



Universidade Estadual da Paraíba  
Centro de Ciências e Tecnologia  
Departamento Estatística

**André Luiz Pinto dos Santos**

# **Estudo de modelos de crescimento via dados simulados**

Campina Grande  
Dezembro de 2011

André Luiz Pinto dos Santos

# **Estudo de modelos de crescimento via dados simulados**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador:

Tiago Almeida de Oliveira

Co-orientadora:

Ana Patricia Bastos Peixoto

Campina Grande

Dezembro de 2011

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

S237e

Santos, André Luiz Pinto dos.

Estudo de modelos de crescimento via dados simulados [manuscrito] / André Luiz Pinto dos Santos. – 2011.

33 f.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2011.

“Orientação: Prof. Dr. Tiago Almeida de Oliveira, Departamento de Estatística”.

1. Estatística. 2. Curvas de crescimento. 3. Seleção de modelos. I. Título.

21. ed. CDD 519.5

André Luiz Pinto dos Santos

## Estudo de modelos de crescimento via dados simulados

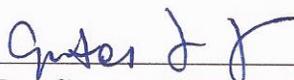
Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 02/12/11

### Banca Examinadora:



Prof. Msc. Tiago Almeida de Oliveira  
Orientador



Prof. Dr. Gustavo Henrique Esteves  
Universidade Estadual da Paraíba



Profa. Msc. Ana Patricia Bastos Peixoto  
Universidade Estadual da Paraíba



Prof. Msc. Ricardo Alves de OLinda  
Universidade Estadual da Paraíba

# Dedicatória

Dedico este trabalho aos meus pais Luiz Carlos e Ilma e as minhas irmãs Karla e Patrícia

# Agradecimentos

Em primeiro lugar gostaria de agradecer a Deus, por ter me dado a capacidade para realizar mais um sonho.

A toda a minha família: meu pai Luiz Carlos, minha mãe Ilma e minhas irmãs Karla e Patrícia, que foram e continuarão sendo o meu porto seguro. Obrigado por tudo o que fizeram e continuam fazendo por mim, nunca teria chegado aqui sem o amor e carinho de todos vocês.

Aos professores, especialmente a minha orientadora Ana Patricia Bastos Peixoto e ao meu co-orientador Tiago Almeida de Oliveira pelo empenho, paciência e credibilidade em tão pouco tempo e pela confiança em mim depositada, obrigado por tudo.

Obrigado a todos os professores responsáveis pelos meus estudos aqui na UEPB: Ana Cristina, Elizangela, Ruth Silveira, Diana Maia, Kátia Elizabete, Juarez, Gil, Ricardo Alves, Edwirde, que sempre estiveram de portas abertas para me ajudar.

Gostaria de agradecer também a todos os meus colegas de graduação: Allana Livia, Bárbara Comboim, Carol Gonçalves, Djair Durand, Hudson, Tamyres Aline, Jaiane, Janeide, Joaquim Braga, Maxwendell Aires e que jamais teria terminado sem vocês do meu lado.

# Resumo

Curvas de crescimento são ferramentas para prever o potencial fenotípico de um animal. Estas permitem a previsão do desempenho de características de importância econômica como peso corporal. Neste trabalho, o objetivo foi ajustar curvas de crescimento em dados que simulam os pesos de suínos ao longo dos dias. Para isso, foram utilizados 2.025 indivíduos de duas famílias em cinco momentos distintos (0, 132, 265, 397, 560) dias. Os dados foram submetidos ao ajuste dos modelos von Bertalanffy, Logístico e Gompertz. O método utilizado para o ajuste foi o iterativo Gauss-Newton, por meio da técnica dos modelos não-lineares utilizando-se o procedimento nls do programa computacional R. Os parâmetros estimados foram  $\alpha$  (valor assintótico),  $\beta$  (maturidade do animal),  $k$  (taxa de maturação) e  $m$  (ponto de inflexão). A interpretação biológica dos parâmetros, o coeficiente de determinação, o quadrado médio dos resíduos, o intervalo de confiança a 95% das estimativas, o AIC (Informação de Akaike) e o BIC (Informação de Bayes), foram utilizados como critérios para a escolha dos modelos que melhor se ajustaram aos dados simulados. Dentre os modelos ajustados o modelo Logístico foi o que apresentou melhor ajuste.

