\beta_0 + \beta_1 x_i}$$

No primeiro caso a resposta possui distribuição binomial com  $n = 20$  e no segundo a distribuição de Poisson. Em ambos casos a média é garantida ter valores compatíveis com o domínio de  $Y$  por uma escolha adequada para função ligação. Nos exemplos utilizamos a função de ligação *logística*  $v(\mu) = \log(\mu/(1 - \mu))$  para regressão binomial e a *logarítmica*  $v(\mu) = \log(\mu)$  para a Poisson.

O modelo binomial com função de ligação logística é também chamado de modelo de regressão logística e não deve ser confundido com o modelo não linear logístico que tem forma  $y = A/(1 + e^{-(\beta_0 + \beta_1 x)}) + z$  e que assume forma análoga a (1.4) com desvios gaussianos.

Modelos lineares e não lineares podem acomodar padrões de variabilidade nos dados como exemplificado na Figura 1.8. No modelo associa-se

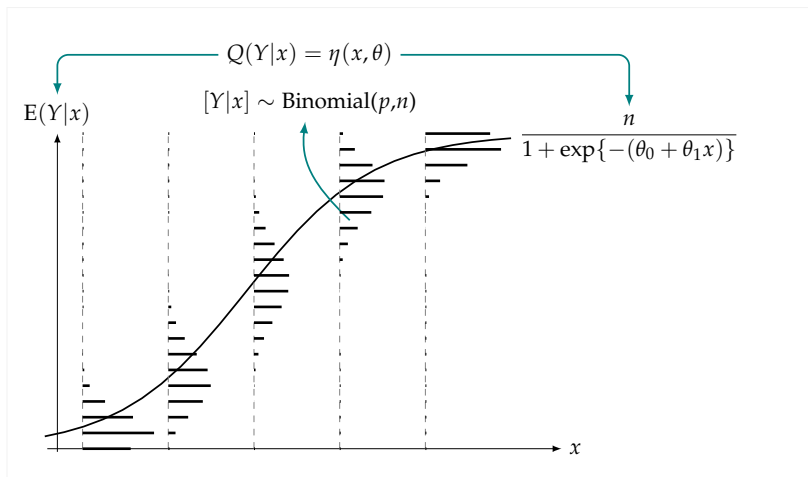


Figura 1.6: Modelo regressão binomial.

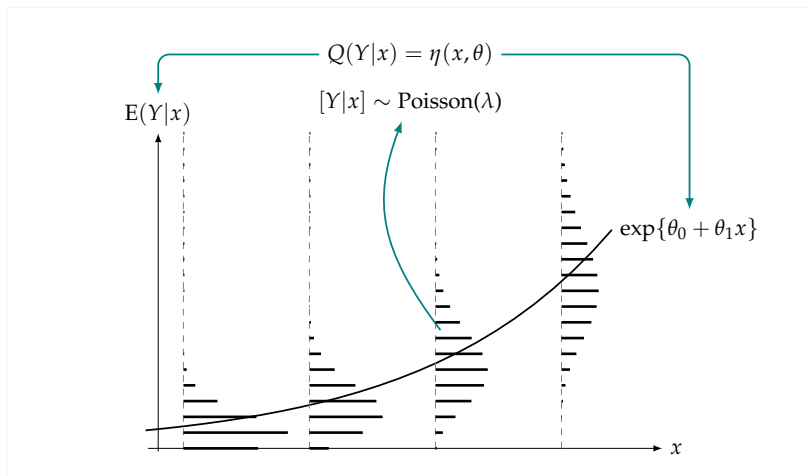


Figura 1.7: Modelo de regressão Poisson.

a um dos parâmetros um termo de flutuação aleatória, usualmente gaus-

siana e o modelo (1.3) se estende a

$$Y_i | (x_i, a_i) \sim N(\mu_i, \sigma^2) \quad (1.10)$$

$$\mu_i = g(\beta_0, \beta_1, x_i) = \frac{\theta_a x_i + a_i}{\theta_v + x_i}$$

$$a_i \sim N(0, \sigma_a^2). \quad (1.11)$$

Modelos desta natureza ditos modelos mistos pelo fato da a resposta são explicada não apenas por uma covariáveis  $x$  que possuem valores conhecidos ou observados, mas também por uma variável aleatória não observada. Tais modelos permitem incorporar fontes de variações conhecidas porém não diretamente observáveis e permitem decompor a variabilidade residual em componentes atribuíveis a tais fontes e não explicáveis.

Considere por exemplo que as curvas descrevam crescimento do valor de um atributo ao longo do tempo. O modelo misto descreve, em sua parte fixa, o comportamento médio dos indivíduos da população. A parte aleatória captura padrões de cada indivíduos gerando curvas que expressam diferentes valores finais, ainda que com a mesma forma funcional. Uma questão essencial em tais modelos é a definição de a qual parâmetro o efeito aleatório deve ser associado para capturar adequadamente as variações individuais. Desta forma a parametrização e interpretação a ela associada dos modelos é de particular relevância. Se por um lado o modelos mistos conferem maior flexibilidade, a estimação e identificabilidade a partir de dados esparsos é em geral mais difícil, reforçando a importância da parsimônia na modelagem.

Nesta introdução procuramos apresentar através de uma série de exemplos diversos modelos, enfatizando e relação entre eles e os aspectos que cada um permite capturar. Na sequência aprofundamos algumas idéias e tópicos de modelagem e inferência em modelos não-lineares.

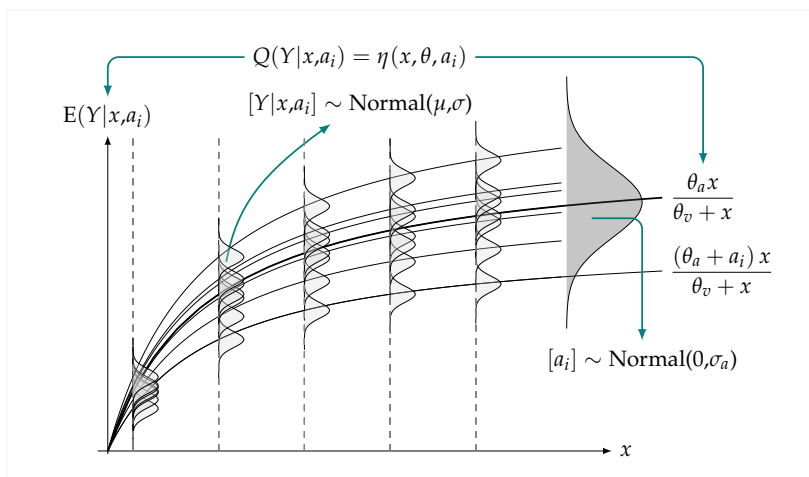


Figura 1.8: Modelo não linear de efeitos aleatórios.

# Capítulo 2

## Modelos não lineares

Modelos de regressão são amplamente usados em diversas aplicações para descrever a relação entre uma variável resposta  $Y$  e uma variável explicativa  $x$ . O modelo estatístico é linear se a quantidade de interesse, geralmente a média de  $Y$ , é função linear dos parâmetros, caso contrário é não linear. Por exemplo, em estudos de adubação a produção média (ton) pode ser modelada como função quadrática da dose de um nutriente (kg)

$$f(x) = a + bx + cx^2. \quad (2.1)$$

O modelo em 2.1 é linear nos parâmetros  $a$ ,  $b$  e  $c$ . No entanto, a produção média também pode ser modelada pela lei de Mitscherlich

$$f(x) = a(1 - \exp\{-b(x - c)\}). \quad (2.2)$$

em que  $a$  (ton) representa a produção máxima teórica possível quando se aumenta indefinidamente a dose do nutriente,  $b$  ( $\text{kg}^{-1}$ ) é interpretado como coeficiente de eficácia do nutriente, e  $c$  (kg) seria o conteúdo do nutriente no solo (Pimentel Gomes, 2009). Todos os parâmetros aparecem de forma não linear em 2.2. Uma vez que pelo menos um dos parâmetros aparece de forma não linear, o modelo é não linear.

A principal vantagem do modelo não linear sobre o linear é que 1) sua escolha está associada à conhecimento prévio sobre a relação a ser modelada e 2) geralmente apresenta interpretação prática para os parâmetros.

## 2.1 Motivações

Modelos não lineares (MNL) usualmente são sustentados por alguma informação sobre a relação entre  $Y$  e  $x$ . Tal informação está vinculada à diferentes graus de conhecimento como 1) uma análise de um diagrama de dispersão de  $y$  contra  $x$ , 2) restrições de forma da função (ser monótona, ser sigmóide), 3) a solução de uma equação diferencial sustentada por algum princípio/teoria, 4) ou a interpretação dos seus parâmetros. Seja qual for o grau de conhecimento, a escolha de um modelo não linear raramente é empírica.

Um número crescente de pesquisadores compartilham o sentimento de que as relações entre variáveis biológicas são melhores descritas por funções não lineares. Processos como crescimento, decaimento, nascimento, mortalidade, abundância, competição e produção raramente são relacionadas linearmente às variáveis explicativas (SCHABENBERGER; PIERCE, 2002, capítulo 5). Nesse sentido, pode-se dizer que os MNL melhor descrevem processos mecânicos e são úteis por acomodarem as restrições referentes à tais processos. Por exemplo, em estudos de produção-adubação sabe-se que os incrementos regulares de adubo ( $x$ ) não geram incrementos regulares, mas incrementos decrescentes na produção ( $Y$ ) e que a produção vai estabilizar após certo nível, pois passa ser limitada por outro fator, mesmo que a adubação seja aumentada indefinidamente. Essa é a lei do mínimo. Mesmo não avaliando a produção sob doses extremas de adubo, pode-se conhecer qual seria o seu valor limite ao considerar um MNL apropriado.

De maneira resumida, os MNL têm as seguintes vantagens sobre os modelos lineares (ML):

- Sua escolha têm sustentação baseada em teoria ou princípios mecânicos (físicos, químicos ou biológicos) ou qualquer outra informação prévia;
- Certos parâmetros são quantidade de interesse para o pesquisador providos de interpretação;
- Podem ser feitas previsões fora do domínio observado de  $x$ ;
- São parsimoniosos pois tipicamente possuem menos parâmetros;
- Partem do conhecimento do pesquisador sobre o fenômeno alvo;

Em contrapartida, as desvantagens são:

- Requerem procedimentos iterativos de estimação baseados no fornecimento de valores iniciais para os parâmetros;
- Métodos de inferência são aproximados;

- Exigem conhecimento do pesquisador sobre o fenômeno alvo.

## 2.2 Propriedades

Matematicamente, um MNL nada mais é que uma função não linear nos parâmetros. Estatisticamente, um modelo de regressão não linear, descreve alguma quantidade relacionada à distribuição de probabilidades de uma variável aleatória  $Y$  como uma função, não linear nos parâmetros. Na grande maioria das vezes, a quantidade de interesse é a média de  $Y$ . Dessa forma, em nossa notação, a partir de agora aprimorada para haver consistência ao longo do texto,

$$E(Y|x) = \eta(x, \theta) \quad (2.3)$$

representa um modelo que descreve, como quantidade de interesse, a média de uma variável aleatória  $Y$  em termos da variável explicativa  $x$  e do vetor de  $p$  parâmetros  $\theta = (\theta_1, \dots, \theta_p)^\top$  por meio da função não linear  $\eta$ .

Antes de seguir, será feita adequação da notação de 2.1 e 2.2 de acordo com 2.9. Dessa forma

$$\eta(x, \theta_0, \theta_1, \theta_2) = \theta_0 + \theta_1 x + \theta_2 x^2 \quad (2.4)$$

$$\eta(x, \theta_a, \theta_e, \theta_c) = \theta_a (1 - \exp\{-\theta_e(x - \theta_c)\}) \quad (2.5)$$

são MNL em que os parâmetros representados por  $\theta$  diferenciam-se pelo índice em subscrito. Em nossa notação, índices numéricos são usados para parâmetros sem, com vaga ou não útil (n.u.), interpretação e letras arábicas minúsculas para os interpretáveis. Por exemplo, em 2.5 os índices  $a$ ,  $e$  e  $c$  correspondem à *assíntota*, *eficiência* e *conteúdo* para levar sua interpretação.

Normalmente existem restrições ao espaço paramétrico  $\Theta$  de  $\theta$ , ao domínio e contra-domínio de  $\eta$  para que a função exista ou tenha interpretação. Por exemplo, em 2.5, os parâmetros  $\theta_a$ ,  $\theta_e$  e  $\theta_c$  devem ser não negativos bem como  $x$  e  $\eta$  por representarem dose e produção média. Além do mais, os parâmetros possuem unidades de medida (u.m.), que dependem de como aparecem em  $\eta$  e tal conhecimento é fundamental para compreensão de seu significado e descrição do modelo.

Por definição, o modelo é não linear se pelo menos uma das derivadas de  $\eta$  com relação à um dos parâmetros em  $\theta$  envolver parâmetros. Ao avaliar-se as derivadas de 2.4,

$$\frac{\partial \eta}{\partial \theta_0} = 1, \quad \frac{\partial \eta}{\partial \theta_1} = x, \quad \frac{\partial \eta}{\partial \theta_2} = x^2, \quad (2.6)$$

verifica-se não envolver parâmetros mas apenas as quantidades conhecidas 1,  $x$  e  $x^2$ . Por isso 2.4 é um modelo linear nos parâmetros. Por outro lado, ao avaliar-se as derivadas de 2.5

$$\begin{aligned}\frac{\partial \eta}{\partial \theta_a} &= 1 - \exp\{-\theta_e(x - \theta_c)\} \\ \frac{\partial \eta}{\partial \theta_e} &= -\theta_a(\theta_c - x) \exp\{-\theta_b(x - \theta_c)\} \\ \frac{\partial \eta}{\partial \theta_c} &= -\theta_a\theta_b \exp\{-\theta_b(x - \theta_c)\},\end{aligned}\tag{2.7}$$

observa-se que pelo dois parâmetros foram envolvidos nas derivadas e por isso 2.5 é modelo não linear.

Alguns autores definem classificações intermediárias. Um modelo é linearizável se por meio de uma transformação se torna um modelo linear. É o caso de

$$\eta(x, \theta) = \exp\{\theta_0 + \theta_1 x\}\tag{2.8}$$

que ao aplicar a transformação log se transforma em um modelo linear

$$\log(\eta(x, \theta)) = \theta_0 + \theta_1 x.\tag{2.9}$$

Modelos em que isso não acontece, como 2.5, são chamados de intrinsecamente não lineares.

Uma classificação útil dos MNL é quanto ao número de parâmetros que ocorrem de forma linear. Observe que nas derivadas de 2.5, dadas em 2.7, que aquela com relação à  $\theta_a$  não envolveu  $\theta_a$ . O mesmo não aconteceu para os demais parâmetros. Por esse fato, diz-se que 2.5 é parcialmente linear em  $\theta_a$  pois se for assumido por um instante que os demais parâmetros são conhecidos, tem-se um modelo linear, neste caso do tipo

$$\eta(x, \theta) = \theta_a z(x)\tag{2.10}$$

em que  $z$  é nada mais que uma transformação não linear de  $x$ , no caso  $1 - \exp\{\theta_e(x - \theta_c)\}$ , pois considerou-se que  $\theta_e$  e  $\theta_c$  são conhecidos. Então, para ser mais preciso, um MNL é linear em um parâmetro quando a segunda derivada com relação a ele for nula. Segundo tal definição tem-se que

$$\begin{aligned}\frac{\partial^2 \eta}{\partial \theta_a^2} &= 0 \\ \frac{\partial^2 \eta}{\partial \theta_e^2} &= -\theta_a(\theta_c - x)^2 \exp\{-\theta_e(x - \theta_c)\} \\ \frac{\partial^2 \eta}{\partial \theta_c^2} &= -\theta_a\theta_e^2 \exp\{-\theta_e(x - \theta_c)\}.\end{aligned}\tag{2.11}$$



Então o modelo 2.5 é parcialmente linear em um parâmetro,  $\theta_a$ . Verifica-se que 2.10 é o modelo linear simples pela origem.

Por essas simples definições, pode-se pensar que os ML são um caso particular dos MNL pois são àqueles nos quais as derivadas de primeira ordem não envolvem parâmetros e as derivadas de segunda ordem são nulas. Então, os modelos de regressão linear múltipla bem como os modelos polinomiais ou que envolvem transformações das variáveis explicativas, são todos lineares. Tanto que esses modelos podem ser escritos de forma matricial

$$\eta(x, \theta) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_p x_p = \mathbf{X}\theta. \quad (2.12)$$

Os MNL, embora o nome seja por ter parâmetros de forma não linear, descrevem relações não lineares entre as variáveis. Essas relações apresentam muitas formas e por isso é fundamental a capacidade de descrevê-las para, ou reconhecê-las a partir de, um MNL. A descrição do formato da função é algo obtido via estudo da função em  $x$  que visa inicialmente 1) limites, 2) derivadas, 3) integrais, 4) inversas e 5) avaliações em certos pontos.

São propriedades relevantes de uma função coisas como ter pontos característicos (crítico, inflexão), seu comportamento (concavidade, monotonicidade) ou aparentar padrões específicos (sigmoide, parabólico). Para referência futura, serão definidas algumas dessas propriedades à começar pelos pontos característicos relacionados ao eixo das coordenadas:

**ASS:** assíntota superior. Quando  $\lim_{x \rightarrow \pm\infty} \eta(x) < \infty$ , ou seja, tem-se um limite superior para  $\eta(x)$  para valores de  $x$  extremos. Produção como função de adubação e curvas de crescimento têm esse padrão. Por exemplo, 2.5 tem ASS com valor limite  $\theta_a$ .

**ASI:** assíntota inferior. Quando  $\lim_{x \rightarrow \pm\infty} \eta(x) > -\infty$ , ou seja, tem-se um limite inferior para  $\eta(x)$  para valores de  $x$  extremos. Curvas de retenção de água do solo e modelos liberação de nutrientes têm essa característica. O modelo  $\exp\{-\theta_0 x\}$  tem ASI com valor limite 0.

**PO:** ponto de origem ou intercepto. Valor de  $\eta(0)$ . É um ponto característico de muitos modelos e por isso é representado por um parâmetro. No modelo  $\theta_0 + \frac{\theta_a x}{\theta_v + x}$  tem-se que  $\theta_0$  é o PO e, para estudos de produção-adubação, é a produção esperada sem adubo. A ASS é  $\theta_a$ .

**PON:** ponto de origem nulo. Quando  $\eta(0) = 0$ . O modelo  $\theta_0(1 - \exp\{-\theta_1 x\})$  tem PON ao passo que 2.5 tem PO dependente dos parâmetros.

Como esses pontos se referem ao eixo das coordenadas, todos carregam a unidade de medida de  $\eta(x)$  que é Y. Pontos característicos no eixo das abscissas carregam a unidade de  $x$  que é X dos quais destacam-se:

**PC:** ponto crítico. Quando para algum  $x$  tem-se  $\eta'(x) = 0$ . Se  $\eta''(x) > 0$  o ponto é de mínimo e se  $\eta''(x) < 0$  é de máximo. O PC é um ponto característico de interesse em estudos de otimização como os de produção-adubação e curvas de produção-densidade. O modelo 2.4 tem PC em  $x = \frac{-\theta_1}{2\theta_2}$ .

**PI:** ponto de inflexão. Quando para algum  $x$  tem-se  $\eta''(x) = 0$ . É um ponto característico quase que exclusivo das curvas sigmóides consideradas em estudos de crescimento/progresso. Representa o valor em  $x$  que corresponde à maior taxa, ou seja, a fase de crescimento/progresso mais acelerado, a idade de mais alta conversão alimentar. O modelo  $1/(1 + \exp\{-\theta_s(x - \theta_i)\})$  tem PI representado por  $\theta_i$ .

**PS:** ponto de simetria. Quando para algum  $x$  tem-se  $|\eta(x + \delta) - \eta(x)| = |\eta(x - \delta) - \eta(x)|$  para todo  $\delta > 0$ . Em outras palavras, PS corresponde a um valor  $x$  que traça um eixo de simetria. Isso acontece, por exemplo, no modelo quadrático 2.4 que tem PS em  $x = \frac{-\theta_1}{2\theta_2}$  e no modelo logístico.

**PD:** ponto de descontinuidade. Quando para algum  $x$ ,  $\delta > 0$  e alguma derivada de  $\eta$  em relação à  $x$  de ordem  $k = 0, 1, \dots$ , tem-se  $\lim_{\delta \rightarrow 0} \eta^{(k)}(x + \delta) \neq \lim_{\delta \rightarrow 0} \eta^{(k)}(x - \delta)$ . É um ponto característico quase que exclusivo de modelos segmentados, como o linear-plato, quadrático-plato, bi-linear. No modelo linear-plato, o PD representa o valor em  $x$  para a troca de regimes, ou seja, para valores menores que  $x$ , a resposta tem taxa diferente de zero e para valores acima, a função é constante. Por isso tal modelo é considerado em estudos de produção-adubação e número ótimo de parcelas.

**PF:** ponto de fração. Suponha que  $\eta(x)$  seja uma função monótona de contra-domínio limitado, ou seja,  $a \leq \eta(x) \leq b$ , para  $a < b$ , então o valor  $x$  para o qual  $\eta(x) = a + q(b - a)$  é um ponto de fração  $q$ ,  $0 < q < 1$ . Funções sigmóides são limitadas pois têm assíntotas e um ponto de fração meio, por exemplo, é valor  $x$  correspondente ao ponto médio entre os limites da função. Em estudos de liberação de nutrientes no solo, o PF é importante para saber o tempo necessário para liberação de  $q = 90\%$  do total e em estudos de mortalidade-resposta é considerado para estabelecer a dose letal para  $q = 50\%$  da

população ( $DL_{50\%}$ ). No modelo  $\frac{\theta_a x}{\theta_v + x}$ , o PF meio ( $q = 0,5$ ) é representado por  $\theta_v$ , uma vez que o contra-domínio é  $[0, \theta_a]$ .

**PZ:** ponto de zero ou raiz da função. São os valores de  $x$  para os quais  $\eta(x) = 0$ . Em alguns modelos, tais pontos característicos são representados por parâmetros como é o caso do modelo beta generalizado,  $\theta_0(x - \theta_1)^{\theta_2}(\theta_3 - x)^{\theta_4}$ , considerado em estudos de crescimento-temperatura de fungos para o qual  $\theta_1$  e  $\theta_3$  são os PZ da função.

**POH:** ponto de offset horizontal. Análogo ao PO só que para o eixo  $x$ , é também um PZ. Em certos contextos, como liberação de nutrientes descrito por  $\theta_a(1 - \exp\{-\theta_c(x - \theta_h)\})$  sendo o POH o parâmetro  $\theta_h$ , é interpretado como período de aquecimento, uma vez que a liberação ocorre para  $x > \theta_h$ , para  $x < \theta_h$  não ocorre liberação pois os microorganismos envolvidos no processo estavam se “aquecendo”, ou melhor, estavam crescendo às custas dos nutrientes não liberados mas por eles consumidos nesse período.

Além dos pontos característicos, aspectos da forma de uma função são úteis para classificá-la. Dentre eles, devem ser destacados os seguintes:

**MC:** monótona crescente. Quando  $\eta(x_1) < \eta(x_2)$  para todo  $x_1 < x_2$  o que implica em  $\eta'(x) > 0$ , ou seja, a função é crescente. O modelo 2.5 é MC.

**MND:** monótona não decrescente. Quando  $\eta(x_1) \leq \eta(x_2)$  para todo  $x_1 < x_2$  o que implica em  $\eta'(x) \geq 0$ , ou seja, a função é crescente ou constante.

**MD:** monótona decrescente. Quando  $\eta(x_1) > \eta(x_2)$  para todo  $x_1 < x_2$  o que implica em  $\eta'(x) < 0$ , ou seja, a função é decrescente.

**MNC:** monótona decrescente. Quando  $\eta(x_1) \geq \eta(x_2)$  para todo  $x_1 < x_2$  o que implica em  $\eta'(x) \leq 0$ , ou seja, a função é decrescente ou constante.

**CCV:** côncava. Quando  $(\eta(x) + \eta(x + \delta))/2 < \eta(x + \delta/2)$  para todo  $x$  e  $\delta > 0$ . Em outras palavras, todo segmento de reta unindo dois pontos sobre a curva,  $(x_1, \eta(x_1))$  e  $(x_2, \eta(x_2))$ , está abaixo da curva no intervalo definido pelas projeções destes pontos na abscissa ( $x_1$  e  $x_2$ ). Tem-se que 2.5 é CCV.

**CVX:** convexa. Quando  $(\eta(x) + \eta(x + \delta))/2 > \eta(x + \delta/2)$  para todo  $x$  e  $\delta > 0$ . Em outras palavras, todo segmento de reta unindo dois pontos sobre a curva,  $(x_1, \eta(x_1))$  e  $(x_2, \eta(x_2))$ , está acima da curva no

intervalo definido pelas projeções destes pontos na abscissa ( $x_1$  e  $x_2$ ). Tem-se que  $\exp\{-\theta_0 x\}$  é CVX.

**SEG:** segmentado. Todo modelo segmentado tem PD com excessão apenas em  $\eta(x)$ . São assim os modelos linear-platô, quadrático-platô, bi-linear, tri-linear, etc.

**AAC:** área abaixo da curva. Quando  $|\int_{\forall x} \eta(x) dx| < \infty$ , em outras palavras, a integral de  $\eta(x)$  em todo domínio de  $x$  tem valor finito. É um caso, por exemplo, dos modelos baseados nas funções de densidade de probabilidade, como o modelo gama para curvas de lactação e o modelo beta para crescimento-temperatura em microbiologia e fitopatologia. Frequentemente se tem AAC tem também PC e ASS e ASI são 0. O modelo gama de lactação  $\theta_0 x^{\theta_1} \exp\{-\theta_2 x\}$  é AAC e o PC é de máximo em  $x = \theta_1 / \theta_2$ .

Certas combinações entre pontos e aspectos característicos resultam em padrões específicos de comportamento de função. Por vezes é mais rápido reconhecer as propriedades de uma função pela sua aparência com algum padrão específico. Dentre os mais comuns estão:

**SIG:** sigmóide. A curva tem formato de S ou S invertido. Toda sigmóide é MC ou MD, tem PI, tem ASS e ASI e eventualmente pode ter PS, além de permitir definir  $PF_q$ .

**SIN:** sinodal. Tem formato de sino. Com isso, toda sinodal tem PC, eventualmente pode ter PS, ASS, ASI e AAC.

**PAR:** parabólica. Tem formato de parabola ou de U, U invertido. Sendo assim, toda parabólica tem PC, é CON ou CVX e eventualmente pode ter PS.

**SEN:** senóide. Tem formato de onda, é oscilatória. São funções consideradas para modelar processos cíclicos como níveis hormonais em função dos períodos reprodutivos, por exemplo.

**EXP:** exponencial ou assintótica. Tem formato de j, valendo quatro orientações resultantes ao coexistir as características monótona crescente/decrecente com concava/convexa. Em termos de aplicação, tem-se que as MC-CON são consideradas para descrever a produção-adubação, as MC-CVX o crescimento populacional explosivo, as MD-CON a produção-desfolha e MC-CVX o decaimento na liberação diária de um nutriente pelo solo.

Todas essas propriedades são facilmente verificadas por meio de gráficos da função. Para melhor compreender o significado de cada item descrito, a coleção funções na Figura 2.1 destaca os pontos característicos e aspectos de forma de curvas variadas. As funções em A e B são monótonas, crescente e decrescente, têm ambas assíntotas e pontos característicos como PO, PI e PF mas somente A é simétrica com relação ao PI. Nas funções C e D destaca-se que são assintóticas, de acúmulo e decaimento, são concava e convexa, têm POH e PZ e são segmentadas. Em E e F têm-se funções com PC, no entanto E tem AAC, não tem PS, tem PON.

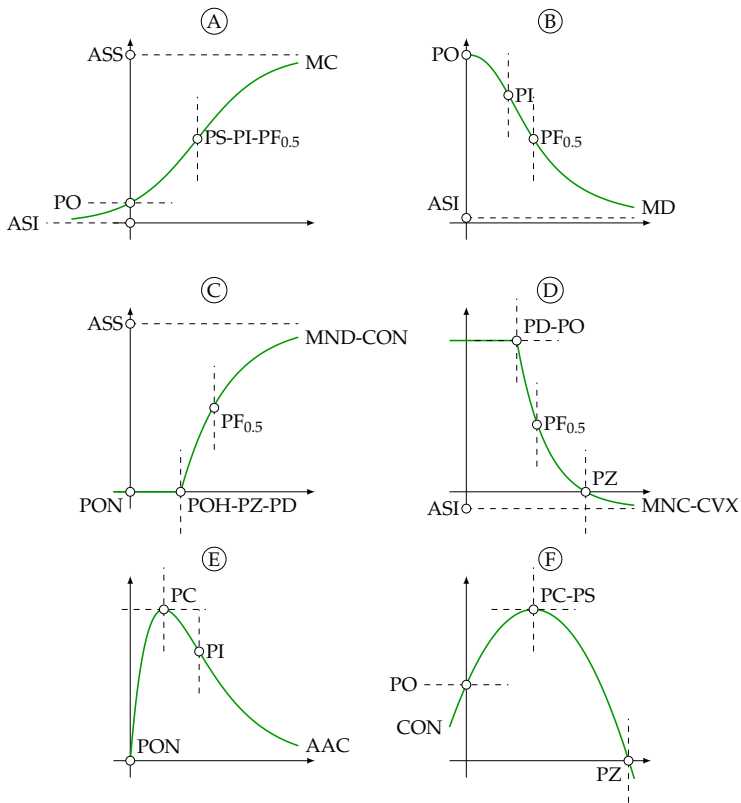


Figura 2.1: Funções não lineares com destaque para os pontos característicos e formas.

## 2.3 Interpretação

Ao assumir MNL para descrever um processo é fundamental conhecer o significado dos seus parâmetros. O estudo de dimensionalidade ou grandeza é o que permite reconhecer a unidade de medida (u.m.), ou dimensão, do parâmetro. Com informações da dimensionalidade é possível entender se o parâmetro corresponde à valores da variável explicativa, se tiver u.m.  $X$ , à valores da variável dependente, se tiver u.m.  $Y$ , à grandezas compostas como uma taxa de variação, se tiver u.m.  $Y X^{-1}$  ou se é adimensional (sem u.m.).

Além da dimensionalidade, o estudo do espaço paramétrico é igualmente importante pois permite reconhecer os valores que um parâmetro assume. Muitos processos envolvem variáveis positivas, como tempo, dose, produção, velocidade, e variáveis limitadas, tipicamente representadas em porcentagem, como teor (%), concentração (%) e rendimento (%). Para descrever adequadamente tais processos, ou seja, para que o domínio e contra-domínio do modelo pertençam ao suporte das variáveis envolvidas, precisa-se definir qual o espaço paramétrico válido para cada um dos parâmetros. São três os tipos de intervalo: intervalo não limitado, limitado de um lado e limitado dos dois lados.

Em muitos MNL os parâmetros são pontos característicos da função ou fatores/coeficientes de forma/escala. Não raramente, alguns parâmetros no modelo não têm interpretação direta, apresentam u.m. vaga, porém são necessários para garantir flexibilidade ao modelo. Independente da interpretação, é sempre vantajoso definir o espaço paramétrico e saber como a função se comporta com relação aos valores dos parâmetros. Determinar a dimensionalidade, o espaço paramétrico e estudar comportamento da função com relação à um parâmetro são as etapas iniciais e indispensáveis na aplicação de modelos de regressão linear. Tais informações vão facilitar a obtenção de valores iniciais, necessários para estimação além de permitir, nos modelos de efeitos mistos, corretamente especificar modelos para os parâmetros como função de covariáveis.

Para conhecer o significado de um parâmetro deve-se 1) determinar a dimensionalidade e 2) estudar a função em termos do parâmetro. A dimensionalidade é obtida por meio de operações com as u.m. dos termos conhecidos do modelo, no caso  $x$  e  $\eta(x)$ . O estudo da função é similar ao que se faz em termos de  $x$ , ou seja, considera os 1) limites, 2) derivadas, 3) integrais, 4) inversas e 5) avaliações em certos pontos mas com foco nos parâmetros.

Considere o modelo linear simples

$$\eta(x) = \theta_0 + \theta_1 x, \quad (2.13)$$

para o qual, a partir das u.m. de  $x$  e  $\eta$ , determinou-se que  $\theta_0$  tem u.m.  $Y$  e  $\theta_1$  tem u.m.  $YX^{-1}$ . Verifica-se pela u.m. que  $\theta_1$  representa uma taxa e  $\theta_0$  um valor no eixo das coordenadas. Ao obter  $\eta'(x) = \theta_1$  certifica-se de que  $\theta_1$  é realmente o parâmetro de taxa e por  $\eta(0) = \theta_0$  tem-se que  $\theta_0$  é o PO ou intercepto do modelo. Se não forem feitas restrições ao domínio e contra-domínio de  $\eta(x)$ , o espaço paramétrico é a reta real para ambos parâmetros.

Para o modelo Michaelis-Menten

$$\eta(x) = \theta_a \frac{x}{\theta_v + x}, \quad (2.14)$$

o emprego do mesmo raciocínio obtém-se que  $\theta_a$  e  $\theta_v$  têm u.m. de  $Y$  e  $X$ , então que podem ser pontos característicos da função  $\eta(x)$ . Pelo estudo da função tem-se que  $\theta_a$  é a ASS uma vez que  $\lim_{x \rightarrow \infty} \eta(x) = \theta_a$ , e  $\theta_v$  é o  $PF_{0.5}$  uma vez que  $\eta(\theta_v) = \theta_a/2$ . Se o modelo for considerado para estudos de produção-adubação, quantidades não negativas, os espaços paramétricos são  $\Theta_a : \{\theta_a \geq 0\}$  e  $\Theta_v : \{\theta_v > 0\}$ .

Ao considerar o modelo logístico

$$\eta(x) = \theta_a \frac{1}{1 + \exp\{\theta_0 + \theta_1 x\}}, \quad (2.15)$$

verifica-se que  $\theta_0$  é adimensional ( $\emptyset$ ) ao passo que  $\theta_a$  e  $\theta_1$  têm  $Y$  e  $X^{-1}$  como u.m.. Pelo estudo da função,  $\theta_a$  é a ASS. Para os demais parâmetros não se tem uma interpretação de ponto característico, no entanto, verifica-se que  $\theta_1$  é proporcional à taxa de variação e  $\theta_0$  é proporcional ao valor do intercepto. Não é raro porém encontrar em artigos que  $\theta_1$  é interpretado como taxa. Para ser uma taxa, deveria apresentar u.m. apropriada





$\theta_1/\theta_2$ . Em termos de espaço paramétrico, como  $\eta(x)$  e  $x$  são não negativos, todos os parâmetros devem ser não negativos para descrever curvas de lactação.

## 2.4 Reparametrização

Foi visto que o modelo logístico tem mais de uma expressão. Embora sejam duas expressões, trata-se do mesmo modelo, que descreve a mesma curva mas com outros parâmetros e por isso se diz que são parametrizações de um mesmo modelo. Veja que os parâmetros em uma parametrização fornecem valores daqueles da outra parametrização,

$$\begin{aligned}\eta(x) &= \theta_a \frac{1}{1 + \exp\{-(x - \theta_i)/\theta_s\}} \\ &= \theta_a \frac{1}{1 + \exp\{\theta_i/\theta_s - x/\theta_s\}}, \\ \theta_0 = \theta_i/\theta_s &\quad \swarrow \quad \searrow \quad \theta_1 = -1/\theta_s\end{aligned}$$

conforme as operações algébricas acima. Então as parametrizações geram modelos iguais. No entanto, os parâmetros ocorrem de forma diferente e portanto têm u.m. e interpretações distintas. Acompanhe a conversão parametrização canônica do modelo quadrático para a polinomial

$$\begin{aligned}\eta(x) &= \theta_y + \theta_c(x - \theta_x)^2 \\ &= \theta_y + \theta_c(x^2 - 2x\theta_x + \theta_x^2) \\ &= (\theta_y + \theta_c\theta_x^2) + (-2\theta_c\theta_x)x + \theta_cx^2, \\ \theta_0 &\quad \swarrow \quad \theta_1 \quad \swarrow \quad \searrow \quad \theta_2\end{aligned}$$

que não tem boa interpretabilidade. Na parametrização canônica,  $\theta_x$  é o PC e PS da função,  $\theta_y$  é o valor correspondente (de mínimo ou máximo) e  $\theta_c$  é proporcional à curvatura ao redor do PC. Essa parametrização é mais interessante em otimização de processos porque os parâmetros de interesse estão no modelo. Considerando a parametrização polinomial, tais pontos característicos teriam que ser calculados após determinação dos parâmetros do modelo, o que envolve mais operações. Vale ressaltar o modelo na

parametrização polinomial é linear e na canônica é parcialmente linear em  $\theta_y$ .

Visando a interpretação do processo pelo uso de um MNL, é mais vantajoso considerar a parametrização mais interpretável dentre as disponíveis. No entanto, nem sempre dispõe-se de mais de uma parametrização. Então, como desenvolver uma parametrização interpretável a partir de uma qualquer?

De acordo com Zeviani (2013), é possível reparametrizar um modelo  $\eta(x, \theta)$  para um parâmetro de interesse  $\vartheta$  caso ele possa ser escrito como função dos parâmetros do modelo,  $\vartheta = g(\theta)$ . O procedimento é simples e envolve três etapas:

1. Escrever o parâmetro de interesse como função dos elementos de  $\theta$ , ou seja,  $\vartheta = g(\theta)$ ;
2. Escolher um dos elementos de  $\theta = (\theta_i, \theta_{-i})$  para ser colocado em função de  $\vartheta$  de tal forma a obter  $\theta_i = h(\theta_{-i}, \vartheta)$ ;
3. Substituir toda ocorrência de  $\theta_i$  em  $\eta(x, \theta)$  pela expressão obtida no passo anterior,  $h(\theta_{-i}, \vartheta)$ , fazendo as simplificações convenientes. Assim o modelo de regressão não linear

$$\eta(x, \theta_{-i}, \vartheta)$$

terá  $\vartheta$  como elemento do vetor de parâmetros  $\theta_\vartheta = (\theta_{-i}, \vartheta)$ .

É necessário que a função  $g$  seja estritamente monótona e diferenciável sobre o domínio relacionado à  $\theta_i$  e dessa forma a função  $h$  é a inversa de  $g$  em  $\theta_i$ .

Recomenda-se no passo 2 priorizar àquele elemento de  $\theta$  com menor significado ou para o qual a obtenção de  $h$  seja menos complicada. Em algumas situações pode ser que  $h$  não tenha solução analítica, por exemplo, quando é impossível isolar  $\theta_i$ , mas quando isso acontecer, uma solução numérica pode ser aplicada. O procedimento também pode ser usado para mais de um parâmetro de interesse, basta repeti-lo até que todos  $(\vartheta_1, \dots, \vartheta_k, k \leq p)$  sejam incorporados.

Uma etapa importante na reparametrização é fazer o estudo de dimensionalidade para determinar as unidades de medida dos parâmetros bem como seus espaços paramétricos. Tais informações são relevantes para descrição do modelo, interpretação dos parâmetros e conversão de valores entre escalas de medida.

Nas subseções a seguir será exemplificado a aplicação do procedimento para reparametrização de modelos de regressão não linear frequentemente adotados em aplicações nas Ciências Agrárias. Casos envolvendo um, dois e mais que  $p$  parâmetros serão considerados, bem como uma particular técnica de reparametrização considerada para polinômios de segundo e terceiro grau.

### 2.4.1 Reparametrização 1:1 - funções MC com ênfase no $PF_q$

O modelo Michaelis-Menten

$$\eta(x, \theta_a, \theta_v) = \frac{\theta_a x}{\theta_v + x}, \quad \begin{cases} \eta : \mathbb{R}^+ \mapsto \mathbb{R}^+ \\ \theta_a \geq 0 (Y), & \text{é a ASS} \\ \theta_v > 0 (X), & \text{é o } PF_{0.5} \end{cases} \quad (2.20)$$

foi inicialmente proposto para descrever a cinética de reações químicas (MICHAELIS; MENTEN, 1913). O estudo da função revela que é MC, tem PON, têm ASS dado por  $\theta_a$  e  $PF_{0.5}$  dado por  $\theta_v$  (meia vida).

Atualmente têm se observado aplicação desse modelo em diversos contextos em que um deles é a descrição do acúmulo de nutrientes liberado do solo (ZEVIANI et al., 2012). Nesse caso,  $\theta_v$  representa o tempo necessário para liberação de metade do conteúdo total de nutrientes do solo  $\theta_a$ , no entanto, os pesquisadores podem estar interessados em outras frações, como a de 95%. Então como obtê-la?

Considere de  $0 < q < 1$  representa a fração para a qual precisa-se determinar o tempo  $\vartheta_q$ . Então,  $\vartheta_q$  é o parâmetro de interesse. Quando  $x = \vartheta_q$  tem-se que  $\eta(\vartheta_q) = q\theta_a$ , então

$$\begin{aligned} q\theta_a &= \frac{\theta_a \vartheta_q}{\theta_v + \vartheta_q} \\ q &= \frac{\vartheta_q}{\theta_v + \vartheta_q} \\ \vartheta_q &= q(\theta_v + \vartheta_q) \\ \vartheta_q - q\vartheta_q &= q\theta_v \\ (1 - q)\vartheta_q &= q\theta_v \\ \vartheta_q &= \left( \frac{q}{1 - q} \right) \theta_v, \end{aligned}$$

verifica-se que o valor de  $\vartheta_q$  é função de  $\theta_v$ . Uma vez definido  $q$  e conhecido  $\theta_v$ , tem-se um valor para  $\vartheta_q$ . Tal desenvolvimento nos levou à conclu-

são da primeira etapa do processo de reparametrização. Agora escreve-se a relação em função de  $\theta_v$ , o único parâmetro do lado direito da equação,

$$\theta_v = \left( \frac{1-q}{q} \right) \vartheta_q.$$

Por fim, toda ocorrência de  $\theta_v$  deve ser substituída pela expressão obtida, fazendo as simplificações convenientes. Por meio do procedimento de reparametrização, chega-se à parametrização para o ponto de fração- $q$  ( $PF_q$ )

$$\eta(x, \theta_a, \vartheta_q) = \frac{\theta_a x}{\left( \frac{1-q}{q} \right) \vartheta_q + x}, \quad \begin{cases} \eta : \mathbb{R}^+ \mapsto \mathbb{R}^+, & 0 < q < 1 \\ \theta_a \geq 0 \text{ (Y)}, & \text{é a ASS} \\ \vartheta_q > 0 \text{ (X)}, & \text{é o } PF_q \end{cases} \quad (2.21)$$

Assim como o Michaelis-Menten, o modelo exponencial assintótico também é uma função MC-CON-PON,

$$\eta(x, \theta_a, \theta_s) = \theta_a (1 - \exp\{-\theta_s x\}), \quad \begin{cases} \eta : \mathbb{R}^+ \mapsto \mathbb{R}^+ \\ \theta_a \geq 0 \text{ (Y)}, & \text{é a ASS} \\ \theta_s > 0 \text{ (X}^{-1}\text{)}, & \propto \eta'(0). \end{cases} \quad (2.22)$$

Por ter as mesmas características, este modelo também tem aplicações semelhantes ao Michaelis-Menten, como liberação de nutrientes (ZEVIANI et al., 2012) além de progresso de doenças (KRANZ, 1990; SEGARRA; JEGGER; BOSCH, 2001). Embora muitos trabalhos atribuam significado equívoco de taxa ao parâmetro  $\theta_s$ , sua dimensionalidade não é de taxa. Pelo estudo de  $\eta'(x)$  verifica-se que  $\theta_s = \eta'(0)/\theta_a$ , ou seja, é a taxa instantânea (slope) na origem dividida pelo conteúdo total  $\theta_a$ .

Então o parâmetro  $\theta_a$  tem interpretação prática mas  $\theta_c$  não possui uma interpretação tão direta ou interessante assim. Da mesma forma como ocorre com o Michaelis-Menten, o tempo  $\vartheta_q$  para a liberação de uma fração  $q$  do conteúdo total  $\theta_a$  têm mais significado. Aplicando o procedimento de reparametrização tem-se que a primeira etapa resulta em

$$\begin{aligned} q\theta_a &= \eta(\vartheta_q) \\ &= \theta_a(1 - \exp\{-\theta_s \vartheta_q\}) \\ q &= 1 - \exp\{-\theta_s \vartheta_q\} \\ 1 - q &= \exp\{-\theta_s \vartheta_q\} \\ \log(1 - q) &= -\theta_s \vartheta_q \\ \vartheta_q &= -\frac{\log(1 - q)}{\theta_s}. \end{aligned}$$

No passo 2 escreve-se  $\theta_s$  como função de  $\vartheta_q$

$$\theta_s = -\frac{\log(1-q)}{\vartheta_q}.$$

Por fim, o passo 3 que é substituir  $\theta_s$  pela expressão obtida chega-se à

$$\eta(x, \theta_a, \vartheta_q) = \theta_a(1 - \exp\{x \log(1-q)/\vartheta_q\}), \quad \begin{cases} \eta : \mathbb{R}^+ \mapsto \mathbb{R}^+, & 0 < q < 1 \\ \theta_a \geq 0 \text{ (Y)}, & \text{é a ASS} \\ \vartheta_q > 0 \text{ (X)}, & \text{é a PF}_q. \end{cases} \quad (2.23)$$

É oportuno enfatizar que os dois modelos até então apresentados possuem formas funcionais diferentes mas que após reparametrização apresentam a mesma interpretação para os parâmetros. Isso torna os modelos diretamente comparáveis em termos de interpretação do fenômeno. A escolha do modelo passa ser mediante qualidade de ajuste aos dados e propriedades inferenciais. Tais aspectos serão esclarecidos no próximo capítulo.

## 2.4.2 Reparametrização 2:2 - Parábolas com ênfase no PC

Já foi vista a correspondência que existe entre parametrizações do modelo quadrático ao obter a polinomial a partir da canônica. Agora será feito o contrário, a partir da polinomial será obtida a canônica usando o processo de reparametrização.

O modelo polinomial de segundo grau ou modelo quadrático (2.4) é um modelo linear utilizado com frequência como uma aproximação local de uma função desconhecida, principalmente em experimentos industriais dedicados à otimização de processos (BAS; BOYACI, 2007), em que existe interesse no PC,  $\vartheta_x$  e no valor correspondente da resposta  $\vartheta_y$ . Tais quantidades são determinadas pelos parâmetros do modelo 2.4 uma vez que

$$\begin{aligned} \left. \frac{\partial \eta(x)}{\partial x} \right|_{x=\vartheta_x} &= \theta_1 + 2\theta_2 \vartheta_x = 0 \\ \vartheta_x &= \frac{-\theta_1}{2\theta_2}, \quad -\infty < \vartheta_x < \infty, \end{aligned}$$

e que

$$\vartheta_y = \eta(\vartheta_x) = \theta_0 + \theta_1 \vartheta_x + \theta_2 \vartheta_x^2, \quad -\infty < \vartheta_y < \infty.$$

Essas definições correspondem à primeira etapa da reparametrização. Considerando que  $\theta_0$  e  $\theta_1$  descrevem propriedades da função na origem ( $x = 0$ ) e que muitas vezes o domínio de interesse em  $x$  está distante da origem, será usado  $\theta_0$  e  $\theta_1$  porque apresentam cada vez menos utilidade longe da origem. No passo 2 substitui-se  $\theta_1$  por  $\vartheta_x$  e  $\theta_0$  por  $\vartheta_y$ . Dessa maneira tem-se

$$\begin{aligned}\theta_1 &= -2\theta_2\vartheta_x \\ \theta_0 &= \vartheta_y - \theta_1\vartheta_x - \theta_2\vartheta_x^2.\end{aligned}$$

Aplicando o passo 3 para substituir as expressões obtidas, após simplificações, chega-se ao modelo na parametrização canônica.

$$\eta(x, \vartheta_x, \vartheta_y, \theta_2) = \vartheta_y + \theta_2(x - \vartheta_x)^2, \quad \begin{cases} \eta : \mathbb{R} \mapsto \mathbb{R} \\ \vartheta_x \in \mathbb{R} (X), & \text{é o PC} \\ \vartheta_y \in \mathbb{R} (Y), & \text{é } \eta(\vartheta_x) \\ \theta_2 \in \mathbb{R} (YX^{-2}), & \text{é a curvatura} \end{cases} \quad (2.24)$$

Essa é portanto a parametrização canônica do modelo quadrático. A ênfase é o PC uma vez que todos os parâmetro são quantidades que descrevem a função no PC. O parâmetro  $\vartheta_x$  é o valor do PC, que pode ser de máximo ou mínimo (depende de  $\theta_2$ ) e  $\vartheta_y$  o valor da função nesse ponto. O parâmetro  $\theta_2$  representa a diferença no valor da função com relação ao  $\vartheta_y$  uma unidade acima ou abaixo do ponto  $\vartheta_x$ . Então que ele pode ser entendido como grau de especificidade à condição ótima uma vez que quanto maior  $\theta_2$  maiores serão as diferenças da função ao se afastar do PC.