**Palavras-chaves:** Curva de crescimento, modelo não-linear, seleção de modelos.

# Abstract

Growth curves are potential tools to predict the phenotype of an animal. These allow the prediction of the performance characteristics of economic importance such as body weight. In this work, the goal was to adjust data in growth curves that simulate the weights of cattle throughout the day. For this, we used 2,025 individuals from two families in five different times (0, 132, 265.397, 560). The data were submitted by the models to fit von Bertalanffy, Gompertz and Logistic. The method was used to adjust the iterative Gauss-Newton, using the technique of nonlinear models using the procedure nls computer program R. The estimated parameters were  $\alpha$  (asymptotic value),  $\beta$  (maturity of the animal),  $k$  (maturing rate) in (inflection point). The biological interpretation of parameters, the coefficient of determination, mean square of the waste, the confidence interval 95% of the estimates, the AIC (Akaike Information) and BIC (Bayesian Information) were used as criteria for the choice of models that best fit the data. Among the models fitted the logistic model was presented the best adjustment.

**Key-words:** Growth curve, non-linear model, model selection.

# Sumário

Lista de Figuras

Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 11
<b>2</b>	<b>Fundamentação Teórica</b>	p. 12
2.1	Modelos de regressão linear . . . . .	p. 12
2.2	Modelos de regressão não-linear . . . . .	p. 14
2.2.1	Modelos de regressão intrinsecamente lineares . . . . .	p. 14
2.2.2	Modelos de regressão intrínsecamente não-lineares . . . . .	p. 15
2.2.2.1	Método de estimação por mínimos quadrados . . . . .	p. 16
2.2.2.2	Aproximação linear . . . . .	p. 18
2.2.2.3	O método de Gauss-Newton . . . . .	p. 19
2.2.2.4	Valores iniciais . . . . .	p. 20
2.3	Curvas de crescimentos . . . . .	p. 21
2.3.1	CrITÉrios para seleção de modelos . . . . .	p. 23
<b>3</b>	<b>Aplicação</b>	p. 25
<b>4</b>	<b>Conclusão</b>	p. 30
<b>5</b>	<b>Referências Bibliográficas</b>	p. 31

# Lista de Figuras

1	Relação entre o peso dos animais ao longo dos dias. . . . .	p. 26
2	Representação gráfica do ajuste do modelo Logístico. . . . .	p. 27
3	Representação gráfica dos resíduos versus dias . . . . .	p. 29
4	Representação gráfica dos resíduos versus dias. . . . .	p. 29

# Lista de Tabelas

1	Derivadas parciais dos modelos (2.19), (2.20) e (2.21) em função do peso assintótico ( $\alpha$ ), da constante de integração ( $\beta$ ), da taxa de maturação (K). . . . .	p. 25
2	Estimativas de peso assintótico ( $\alpha$ ), maturidade do animal ao nascimento ( $\beta$ ), taxa de maturação (k), o erro padrão (E. Pad), Parâmetro (Par.), Estimativas (Estim.) e seus respectivos intervalos de confiança (Min. IC e Max. IC) . . . . .	p. 28
3	Estatísticas da qualidade de informação de ajuste, valores da soma de quadrado dos resíduos (SQR) e coeficiente de determinação ( $R^2$ ), AIC (Informação de Akaike) e BIC (Informação de Bayes). . . . .	p. 28

# 1 Introdução

A análise de regressão tem o objetivo de verificar a existência de uma relação funcional significativa entre uma variável com uma ou mais variáveis, obtendo-se uma equação que explique a variação da variável dependente em relação aos níveis das variáveis independentes. Gujarati (2006) afirma que alguns modelos podem parecer não-lineares nos parâmetros, mas são inerente ou intrinsecamente lineares pois com as devidas transformações, podem-se tornar modelos de regressão lineares com novos parâmetros.