Outro modelo cujo interesse está no ponto crítico é o de Bleasdale e Nelder (1960),

$$\eta(x, \theta_0, \theta_1, \theta_2) = x(\theta_0 + \theta_1 x)^{-1/\theta_2}, \quad \begin{cases} \eta : \mathbb{R}^+ \mapsto \mathbb{R}^+ \\ \theta_0 \text{ (n.u.)}, & \text{vago} \\ \theta_1 \text{ (n.u.)}, & \text{vago} \\ \theta_2 \in \mathbb{R}_*^+ (\emptyset), & \text{forma,} \end{cases} \quad (2.25)$$

aplicado para representar a relação entre produção em função da densidade de plantas por unidade de área. Os parâmetros não têm significado prático, com exceção de  $\theta_2$  que tem relação com a forma, sendo a função assintótica e parabólica para  $\theta_2 > 1$  e  $0 < \theta_2 < 1$ , portanto, só têm ponto de máximo para  $0 < \theta_2 < 1$  que é o espaço paramétrico para estudos de produção-densidade. Para  $\theta_0$  e  $\theta_1$  as u.m. são complicadas ao ponto de serem não úteis (n.u.) e os espaços paramétricos são indeterminados.

Segundo Schabenberger e Pierce (2001), a densidade de plantas  $\vartheta_x$  correspondente à máxima produção por unidade de área  $\vartheta_y$  são dados por

$$\begin{aligned}\vartheta_x &= \frac{\theta_0}{\theta_1} \left( \frac{\theta_2}{1 - \theta_2} \right), & \vartheta_x > 0 \text{ (X)} \\ \vartheta_y &= \frac{\theta_2}{\theta_1} \left( \frac{1 - \theta_2}{\theta_0} \right)^{(1 - \theta_2)/\theta_2} = \vartheta_x \left( \frac{1 - \theta_2}{\theta_0} \right)^{1/\theta_2}, & \vartheta_y \geq 0 \text{ (Y)},\end{aligned}$$

são parâmetros de fundamental interesse em estudos dessa natureza.

As duas expressões acima definem os parâmetros de interesse como função dos parâmetros do modelo e portanto representam o passo 1 do procedimento de reparametrização.

Pela análise dessas expressões conclui-se que é mais simples escrever  $\theta_0$  como função de  $\vartheta_y$  e  $\theta_1$  como função de  $\vartheta_x$ . Se considerar o parâmetro  $\theta_2$  em alguma substituição o resultado seria uma expressão sem solução analítica. Dessa forma, para o passo 2 tem-se

$$\begin{aligned}\theta_0 &= (1 - \theta_2) \left( \frac{\vartheta_y}{\vartheta_x} \right)^{-\theta_2} \\ \theta_1 &= \frac{\theta_0}{\vartheta_x} \left( \frac{\theta_2}{1 - \theta_2} \right).\end{aligned}$$

Finalmente, no passo 3 substituí-se tais expressões dentro do modelo e simplifica-se convenientemente de forma que a versão reparametrizada do modelo é

$$\begin{aligned}\eta(x, \vartheta_x, \vartheta_y, \theta_2) &= \vartheta_y \frac{x}{\vartheta_x} \left( 1 - \theta_2 \left( 1 - \frac{x}{\vartheta_x} \right) \right)^{-1/\theta_2}, & (2.26) \\ \left\{ \begin{array}{ll} \eta : \mathbb{R}^+ \mapsto \mathbb{R}^+ & \\ \vartheta_x \in \mathbb{R}_*^+ \text{ (X)}, & \text{PC} \\ \vartheta_y \in \mathbb{R}^+ \text{ (Y)}, & \eta(\vartheta_x) \\ \theta_2 \in \mathbb{R}_*^+ (\emptyset), & \text{curvatura}, \end{array} \right.\end{aligned}$$

Na versão reparametrizada do modelo Bleasdale-Nelder, verifica-se que o parâmetro  $\theta_2$  está relacionado a curvatura da função ao redor do ponto crítico (Figura 2.2) para valores fixos dos demais parâmetros. Em oposição ao modelo original, o modelo reparametrizado possui interpretação e dimensionalidade em  $\vartheta_x$  e  $\vartheta_y$  além de esclarecer o significado sobre o parâmetro  $\theta_2$  para o qual  $1/\theta_2$  pode ser interpretado como coeficiente

de especificidade à condição de ótimo. Tal conhecimento com relação ao modelo e seus parâmetros facilita a obtenção de valores iniciais em procedimentos de estimação, a definição de hipóteses e atribuição de prioris. Além do mais, permite que se estude alterações fisiológicas e produtivas por meio de tais parâmetros como função, por exemplo, dos materiais genéticos, dos métodos de manejo, dos níveis de adubação e irrigação pelo emprego de modelos não lineares de efeitos mistos (PEEK et al., 2002).

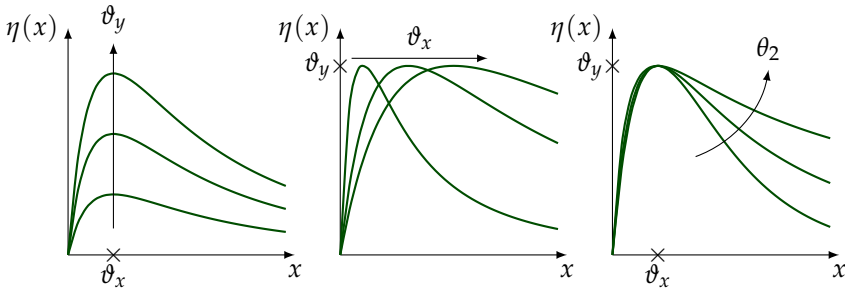


Figura 2.2: Modelo de Bleasdale-Nelder reparametrizado para o ponto crítico com destaque para o significado e comportamento com relação aos parâmetros.

Mais sobre reparametrização visando interpretação em modelos não lineares estão disponíveis em Zeviani (2013). Essa obra aborda casos com excesso de parâmetros de interesse, emprego de soluções numéricas e casos envolvendo distribuições de probabilidade além de fornecer um catálogo com 17 modelos não lineares reparametrizados em termos de parâmetros de significado para aplicações em Ciências Agrárias. As vantagens da reparametrização, do ponto de vista estatístico, são comentadas e é feito estudo de caso de modelos para produção do algodão em função da desfolha e de curvas de retenção de água do solo.



# Capítulo 3

## Verossimilhança

Os modelos de regressão não linear normalmente são utilizados para respostas contínuas como o tamanho de um indivíduo em certa idade, o conteúdo liberado de uma substância no tempo e a produção de uma planta para uma dose de fertilizante. Dessa forma, a distribuição de probabilidades considerada na grande maioria das vezes é a distribuição Gaussiana. No entanto, como foi visto no capítulo 1, outras distribuições podem ser consideradas.

Se for considerada a distribuição Gaussiana para a resposta da variável aleatória  $Y$ , supondo que a variância é constante, o modelo pode ser representado por

$$Y|x \stackrel{iid}{\sim} \text{Gaussiana}(\mu = \eta(x, \theta), \sigma^2). \quad (3.1)$$

Ou seja, a notação representa que a distribuição de probabilidades da variável aleatória  $Y$  condicional a um valor fixo da variável explicativa  $x$  é Gaussiana, sua média depende de  $x$  por meio de uma função  $\eta$  e sua variância é constante. Uma vez observada uma amostra aleatória de  $Y$  em correspondência com valores de  $x$ , deseja-se determinar os valores das quantidades desconhecidas  $\theta$  e  $\sigma^2$ , que são os parâmetros do modelo.

Em modelos não lineares como (3.1), que tem apenas termos de efeito fixo e distribuição Gaussiana de variância constante, sob a suposição de independência, a estimação dos parâmetros geralmente é feita pelo método dos mínimos quadrados. O objetivo é minimizar a soma do quadrado dos desvios entre os valores observados ( $y$ ) e os valores preditos pelo modelo,

$$(\hat{y} = \eta(x, \hat{\theta})),$$

$$\text{SQD}(\theta) = \sum_i (y_i - \eta(x_i, \theta))^2. \quad (3.2)$$

Quando o modelo é linear, os estimadores dos parâmetros são

$$\hat{\theta} = (X^\top X)^{-1} X^\top y \quad (3.3)$$

$$\hat{\sigma}^2 = \frac{\text{SQD}(\hat{\theta})}{n - p}, \quad (3.4)$$

em que  $X$  é a matriz com a derivada primeira do modelo em relação a cada parâmetro,  $n$  é o número de observações e  $p$  é o número de parâmetros em  $\theta$ .

Para os modelos não lineares, a matrix  $X$  não é completamente conhecida pois a primeira derivada de  $\eta(x, \theta)$  em relação a  $\theta$  envolve parâmetros que são desconhecidos. Então tem-se quantidades desconhecidas em ambos os lados da igualdade da expressão (3.3). Nesse caso, considera-se um procedimento numérico para estimação de  $\theta$ . Este procedimento nada mais é que a aplicação recursiva da expressão (3.3) partindo de um valor inicial para  $\theta$ ,  $\theta^{(0)}$ . Esse valor inicial é o usuário que fornece e serve para tornar conhecida a matrix  $X^{(0)}$  e com isso calcular  $\theta^{(1)}$  por meio de (3.3). Então  $\theta^{(1)}$  agora é usado para obter  $X^{(1)}$ . Esse procedimento é repetido até que se alcance uma convergência de acordo com algum critério. Esse procedimento é o chamado método de Gauss-Newton,

$$\theta^{(u+1)} = \theta^{(u)} + (X^{(u)\top} X^{(u)})^{-1} X^{(u)\top} (y - f(x, \theta^{(u)})), \quad (3.5)$$

em que  $u$  é o índice de iterações. Um critério de convergência adotado é baseado na diferença entre valores consecutivos de  $\theta$ ,  $|\theta^{(u+1)} - \theta^{(u)}|/\theta^{(u)}$ , ou de  $\text{SQD}(\theta)$ ,  $\text{SQD}(\theta^{(u+1)}) - \text{SQD}(\theta^{(u)})/\text{SQD}(\theta^{(u)})$ , que quando for inferior a um valor  $\delta > 0$  considera-se que houve convergência.

O código abaixo mostra uma implementação do método Gauss-Newton. Serão usados dados fictícios como exemplo e o modelo Michaelis-Menten com o valor do parâmetro assíntota superior fixo em 1,8. Assim, está sendo estimado apenas o parâmetro tempo de meia-vida.

```
> #-----
> y <- matrix(c(0.1, 0.4, 0.6, 0.9))
> x <- matrix(c(0.2, 0.5, 0.7, 1.8))
> plot(y~x) # dispersão
> curve(1.8*x/(1.6+x), add=TRUE) # curva inicial
> #-----
> # deriv3: derivadas analíticas,
> # 1) f em theta, 2) f' em theta "gradient", 3) f'' "hessian" em theta
> d3 <- deriv3(~1.8*x/(V+x), c("V"), function(x, V){ NULL })
> d3(x, V=1.6)
```

```

      [,1]
[1,] 0.2000000
[2,] 0.4285714
[3,] 0.5478261
[4,] 0.9529412
attr(,"gradient")
      V
[1,] -0.1111111
[2,] -0.2040816
[3,] -0.2381853
[4,] -0.2802768
attr(,"hessian")
, , V

```

```

      V
[1,] 0.1234568
[2,] 0.1943635
[3,] 0.2071176
[4,] 0.1648687

```

```

> #-----
> theta0 <- c(V=1.6)           # valor inicial
> sqr0 <- crossprod(y)         # soma de quadrados total
> maxiter <- 50                 # número máximo de iterações
> u <- 1                        # inicia contador de iterações
> delta <- 1e-8; dif <- 2*delta # delta
> #-----
> while(u<=50 & dif>delta){
+   M <- d3(x, V=theta0)
+   f <- c(M)                   # preditos, f(x, theta0)
+   X <- attr(M, "gradient")    # gradiente, matrix X
+   r <- (y-f)                  # resíduos
+   sqr1 <- crossprod(r)        # SQD
+   dif <- abs(sqr0-sqr1)/sqr0  # diferença relativa em SQD
+   theta1 <- theta0+           # "ajusta" theta
+     crossprod(solve(crossprod(X)), crossprod(X, r))
+   #solve(t(X)%*%X)%*%t(X)%*%r
+   cat(paste(formatC(c(SQD=sqr1, theta=theta0),
+     digits=6, format="f"),
+     collapse="\t", "\n")    # histórico
+   theta0 <- c(theta1) # atualiza theta
+   sqr0 <- sqr1        # atualiza SQD
+   u <- u+1            # incrementa contador
+ }

```

```

0.016341      1.600000
0.014371      1.702245
0.014368      1.706213
0.014368      1.706209

```

```

> #-----
> theta0

```

```

[1] 1.706209

```

```

> s2 <- c(sqr0)/(length(y)-1); s2

```

```
[1] 0.004789388
```

```
> #-----
```

Esse código é válido para estimar modelos com mais de um parâmetro. Para exemplificar será considerada a estimação do parâmetro assíntota superior e tempo de meia-vida. Para isso precisa-se trocar a expressão do modelo, adequar o vetor de valores iniciais que o método pode ser aplicado.

```
> #-----
> d3 <- deriv3(~A*x/(V+x), c("A","V"), function(x, A, V){ NULL })
> theta0 <- c(A=1.8, V=1.6) # valor inicial
> #-----
> sqr0 <- crossprod(y) # soma de quadrados total
> maxiter <- 50 # número máximo de iterações
> u <- 1 # inicia contador de iterações
> delta <- 1e-8; dif <- 2*delta # delta
> #-----
> while(u<=50 & dif>delta){
+ M <- d3(x, A=theta0[1], V=theta0[2])
+ f <- c(M) # preditos, f(x, theta0)
+ X <- attr(M, "gradient") # gradiente, matrix X
+ r <- (y-f) # resíduos
+ sqr1 <- crossprod(r) # SQD
+ dif <- abs(sqr0-sqr1)/sqr0 # diferença relativa em SQD
+ thetal <- theta0+ # "ajusta" theta
+ crossprod(solve(crossprod(X)), crossprod(X, r))
+ #solve(t(X)%*%X)%*%t(X)%*%r
+ cat(paste(formatC(c(SQD=sqr1, theta=theta0),
+ digits=6, format="f"),
+ collapse="\t"), "\n") # histórico
+ theta0 <- c(thetal) # atualiza theta
+ sqr0 <- sqr1 # atualiza SQD
+ u <- u+1 # incrementa contador
+ }
```

0.016341	1.800000	1.600000
0.014293	1.744176	1.619132
0.014293	1.742698	1.617481
0.014293	1.742834	1.617701
0.014293	1.742816	1.617672

```
> #-----
```

```
> theta0
```

```
[1] 1.742818 1.617676
```

```
> s2 <- c(sqr0)/(length(y)-1); s2
```

```
[1] 0.004764382
```

```
> #-----
```

Apesar de útil para todos os modelos não lineares como o (3.1), será considerada a função de verossimilhança para estimação dos parâmetros. Nesse caso o objetivo é maximizar a função de verossimilhança,  $L(\theta, \sigma^2)$ . Considerando a distribuição Gaussiana para  $Y|x$ , a função objetivo é

$$\begin{aligned} L(\theta, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_i - \eta(x_i, \theta))^2}{2\sigma^2} \right\} \\ &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \eta(x_i, \theta))^2 \right\}. \end{aligned} \quad (3.6)$$

Para o propósito de estimação, maximizar a função de log-verossimilhança,  $\ell(\theta, \sigma^2) = \log(L(\theta, \sigma^2))$ , irá retornar as mesmas estimativas para os parâmetros. Então, a função objetivo passa a ser

$$\ell(\theta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_i^n (y_i - \eta(x_i, \theta))^2. \quad (3.7)$$

Pela expressão (3.7) percebe-se a equivalência entre minimizar a SQD e maximizar  $\ell$  para estimar  $\theta$  pois ambas envolvem a quantidade  $\sum_i^n (y_i - \eta(x_i, \theta))^2$ . Essa é a função de log-verossimilhança correspondente o modelo representado na figura 1.3 na página 6.

Em modelos não lineares, maximizar (3.7) envolve métodos numéricos. Uma das possibilidades é usar o método Newton-Raphson.

Em algumas situações a suposição de variância constante não é atendida. Isso é uma característica típica de curvas de crescimento na quais a variabilidade é maior a medida que os indivíduos são maiores. Não se deve ajustar o modelo (3.1) nessa situação pois isso iria retornar resultados viesados. Deve-se procurar acomodar no modelo o fato da variância ser não constante. Pode-se então atribuir um modelo para a variância assim como se faz para a média. Considere que  $\mu = \eta_\mu(x, \theta)$  representa um modelo para a média de  $Y|x$  e  $\sigma^2 = \eta_{\sigma^2}(z, \varphi)$  um modelo para a variância de  $Y|x$ . A variável  $z$  pode ser outra variável explicativa mas normalmente considera-se  $z = x$  ou  $z = \mu$ . Sendo assim, a função de log-verossimilhança é

$$\ell(\theta, \varphi) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \left\{ \log(\eta_{\sigma^2}(z, \varphi)) + \frac{(y_i - \eta_\mu(x_i, \theta))^2}{\eta_{\sigma^2}(z, \varphi)} \right\}. \quad (3.8)$$

A função  $\sigma^2 = \eta_{\sigma^2}(z, \varphi)$  deve ter contra-domínio positivo porque a variância é maior que zero. Essa é a função de log-verossimilhança correspondente ao modelo representado na Figura 1.4 na página 7.

Verossimilhança de um modelo não linear com um termo de efeito aleatório.

$$\begin{aligned}
 L(\theta, \sigma_a^2, \sigma^2) &= \prod_{i=1}^I \int \prod_{j=1}^{n_i} \phi(y_{ij}, \eta(x_{ij}, \theta, a_i), \sigma^2) \phi(a_i, 0, \sigma_a^2) da_i \\
 &= \prod_{i=1}^I \int \prod_{j=1}^{n_i} \phi\left(y_{ij}, \frac{(\theta_a + a_i) x_{ij}}{\theta_v + x_{ij}}, \sigma^2\right) \phi(a_i, 0, \sigma_a^2) da_i
 \end{aligned} \tag{3.9}$$

em que  $\phi$  é função densidade da distribuição Gaussiana. Essa é a função de verossimilhança correspondente o modelo representado na figura 1.8 na página 12.

# Capítulo 4

## Uma única curva

### 4.1 Ajuste e diagnóstico

O objetivo com essa seção é apresentar as funções para trabalhar com inferência em modelos de regressão não linear. Vamos considerar parte dos dados estudados por Zeviani et al. (2012). Os dados se referem ao conteúdo liberado de potássio em função do tempo de incubação de Esterco ... em latossolo ...

```
> da <- data.frame(y=c(51.03, 57.76, 26.60, 60.65, 87.07, 64.67,  
+ 91.28, 105.22, 72.74, 81.88, 97.62, 90.14,  
+ 89.88, 113.22, 90.91, 115.39, 112.63, 87.51,  
+ 104.69, 120.58, 114.32, 130.07, 117.65, 111.69,  
+ 128.54, 126.88, 127.00, 134.17, 149.66, 118.25,  
+ 132.67, 154.48, 129.11, 151.83, 147.66, 127.30),  
+ x=rep(c(15, 30, 45, 60, 75, 90,  
+ 120, 150, 180, 210, 240, 270), each=3),  
+ r=rep(1:3, 12))
```

A função básica para se trabalhar com o modelo de regressão não linear é a `nls()` ((nonlinear least squares)). Os argumentos básicos são:

**formula:** expressão do modelo não linear;

**data:** data.frame contendo os valores observados das variáveis envolvidas;

**start:** lista nomeada com valores iniciais para os parâmetros do modelo.

```
> args(nls)
```

Tabela 4.1: Conteúdo acumulado de potássio liberado (mg) dias após a incubação de esterco de codorna.

	x	1	2	3
1	15.00	51.03	57.76	26.60
2	30.00	60.65	87.07	64.67
3	45.00	91.28	105.22	72.74
4	60.00	81.88	97.62	90.14
5	75.00	89.88	113.22	90.91
6	90.00	115.39	112.63	87.51
7	120.00	104.69	120.58	114.32
8	150.00	130.07	117.65	111.69
9	180.00	128.54	126.88	127.00
10	210.00	134.17	149.66	118.25
11	240.00	132.67	154.48	129.11
12	270.00	151.83	147.66	127.30

```
function (formula, data = parent.frame(), start, control = nls.control(),
  algorithm = c("default", "plinear", "port"), trace = FALSE,
  subset, weights, na.action, model = FALSE, lower = -Inf,
  upper = Inf, ...)
NULL
```

Uma boa prática antes de ir para o ajuste do modelo é explorar os dados e o modelo em busca de valores iniciais apropriados. Pode-se fazer isso por meio de gráficos de dispersão como na figura 4.1. Conhecimento preliminar sobre o fenômeno, uma vez que trata-se do conteúdo acumulado liberado, aponta que uma função monótona não decrescente a partir da origem deve ser considerada. Sendo assim, considerou-se o modelo Michaelis-Menten

$$\eta(x, \theta) = \frac{\theta_a x}{\theta_v + x}. \quad (4.1)$$

No gráfico a curva vermelha é o traçado para o modelo (4.1) com valores iniciais obtidos por inspeção visual. Como  $\theta_a$  é assintota considerou-se o valor 140 e como  $\theta_v$  é o tempo de meia-vida considerou-se 40.

Para ajustar o modelo aos dados usa-se a função `nls()` passando os argumentos `formula=`, `data=` e `start=`.

```
> n0 <- nls(formula=y~A*x/(V+x), data=da,
+           start=list(A=140, V=40), trace=TRUE)
```

```
9674.144 : 140 40
4547.413 : 157.59100 40.59483
4547.345 : 157.63475 40.57603
4547.345 : 157.63346 40.57469
```



```

> plot(y~x, data=da,
+       xlab="Dias após incubação",
+       ylab="Conteúdo acumulado liberado de potássio")
> A <- 140; V <- 40
> abline(v=V, h=A/c(1,2), lty=2)
> curve(A*x/(V+x), add=TRUE, col=2, lty=2)

```

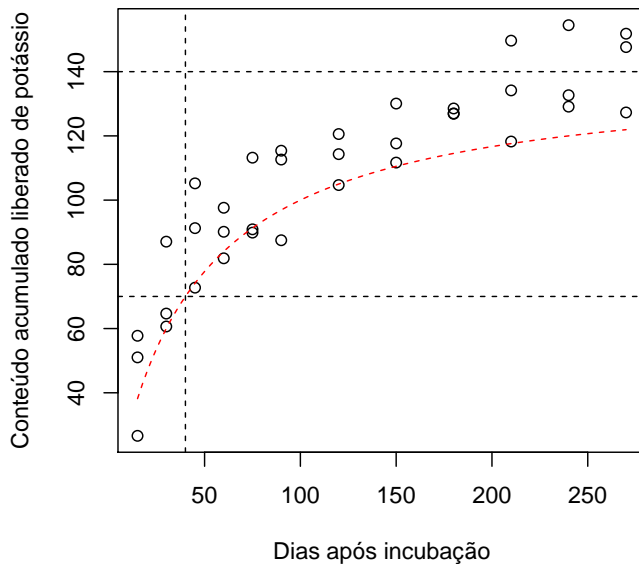


Figura 4.1: Conteúdo acumulado de potássio liberado no solo em função dos dias após a incubação de esteco de codorna. A linha pontilhada vermelha é uma curva obtida pelo gráfico da função para valores iniciais dos parâmetros.

```
> summary(n0)
```

```
Formula: y ~ A * x/(V + x)
```

```
Parameters:
```

	Estimate	Std. Error	t value	Pr(> t )
A	157.633	6.110	25.800	< 2e-16 ***
V	40.575	5.658	7.171	2.71e-08 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 11.56 on 34 degrees of freedom
```

```
Number of iterations to convergence: 3
```

```
Achieved convergence tolerance: 2.906e-06
```

O argumento `trace=TRUE` é para patrulhar a convergência mostrando os valores de cada iteração até chegar às estimativas. As funções-método como a `print()` e `summary()` retornam números relacionados ao ajuste, como as estimativas dos parâmetros, erro padrão residual e número de passos até a convergência. Após o ajuste é boa prática sobrepor o modelo ajustado aos dados para certificar-se de que o modelo está representando os dados e que a convergência não foi para um mínimo local (figura ??).

Além de checar a adequação do ajuste pela figura 4.2, é importante fazer uma análise dos resíduos para verificar o atendimento dos pressupostos do modelo. No caso, o modelo estatístico considerado supõe que a função não linear representa a média de  $y$  como função de  $x$ , que existe independência entre observações e que os resíduos têm distribuição normal e variância constante. Em certos casos a independência é garantida pelo processo de casualização. Para os outros quesitos, três gráficos são úteis: diagrama de dispersão dos resíduos em função dos valores ajustados, do valor absoluto dos resíduos em função dos valores ajustados e o gráfico quantil-quantil normal dos resíduos (figura 4.3). Pelos gráficos da figura 4.3 não se têm razões para negar o atendimento dos pressupostos.

Muitos pesquisadores optam por quantificar a qualidade de um ajuste por meio do coeficiente de determinação,  $R^2$ . Em nossa opinião, o  $R^2$  isoladamente não quer dizer nada, muito embora seja uma medida entre 0 e 1 onde 1 indica um perfeito ajuste do modelo aos dados, ou seja, os valores preditos são iguais aos valores observados. O ponto é que um  $R^2$  baixo não indica afastamento dos pressupostos da mesma forma que um  $R^2$  não indica que os pressupostos foram verificados. No entanto, para comparação de dois modelos o  $R^2$  é interessante, bem como o valor da log-verossimilhança, AIC e BIC.

Existem muitas formas de calcular o  $R^2$ . Duas indicadas para modelos não lineares são: o quadrado da correlação entre valores ajustados e

```

> plot(y~x, data=da,
+       xlab="Dias após incubação",
+       ylab="Conteúdo acumulado liberado de potássio")
> with(as.list(coef(n0)), curve(A*x/(V+x), add=TRUE, col=2))

```

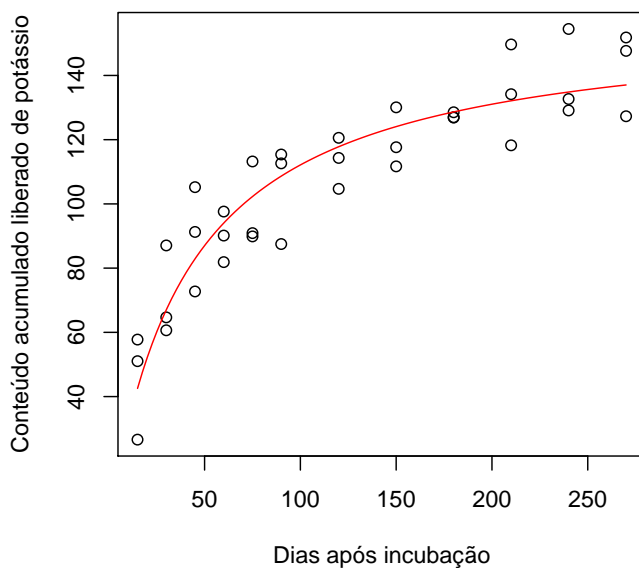


Figura 4.2: Conteúdo acumulado de potássio liberado no solo em função dos dias após a incubação de esteco de codorna. A linha contínua vermelha é o a função do modelo ajustado aos dados.

```

> r <- residuals(n0)
> f <- fitted(n0)
> par(mfrow=c(1,3), mar=c(5,4,4,2))
> plot(r~f, xlab="Valores ajustados", ylab="Resíduos",
+      main=expression("Adequação de "eta(x,theta)))
> lines(smooth.spline(r~f), col=2)
> plot(sqrt(abs(r))~f, xlab="Valores ajustados",
+      ylab="Raiz dos valores absolutos",
+      main=list("Relação média-variância", font=1))
> lines(smooth.spline(sqrt(abs(r))~f), col=2)
> qqnorm(r, xlab="Quantis teóricos", ylab="Quantis amostrais",
+      main=list("Normalidade", font=1)); qqline(r, col=2)
> layout(1)

```

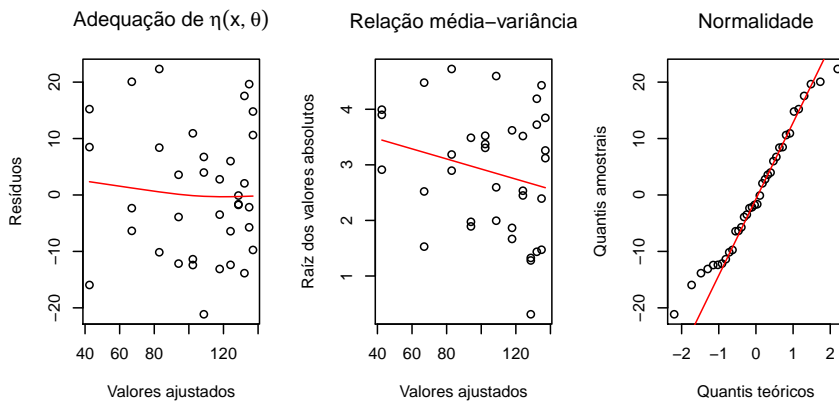


Figura 4.3: Gráficos de diagnóstico do modelo estatístico por meio dos resíduos.

observados e o  $R^2$  ajustado para a média geral

$$R^2 = 1 - \frac{SQR_{\hat{y}}}{SQR_{\bar{y}}}, \quad (4.2)$$

em que  $SQR_{\hat{y}}$  é a soma de quadrados dos resíduos do modelo não linear e  $SQR_{\bar{y}} = (n - 1)s^2$  é a soma de quadrados de um modelo com apenas o intercepto (média). Essas duas definições de  $R^2$  independem da forma do modelo não linear. Algumas definições supõem que o modelo não linear se reduza a um modelo com apenas intercepto por alguma restrição paramétrica.

```
> cor(da$y, f)^2
[1] 0.8592582

> 1-deviance(n0)/deviance(lm(y~1, da))
[1] 0.8586328
```

## 4.2 Extendendo o modelo não linear

Alguns nutrientes são liberados por um processo de duas fases: uma fase rápida e uma fase lenta. A porção rápida está relacionada ao nutriente livre na solução e a fase lenta ao nutriente preso a matriz do solo. A fase rápida libera com taxa de liberação que vai se reduzindo à medida que o nutriente vai sendo liberado. A fase lenta por sua vez libera com taxa constante. Esse processo hipotético pode ser representado pelo modelo de regressão não linear

$$\eta(x, \theta) = \frac{\theta_a x}{\theta_v + x} + \theta_d x \quad (4.3)$$

em que  $\theta_d x$  é o termo do modelo que representa a porção de lenta liberação do nutriente, cuja taxa é  $\theta_d$ . Não havendo porção de lenta liberação ( $\theta_d = 0$ ) o modelo se reduz ao (4.1). Não havendo porção de rápida liberação ( $\theta_v \rightarrow 0$ ), o modelo se reduz à equação da reta,  $\theta_a + \theta_d x$ .

Para ajustar o modelo (4.3) usamos

```
> n1 <- nls(formula=y~A*x/(V+x)+D*x, data=da,
+           start=list(A=140, V=40, D=0.05), trace=TRUE)

5395.787 : 140.00  40.00  0.05
4333.216 : 106.0235822  16.2393614  0.1544494
4131.034 : 116.8665492  22.7029728  0.1292651
```

```

4129.249 : 118.0527086 23.4083064 0.1268105
4129.247 : 117.9771585 23.3712846 0.1270626
4129.247 : 117.9820786 23.3737476 0.1270471

> summary(n1)

Formula: y ~ A * x/(V + x) + D * x

Parameters:
      Estimate Std. Error t value Pr(>|t|)
A 117.98208    17.64509    6.686 1.3e-07 ***
V 23.37375     8.35389    2.798 0.00852 **
D 0.12705      0.06085    2.088 0.04459 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.19 on 33 degrees of freedom

Number of iterations to convergence: 5
Achieved convergence tolerance: 3.405e-06

> anova(n0, n1)

Analysis of Variance Table

Model 1: y ~ A * x/(V + x)
Model 2: y ~ A * x/(V + x) + D * x
  Res.Df Res.Sum Sq Df Sum Sq F value
1      34    4547.3
2      33    4129.2  1  418.1  3.3413
  Pr(>F)
1
2 0.07661 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

A função-método `anova()` quando recebe mais de um modelo realiza o teste de razão de verossimilhanças. Os modelos fornecidos devem obrigatoriamente serem modelos aninhados para o teste ser interpretado. No caso, verifica-se que não existe evidência de uma porção de lenta liberação do nutriente no solo uma vez que  $\theta_d = 0$  à 5%. Por outro lado, o teste  $t$  para  $\theta_d$  deu  $p$ -valor inferior à 5%. Então, qual teste considerar? Considere o teste de razão de verossimilhanças pois o teste  $t$  é influenciado pelas medidas de não linearidade do modelo.

## 4.3 Comparando parametrizações

Outro modelo capaz de representar a liberação de nutrientes é

$$\eta(x, \theta) = \theta_a(1 - \exp\{-\theta_c x\}). \quad (4.4)$$

Esse modelo é conhecido como monomolecular e é uma simplificação do modelo Mitscherlich. A assíntota é representada por  $\theta_a$  (Y) e  $\theta_c$  ( $X^{-1}$ ) é o valor da taxa de liberação relativa ao conteúdo total na origem, ou seja

$$\left. \frac{\partial \eta(x, \theta)}{\partial x} \right|_{x=0} = \theta_a \theta_c.$$

### 4.3.1 Reparametrização

Apesar da mencionada interpretação de  $\theta_c$ , pesquisadores consideram mais útil ter um modelo com o tempo de meia-vida,  $\vartheta_v$  (X), como parâmetro. Dessa forma, uma reparametrização pode ser aplicada. Tem-se que o tempo de meia-vida é

$$\vartheta_v = -\frac{\log(0,5)}{\theta_c} \Rightarrow \theta_c = -\frac{\log(0,5)}{\vartheta_v}. \quad (4.5)$$

Então o modelo reparametrizado é

$$\eta(x, \theta) = \theta_a (1 - \exp\{-x \log(2)/\vartheta_v\}), \quad (4.6)$$

uma vez que  $-\log(0,5) = \log(2)$ .

### 4.3.2 Medidas de curvatura

Considerando a interpretação, existe preferência pelo modelo (4.6). No entanto, as parametrizações de um mesmo modelo têm propriedades estatísticas diferentes e estas devem ser avaliadas. Será feito o ajuste das duas parametrizações e obtidas as medidas de curvatura de Bates e Watts (1980). Para obtê-las é necessário representar os modelo de uma forma especial para que se tenha a matriz gradiente ( $\mathbf{F}_{(n \times p)}$ ) e o arranjo hessiano do ajuste ( $\mathbf{H}_{(n \times p \times p)}$ ). Para isso declaramos o modelo por meio da função `deriv3()` que retorna três quantidades:  $\eta(x, \theta)$ ,  $\mathbf{F}(x, \theta)$  e  $\mathbf{H}(x, \theta)$ .

```
> prmt0 <- deriv3(~A*(1-exp(-C*x)), c("A","C"),
+               function(x, A, C){NULL})
> prmt1 <- deriv3(~A*(1-exp(-x*log(2)/V)), c("A","V"),
+               function(x, A, V){NULL})
> prmt0(x=1:3, A=3, C=1)
```

```
[1] 1.896362 2.593994 2.850639
attr(,"gradient")
      A      C
[1,] 0.6321206 1.1036383
```

```
[2,] 0.8646647 0.8120117
[3,] 0.9502129 0.4480836
attr(,"hessian")
, , A
```

```
      A      C
[1,] 0 0.3678794
[2,] 0 0.2706706
[3,] 0 0.1493612
```

```
, , C
```

```
      A      C
[1,] 0.3678794 -1.103638
[2,] 0.2706706 -1.624023
[3,] 0.1493612 -1.344251
```

Depois disso basta usar-se a `nls()`.

```
> n2 <- nls(y-prmt0(x, A, C), data=da, start=list(A=140, C=log(2)/40))
> coef(n2)
```

```
      A      C
132.6973827 0.0208804
```

```
> deviance(n2)
```

```
[1] 5929.083
```

```
> n3 <- nls(y-prmt1(x, A, V), data=da, start=list(A=140, V=40))
> coef(n3)
```

```
      A      V
132.6973 33.1960
```

```
> deviance(n3)
```

```
[1] 5929.083
```

Verifica-se que os modelos apresentam a mesma deviance pois de fato são duas parametrizações do mesmo modelo. O mesmo valor de valores preditos pode ser obtido por eles. Portanto, as medidas de qualidade de ajuste (deviance,  $R^2$ , log-verossimilhança, AIC, BIC) serão iguais para ambas as parametrizações. Então, baseadas nessas medidas não tem-se como selecionar o modelo mais adequado. Para isso considera-se a interpretação dos parâmetros e as medidas de curvatura. As medidas de curvatura são calculadas com a função `rms.curv()` do pacote MASS.

```
> require(MASS)
> rms.curv(n2)
```

```
Parameter effects: c^theta x sqrt(F) = 0.4163
Intrinsic: c^iota x sqrt(F) = 0.1001
```



```
> rms.curv(n3)
```

```
Parameter effects: c^theta x sqrt(F) = 0.1546  
Intrinsic: c^iota x sqrt(F) = 0.1001
```

São duas medidas de curvatura. A curvatura intrínseca está relacionada à suposição de planicidade da superfície esperada gerada pelo modelo. A curvatura devido ao efeito de parâmetros está relacionada a suposição de coordenadas de espaçamento regular e reto sobre a superfície esperada. Quando um modelo é linear essas duas medidas são zero pois a superfície esperada é plana e as coordenadas são regulares sobre ela. A superfície esperada nada mais é que a superfície gerada pelo modelo como função do vetor  $\theta$ , ou seja, são os valores preditos pelo modelo. Os valores preditos formam um vetor de dimensão  $n$  que dependem do vetor de dimensão  $p$ ,  $\theta$ .

A curvatura intrínseca não depende da parametrização, por isso que o valor foi igual para os dois modelos. A parametrização devido ao efeito de parâmetros, como o nome já indica, mede o efeito da parametrização. No caso, o modelo (4.6) apresentou menor curvatura e por isso deve ser escolhido. Além do mais, possui interpretação mais conveniente. Essas medidas de curvatura se podem ser visualizadas nos gráficos de perfil de log-verossimilhança ou deviance. É mais interessante usar a escala deviance pois fica fácil definir intervalos de confiança diretamente pelos valores da deviance.

## 4.4 Intervalos de confiança

Os intervalos de confiança baseados em perfil de log-verossimilhança são definidos por

$$D(\theta_i) = 2 (\ell(\hat{\theta}_i, \hat{\theta}_{-i}) - \ell(\theta_i, \hat{\theta}_{-i})) \leq c, \quad (4.7)$$

em que  $D(\theta_i)$  é a deviance para o parâmetro  $\theta_i$ . A deviance é o dobro da diferença entre a log-verossimilhança maximizada para um valor fixo (ou condicional à) de  $\theta_i$  e a log-verossimilhança maximizada sem restrições. Todos os valores de  $\theta_i$  correspondentes à  $D(\theta_i) \leq c$  formam um intervalo de confiança. Intervalos de perfil, em modelos de regressão não linear, envolvem sucessivas maximizações da log-verossimilhança e dessa forma têm custo computacional. Para ter interpretação probabilística para o intervalo de confiança, o valor de  $c$  é baseado na sua distribuição amostral, que é chi-quadrado. Para um único parâmetro e confiança de 0,95 tem-se que  $c = 3,84$ .

Os intervalos de Wald são baseados na distribuição amostral assintótica do estimador, que normal, e são obtidos por

$$\hat{\theta} \pm z_{\alpha/2} \sqrt{V(\hat{\theta})}, \quad (4.8)$$

em que  $z_{\alpha/2}$  é quantil da distribuição normal padrão para um nível de confiança  $1 - \alpha$  e  $V(\hat{\theta})$  é a variância do estimador. Os intervalos de Wald têm menor custo computacional. Por outro lado são intervalos assintóticos, ou seja, supõem que o estimador tem distribuição normal mas isso em pequenas amostras pode ser não verificado.

Pela figura 4.4 verifica-se que o perfil de deviance para o parâmetro A é o mesmo em ambas parametrizações, pois de fato, a informação da amostra sobre esse parâmetro não muda com a parametrização do modelo. Para o parâmetro C e V observa-se assimetria a direita para os intervalos de deviance. Isso acontece porque a função de log-verossimilhança é uma função não simétrica em relação aos parâmetro que ocorrem de forma não linear no modelo, no caso, C e V. Mesmo para parâmetros que ocorrem de forma linear, como A, o perfil de verossimilhança pode ser assimétrico. Isso porque o perfil de verossimilhança envolve a maximização da log-verossimilhança em V para um valor fixo de A, por exemplo. A olho nu é difícil perceber mas o intervalo de confiança para V é mais simétrico e as linhas que formam o “V” no gráfico são menos curvadas. Para uma comparação mais objetiva pode-se verificar a assimetria dos intervalos de confiança baseados na deviance, calculando o tamanho relativo da margem de erro a esquerda e a direita da estimativa pontual.

```
> # intervalos de confiança baseados na log-verossimilhança
> c2 <- confint(n2); c2

      2.5%      97.5%
A 124.76890507 141.54495470
C  0.01667767  0.02639575

> c3 <- confint(n3); c3

      2.5%      97.5%
A 124.76891 141.54495
V 26.26158 41.56338

> sweep(c2, 1, coef(n2), function(lim, est) abs(lim-est)/est)

      2.5%      97.5%
A 0.05974856 0.0666748
C 0.20127600 0.2641404

> sweep(c3, 1, coef(n3), function(lim, est) abs(lim-est)/est)
```

```

> par(mfrow=c(2,2))
> plot(profile(n2), main=expression(eta(x,A,C)))
> plot(profile(n3), main=expression(eta(x,A,V)))
> layout(1)

```

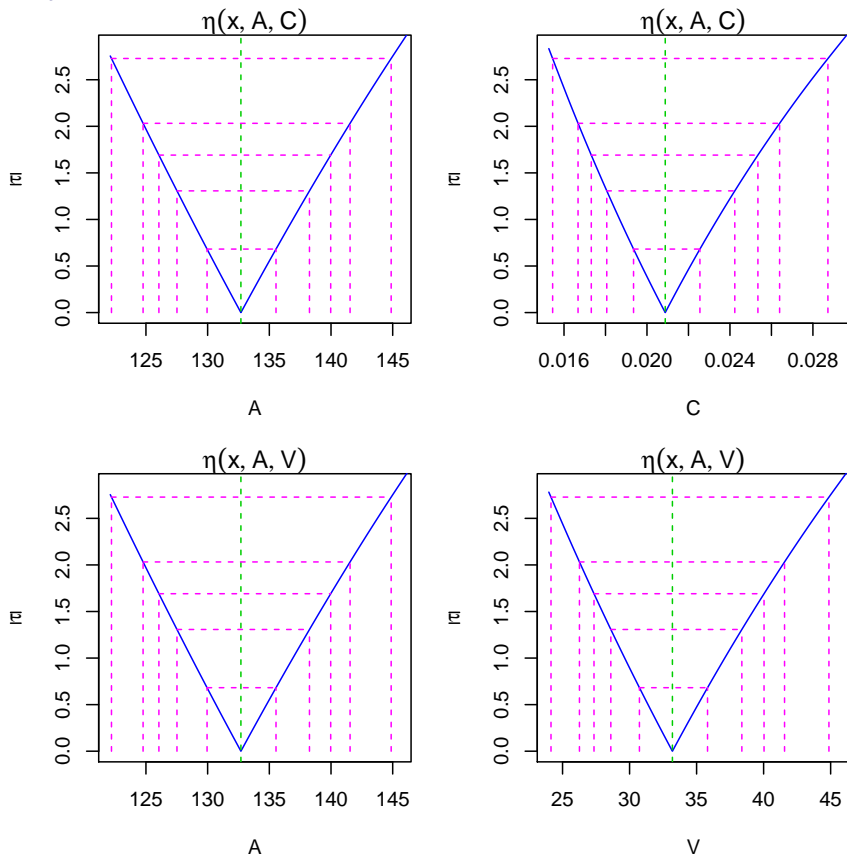


Figura 4.4: Gráficos de perfil de log-verossimilhança na escala da deviance para duas parametrizações do modelo monomolecular.

```

      2.5%      97.5%
A 0.05974812 0.0666753
V 0.20889334 0.2520599

```

```

> # intervalos de confiança de Wald
> confint.default(n2)

```

```

      2.5 %      97.5 %
A 125.16936817 140.22539718
C  0.01680012  0.02496068

```

```

> confint.default(n3)

```

```

      2.5 %      97.5 %
A 125.1693 140.22533
V 26.7091 39.68289

```

## 4.5 Bandas de confiança

Além de intervalos de confiança para os parâmetros pode-se obter intervalos de confiança para o valor predito pelo modelo, ou seja, intervalos de confiança para  $\mu|x$ . Para obter o intervalo de confiança para  $\mu|x$  consideramos o método delta e obtém-se por

$$f(x_0, \hat{\theta}) \pm z_{\alpha/2} \sqrt{\mathbf{F}_0 V(\hat{\theta}) \mathbf{F}_0^\top}, \quad (4.9)$$

em que  $x_0$  é o vetor de valores de  $x$  para predição,  $\mathbf{F}_0$  é a matriz gradiente avaliada em  $x_0$ ,  $V(\hat{\theta})$  é a matriz de covariância das estimativas dos parâmetros e  $z_{\alpha/2}$  é um quantil da distribuição normal padrão correspondente a um intervalo de confiança  $1 - \alpha$  (distribuição  $t$  de Student também pode ser considerada).

```

> V <- vcov(n2); V

```

```

      A      C
A 14.752469056 -5.756942e-03
C -0.005756942  4.333944e-06

```

```

> xnew <- c(10,50,100,150)
> m <- prmt0(x=xnew, A=coef(n2)["A"], C=coef(n2)["C"])
> F0 <- attr(m, "gradient"); F0

```

```

      A      C
[1,] 0.1884457 1076.9113
[2,] 0.6479633 2335.7174
[3,] 0.8760702 1644.5165
[4,] 0.9563722  868.3952

```

```

> # calculo pela algebra matricial didática
> sqrt(diag(F0%*%V%*%t(F0)))

[1] 1.792625 3.523107 2.540705 2.683132

> # forma computacionalmente mais apropriada
> U <- chol(vcov(n2)); U

      A      C
A 3.840894 -0.001498855
C 0.000000 0.001444776

> sqrt(apply(F0%*%t(U), 1, function(x) sum(x^2)))

[1] 1.792625 3.523107 2.540705 2.683132

> pred <- data.frame(x=0:300)
> m <- prmt0(x=pred$x, A=coef(n2)["A"], C=coef(n2)["C"])
> pred$y <- c(m)
> F0 <- attr(m, "gradient")
> pred$se <- sqrt(apply(F0%*%t(U), 1, function(x) sum(x^2)))
> z <- qnorm(0.975)
> pred <- transform(pred, lwr=y-z*se, upr=y+z*se)
> str(pred)

'data.frame':      301 obs. of  5 variables:
 $ x : int  0 1 2 3 4 5 6 7 8 9 ...
 $ y : num  0 2.74 5.43 8.06 10.63 ...
 $ se : num  0 0.22 0.431 0.632 0.823 ...
 $ lwr : num  0 2.31 4.58 6.82 9.02 ...
 $ upr : num  0 3.17 6.27 9.3 12.25 ...

> require(latticeExtra)
> xyplot(y~x, data=da, col=1, ylim=c(0,NA), xlim=c(0,NA),
+       ylab="Conteúdo liberado acumulado de potássio",
+       xlab="Dias após incubação")+
+   as.layer(xyplot(y+lwr+upr~x, data=pred,
+       type="l", col=1, lty=c(1,2,2)))
>

```

## 4.6 Comparando modelos

Até agora considerou-se dois modelos, (4.1) e (4.6), para a liberação de potássio no solo. O modelo (4.6) é uma reparametrização do modelo monomolecular que apresenta interpretação para os parâmetros igual ao modelo Michaelis-Menten. Sendo assim, a escolha do modelo mais adequado se baseia apenas na qualidade do ajuste.