Na prática, um modelo não-linear é linearizado para facilitar a obtenção das estimativas dos parâmetros (BATES e WATTS, 1988). O inconveniente de uma transformação é que além do parâmetro perder sua interpretação intrínseca, pode-se alterar a estrutura e a distribuição do erro. Ou seja, se os erros do modelo original satisfazem às suposições usuais de normalidade, independência e homogeneidade da variância, os erros do novo modelo, em geral, não satisfarão tais suposições porque tais transformações podem mudar a forma da distribuição dos mesmos ou fazer com que a variância desta distribuição deixe de ser constante (CURRIE, 1982).

Segundo Oliveira (1995), modelos matemáticos não-lineares, desenvolvidos empiricamente, têm se mostrado adequados para descrever curvas de crescimento. Em bovinos, os cinco mais utilizados são: Brody, Von Bertalanffy, Richards, Logístico e Gompertz. Além dos parâmetros que são estimados diretamente do modelo e que apresentam interpretação biológica, podem ser obtidos também os parâmetros genéticos referentes aos indivíduos, como as herdabilidades e correlações, as quais podem ser utilizadas em programas de melhoramento animal.

O objetivo deste trabalho foi obter estimativas para os parâmetros dos modelos não-lineares de curva de crescimento Logístico, Gompertz e von Bertalanffy através de métodos numéricos, compará-los entre si e identificar o modelo que melhor se ajusta ao padrão de crescimento utilizando para isso dados simulados.

## 2 Fundamentação Teórica

### 2.1 Modelos de regressão linear

A análise de regressão é uma das técnicas de estimação mais usadas na estatística, com aplicações na indústria, economia, estudos biológicos, entre outros. Um dos objetivos do modelo de regressão é explicar a existência da relação funcional de uma ou mais variáveis de interesse em função de outras variáveis explicativas. Sendo  $X$  a variável explicativa (independente) e  $Y$  a variável resposta (dependente). Outros objetivos da análise de regressão são estimação de parâmetros, predição e controle, assim é possível determinar como duas ou mais variáveis se relacionam, além de estimar a função que determina a relação entre as variáveis e usar a equação ajustada para prever valores da variável dependente.

Com base numa amostra, a equação estimada não responde ao pesquisador se a variação da variável independente  $X$  influencia significativamente na variação da variável dependente  $Y$ . Ela apenas estabelece uma relação funcional não determinística entre a variável  $Y$  e  $X$ , para mostrar claramente o fenômeno em estudo (SOUZA, 2008).

É conveniente realizar um teste estatístico para estimativas dos coeficientes da equação de regressão estimada, por meio deste teste é possível quantificar quanto da variação da variável independente influencia significativamente a variação da variável dependente, neste caso, o teste  $F$  da análise da variância proposto por Fisher e Snedecor é o mais utilizado (BATES e WATTS, 1988).

A regressão linear fornece estimativas para o vetor de parâmetros  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  no modelo:

$$\mathbf{Y}_n = \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \varepsilon_n = (x_{n1}, x_{n2}, \dots, x_{np})\boldsymbol{\beta} + \varepsilon_n \quad (2.1)$$

em que a variável aleatória  $Y_n$ , representa a resposta  $n$ ,  $n = 1, 2, \dots, N$ , tem uma parte determinística,  $(x_{n1}, x_{n2}, \dots, x_{np})\boldsymbol{\beta}$ , depende das variáveis regressoras e dos parâmetros e

uma parte estocástica, representada pela variável aleatória  $\varepsilon_n$ , que é o erro aleatório. O modelo linear (2.1) pode ser escrito na forma matricial como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.2)$$

na qual  $\mathbf{Y}$  é o vetor das variáveis aleatórias representando os dados observados,  $\mathbf{X}$  é a matriz de dimensões  $N \times P$ , das variáveis regressoras e tem posto coluna completo, sendo  $P$  o número de parâmetros e  $N$  o número de observações do modelo

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1P} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2P} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{NP} \end{bmatrix}$$