Para avaliar a qualidade do ajuste podemos acessar diversas medidas, muito embora todas elas sejam função da soma de quadrados dos desvios. Pode-se considerar à própria soma de quadrados dos resíduos, o

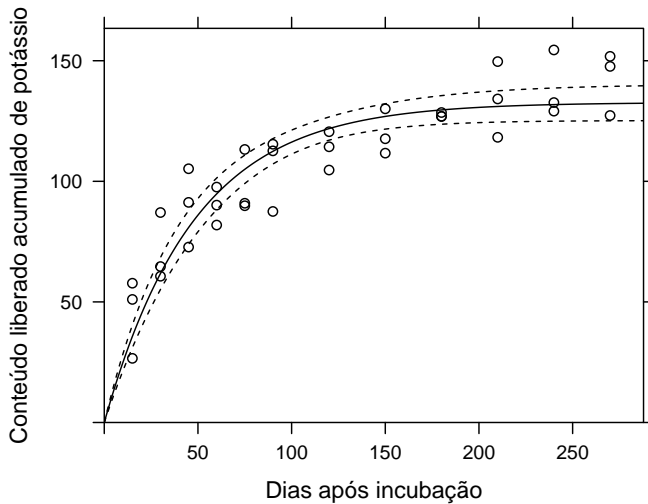


Figura 4.5: Bandas de confiança.

desvio-padrão residual, o coeficiente de determinação ( $R^2$ ), o erro quadrático médio de predição, a log-verossimilhança, o AIC ou o BIC. Estes dois últimos ponderam pelo número de parâmetros do modelo, o que no caso não têm importância pois ambos os modelos têm dois parâmetros.

```
> # soma de quadrado dos resíduos (menor melhor)
> sqr <- c(deviance(n0), deviance(n3)); sqr
```

```
[1] 4547.345 5929.083
```

```
> # desvio padrão residual (menor melhor)
> sqrt(sqr/df.residual(n0))
```

```
[1] 11.56484 13.20548
```

```
> # log-verossimilhança (maior melhor)
> c(logLik(n0), logLik(n3))
```

```
[1] -138.1798 -142.9557
```

```
> # AIC (menor melhor)
> c(AIC(n0), AIC(n3))
```

```
[1] 282.3597 291.9114
```

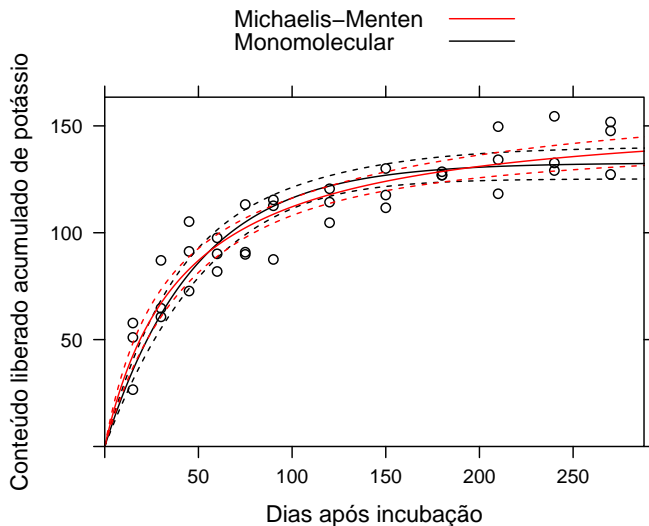


Figura 4.6: Bandas de confiança.

```
> # BIC (menor melhor)
> c(BIC(n0), BIC(n3))

[1] 287.1102 296.6619

> # R² (maior melhor)
> sqrm <- deviance(lm(y~1, data=da))
> 1-sqr/sqrm

[1] 0.8586328 0.8156775
```

Considerando qualquer das medidas tem-se que o modelo Michaelis-Menten ajustou-se melhor aos dados de liberação de potássio. Com finalidade de comparação gráfica pode-se sobrepor as curvas ajustadas com as bandas e os valores observados, tal como apresentado na figura 4.6.

As medidas de curvatura também podem ser consideradas na escolha do melhor modelo. As inferências são mais seguras para o modelo que apresentar menor valor para as curvaturas.

```
> prmt.mm <- deriv3(-A*x/(V+x), c("A","V"), function(x, A, V){NULL})
> n0 <- nls(y~prmt.mm(x, A, V), data=da, start=coef(n0))
> rms.curv(n0)
```

```
Parameter effects: c^theta x sqrt(F) = 0.2087  
Intrinsic: c^iota x sqrt(F) = 0.0747
```

```
> rms.curv(n3)
```

```
Parameter effects: c^theta x sqrt(F) = 0.1546  
Intrinsic: c^iota x sqrt(F) = 0.1001
```

Com relação a curvatura intrínseca, o modelo Michaelis é melhor mas com relação à devido ao efeito de parâmetros, o modelo monomolecular reparametrizado é melhor. Considerando o valor de corte 0,3 para as curvaturas, conforme Ritz e Streibig (2008), ambos modelos são adequados pois apresentaram curvaturas abaixo de 0,3. A curvatura que é mais preocupante é a devido ao efeito de parâmetros, já que a curvatura intrínseca não se pode amenizar. Por meio de reparametrizações pode-se diminuir a curvatura devido ao efeito de parâmetros.



# Capítulo 5

## Curvas em experimentos planejados

Em muitos experimentos é comum estudar fatores de níveis quantitativos em combinação com fatores de níveis categóricos. Para os fatores quantitativos, quando o número de níveis é superior à quatro, é possível ajustar modelos de regressão. Os normalmente empregados são os polinômios pois são flexíveis e disponíveis em muitos aplicativos. No entanto, os parâmetros das funções como polinômios não têm interpretação prática. Os modelos não lineares, como já discutido, possuem a vantagem de serem interpretáveis. O objetivo com esse capítulo é apresentar os passos para ajuste de modelos não lineares dentro de uma estrutura experimental. Ênfase será dada na formulação do modelo, na formulação das hipóteses até a seleção de um modelo.

Na literatura de nutrição de plantas é conhecido que o potássio está envolvido na abertura dos estômatos da planta atuando assim no controle da transpiração. Quando a planta está bem nutrida desse nutriente, espera-se uma menor perda de água devido à transpiração e por isso uma maior capacidade de suportar períodos mais longo sem água. Sendo assim, em regiões cujo risco de déficit hídrico é alto, uma adubação com potássio nas culturas poderia ser uma forma de minimizar os prejuízos da escassez de água.

Para avaliar os potenciais benefícios da nutrição com potássio em com-

binação com níveis de água, um experimento foi realizado com a cultura da soja em casa de vegetação (SERAFIM et al., 2012). Foram combinados 5 níveis de potássio (0, 30, 60, 120, 180 mg dm<sup>-3</sup>) com três faixas de umidade do solo (35 à 40, 47,5 à 52,5 e 60 à 65% da capacidade de campo). Foram feitas 5 repetições e usadas duas plantas por vaso (Tabela 5.1). Ao final do experimento mediu-se o rendimento de grãos (g vaso<sup>-1</sup>).

```
> link <- "http://www.leg.ufpr.br/~walmes/data/soja.txt"
> da <- read.table(link, header=TRUE, sep="\t", dec=",")
> str(da)

'data.frame':      75 obs. of  10 variables:
 $ potassio: int  0 30 60 120 180 0 30 60 120 180 ...
 $ agua    : num  37.5 37.5 37.5 37.5 37.5 50 50 50 50 50 ...
 $ bloco   : Factor w/ 5 levels "I","II","III",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ rengrao : num  14.6 21.5 24.6 21.9 28.1 ...
 $ pesograo: num  10.7 13.5 15.8 12.8 14.8 ...
 $ kgrao   : num  15.1 17.1 19.1 18.1 19.1 ...
 $ pgrao   : num  1.18 0.99 0.82 0.85 0.88 1.05 1.08 0.74 1.01 1.01 ...
 $ ts      : int  136 159 156 171 190 140 193 200 208 237 ...
 $ nvi     : int  22 2 0 2 0 20 6 6 7 10 ...
 $ nv      : int  56 62 66 68 82 63 86 94 86 97 ...
```

Tabela 5.1: Rendimento de grãos (g vaso<sup>-1</sup>) de soja em função do conteúdo de potássio do solo (mg dm<sup>-3</sup>) e da faixa de umidade do solo (%).

potassio	agua	I	II	III	IV	V
0	37.5	14.55	15.72	12.77	14.26	10.30
30	37.5	21.51	19.72	20.45	23.71	16.28
60	37.5	24.62	24.29	24.35	22.76	23.61
120	37.5	21.88	25.39	27.15	22.46	29.66
180	37.5	28.11	28.45	24.08	22.97	23.34
0	50	17.16	15.44	16.01	15.54	14.41
30	50	24.44	28.12	24.57	18.48	17.24
60	50	28.50	31.49	28.85	25.54	29.08
120	50	32.94	29.95	26.23	30.29	29.52
180	50	28.76	30.06	26.36	30.64	27.98
0	62.5	14.13	15.42	16.08	17.26	17.56
30	62.5	19.71	23.68	20.99	20.51	20.05
60	62.5	31.05	33.27	30.45	24.47	29.90
120	62.5	38.35	36.63	36.83	34.52	20.86
180	62.5	36.97	41.04	34.60	37.62	35.50

A grande maioria dos experimentos planejados possui um modelo estatístico associado. No caso deste experimento que é em arranjo fatorial

3×5 em cinco blocos o modelo estatístico associado é

$$E(Y_{ijk}) = \mu + \gamma_k + \tau_i + \kappa_j + \lambda_{ij} \quad (5.1)$$

em que  $\mu$  é uma constante geral e  $\gamma$ ,  $\tau$ ,  $\kappa$  e  $\lambda$  representam os efeitos dos níveis de bloco, faixa de água, nível de potássio e interação faixa de água com nível de potássio. Esse modelo trata ambos fatores como qualitativos. No entanto, o fator potássio é quantitativo. Dessa forma é razoável pensar que o rendimento de grãos seja uma função contínua do nível de potássio. Além do mais, conhecimento prévio sobre respostas nutricionais de plantas indica que é uma função monótona não decrescente do fertilizante, pois a medida que se aumenta o conteúdo de fertilizante o incremento na produção é cada vez menor.

Pela figura 5.1 verifica-se que o rendimento de grãos aumenta com o nível de potássio e com a faixa de água. Para potássio entre 0 e 60 tem-se pouca diferença entre as faixas de água; as linhas médias parecem ter mesma inclinação e mesmo ponto de partida (intercepto). Para potássio acima de 60 o incremento em rendimento se dá a uma taxa menor e existe maior efeito da faixa de água. Então o que se pode supor é que quando o potássio está abaixo de 60 não existe pouco efeito de faixa de água talvez porque o potássio seja o fator limitante do rendimento mas se o potássio é superior à 60 então confere a planta melhor capacidade de aproveitar a água disponível. Dessa maneira pode-se antecipar uma interação entre faixa de água e nível de potássio. A observação 74 será removida visto que foi um vaso onde as plantas morreram por causas externas.

Pode-se analisar os dados considerando o modelo (5.1), ou seja, fazendo uma análise de variância e em seguida um teste de médias. Esse é tipo de análise é muito comum embora não seja forma mais simples de interpretar esses dados como será visto.

Pela análise de variância verifica-se interação entre faixa de água e nível de potássio. Pelo gráfico de diagnóstico dos resíduos tem-se que os pressupostos do modelo foram atendidos (figura 5.2). O resultado do teste de médias para as comparações duas a duas está apresentado na tabela 5.2. Por remover uma observação causou-se o desbalanceamento do experimento. Com isso tem-se que os valores preditos apresentam erros-padrões diferentes. Não é o objetivo com esse livro tratar de análise de dados desbalanceados então justificativas dos procedimento aqui adotados serão omitidos. Desse resultado percebe-se que o rendimento médio não é diferente entre os últimos dois níveis de potássio em qualquer faixa de água e que do primeiro até o terceiro nível as médias diferem entre si. Uma pergunta natural ao inspecionar esse resultado é: qual o valor de potássio para o qual a produção estabiliza?

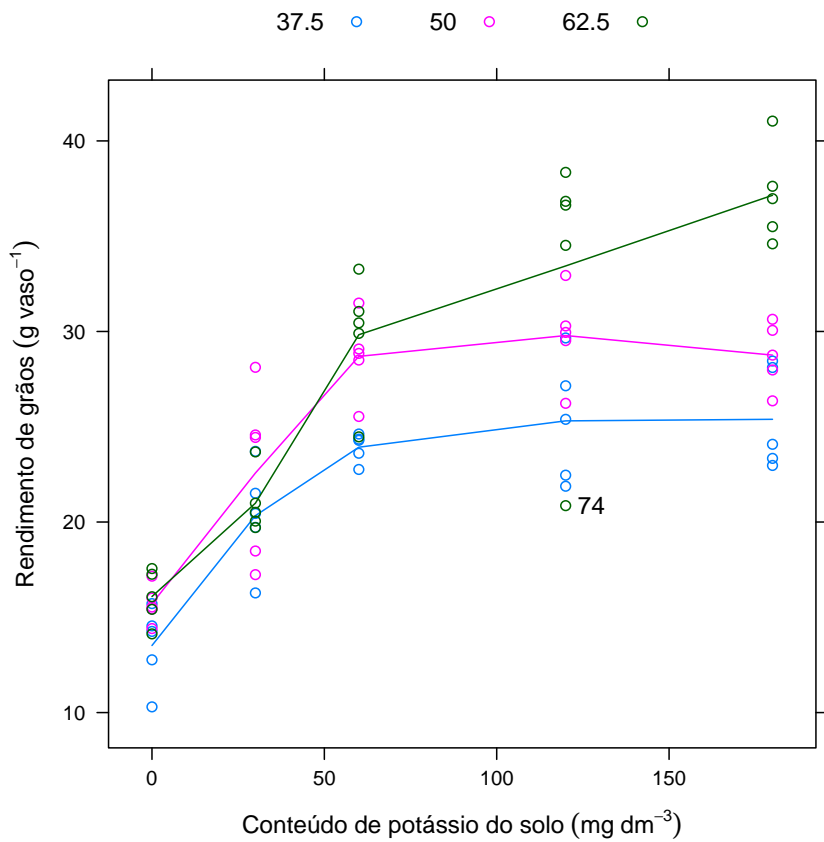


Figura 5.1: Rendimento de grãos em função do conteúdo de potássio do solo e faixa de umidade. As linhas passam pelas médias dos valores em cada nível de potássio.

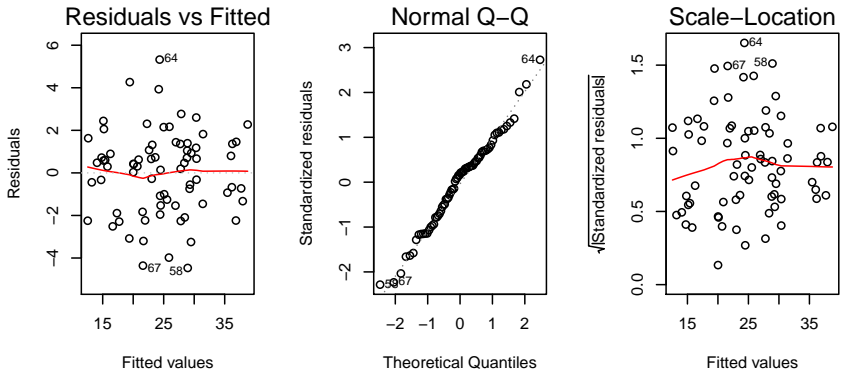


Figura 5.2: Gráfico de diagnósticos dos resíduos considerados para verificar os pressupostos do modelo.

```
> da <- da[-74,]
> da <- transform(da, POT=factor(potassio), AG=factor(agua))
> m0 <- lm(reengrao~bloco+AG*POT, data=da)
> anova(m0)
```

#### Analysis of Variance Table

Response: reengrao

	Df	Sum Sq	Mean Sq	F value
bloco	4	100.39	25.10	4.8948
AG	2	446.53	223.26	43.5456
POT	4	2592.13	648.03	126.3925
AG:POT	8	286.00	35.75	6.9726
Residuals	55	281.99	5.13	

	Pr(>F)
bloco	0.001896 **
AG	4.617e-12 ***
POT	< 2.2e-16 ***
AG:POT	2.629e-06 ***
Residuals	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Para responder essa pergunta modelos não lineares serão considerados na análise dos dados. O modelo linear-platô pode representar o comportamento destes dados

$$\eta(x, \theta) = \begin{cases} \theta_0 + \theta_1 x & , x \leq \theta_b \\ \theta_0 + \theta_1 \theta_b & , x > \theta_b \end{cases} \quad (5.2)$$

Tabela 5.2: Teste de médias para o rendimento de grãos em função dos níveis de potássio dentro de cada nível de faixa de água.

	37.5	50	62.5
0	13.52a	15.71a	16.09a
30	20.33b	22.57b	20.99b
60	23.93c	28.69c	29.83c
120	25.31c	29.79c	36.34d
180	25.39c	28.76c	37.15d

em que  $\theta_0$  é o intercepto (Y),  $\theta_1$  é taxa de incremento no rendimento por unidade de potássio ( $Y\ X^{-1}$ ) e  $\theta_b$  é o ponto de quebra (*break*); para valores de  $x$  maiores que  $\theta$  a função é constante refletindo a estabilização do rendimento (figura 5.3).

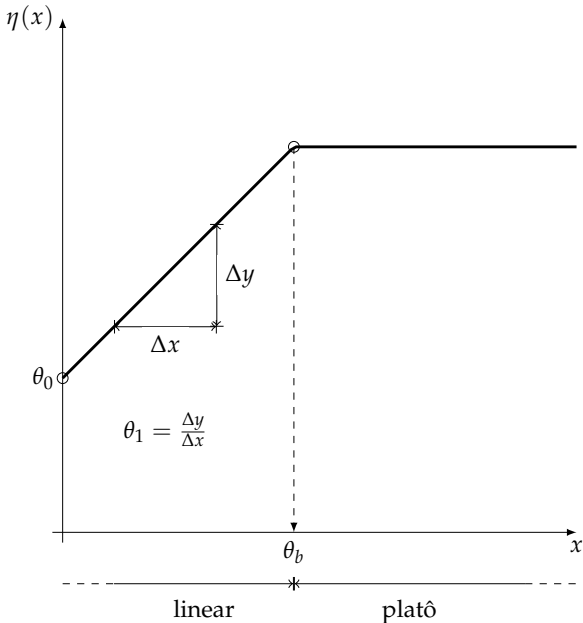


Figura 5.3: Esboço do modelo linear platô com destaque para interpretação cartesiana dos parâmetros.

Pode-se ajustar o modelo (5.2) separado para cada nível de faixa de água, ou seja, usando a `nls()` três vezes. No entando, se o interesse é comparar as curvas ajustadas por meio de hipóteses sobre os parâmetros,

é mais adequado ajustar as três curvas sob um único modelo. Valores iniciais para os parâmetros são obtidos visualmente a partir da figura 5.1.

```
> n37 <- nls(rengrao~th0+th1*potassio*(potassio<=thb)+
+           th1*thb*(potassio>thb),
+           data=subset(da, agua==37.5),
+           start=list(th0=15, th1=15/60, thb=60))
> n50 <- nls(rengrao~th0+th1*potassio*(potassio<=thb)+
+           th1*thb*(potassio>thb),
+           data=subset(da, agua==50),
+           start=list(th0=15, th1=15/60, thb=70))
> n62 <- nls(rengrao~th0+th1*potassio*(potassio<=thb)+
+           th1*thb*(potassio>thb),
+           data=subset(da, agua==62.5),
+           start=list(th0=15, th1=15/60, thb=80))
> rbind(coef(n37), coef(n50), coef(n62))
```

```
      th0      th1      thb
[1,] 14.05700 0.1734333 65.10859
[2,] 15.83467 0.2163333 62.11864
[3,] 15.43300 0.2289667 93.73659
```

Comparando-se as estimativas pontuais verifica-se pouca diferença entre os valores de  $\theta_0$  o que indica que para o nível zero de potássio o rendimento de grãos. Então a primeira hipótese a ser avaliada é

$$H_0 : \theta_0 \sim 1 \quad \text{versus} \quad H_a : \theta_0 \sim \text{faixa de água}, \quad (5.3)$$

em palavras esta-se testando se  $\theta_0$  é o mesmo (ou único) para todas as curvas ( $H_0$ ) ou se cada faixa de água tem um  $\theta_0$  diferente. Pode-se considerar a mesma estrutura de hipóteses para os parâmetros  $\theta_1$  e  $\theta_b$ . Pode-se começar ajustando o modelo irrestrito, ou seja, onde todos os parâmetros estão sob  $H_a$  e ir testando os modelos sob as restrições previstas por  $H_0$ . Uma vez que ajustou-se os modelos separados pode-se usar as estimativas dos parâmetros como valores iniciais para o ajuste conjunto.

```
> nHa <- nls(rengrao~th0[AG]+th1[AG]*potassio*(potassio<=thb[AG])+
+           th1[AG]*thb[AG]*(potassio>thb[AG]), data=da,
+           start=list(th0=c(14,15,15),
+                       th1=c(0.17,0.21,0.22),
+                       thb=c(65,62,93)))
> confint.default(nHa)
```

```
      2.5 %      97.5 %
th01 12.1283970 15.9856030
th02 13.9060636 17.7632697
th03 13.5043970 17.3616030
th11 0.1236370 0.2232296
th12 0.1665370 0.2661296
th13 0.1791704 0.2787630
thb1 50.0995234 80.1176593
thb2 50.5369540 73.7003342
thb3 77.3713144 110.1018743
```

Os intervalos de confiança para os parâmetros sugerem que os valores de  $\theta_0$  e  $\theta_1$  parecem não depender da faixa de água uma vez que para cada parâmetro os intervalos apresentam sobreposição. Então pode-se testar  $H_0$  para  $\theta_0$  e  $\theta_1$  simultaneamente. Para isso ajusta-se um modelo sob as restrições previstas em  $H_0$  e testa-se essa redução de modelos pelo teste de razão de verossimilhanças.

```
> nH01 <- nls(rengrao~th0+th1*potassio*(potassio<=thb[AG])+
+           th1*thb[AG]*(potassio>thb[AG]),
+           data=da, start=list(th0=c(15),
+                               th1=c(0.21),
+                               thb=c(65,62,93)))
> confint.default(nH01)
```

```
      2.5 %      97.5 %
th0  13.706450  15.9627807
th1   0.201535   0.2656752
thb1 36.408050  49.5494065
thb2 53.117214  70.4963787
thb3 82.440969 106.4327972
```

```
> anova(nH01, nHa)
```

#### *Analysis of Variance Table*

```
Model 1: rengrao ~ th0 + th1 * potassio * (potassio <= thb[AG]) + th1 * thb[AG] * (p
Model 2: rengrao ~ th0[AG] + th1[AG] * potassio * (potassio <= thb[AG]) + th1[AG] *
Res.Df Res.Sum Sq Df Sum Sq F value Pr(>F)
1      69      405.27
2      65      377.62  4 27.647  1.1897 0.3237
```

O teste de razão de verossimilhanças apontou que o modelo menor pode ser considerado. Então conclui-se até aqui que  $\theta_0$  e  $\theta_1$  não dependem do nível de água. Dessa forma conclui-se que o rendimento para o potássio no nível zero e o incremento até a estabilização não dependem da faixa de água. No entanto, o aumento da faixa de água permite que a planta aproveite mais o potássio disponível no solo pois o nível de potássio para estabilização do rendimento aumenta com a faixa de água. Ao considerar um modelo não linear apropriado chegou-se a uma interpretação para o fenômeno muito mais clara, parsimoniosa e funcional que os resultados obtidos pelo modelo (5.1). Vale ressaltar que a análise pelo modelo (5.1) não está errada e sim pouco adequada pois considera-se como fator de níveis qualitativos um que é de níveis quantitativos. Ao fazer isso estimativa-se 15 médias para explicar o comportamento dos dados (sem considerar o efeito dos blocos) enquanto que o modelo não linear final fez isso com apenas cinco parâmetros.

Para ser didático, nessa análise não se considerou o efeito dos blocos que não podem ser ignorados. Nos modelos que representam os



experimentos, como (5.1), o efeito dos blocos na grande maioria é sobre o intercepto. Isso não é arbitrário e sim porque é a única opção possível para representar o efeito dos blocos. No modelo (5.2) pode-se associar o efeito dos blocos a qualquer um dos três parâmetros mas pela força da simplicidade será feito sobre o intercepto também.

```
> nH01b <- nls(rengrao~th0[bloco]+th1*potassio*(potassio<=thb[AG])+
+             th1*thb[AG]*(potassio>thb[AG]),
+             data=da, start=list(th0=c(15,15,15,15,15),
+                                 th1=c(0.21), thb=c(65,62,93)))
> anova(nH01, nH01b)
```

#### Analysis of Variance Table

```
Model 1: rengrao ~ th0 + th1 * potassio * (potassio <= thb[AG]) + th1 * thb[AG]
Model 2: rengrao ~ th0[bloco] + th1 * potassio * (potassio <= thb[AG]) + th1 *
  Res.Df Res.Sum Sq Df Sum Sq F value
1      69      405.27
2      65      335.78  4 69.491  3.3631
  Pr(>F)
1
2 0.01455 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Assim como bloco foi significativo pelo modelo (5.1) também foi pelo modelo (5.2). Isso não necessariamente vai acontecer para todos os casos semelhantes a este. Além do mais, o efeito de blocos deveria ter sido declarado desde o início mas optou-se por coloca-lo no final para ser mais didático.

Vale comentar que dentre todos os modelos que poderiam ser considerados para esses dados o modelo (5.1) é o maior (modelo saturado). Este é um modelo de médias de caselas onde todos os fatores têm níveis qualitativos de efeito fixo. Sendo assim é o modelo que apresenta a menor soma de quadrados residual possível devido ao grande número de parâmetros presentes. O menor modelo seria aquele com apenas um intercepto (modelo nulo), ou seja, que não considera o efeito de nenhuma fonte de variação. O modelo (5.2) é um modelo intermediário aos dois e portanto a soma de quadrados residual será por lógica maior que o modelo de médias de caselas. Por outro lado, este têm menos parâmetros, portanto é mais parcimonioso e pelo critério de informação de Akaike é um modelo melhor que o modelo saturado.

```
> c(deviance(m0), deviance(nH01b)) # SQR
```

```
[1] 281.9930 335.7752
```

```
> c(AIC(m0), AIC(nH01b))
```

```
# critério de informação de Akaike
```

```
[1] 349.0014 341.9188
```

Para concluir a análise pode-se representar os resultados em um gráfico em associação as estimativas intervalares e tabelas com níveis descritivos dos testes de hipótese.