$\boldsymbol{\beta}$  é o vetor dos parâmetros com dimensão  $P \times 1$ , e  $\boldsymbol{\varepsilon}$  é o vetor das variáveis aleatórias não observáveis de dimensão  $N \times 1$ . O modelo matemático para as respostas pode ser obtido através da parte determinística,  $\mathbf{X}\boldsymbol{\beta}$ , juntamente com uma função de parâmetros e variáveis regressoras. Para o modelo (2.2) supõe-se em geral, que:

$$E[\boldsymbol{\varepsilon}] = 0$$

ou

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta} \quad (2.3)$$

Portanto, a função esperança para o modelo de regressão é chamada de  $\mathbf{X}\boldsymbol{\beta}$  e  $\mathbf{X}$  de matriz derivada, que se adquire derivando-se o modelo em relação a  $\boldsymbol{\beta}$ .

Além dos pressupostos gerais acerca da correta especificação do modelo e da medição sem erros das variáveis observadas, um pressuposto importante para todo o desenvolvimento é de que os erros do modelo têm média zero (2.3), não estão correlacionados e têm variância constante  $\sigma^2$ . Uma condição adicional para os erros do modelo é de que estejam normalmente distribuídos. Admitindo-se que o  $\boldsymbol{\varepsilon}$  é normalmente distribuído com:

$$Var[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = I\sigma^2, \quad (2.4)$$

em que a matriz identidade  $N \times N$  é denotada por  $I$ , assim, a função densidade de probabilidade conjunta para o  $\mathbf{y}$ , dado o  $\boldsymbol{\beta}$  e a  $\sigma^2$ , é:

$$P(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(\frac{-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right) \quad (2.5)$$

A ideia principal aqui é fazer inferência sobre  $\sigma^2$  e os  $P$  parâmetros de  $\beta$ . Para as inferências sobre os parâmetros é necessário a utilização de métodos de estimação, dentre os métodos de estimação mais utilizados na regressão estão o método dos mínimos quadrados e de máxima verossimilhança.

## 2.2 Modelos de regressão não-linear

A definição de um modelo não-linear apresentada pela maioria dos autores, como Draper e Smith (1998), Bates e Watts (1988), Ratkowsky (1983) e outros, é que pelo menos uma derivada parcial da variável dependente, com relação a algum parâmetro presente no modelo, depende de algum parâmetro.

### 2.2.1 Modelos de regressão intrinsecamente lineares

Segundo Gujarati (2006) alguns modelos podem parecer não-lineares nos parâmetros, mas são intrinsecamente lineares, pois, as devidas transformações, podem se tornar modelos de regressão lineares em novos parâmetros. Gujarati (2006) utilizando-se a função de produção de Cobb-Douglas, detalha esse conceito fazendo-se  $Y$  = produção,  $X_2$  = mão-de-obra,  $X_3$  = capital e  $u_i$  o componente do erro aleatório. Pode-se escrever esta função de três maneiras:

A primeira:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \quad (2.6)$$

Aplicando-se o logaritmo neperiano em ambos os membros da equação (2.6), obtém-se o modelo:

$$W_i = \alpha + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \quad (2.7)$$

Com  $\alpha = \ln \beta_1$  e  $W_i = \ln Y_i$  e as seguintes derivadas parciais em relação aos novos parâmetros

$$\frac{\partial W_i}{\partial \alpha} = 1; \quad \frac{\partial W_i}{\partial \beta_2} = \ln X_{2i} \quad e \quad \frac{\partial W_i}{\partial \beta_3} = \ln X_{3i}$$

não dependem dos parâmetros, o que de acordo com a definição de linearidade, indica que o modelo (2.6) é intrinsecamente linear e (2.7) é linear segundo os parâmetros  $\alpha, \beta_2$  e  $\beta_3$

A segunda:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} u_i \quad (2.8)$$

Aplicando-se o logaritmo neperiano em ambos os membros da igualdade (2.8), o modelo passou-se a:

$$W_i = \alpha + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \ln u_i \quad (2.9)$$

com  $\alpha = \ln \beta_1$  e  $W_i = \ln Y_i$  as seguintes derivadas parciais em relação aos novos parâmetros são:

$$\frac{\partial W_i}{\partial \alpha} = 1; \quad \frac{\partial W_i}{\partial \beta_2} = \ln X_{2i} \quad e \quad \frac{\partial W_i}{\partial \beta_3} = \ln X_{3i}$$

Pode-se concluir de acordo com a definição de linearidade, o modelo (2.8) é intrinsecamente linear e que o modelo (2.9) é linear segundo os parâmetros  $\alpha, \beta_2$  e  $\beta_3$ .