```
> require(rootSolve)
> numF <- function(pars, k, bloco, agua){
+   bloco%%pars[1:5]+pars[6]*k*(k<agua%%pars[7:9])+
+   pars[6]*agua%%pars[7:9]*(k>=agua%%pars[7:9])
+ }
> pred <- rbind(expand.grid(bloco=levels(da$bloco)[1],
+   AG=levels(da$AG), potassio=seq(0,180,1)),
+   expand.grid(bloco=levels(da$bloco)[-1],
+   AG=levels(da$AG)[1], potassio=0))
> c0 <- coef(nH01b)
> F <- gradient(numF, x=c0, k=pred$potassio,
+   bloco=matrix(1/nlevels(da$bloco),
+   nrow=nrow(pred),
+   ncol=nlevels(da$bloco)),
+   agua=model.matrix(~-1+AG, pred))
> dim(F)
```

```
[1] 547 9
```

```
> U <- chol(vcov(nH01b))
> pred$se <- sqrt(apply(F%*%t(U), 1, function(x) sum(x^2)))
> tval <- qt(p=c(lwr=0.025, fit=0.5, upr=0.975), df=df.residual(nH01b))
> me <- outer(pred$se, tval, "*")
> pred$rengrao <- predict(nH01b, newdata=pred)
> pred <- cbind(pred, sweep(me, 1, pred$rengrao, "+"))
> pred <- subset(pred, bloco=="I")
>
> # xyplot(rengrao-potassio|AG, data=da, col=1)+
> #   as.layer(xyplot(fit+lwr+upr~potassio|AG, data=pred,
> #     type="l", col=1, lty=c(1,2,2)))
```

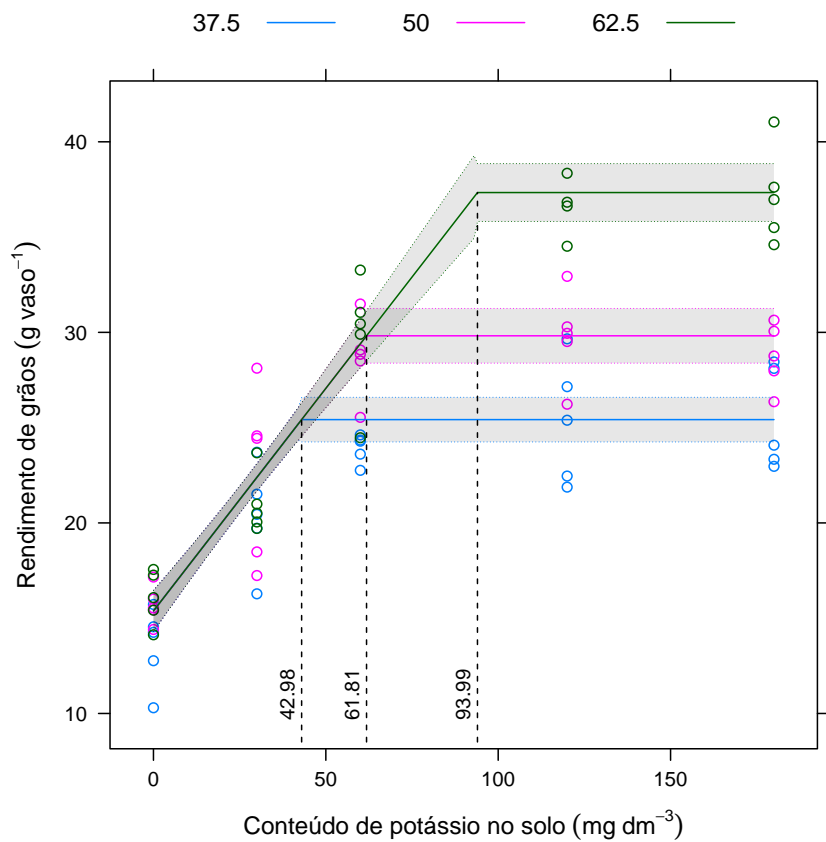


Figura 5.4: Valores ajustados seguidos das bandas de confiança.



## Efeito aleatório

### 6.1 Liberação de potássio

Nesse capítulo serão considerados modelos não lineares com termos de efeito aleatório. No exemplo de liberação de potássio no solo (tabela 4.1) tem-se o efeito das três repetições ou unidades experimentais que não foi considerado no modelo. Decidiu-se não mencionar sobre o efeito das unidades experimentais por razões didáticas, no entanto, na prática não se deve negligenciar o seu efeito. As inferências sob a negligência estão comprometidas.

Nesse experimento o conteúdo liberado de nutrientes foi observado em três unidades experimentais ao longo do tempo. Sendo assim tem-se um típico caso de medidas repetidas ou estudo longitudinal. Conforme ilustrado na figura 1.8 espera-se que exista o efeito das unidades experimentais sobre a resposta. Essas unidades amostrais não são exatamente iguais. Por exemplo, durante a instalação do experimento forças do acaso fazem com que a quantidade de solo em cada recipiente seja diferente, ou que a quantidade incubada de esterco tenha um pouco de variação, ou que a concentração de potássio em cada amostra de esterco incorporada seja diferente. Mesmo considerando balanças de alta precisão, tendo rigor durante o processo de mistura do esterco no solo e demais medidas de controle local, ainda assim haverá diferença entre unidades experimentais, mesmo em pequena escala. Tal variação em termos de unidade experimental pode ser acomodada na modelagem associando termos de efeito

aleatório aos parâmetros do modelo não linear.

No exemplo da produção de soja em função da adubação com potássio (tabela 5.1) o efeito dos blocos foi considerado como fixo. Normalmente prefere-se declarar o efeito dos blocos como aleatório. Como saber se um fator é de efeito fixo ou aleatório não é escopo desse material.

O pacote `nlme` possui funções destinadas à estimação de modelo lineares e não lineares com termos de efeito aleatório. No caso do modelo ser linear a função é a `lme()` (*linear mixed effects*) e se é não linear é a `nlme()` (*nonlinear mixed effects*). Por meio do argumento `random=` dessas funções é que se especifica o termo de efeito aleatório.

A função `nlme()` tem argumentos iguais aos da `nlm()` além de seus argumentos exclusivos `random=` para especificar termos de efeito aleatório, `weights=` para modelar a variância juntamente com a média e `correlation=` para associar estrutura de correlação aos erros.

```
> require(nlme)
> require(lattice)
> require(latticeExtra)
> args(nlme)
```

```
function (model, data = sys.frame(sys.parent()), fixed, random = fixed,
  groups, start, correlation = NULL, weights = NULL, subset,
  method = c("ML", "REML"), na.action = na.fail, naPattern,
  control = list(), verbose = FALSE)
NULL
```

Para ajustar um modelo de efeitos aleatórios antes é necessário que se declare qual o fator de efeitos aleatórios. No exemplo da liberação de potássio tem-se que a unidade experimental é um fator de efeito aleatórios. Então cria-se um objeto de classe `groupedData` onde representa-se o fator de efeitos aleatórios na fórmula.

```
> db <- groupedData(formula=y~x|r, data=da, order.groups=FALSE)
> str(db)
```

```
Classes 'nfnGroupedData', 'nfGroupedData', 'groupedData' and 'data.frame':
 $ y: num  51 57.8 26.6 60.6 87.1 ...
 $ x: num  15 15 15 30 30 30 45 45 45 60 ...
 $ r: Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3 1 2 3 1 ...
 - attr(*, "formula")=Class 'formula' length 3 y ~ x | r
 .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
 - attr(*, "FUN")=function (x)
 - attr(*, "order.groups")= logi FALSE
```

Uma vez criado o objeto `groupedData` só é necessário declarar o modelo para a função `nlme()`. No argumento `random=` se especifica quais os parâmetros do modelo estão sujeitos ao fator de efeitos aleatórios. A es-

colha é o usuário que faz porém não é totalmente arbitrária. No caso da liberação de potássio tem-se a premissa de que variações nas unidades experimentais em termos de quantidade incubada de resíduo orgânico vão implicar em diferenças nas estimativas do conteúdo total de potássio,  $\theta_a$ . O tempo de meia vida,  $\theta_v$ , também pode depender da unidade experimental. De forma geral, pode-se associar efeito aleatório à todos os parâmetros do modelo. No entanto, na prática isso não é usual pois 1) o modelo acaba tendo parâmetros demais, 2) a variação devido ao fator aleatório pode ser acomodada associando-se à poucos parâmetros e 3) pode-se não alcançar convergência quando se declara efeito aleatório sobre todos os parâmetros.

Como ponto de partida será considerado efeito aleatório sobre o parâmetro  $\theta_a$ , conforme já justificado. Os argumentos `fixed=` e `random=` permitem especificar modelos para a parte fixa e aleatória. Na parte fixa pode-se incluir efeito de covariáveis que no caso da produção de soja foram os níveis de umidade. Na parte aleatória declarou-se que o efeito de unidade experimental está associado ao parâmetro  $\theta_a$ . Ou seja, as diferenças entre unidades experimentais serão representados por desvios com relação à um valor fixo, o parâmetro  $\theta_a$ , e esses desvios têm média zero e variância  $\sigma_a^2$  que será estimada.

```
> # modelo que ignora o efeito de unidade experimental
> n0 <- nls(y~tha*x/(thv+x), data=db, start=c(tha=140, thv=40))
> summary(n0)
```

*Formula: y ~ tha \* x/(thv + x)*

*Parameters:*

	Estimate	Std. Error	t value	Pr(> t )
tha	157.633	6.110	25.800	< 2e-16 ***
thv	40.575	5.658	7.171	2.71e-08 ***

*Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1*

*Residual standard error: 11.56 on 34 degrees of freedom*

*Number of iterations to convergence: 3*

*Achieved convergence tolerance: 2.906e-06*

```
> # efeito de unidade experimental em tha
> nn0 <- nlme(y~tha*x/(thv+x), data=db, method="ML",
+           fixed=tha+thv~1, random=tha~1,
+           start=c(160, 40))
> summary(nn0)
```

*Nonlinear mixed-effects model fit by maximum likelihood*

*Model: y ~ tha \* x/(thv + x)*

*Data: db*

	AIC	BIC	logLik
	274.8572	281.1913	-133.4286

```

Random effects:
Formula: tha ~ 1 | r
          tha Residual
StdDev: 9.518301 9.056731

Fixed effects: tha + thv ~ 1
          Value Std.Error DF   t-value p-value
tha 156.81823   7.452555 32 21.042210      0
thv 39.72996   4.466827 32  8.894447      0
Correlation:
  tha
thv 0.579

Standardized Within-Group Residuals:
          Min          Q1          Med
-1.64046950 -0.54059037  0.09200759
          Q3          Max
 0.74537408  1.78935044

Number of Observations: 36
Number of Groups: 3

> anova(nn0, n0)

```

```

      Model df      AIC      BIC    logLik
nn0      1  4 274.8572 281.1913 -133.4286
n0       2  3 282.3596 287.1102 -138.1798
      Test  L.Ratio p-value
nn0
n0  1 vs 2 9.502467  0.0021

```

Modelos com (nn0) e sem (n0) o termo de efeito aleatório foram ajustados. Pelas saídas verifica-se que o modelo de efeitos aleatórios teve um erro padrão residual menor o que indica que o termo de efeito aleatório está explicando parte da variação. Pelo teste de razão de verossimilhanças verificou-se que existe efeito da unidade experimental associada ao parâmetro assíntota, ou seja, as unidades experimentais têm conteúdo total de potássio diferente. O desvio padrão do termo de efeito aleatório está na saída do ajuste, 9.51. As estimativas dos parâmetros foram diferentes e os erros padrões também.

Pode-se verificar quão melhor é o modelo com efeito aleatório associado ao parâmetro  $\theta_v$ . Ao declarar tal modelo verifica-se que não há necessidade de associar efeito aleatório ao parâmetro  $\theta_v$ .

```

> # efeito de unidade experimental tem tha e thv
> nn1 <- nlme(y~tha*x/(thv+x), data=db, method="ML",
+           fixed=tha+thv~1, random=tha+thv~1,
+           start=c(160, 40))
> summary(nn1)

```

```

Nonlinear mixed-effects model fit by maximum likelihood
Model: y ~ tha * x/(thv + x)

```



```

Data: db
      AIC      BIC    logLik
274.7412 284.2424 -131.3706

Random effects:
Formula: list(thu ~ 1, thv ~ 1)
Level: r
Structure: General positive-definite, Log-Cholesky parametrization
           StdDev  Corr
thu      1.085483  thu
thv      8.938301 -1
Residual 8.437085

Fixed effects: thu + thv ~ 1
           Value Std.Error DF  t-value p-value
thu 157.29841  4.522485 32 34.78141      0
thv  40.78159  6.743684 32  6.04738      0
Correlation:
  thu
thv 0.427

Standardized Within-Group Residuals:
           Min           Q1           Med
-1.70124520 -0.76092727  0.03243357
           Q3           Max
 0.79753600  1.82401317

Number of Observations: 36
Number of Groups: 3

> anova(nn0, nn1)

      Model df      AIC      BIC    logLik
nn0      1  4 274.8572 281.1913 -133.4286
nn1      2  6 274.7412 284.2423 -131.3706
      Test  L.Ratio p-value
nn0
nn1 1 vs 2 4.115944 0.1277

```

As previsões dos efeitos aleatórios podem ser acessadas com a função `ranef()` e as estimativas dos parâmetros do modelo não linear com a função `fixef()`. Com a função `augPred()` pode-se acessar e plotar os valores preditos de forma automática.

```

> fixef(nn0)

      thu      thv
156.81823 39.72996

> ranef(nn0)

      thu
1  0.1284499
2 10.7857930
3 -10.9142429

```

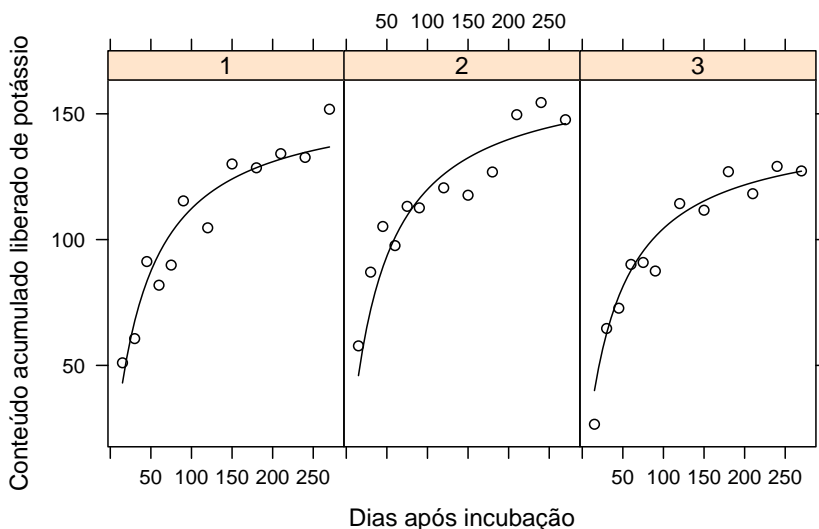


Figura 6.1: Valores observados e preditos para o conteúdo liberado acumulado de potássio em função dos dias após a incubação do estercor de codorna em LVdf.

```
> plot(augPred(nn0), layout=c(3,1), col=1,
+       xlab="Dias após incubação",
+       ylab="Conteúdo acumulado liberado de potássio")
```

Com a finalidade de representar os resultados tal como ilustrado na figura 1.8 pode-se obter as predições a nível populacional, sem considerar os termos de efeito aleatório, e as predições a nível de unidade experimental. Com a função `predict()` obtém-se as predições e o argumento `level=` controla o nível de predição.

```
> range(db$x)

[1] 15 270

> xseq <- seq(0,270,3)
> pred.pop <- data.frame(x=xseq)
> pred.pop$y <- predict(nn0, newdata=pred.pop, level=0)
> pred.ue <- expand.grid(x=xseq, r=levels(db$r))
> pred.ue$y <- predict(nn0, newdata=pred.ue, level=1)
>
> # xyplot(y~x, groups=r, data=db,
> #       xlab="Dias após incubação",
```

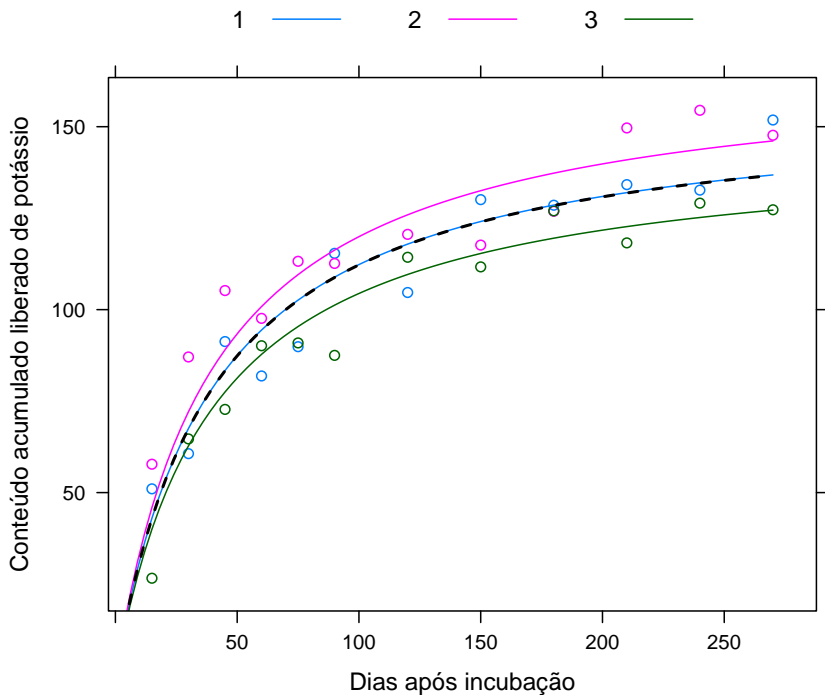


Figura 6.2: Valores observados e preditos para o conteúdo liberado acumulado de potássio em função dos dias após a incubação do esterco de codorna em LVdf com destaque para a predição a nível populacional e de unidade experimental.

```
> #      ylab="Conteúdo acumulado liberado de potássio",
> #      auto.key=list(columns=3, lines=TRUE, points=FALSE))+
> #      as.layer(xyplot(y~x, groups=r, data=pred.ue, type="l"))+
> #      as.layer(xyplot(y~x, data=pred.pop, type="l", lwd=2, col=1, lty=2))

> pred <- data.frame(x=xseq)
> model <- deriv3(~tha*x/(thv+x), c("tha","thv"),
+               function(x, tha, thv){NULL})
> z <- qnorm(0.975)
> V <- vcov(n0); V

      tha      thv
tha 37.32882 30.81582
thv 30.81582 32.01306

> U <- chol(V)
```

```

> m <- model(x=xseq, tha=coef(n0)["tha"], thv=coef(n0)["thv"])
> F <- attr(m, "gradient")
> pred$ynls <- c(m)
> pred$senls <- sqrt(apply(F%*%t(U), 1, function(x) sum(x^2)))
> pred <- transform(pred,
+                     lwrnls=ynls-z*senls,
+                     uprnls=ynls+z*senls)
> V <- vcov(nn0); V

      tha      thv
tha 52.45499 18.21295
thv 18.21295 18.84407

> U <- chol(V)
> m <- model(x=xseq, tha=fixef(nn0)["tha"], thv=fixef(nn0)["thv"])
> F <- attr(m, "gradient")
> pred$ynlme <- c(m)
> pred$senlme <- sqrt(apply(F%*%t(U), 1, function(x) sum(x^2)))
> pred <- transform(pred,
+                     lwrnlme=ynlme-z*senlme,
+                     uprnlme=ynlme+z*senlme)
> str(pred)

'data.frame':      91 obs. of  9 variables:
 $ x      : num  0 3 6 9 12 15 18 21 24 27 ...
 $ ynls   : num  0 10.9 20.3 28.6 36 ...
 $ senls  : num  0 1.05 1.8 2.33 2.7 ...
 $ lwrnls : num  0 8.79 16.78 24.05 30.68 ...
 $ uprnls : num  0 12.9 23.8 33.2 41.3 ...
 $ ynlme  : num  0 11 20.6 29 36.4 ...
 $ senlme : num  0 0.922 1.602 2.109 2.49 ...
 $ lwrnlme: num  0 9.2 17.4 24.8 31.5 ...
 $ uprnlme: num  0 12.8 23.7 33.1 41.3 ...

>
> # xyplot(y~x, data=db, col=1, ylim=c(0,NA), xlim=c(0,NA),
> #       ylab="Conteúdo liberado acumulado de potássio",
> #       xlab="Dias após incubação",
> #       key=list(columns=2, text=list(c("nls","nlme")),
> #       lines=list(lty=2:3)))+
> #   as.layer(xyplot(ynls+lwrnls+uprnls~x, data=pred,
> #       type="l", col=1, lty=c(1,2,2)))+
> #   as.layer(xyplot(ynlme+lwrnlme+uprnlme~x, data=pred,
> #       type="l", col=1, lty=c(1,3,3)))

```

Os valores preditos pelos dois modelos, com e sem efeito aleatório, foram praticamente iguais (figura 6.3). Isso é em parte devido à estrutura balanceada do experimento e ao fato do efeito aleatório estar associado a um parâmetro que aparece linearmente no modelo. A amplitude das bandas de confiança foram diferentes e aumentam ao afastar do dia zero. A assíntota representa o valor predito para a resposta  $Y$  quando  $x$  vai para o infinito, ou seja,  $\theta_a = \lim_{x \rightarrow \infty} E(Y|x)$ . Como no modelo de efeito aleatórios associou-se efeito aleatório ao parâmetro  $\theta_a$  isso implica em maior incerteza sobre seu parâmetro e por isso as bandas de confiança abrem ao

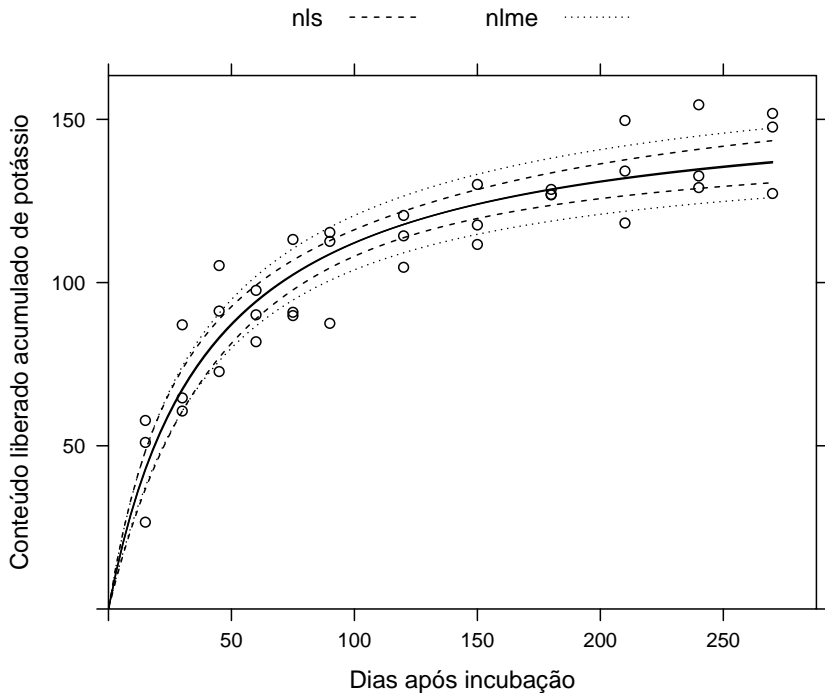


Figura 6.3: Predição intervalar para o conteúdo liberado acumulado de potássio em função dos dias após incubação pelo modelo não linear sem e com efeito aleatório de unidade experimental.

distanciar de zero. A variância da estimativa de  $\theta_a$  foi maior no modelo de efeitos aleatórios ao passo que  $\theta_v$  teve menor variância. Tal resultado indica que houve um efeito compensatório com a inclusão dos efeitos aleatórios que aumentaram a incerteza de  $\theta_a$  por estarem associados à ele e diminuíram a incerteza sobre  $\theta_v$  por acomodar mais variação com os termos de efeito aleatório. No entanto, não se pode considerar esse resultado como geral pois depende do tamanho da variância dos efeitos aleatórios e do número de níveis.

## 6.2 Produção de soja

A análise dos dados de produção de soja em função do conteúdo de potássio do solo e nível de umidade foi feita considerando o efeito de blocos como fixo. Agora a análise será refeita considerando efeito aleatório para os blocos. Primeiro declara-se um objeto de classe `groupedData` que informa a estrutura de efeitos aleatórios dos dados.