Em uma terceira possibilidade, considerando-se o modelo (2.8)

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} u_i$$

Sem o uso de transformações no modelo (2.8) tem-se que as derivadas parciais de  $Y_i$  em relação aos parâmetros dependem dos mesmos:

$$\frac{\partial Y_i}{\partial \beta_1} = X_{2i}^{\beta_2} X_{3i}^{\beta_3}; \quad \frac{\partial Y_i}{\partial \beta_2} = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} \ln(X_{2i}) \quad e \quad \frac{\partial Y_i}{\partial \beta_3} = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} \ln(X_{3i})$$

neste sentido não existe possibilidades de transformá-la em um modelo linear nos parâmetros e nem satisfaz a definição de um modelo de regressão linear o modelo (2.2.1).

Os modelos não-lineares são linearizados para facilitar a obtenção das estimativas dos parâmetros (BATES e WATTS, 1988). Segundo CURRIE (1982), se por um lado os erros do modelo original satisfazem às suposições usuais de normalidade, independência e homogeneidade da variância, por outro, os erros do novo modelo, não satisfarão tais suposições e as transformações farão com que os parâmetros percam sua explicação intrínseca, e além do mais, poderá alterar a estrutura e as distribuições dos erros.

## 2.2.2 Modelos de regressão intrinsecamente não-lineares

Diz-se que um modelo de regressão é intrinsecamente não-linear nos parâmetros, quando ele não é linear e nem intrinsecamente linear. Segundo Gallant (1987) uma das mais comuns circunstâncias em análise estatística é aquela de dados observados que consiste em respostas univariadas  $Y_t$ , conhecidas sendo dependentes com contribuição  $X_t$  que pode ser representada pela seguinte equação de regressão:

$$y_t = f(x_t, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, N \quad (2.10)$$

em que  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_P)$  é um vetor P-dimensional de parâmetros desconhecidos e  $\boldsymbol{\varepsilon}_t$  é o erro aleatório aditivo, em que os erros são independentes identicamente distribuídos com média zero e variância  $\sigma^2$ .

A sequência de valores de  $x_t$  também é chamada de ajuste experimental. A função modelo  $f = (x_t, \boldsymbol{\theta})$  é determinada pelo vetor de parâmetros e o ajuste experimental sobre as suposições de que  $E[\boldsymbol{\varepsilon}_t] = 0$ , é o valor esperado de  $\mathbf{y}_t$  condicional a  $\mathbf{x}_t$  e  $\boldsymbol{\theta}$  (BATES e WATTS, 1980).

$$E[y_t | x_t, \boldsymbol{\theta}] = f(x_t, \boldsymbol{\theta})$$

Como exemplo de uma função não-linear têm-se:

$$f(X, \theta_0, \theta_1, \theta_2) = \theta_0 + \theta_1 e^{\theta_2 X}$$

Se pelo menos uma das derivadas parciais  $\frac{\partial f}{\partial \theta_i}$  depende de pelo menos um dos parâmetros, então,  $f(\cdot)$  é não-linear nos parâmetros.

$$\frac{\partial f}{\partial \theta_0} = 1, \quad \frac{\partial f}{\partial \theta_1} = e^{\theta_2 X} \quad e \quad \frac{\partial f}{\partial \theta_2} = \theta_1 X e^{\theta_2 X}$$

Sendo, assim, a linearidade ou a não linearidade de uma função está ligada aos parâmetros, ou seja, uma função é dita linear nos parâmetros se  $y$  depende apenas de uma variável  $X$ , isto é,

$$Y = f(X, \theta_0, \theta_1, \dots, \theta_k) + \varepsilon$$

Tem-se que  $f(\cdot)$  é linear nos parâmetros  $\theta_0, \theta_1, \dots, \theta_k$  se

$$\frac{\partial f}{\partial \theta_i} = g(X), \quad i = 0, 1, \dots, k,$$

sendo  $g(X)$  dependente apenas de  $X$ .