```
> link <- "http://www.leg.ufpr.br/~walmes/data/soja.txt"
> da <- read.table(link, header=TRUE, sep="\t", dec=",")
> da <- subset(da, select=c(agua, potassio, rengrao, bloco))
> da <- da[-74,]
> db <- groupedData(rengrao~potassio|bloco, data=da, order=FALSE)
> db$AG <- factor(db$agua)
> str(db)
```

```
Classes 'nfnGroupedData', 'nfGroupedData', 'groupedData' and 'data.frame':
 $ agua      : num  37.5 37.5 37.5 37.5 37.5 50 50 50 50 ...
 $ potassio  : int   0 30 60 120 180 0 30 60 120 180 ...
 $ rengrao   : num  14.6 21.5 24.6 21.9 28.1 ...
 $ bloco     : Factor w/ 5 levels "I","II","III",...: 1 1 1 1 1 1 1 1 1 ...
 $ AG        : Factor w/ 3 levels "37.5","50","62.5": 1 1 1 1 1 2 2 2 2 ...
- attr(*, "formula")=Class 'formula' length 3 rengrao ~ potassio | bloco
.. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
- attr(*, "FUN")=function (x)
- attr(*, "order.groups")= logi FALSE
```

Com a função `nlme()` declara-se a expressão do modelo não linear e os termos de efeito fixo e efeito aleatório são passados para os argumentos `fixed=` e `random=`. Os blocos serão declarados como de efeito aleatório enquanto que o efeito dos níveis de umidade serão declarados como de efeito fixo.

Na hora de passar valores iniciais para os parâmetros deve-se estar atento a parametrização de efeitos considerada. Por padrão o R assume como nulo o efeito da primeira categoria (ou nível) de um fator. Sendo assim, deve-se passar como valor inicial um número que corresponde a estimativa para a primeira categoria e para as demais a diferença com relação a primeira categoria. Além disso, os valores iniciais são passados em um vetor sem nomes e deve-se respeitar a ordem com que os parâmetros foram declarados no argumento `fixed=`.

```
> nn1 <- nlme(rengrao~th0+th1*potassio*(potassio<=thb)+
+             th1*thb*(potassio>thb),
+             data=db, method="ML",
+             fixed=th0+th1+thb~AG, random=th0~1,
+             start=c(15,0,0, # th0
+                     0.22,0,0, # th1
+                     40,20,50)) # thb
> summary(nn1)
```

Nonlinear mixed-effects model fit by maximum likelihood

Model:  $\text{rengrao} \sim \text{th0} + \text{th1} * \text{potassio} * (\text{potassio} \leq \text{thb}) + \text{th1} * \text{thb} *$

Data: db

AIC	BIC	logLik
347.9881	373.3328	-162.994

Random effects:

Formula:  $\text{th0} \sim 1 \mid \text{bloco}$

	$\text{th0. (Intercept)}$	Residual
StdDev:	0.7996375	2.106687

Fixed effects:  $\text{th0} + \text{th1} + \text{thb} \sim \text{AG}$

	Value	Std.Error	DF
$\text{th0. (Intercept)}$	13.52000	1.075230	61
$\text{th0. AG50}$	2.31467	1.361115	61
$\text{th0. AG62.5}$	1.91300	1.361115	61
$\text{th1. (Intercept)}$	0.22713	0.047388	61
$\text{th1. AG50}$	-0.01080	0.052981	61
$\text{th1. AG62.5}$	0.00183	0.052981	61
$\text{thb. (Intercept)}$	49.99119	8.343789	61
$\text{thb. AG50}$	12.12745	9.999372	61
$\text{thb. AG62.5}$	43.43722	11.396492	61

	t-value	p-value
$\text{th0. (Intercept)}$	12.574059	0.0000
$\text{th0. AG50}$	1.700567	0.0941
$\text{th0. AG62.5}$	1.405465	0.1650
$\text{th1. (Intercept)}$	4.793062	0.0000
$\text{th1. AG50}$	-0.203845	0.8392
$\text{th1. AG62.5}$	0.034603	0.9725
$\text{thb. (Intercept)}$	5.991426	0.0000
$\text{thb. AG50}$	1.212821	0.2299
$\text{thb. AG62.5}$	3.811455	0.0003

Correlation:

	$t0. (I)$	$t0. AG5$	$t0. AG6$	$t1. (I)$
$\text{th0. AG50}$	-0.690			
$\text{th0. AG62.5}$	-0.690	0.545		
$\text{th1. (Intercept)}$	-0.661	0.522	0.522	
$\text{th1. AG50}$	0.591	-0.701	-0.467	-0.894
$\text{th1. AG62.5}$	0.591	-0.467	-0.701	-0.894
$\text{thb. (Intercept)}$	0.330	-0.261	-0.261	-0.875
$\text{thb. AG50}$	-0.276	0.287	0.218	0.730
$\text{thb. AG62.5}$	-0.242	0.191	0.397	0.641
	$t1. AG5$	$t1. AG6$	$\text{th. (I)}$	$t. AG50$

$\text{th0. AG50}$				
$\text{th0. AG62.5}$				
$\text{th1. (Intercept)}$				
$\text{th1. AG50}$				
$\text{th1. AG62.5}$	0.800			
$\text{thb. (Intercept)}$	0.783	0.783		
$\text{thb. AG50}$	-0.810	-0.653	-0.834	
$\text{thb. AG62.5}$	-0.573	-0.831	-0.732	0.611

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3
	-2.1121286	-0.6853820	-0.1468571	0.6643428
	Max			
	2.5729537			

Number of Observations: 74  
 Number of Groups: 5

No `summary()` do objeto tem-se as estimativas dos parâmetro de efeito fixo e de efeito aleatório, como o desvio padrão do efeito de blocos. Estudar as estimativas dos parâmetros sob a parametrização considerada é de certa forma complicada. Mais importante que isso é antes saber se existe efeito do nível de umidade em cada parâmetro. Acessa-se essa hipótese aplicando teste de Wald marginal por meio da função genérica `anova()`.

```
> anova(nn1, type="marginal")
```

	numDF	denDF	F-value
th0.(Intercept)	1	61	158.10695
th0.AG	2	61	1.60851
th1.(Intercept)	1	61	22.97344
th1.AG	2	61	0.07505
thb.(Intercept)	1	61	35.89719
thb.AG	2	61	8.25653

	p-value
th0.(Intercept)	<.0001
th0.AG	0.2086
th1.(Intercept)	<.0001
th1.AG	0.9278
thb.(Intercept)	<.0001
thb.AG	0.0007

Os resultados do teste de Wald apontam que os parâmetros  $\theta_0$  e  $\theta_1$  não dependem do nível de umidade. Sendo assim, pode-se reduzir o modelo e avaliar o impacto da redução pelo teste de razão de verossimilhanças. Deve-se estar ciente de que o teste de Wald tem comportamento dependente da parametrização usada e que o teste de razão de verossimilhanças é mais indicado. A seguinte regra pode ser adotada: usar os teste de Wald como indicador do modelo reduzido e testar a redução pelo teste de razão de verossimilhanças.

```
> nn2 <- nlme(rengrao~th0+th1*potassio*(potassio<=thb)+
+             th1*thb*(potassio>thb),
+             data=db, method="ML",
+             fixed=list(th0+th1~1, thb~AG), random=th0~1,
+             start=c(13, # th0
+                   0.22, # th1
+                   50,12,43)) # thb
> anova(nn1, nn2)
```

	Model	df	AIC	BIC	logLik
nn1	1	11	347.9881	373.3328	-162.994
nn2	2	7	346.3440	362.4724	-166.172

	Test	L.Ratio	p-value
nn1			
nn2	1 vs 2	6.355875	0.1741



Verificou-se pelo teste do modelo reduzido em relação ao modelo maior que não existe efeito de umidade sobre os parâmetros  $\theta_0$  e  $\theta_1$  e essa foi a mesma conclusão obtida na análise com efeito fixo de blocos. A título de comparar as abordagens pode-se verificar os resultados na predição. Para facilitar os cálculos a seguir será declarado o modelo com os efeitos sem a restrição de zerar o primeiro nível mas estimando os valores para categoria conforme foi feito com a `nls()`. É preciso obter a matriz gradiente, `F`, usada na obtenção das bandas de confiança. O intervalo de confiança para os valores preditos são baseados no método delta. A função `rootSolve::gradiente()` é usada para obter a matriz gradiente numericamente.

```
> fixef(nn2)
```

```

          th0          th1
14.8346154    0.2336051
thb.(Intercept)  thb.AG50
42.9787281    18.8280684
thb.AG62.5
51.1701934
```

```

> nn2 <- nlme(rengrao~th0+th1*potassio*(potassio<=thb)+
+           th1*thb*(potassio>thb),
+           data=db, method="ML",
+           fixed=list(th0+th1~1, thb~AG-1), random=th0~1,
+           start=c(15, # th0
+                 0.22, # th1
+                 30,50+12,50+43)) # thb
> fixef(nn2)
```

```

          th0          th1 thb.AG37.5 thb.AG50
14.8346154    0.2336051 42.9787281 61.8067966
thb.AG62.5
94.1489215
```

```

> numF <- function(pars, k, agua){
+   pars[1]+pars[2]*k*(k<=agua**pars[3:5])+
+   pars[2]*agua**pars[3:5]*(k>agua**pars[3:5])
+ }
> pred2 <- expand.grid(AG=levels(db$AG), potassio=seq(0,180,1))
> require(rootSolve)
> c2 <- fixef(nn2)
> F <- gradient(numF, x=c2, k=pred2$potassio,
+               agua=model.matrix(~-1+AG, pred2))
> dim(F)
```

```
[1] 543 5
```

```
> colnames(F)==names(c2)
```

```
[1] TRUE TRUE TRUE TRUE TRUE
```

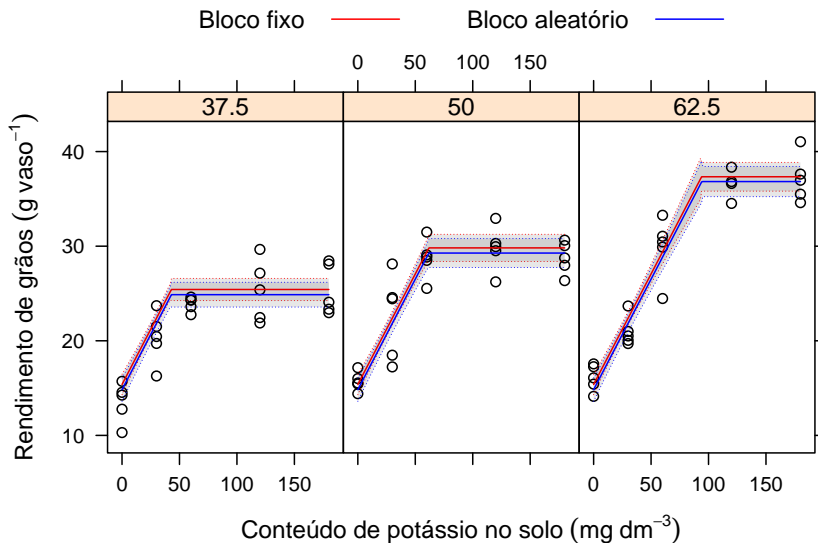


Figura 6.4: Valores observados e ajustados para a produção de soja como função da adubação com potássio e nível de umidade seguidos das bandas de confiança (95%).

```
> U <- chol(vcov(nn2))
> pred2$se <- sqrt(apply(F%*%t(U), 1, function(x) sum(x^2)))
> zval <- qnorm(p=c(lwr=0.025, fit=0.5, upr=0.975))
> me <- outer(pred2$se, zval, "*")
> pred2$rengrao <- predict(nn2, newdata=pred2, level=0)
> pred2 <- cbind(pred2, sweep(me, 1, pred2$rengrao, "+"))
> str(pred2)

'data.frame':      543 obs. of  7 variables:
 $ AG      : Factor w/ 3 levels "37.5","50","62.5": 1 2 3 1 2 3 1 2 3 1 ...
 $ potassio: num  0 0 0 1 1 1 2 2 2 3 ...
 $ se      : num  0.63 0.63 0.63 0.621 0.621 ...
 $ rengrao : atomic 14.8 14.8 14.8 15.1 15.1 ...
 ..- attr(*, "label")= chr "Predicted values"
 $ lwr     : num  13.6 13.6 13.6 13.9 13.9 ...
 $ fit     : num  14.8 14.8 14.8 15.1 15.1 ...
 $ upr     : num  16.1 16.1 16.1 16.3 16.3 ...
```

Pela figura 6.4 verifica-se diferença no valor predito entre os modelos com efeito fixo e aleatório para blocos. A amplitude das bandas de confiança foi praticamente a mesma nos dois modelos.

## Efeito da parametrização

O modelo logístico é um modelo de regressão não linear largamente usado em diversas áreas. Na área biológica é empregado principalmente para descrever o crescimento ou evolução, como crescimento de seres vivos e evolução de doenças. Devido à grande aplicação por diversas áreas, o modelo apresenta diferentes parametrizações, motivadas pela conveniente interpretação que cada uma delas dá ao fenômeno. Entretanto, a parametrização de um modelo têm um importante papel com relação a estabilidade numérica do processo de estimação e a qualidade das inferências estatísticas.

De forma simplificada, o modelo logístico tem a seguinte expressão geral

$$\text{logis}(x, \theta_a, \beta) = \frac{\theta_a}{1 + f(\exp\{x\}, \beta)} \quad (7.1)$$

em que  $x$  é a variável independente,  $\theta_a$  é assíntota do modelo que representa o valor limite da função quando  $x$  tende ao infinito, e  $f(\cdot)$  é uma função do vetor de parâmetros  $\beta$  e do exponencial de  $x$ . As reparametrizações desse modelo surgem das diferentes formas da função  $f(\cdot)$ . Será considerado aqui quatro parametrizações, a saber

- $f_a(x) = \exp\{(a_1 - x)/a_2\}$ ;
- $f_b(x) = b_1 \exp\{b_2 x\}$ ;
- $f_c(x) = \exp\{c_1 + c_2 x\}$ ;

- $f_d(x) = (-1 + 1/d_1) \exp\{-d_2x\}$ .

Em todas essas parametrizações consideramos  $\theta$  conhecido e igual a 1 simplesmente para facilitar a ilustração dos resultados e superfícies de verossimilhança. O vetor  $\beta$  com dois elementos representados por letras arábicas. Todas as expressões são reparametrizações de um mesmo modelo. Para ilustrar, considere o modelo  $f_a(\cdot)$  de onde obtemos que

$$\begin{array}{lll} b_1 = \exp\{a_1/a_2\} & c_1 = a_1/a_2 & d_1 = 1/(1 + \exp\{a_1/a_2\}) \\ b_2 = -1/a_2 & c_2 = -1/a_2 & d_2 = 1/a_2. \end{array}$$

Dessa maneira, a partir as estimativas pontuais sob qualquer parametrização podemos obter as de qualquer outra. Este é um fato importante pois podemos ajustar aos dados apenas o modelo que seja mais adequado para estimação e inferência e a partir deste obter estimativas em outras parametrizações que podem ser mais interessantes para interpretações práticas.

Para avaliar a inferência sob diferentes parametrizações, faremos o ajuste de cada uma delas à dados de incidência de ferrugem do pessegueiro, causada pelo fungo *Tranzshcelia discolor*, em função do tempo em plantas da cultivar Chimarrita (figura 7.1). O pomar foi implantado em 2004 em Curitiba (25°, 55' 10" S, 49° 57' 26" W, 945m), com espaçamento de 1m entre plantas de 2,5m entre linhas. Cada planta foi considerada como uma unidade experimental onde foram avaliadas o número de folhas com presença de ferrugem em relação ao total. Isso foi feito em 6 ramos marcados por planta durante 14 avaliações com intervalo de 10 à 16 dias entre avaliações. Para análise considerou-se a proporção observada de folhas atacadas.

Considerou-se modelo gaussiano para as observações e a função de log-verossimilhança foi implementada de forma a receber qualquer modelo de regressão não linear. Esta suposição é usual na prática embora, neste contexto uma resposta binomial poderia ser considerada. Os funções em R para os modelos foram criadas e a estimação foi feita pela `optim()`. Após estimação foram obtidos os contornos de log-verossimilhança para os pares de parâmetros mostrados na figura 7.2.

```
> ll <- function(th, y, x, model){
+   ex <- do.call(model, list(x=x, th=th))
+   sd <- sqrt(crossprod(y-ex)/length(x))
+   ll <- sum(dnorm(y, mean=ex, sd=sd, log=TRUE))
+   ll
+ }

> # parametrizações
> f.a <- function(x, th){ 1/(1+exp((th[1]-x)/th[2])) }
```

```

> f.b <- function(x, th){ 1/(1+th[1]*exp(th[2]*x)) }
> f.c <- function(x, th){ 1/(1+exp(th[1]+th[2]*x)) }
> f.d <- function(x, th){ 1/(1+(-1+1/th[1])*exp(-th[2]*x)) }
> # dados
> y <- dados$inc2; x <- dados$dia
> # lista com valores iniciais e modelo
> init.list <- list(A=list(par=c(80,13), model=f.a),
+                  B=list(par=c(120,-0.06), model=f.b),
+                  C=list(par=c(5,-0.06), model=f.c),
+                  D=list(par=c(0.008, 0.065), model=f.d))
> # lista com termos fixos durante otimização
> fixed.list <- list(fn=ll, x=x, y=y, method="BFGS",
+                  control=list(fnscale=-1))
> # otimização em série dos modelos
> op.all <-
+   lapply(init.list,
+         function(i){
+           op <- do.call(optim, c(i, fixed.list)); op
+         })
> # estimativas dos parâmetros
> pars <- sapply(op.all, "[", "par"); pars

           A           B           C
[1,] 73.12710 120.01001195  4.81592735
[2,] 15.23597 -0.06548208 -0.06584788
           D
[1,] 0.008455968
[2,] 0.065166770

> # log-verossimilhança dos modelos
> ll0 <- sapply(op.all, "[", "value"); ll0

           A           B           C           D
57.13773 57.13735 57.13700 57.13430

> # contornos de log-verossimilhança
> leng <- 100
> grid.a <- expand.grid(th1=seq(67,79,l=leng), th2=seq(11,20,l=leng))
> grid.a$ll <- apply(grid.a, 1, ll, y=y, x=x, model=f.a)
> grid.a$dev <- with(grid.a, -2*(ll-ll0["A"]))
> levels <- c(0.05,0.5,0.75,0.9,0.95,0.99)
> qchi <- qchisq(levels, df=2)
> require(lattice)
> contourplot(dev~th1+th2, data=grid.a, at=qchi,
+            labels=list(labels=as.character(levels), cex=0.8))
>

```

Verifica-se que na parametrização A os contornos de deviance são praticamente elípticos com eixos paralelos aos eixos cartesianos. Do ponto de vista de otimização, o procedimento baseado em avaliação de gradiente tem melhor taxa de convergência quando a função objetivo apresenta simetria e ortogonalidade. Do ponto de vista de inferência estatística, por exemplo na obtenção de intervalos de confiança, a ortogonalidade permite fazer inferência para um parâmetro sem a necessidade de correções considerando a curvatura na direção dos demais parâmetros.

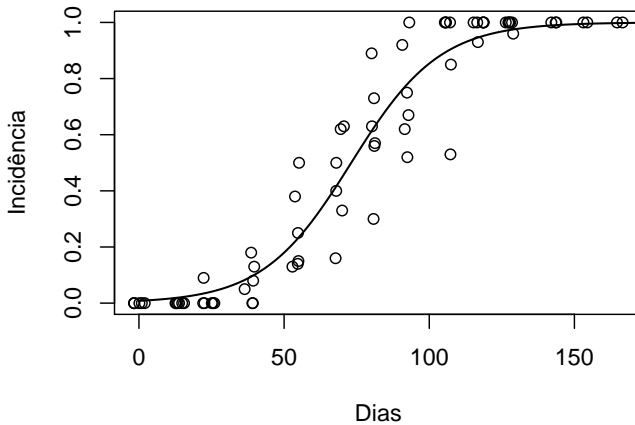


Figura 7.1: Valores observados e ajustados de incidência em função do tempo. As parametrizações são iguais em termos de valores ajustados.

A parametrização *C*, em comparação com *A*, apresentou rotação dos eixos do contorno com inclinação negativa e uma leve assimetria dos contornos. A rotação deve-se ao fato de que as funções de reparametrização entre *A* e *C* são praticamente lineares e a assimetria devido uma delas ser um quociente. As parametrizações *B* e *D* apresentaram forte desvio da forma elíptica com contornos em formato de "banana". Dentre esses últimos, o modelo *B* parece ser o mais susceptível a problemas numéricos pois apresenta assimetria acentuada. Dessa forma, as parametrizações em termos de conveniência para estimação e inferência estatística seguem a ordem de preferência: *A*, *C*, *D* e *B*. Vale relembrar que tanto as estimativas pontuais quanto intervalares do modelo em uma parametrização podem levar aos seus equivalentes em outra para metrização pelo princípio da invariância do estimador de máxima verossimilhança.

```
> grid.b <- expand.grid(th1=seq(30,700,l=leng),
+                       th2=seq(-0.09,-0.05,l=leng))
> grid.b$ll <- apply(grid.b, 1, ll, y=y, x=x, model=f.b)
> grid.b$dev <- with(grid.b, -2*(ll-ll0["B"]))
> grid.c <- expand.grid(th1=seq(3.5,6.7,l=leng),
+                       th2=seq(-0.09,-0.05,l=leng))
> grid.c$ll <- apply(grid.c, 1, ll, y=y, x=x, model=f.c)
> grid.c$dev <- with(grid.c, -2*(ll-ll0["C"]))
```

```

> grid.d <- expand.grid(th1=seq(0.0005,0.027,l=leng),
+                       th2=seq(0.049,0.09,l=leng))
> grid.d$ll <- apply(grid.d, 1, ll, y=y, x=x, model=f.d)
> grid.d$dev <- with(grid.d, -2*(ll-ll0["D"]))
> grid.full <- rbind(grid.a, grid.b, grid.c, grid.d)
> grid.full$para <- rep(c("A","B","C","D"), each=leng^2)
> expr <- c(expression(f[a](x) == 1/(1+exp((a[1]-x)/a[2]))),
+           expression(f[b](x) == 1/(1+b[1]*exp(b[2]*x))),
+           expression(f[c](x) == 1/(1+exp(c[1]+c[2]*x))),
+           expression(f[d](x) == 1/(1+(-1+1/d[1])*exp(-d[2]*x))))

```

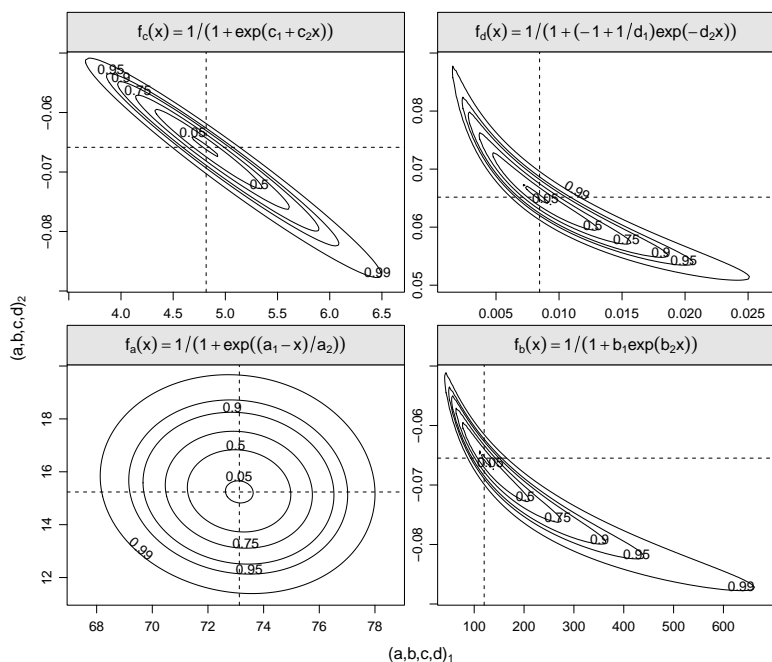


Figura 7.2: Contornos de *deviance* para os parâmetros do modelo logístico em cada uma das parametrizações.





## Modelagem da variância

A goiabeira pertence à família Mirtaceae, gênero *Psidium*, espécie *P. guajava*, tendo como origem a América Tropical. Assim como a maioria dos frutos carnosos a goiaba apresenta uma fase de crescimento inicial caracterizada por intensa divisão celular, seguida por um período de expansão celular. Após esta fase, ocorre redução drástica na taxa de aumento do volume ou massa dos frutos, seguindo-se a fase final, na qual a taxa de crescimento praticamente cessa. Ao final da fase de crescimento, inicia-se a fase de maturação, os frutos atingem o estágio de maturidade fisiológica. A senescência corresponde aos processos que ocorrem após a maturidade fisiológica.

No ajuste de modelos em geral, o método dos mínimos quadrados ordinários (MQO) é a técnica de estimação mais comumente aplicada em situações de distribuição normal independente e homogênea dos erros aleatórios. Entretanto, o estimador de mínimos quadrados, apesar de permanecer não viesado, pode se tornar muito ineficiente sob condições nas quais os erros não apresentam homogeneidade de variância. Sob tais circunstâncias, a estimação pelo método dos mínimos quadrados ponderados (MQP) deve ser considerada. Contudo, a maior dificuldade associada a aplicação do MQP é, na prática, a determinação da ponderação, ou função, que represente o comportamento da variância dos erros. Normalmente se torna conveniente associar a variância à mesma variável independente usada para descrever a média. Essas funções para variância são usadas para contemplar as estruturas de heterocedasticidade por associação às covariáveis observadas, resultando em maior esclarecimento sobre o fenômeno

observado.

Nesse capítulo o objetivo foi modelar o crescimento em massa (g) de frutos de goiaba por meio de modelos não lineares com modelagem da variância.

Frutos de goiabeira “Pedro Sato” (*Psidium guajava* L.) foram colhidos em um pomar comercial em plena produção, com 9 anos de idade, no município de Paula Cândido, MG. Flores foram marcadas após a antese e o desenvolvimento dos frutos foi acompanhado até a sua senescência. Durante a fase de formação e maior crescimento (entre 06/09/2007 a 05/12/2007), os frutos foram colhidos a cada 15 dias e submetidos às análises físicas feitas no laboratório de fruticultura, do Departamento de Fitotecnia, da Universidade Federal de Viçosa. Nesse estudo foi avaliado a massa da matéria fresca (g) dos frutos obtido pela determinação em balança de 15 frutos colhidos, a cada data, aleatoriamente de quatro plantas clones, sendo que a última coleta teve apenas 9 frutos.

Para a modelar a massa fresca ( $mf$ ) de frutos de goiaba em função dos dias após a antese ( $daa$ ) considerou-se, após avaliar outros modelos e parametrizações, aquele obtido a partir da manipulação do modelo Gompertz. O modelo final pode ser escrito como

$$mf = A - (A - B) * \exp(-\exp(T \cdot (daa - I))) \quad (8.1)$$

no qual  $A$  representa a assíntota (massa final do fruto),  $B$  é o intercepto,  $T$  é proporcional a taxa de crescimento e  $I$  é o tempo ( $daa$ ) em que ocorre a taxa máxima de crescimento (inflexão).

Para modelar a variância dos erros aleatórios ( $\text{Var}(\epsilon)$ ) considerou-se funções potência e exponencial da covariável, sendo a esta inserida na escala original e na escala logarítmica, resultando em quatro modelos

$$\text{Var}(\epsilon) = \begin{cases} g_0(\sigma) = \sigma^2 & (\text{constante}) \\ g_1(daa, \delta, \sigma) = \sigma^2 \cdot \exp\{2\delta \cdot daa\} & (\text{exponencial}) \\ g_2(daa, \delta, \sigma) = \sigma^2 \cdot |daa|^{2\delta} & (\text{potncia}) \\ g_3(daa, \delta, \sigma) = \sigma^2 \cdot |\ln daa|^{2\delta} & (\text{potncia do logaritmo}) \end{cases} \quad (8.2)$$

Trabalhou-se no programa R por meio da função *gnls* do pacote *nlme*, em que a especificação da função de variância foi feita por meio do argumento *weights*. Foi empregado o método de estimação por máxima verossimilhança restrita (REML) e após obtenção das estimativas foram realizados os diagnósticos sobre os resíduos padronizados e calculada a razão de verossimilhança entre os modelos encaixados para testar a significância da inclusão da função de variância ( $H_0 : \delta = 0$ ).

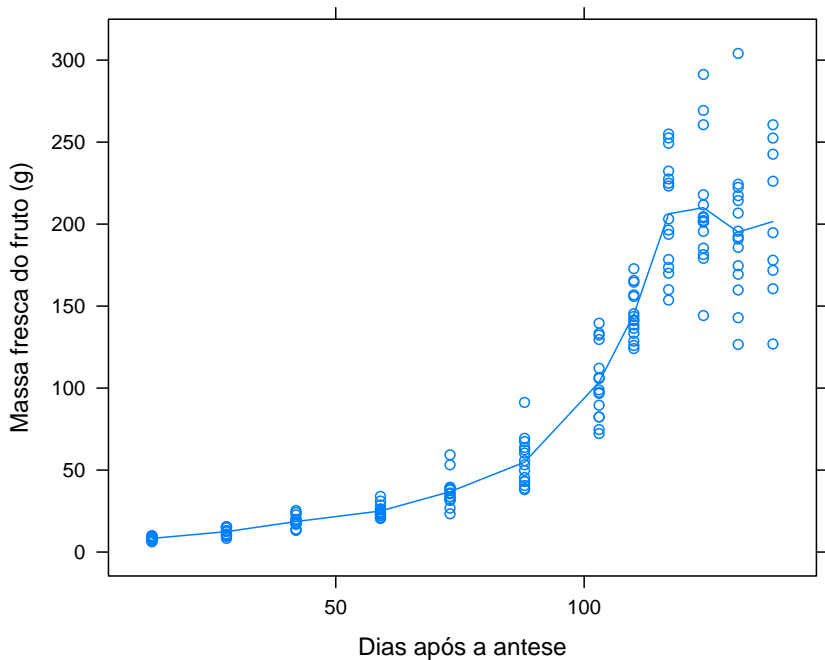


Figura 8.1: Massa fresca de frutos de goiaba em função dos dias após antese.

```
> require(lattice)
> require(latticeExtra)
> goi <- read.table("http://www.leg.ufpr.br/~walmes/data/goiaba.txt",
+                  header=TRUE, sep="\t")
> str(goi)

'data.frame':      174 obs. of  7 variables:
 $ daa   : int  13 13 13 13 13 13 13 13 13 13 ...
 $ coleta: int  1 1 1 1 1 1 1 1 1 1 ...
 $ rep   : int  1 2 3 4 5 6 7 8 9 10 ...
 $ long  : num  29.1 32.9 36.3 34.7 33.3 ...
 $ trans : num  20.5 22.8 22.9 22.7 21.4 ...
 $ peso  : num  6.66 8.51 9.93 9.14 7.6 7.54 6.27 9.94 7.95 7.79 ...
 $ volume: int  6 9 11 10 7 8 6 9 8 8 ...

> xyplot(peso-daa, data=goi, type=c("p", "a"))
```

Por meio da função `nls()` ajusta-se o modelo ordinário, ou seja, que

supoe variância homogênea. Com a função `gnls()` (*generalized nonlinear least squares*) modelos a variância por meio de funções passadas para o argumento `weights =`. A `nls()` e `lm()` possuem esse argumento que nestas recebem um vetor de pesos enquanto que na `gnls()` é passado uma função que representa o comportamento da variância como função de covariáveis ou da média.

```
> # modelo ordinário
> m0 <- nls(peso~A-(A-D)*exp(-exp(C*(daa-B))), data=goi,
+          start=c(A=200, C=0.09, B=105, D=7))
> summary(m0)
> require(nlme)
> # var(e)=sigma^2*|daa|^{2*delta}
> m1 <- gnls(peso~A-(A-D)*exp(-exp(C*(daa-B))), data=goi,
+           start=c(A=225, C=0.05, B=109, D=7),
+           weights=varPower(0.1, form=~daa))
> summary(m1)
> # var(e)=sigma^2*|ln{daa}|^{2*delta}
> m2 <- gnls(peso~A-(A-D)*exp(-exp(C*(daa-B))), data=goi,
+           start=c(A=225, C=0.05, B=109, D=7),
+           weights=varPower(0.1, form=~log(daa)))
> summary(m2)
> # var(e)=sigma^2*exp(2*delta*daa)
> m3 <- gnls(peso~A-(A-D)*exp(-exp(C*(daa-B))), data=goi,
+           start=c(A=225, C=0.05, B=109, D=7),
+           weights=varExp(0.03, form=~daa))
> summary(m3)
```

Observou-se que o modelo com variância constante ( $g_0$ ) estimou  $A$  e  $B$  com maior precisão enquanto que estimou com baixa precisão  $C$  e  $D$ . Alta precisão de  $g_0$  para  $A$  parece ser um equívoco resultante da negligência de heterocedasticidade sugerida por esse modelo, uma vez que a análise gráfica exploratória apontou que erro aleatório tem alta variância a medida que se aproxima da assíntota. Para os modelos que consideram a heterogeneidade dos erros a precisão de  $A$  foi menor justamente porque considera a variabilidade na região da curva associada a tal parâmetro.

Em termos gerais a modelagem da variância tornou mais precisa a estimativa dos parâmetros associados ao início da curva ( $C$  e  $D$ ) enquanto que tornou menos precisas as estimativas dos parâmetros associados ao final da curva ( $A$  e  $B$ ). Tal fato está relacionado a heterogeneidade de variância entre essas regiões, que por ser modelada influenciou a precisão dos parâmetros de uma forma proporcional à variância na região em que o parâmetro é estimado.

```
> summary(m0)$coeff[,1:2]
```

```
      Estimate Std. Error
A 205.76226746  4.19618287
C   0.09366616  0.01014825
```

```
B 107.35178155 0.87438902
D 18.70608429 3.17907984
```

```
> summary(m1)$tTable[,1:2]
```

```
      Value      Std.Error
A 225.63349158 12.091087071
C  0.04936564  0.002928354
B 109.96894519  2.498713907
D   6.74211939  0.529455595
```

```
> summary(m2)$tTable[,1:2]
```

```
      Value      Std.Error
A 218.30945857  9.544151359
C  0.05236619  0.003336532
B 108.28252562  2.078244660
D   7.10495060  0.485459609
```

```
> summary(m3)$tTable[,1:2]
```

```
      Value      Std.Error
A 262.51047190 28.861488254
C  0.04446163  0.002576577
B 117.32301180  4.541362028
D   6.64837934  0.721373801
```

```
> # valor das log verossimilhanças
```

```
> c(logLik(m0), logLik(m1), logLik(m2), logLik(m3))
```

```
[1] -804.5465 -703.0136 -710.1183 -695.1840
```

```
> # comparações de log verossimilhança
```

```
> anova(m1, m0); anova(m2, m0); anova(m3, m0)
```

```
      Model df      AIC      BIC    logLik
m1      1  6 1418.027 1436.982 -703.0136
m0      2  5 1619.093 1634.888 -804.5465
      Test L.Ratio p-value
m1
m0 1 vs 2 203.0659 <.0001
```

```
      Model df      AIC      BIC    logLik
m2      1  6 1432.237 1451.191 -710.1183
m0      2  5 1619.093 1634.888 -804.5465
      Test L.Ratio p-value
m2
m0 1 vs 2 188.8564 <.0001
```

```
      Model df      AIC      BIC    logLik
m3      1  6 1402.368 1421.322 -695.1840
m0      2  5 1619.093 1634.888 -804.5465
      Test L.Ratio p-value
m3
m0 1 vs 2 218.725 <.0001
```

Dentre as funções para a variância o modelo  $g_3$  foi o que apresentou maior razão de verossimilhança quando comparado com o modelo  $g_0$ . Tal resultado é indicativo de que a heterogeneidade de variância não foi satisfatoriamente corrigida. O gráfico de quantil quantil apontou forte evidência sobre a não adequabilidade do ajuste com  $g_0$ , enquanto que não apresentou evidências contra a hipótese de normalidade para os resíduos dos modelos  $g_2$  e  $g_3$  (figura 8.2). Pelos gráficos dos resíduos padronizados em função dos valores ajustados verifica-se uma disposição de pontos na forma de corneta para os modelo  $g_0$  e uma distribuição mais constante, apontando não haver relação entre média e variância (figura 8.3)

```
> # gráfico dos resíduos
> residuos <- expand.grid(daa=goi$daa,
+                         modelo=c("Constante", "Potência",
+                                 "Potência do logaritmo", "Exponencial"))
> residuos$respad=c(residuals(m0, type="p"), residuals(m1, type="p"),
+                  residuals(m2, type="p"), residuals(m3, type="p"))
> residuos$fitted=c(fitted(m0), fitted(m1), fitted(m2), fitted(m3))
> qqmath(~respad|modelo, data=residuos)
> xyplot(respad~daa|modelo, data=residuos)

> model <- deriv3(~ A-(A-D)*exp(-exp(C*(daa-B))), c("A","C","B","D"),
+               function(daa, A, C, B, D){NULL})
> U <- chol(vcov(m0))
> pred <- data.frame(daa=seq(0,140,1))
> m <- model(daa=pred$daa, A=coef(m0)["A"], C=coef(m0)["C"],
+          B=coef(m0)["B"], D=coef(m0)["D"])
> F0 <- attr(m, "gradient")
> pred$y <- c(m)
> pred$se <- sqrt(apply(F0%*t(U), 1, function(x) sum(x^2)))
> z <- qnorm(0.975)
> pred <- transform(pred, lwr=y-z*se, upr=y+z*se, model="Constante")
> str(pred)

'data.frame':      141 obs. of  6 variables:
 $ daa  : num  0 1 2 3 4 5 6 7 8 9 ...
 $ y    : num  18.7 18.7 18.7 18.7 18.7 ...
 $ se   : num  3.17 3.17 3.17 3.17 3.17 ...
 $ lwr  : num  12.5 12.5 12.5 12.5 12.5 ...
 $ upr  : num  24.9 24.9 24.9 24.9 24.9 ...
 $ model: Factor w/ 1 level "Constante": 1 1 1 1 1 1 1 1 1 1 ...

> # fazer o mesmo para os demais modelos...
```

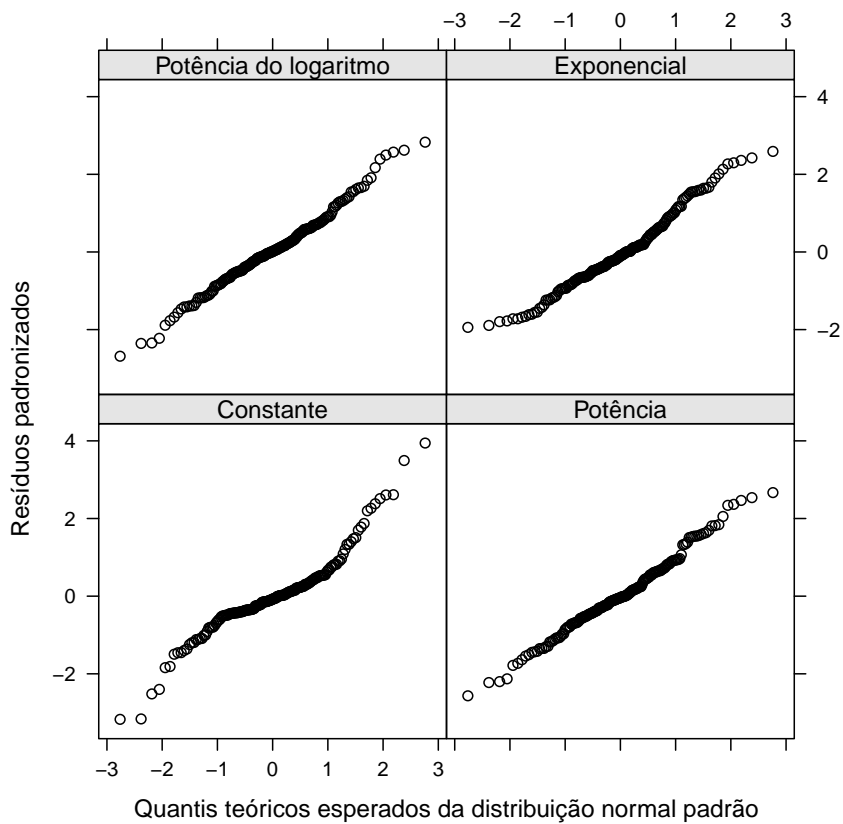


Figura 8.2: Gráfico de quantil-quantil normal para os resíduos dos modelos com diferentes funções para a variância.

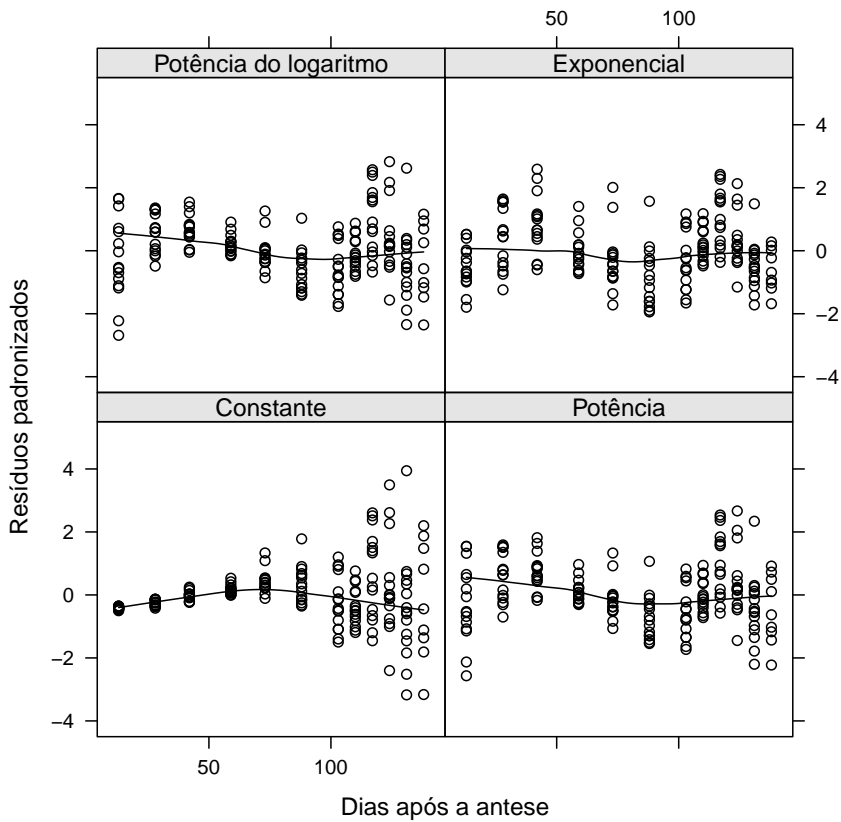


Figura 8.3: Resíduos padronizados em função dos valores ajustados para os modelos com diferentes funções para a variância.



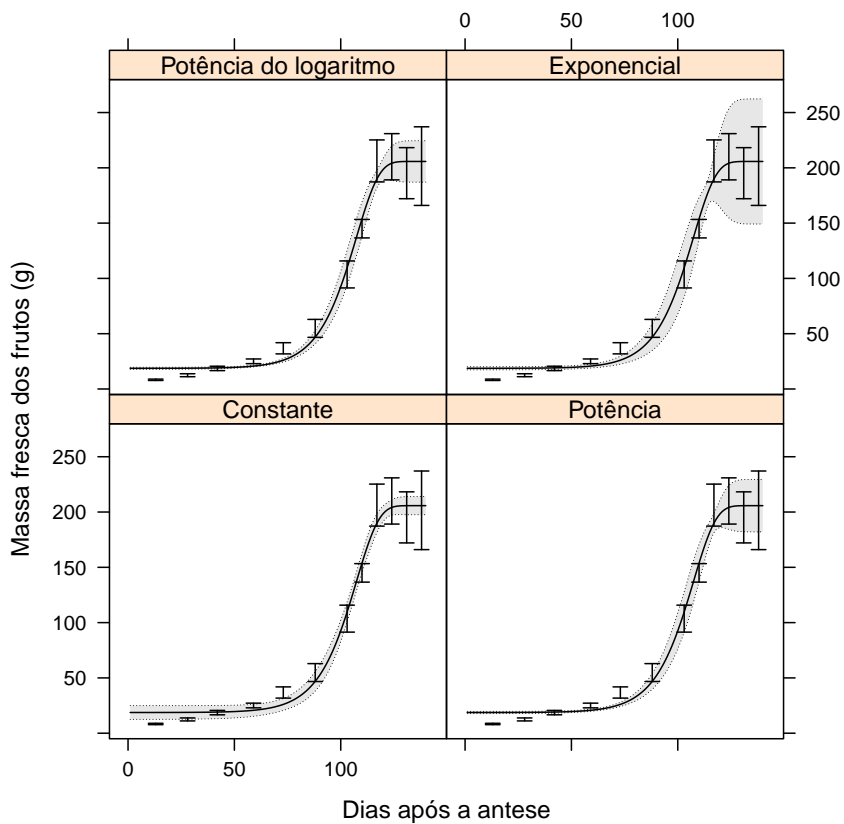


Figura 8.4: Valores ajustados acompanhados da banda de confiança para o valor predito. Segmentos representam o intervalo de confiança para a média amostral para cada dia de coleta.



# Bibliografia

BAS, D.; BOYACI, I. H. Modeling and optimization I: Usability of response surface methodology. **Journal of Food Engineering**, v. 78, n. 3, p. 836–845, fev. 2007. ISSN 02608774. Disponível em: <<http://dx.doi.org/10.1016/j.jfoodeng.2005.11.024>>.

BATES, D.; WATTS, D. Relative curvature measures of non-linearity. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 42, n. 1, p. 1–25, jan. 1980. Disponível em: <<http://www.jstor.org/stable/10.2307/2984733>>.

BLEASDALE, J. K. A.; NELDER, J. A. Plant Population and Crop Yield. **Nature**, v. 188, n. 4747, p. 342–342, out. 1960. ISSN 0028-0836. Disponível em: <<http://dx.doi.org/10.1038/188342a0>>.

BONAT, W. H.; Ribeiro Jr, P. J.; ZEVIANI, W. M. Regression models with responses on the unity interval: specification, estimation and comparison. **Biometrical Brazilian Journal**, v. 30, p. 415–431, 2012.

FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v. 31, n. 7, p. 799–815, 2004.

KRANZ, J. (Ed.). **Epidemics of Plant Diseases**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990. (Ecological Studies, v. 13). ISBN 978-3-642-75400-5. Disponível em: <<http://www.springerlink.com/index/10.1007/978-3-642-75398-5>>.

MICHAELIS, L.; MENTEN, M. Die kinetik der invertinwirkung. **Biochemische zeitung**, v. 49, p. 333–369, 1913. Disponível em: <<http://www.citeulike.org/user/oliverguntli/article/5936552>>.

PEEK, M.; RUSSEK-COHEN, E.; WAIT, A.; FORSETH, I. Physiological response curve analysis using nonlinear mixed models. **Oecologia**, v. 132, n. 2, p. 175–180, jul. 2002. ISSN 0029-8549. Disponível em: <<http://link.springer.com/article/10.1007/s00442-002-0954-0>>.

Pimentel Gomes, F. **Curso de estatística experimental**. 15. ed. Piracicaba: FEALQ, 2009. 451 p. ISBN 978-85-7133-055-9.

RITZ, C.; STREIBIG, J. C. **Nonlinear Regression with R**. 1. ed. Springer, 2008. ISBN 0387096159. Disponível em: <<http://www.amazon.com/dp/0387096159>>.

SCHABENBERGER, O.; PIERCE, F. J. **Contemporary Statistical Models for the Plant and Soil Sciences**. CRC Press, 2001. 738 p. ISBN 1584881119. Disponível em: <<http://www.amazon.com/Contemporary-Statistical-Models-Plant-Sciences/dp/1584881119>>.

SCHABENBERGER, O.; PIERCE, F. J. **Contemporary Statistical Models for the Plant and Soil Sciences**. Boca Raton: CRC Press, 2002. 738 p. ISBN 1584881119. Disponível em: <<http://www.amazon.com/Contemporary-Statistical-Models-Plant-Sciences/dp/1584881119>>.

SEGARRA, J.; JEGER, M. J.; BOSCH, F. van den. Epidemic dynamics and patterns of plant diseases. **Phytopathology**, The American Phytopathological Society, v. 91, n. 10, p. 1001–10, out. 2001. ISSN 0031-949X. Disponível em: <<http://apsjournals.apsnet.org/doi/abs/10.1094/PHYTO.2001.91.10.1001>>.

SERAFIM, M. E.; ONO, F. B.; ZEVIANI, W. M.; NOVELINO, J. O.; SILVA, J. V. Umidade do solo e doses de potássio na cultura da soja. **Revista Ciência Agronômica**, Fortaleza, v. 43, n. 2, p. 222–227, jul. 2012. ISSN 1806-6690. Disponível em: <<http://ccarevista.ufc.br/seer/index.php/ccarevista/article/view/1454>>.

VENABLES, W. N.; DICHMONT, C. M. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. **Fisheries research**, v. 70, p. 319–337, 2004.

ZEVIANI, W. M. **Parametrizações interpretáveis em modelos não linear**. 146 p. Tese (Tese de Doutorado) — Universidade Federal de Lavras, 2013.

ZEVIANI, W. M.; SILVA, C. A.; CARNEIRO, W. J. d. O.; MUNIZ, J. A. Modelos não lineares para a liberação de potássio de esterco animais em latossolos. **Ciência Rural**, v. 42, n. 10, p. 1789–1796, out. 2012. ISSN 0103-8478. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-84782012001000012&lng=en&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782012001000012&lng=en&nrm=iso&tlng=pt)>.