### 2.2.2.1 Método de estimação por mínimos quadrados

O desenvolvimento dos métodos dos mínimos quadrados, de acordo com Memória (2004), está relacionado à aplicações na astronomia para a determinação das órbitas de cometas e planetas. Atualmente é utilizado nas mais variadas ciências, sendo uma técnica de otimização matemática (“minimizar ou maximizar”), utilizada quando tem-se uma distribuição de pontos e se quer ajustar a melhor curva a este conjunto de dados tentando-se minimizar os erros ou desvios entre os valores observados e os valores projetados a partir do modelo estimado.

Define-se a soma de quadrados dos erros para o modelo não-linear fornecido pelos dados como (DRAPER e SMITH, 1998):

$$S(\boldsymbol{\theta}) = \sum_{t=1}^N \{y_t - f(x_t, \boldsymbol{\theta})\}^2 \quad (2.11)$$

Desde que  $y_t$  e  $x_t$  sejam observações fixas, a soma dos quadrados é uma função de  $\boldsymbol{\theta}$ . A estimativa de mínimos quadrados  $\hat{\boldsymbol{\theta}}$  de  $\boldsymbol{\theta}$  é o valor que minimiza  $S(\boldsymbol{\theta})$ . Pode ser mostrado que a estimativa de mínimos quadrados de  $\boldsymbol{\theta}$  é a mesma estimativa obtida pela máxima verossimilhança (DRAPER e SMITH, 1998). Para encontrar o estimador de mínimos quadrados  $\hat{\boldsymbol{\theta}}$  é necessário diferenciar a Equação (2.12) com relação a  $\boldsymbol{\theta}$ . Isso gera  $P$  equações normais que precisam ser resolvidas para  $\hat{\boldsymbol{\theta}}$ . As equações normais podem ser tomadas na forma:

$$\sum_{t=1}^N \{y_t - f(x_t, \boldsymbol{\theta})\} \left[ \frac{\partial f(x_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \quad (2.12)$$

para  $t = 1, 2, \dots, P$  em que a quantidade denotada pelos colchetes são as derivadas de  $f(x_t, \boldsymbol{\theta})$  com os respectivos  $\boldsymbol{\theta}_t$  com todos os  $\boldsymbol{\theta}'$ s substituídos pelos correspondentes  $\hat{\boldsymbol{\theta}}'$ s, que têm o mesmo subscrito.

Para cada ajuste experimental considerando  $x_t$ , a resposta condicional será:

$$\boldsymbol{\eta}_t(\boldsymbol{\theta}) = E[y_t | \boldsymbol{\theta}] = f(x_t, \boldsymbol{\theta}) \quad (2.13)$$

e os vetores

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), \eta_2(\boldsymbol{\theta}), \dots, \eta_N(\boldsymbol{\theta}))^T, \quad \mathbf{y} = (y_1, y_2, \dots, y_N)^T \text{ e } \boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^T,$$

logo, pode-se escrever o modelo em forma vetorial como:

$$\mathbf{y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \quad (2.14)$$

e a função de quadrados (2.12) como:

$$S(\boldsymbol{\theta}) = \|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\|^2 \quad (2.15)$$

em que a norma de um vetor é indicador por  $\|\cdot\|$ .

O método dos mínimos quadrados é utilizado na estimação dos parâmetros em modelos não-lineares, da mesma maneira que em modelos lineares. Porém, para o Sistema

de Equação Normais não-linear, não existe uma solução explícita a qual deve ser obtida através de processos iterativos por aproximação em série de Taylor, usando-se um *software* (GALLANT, 1987). A maioria dos algoritmos para computação das estimativas de mínimos quadrados  $\hat{\boldsymbol{\theta}}$  e a maioria dos métodos inferenciais para modelos não-lineares são baseados em métodos iterativos que consideram uma aproximação linear local para o modelo (BATES e WATTS, 1980).

### 2.2.2.2 Aproximação linear

A função modelo do valor de um parâmetro  $\boldsymbol{\theta}^0$  é aproximada por:

$$f(x_t, \boldsymbol{\theta}^0) \cong f(x_t, \boldsymbol{\theta}^0) + \sum_{i=1}^P (\theta_i - \theta_i^0) v_i(x_i)$$

em que:

$$v_i(x_i) = \left[ \frac{\partial f(x_t, \boldsymbol{\theta})}{\partial \theta_i} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

Agrupando-se os componentes dentro dos vetores  $\boldsymbol{\eta}(\boldsymbol{\theta})$ ,  $\boldsymbol{\eta}(\boldsymbol{\theta}^0)$  e  $\mathbf{v}_i$ , em que:

$$\mathbf{v}_i = (v_i(x_1), v_i(x_2), \dots, v_i(x_n))^T = \left[ \frac{\partial \boldsymbol{\eta}}{\partial \theta_i} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}, \quad i = 1, 2, \dots, P$$

tem-se:

$$\begin{aligned} (\boldsymbol{\eta}\boldsymbol{\theta}) &\cong \boldsymbol{\eta}(\boldsymbol{\theta}^0) + \sum_{i=1}^P (\theta_i - \theta_i^0) \mathbf{v}_i \\ (\boldsymbol{\eta}\boldsymbol{\theta}) &\cong \boldsymbol{\eta}(\boldsymbol{\theta}^0) + \mathbf{V}^0 (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \end{aligned} \quad (2.16)$$

em que  $\mathbf{V}^0$  é a matriz de derivadas avaliada em  $\boldsymbol{\theta}^0$ , tendo como colunas os vetores  $\mathbf{v}_i$ s ordenadamente. A matriz  $\mathbf{V}^0$  também chamada de matriz jacobiana de ordem  $N \times P$  de  $(\boldsymbol{\eta}\boldsymbol{\theta})$  avaliada em  $\boldsymbol{\theta}^0$ :

$$\mathbf{V}^0 = \left[ \begin{array}{cccc} \frac{\partial \eta_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_1(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial \eta_1(\boldsymbol{\theta})}{\partial \theta_P} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \eta_n(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_n(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial \eta_n(\boldsymbol{\theta})}{\partial \theta_P} \end{array} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

O efeito dessa aproximação linear consiste juntamente segundo Bates & Watts (1980), em substituir o local da solução pelo seu plano tangente em  $\boldsymbol{\eta}(\boldsymbol{\theta})$ . A aproximação da linearidade do modelo afeta diretamente a efetividade do algoritmo de mínimos quadrados e a validade das inferências feitas no que se refere aos parâmetros de um modelo não linear.

### 2.2.2.3 O método de Gauss-Newton

Gauss sugere uma aproximação linear à função esperança, como em (2.16), para melhorar o  $\boldsymbol{\theta}^0$  como valor inicial para  $\boldsymbol{\theta}$  e continuar melhorando as estimativas, até que as mudanças não mais existam. Esse método expande uma série de Taylor sobre  $\boldsymbol{\theta}_i$  (o subscrito  $i$ -ésima interação), segundo Ratkowsky (1983), como em (2.16). Rearranjando-se (2.17) com (2.16), tem-se:

$$\begin{aligned} \mathbf{S}(\boldsymbol{\theta}) &= \|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\|^2 = [\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})]^T [\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})] = \\ &\cong [\mathbf{y} - (\boldsymbol{\eta}(\boldsymbol{\theta}_i) + \mathbf{v}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_i))]^T [\mathbf{y} - (\boldsymbol{\eta}(\boldsymbol{\theta}_i) + \mathbf{v}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_i))] = \\ &= [(\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}_i)) - \mathbf{v}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_i)]^T [(\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}_i)) - \mathbf{v}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_i)] = \\ &= [\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}_i)]^T [\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}_i)] \Rightarrow \\ &-2 [\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}_i)]^T \mathbf{v}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_i) + (\boldsymbol{\theta} - \boldsymbol{\theta}_i)^T \mathbf{v}_i^T \mathbf{v}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_i) \end{aligned}$$

O vetor gradiente  $g(\boldsymbol{\theta}) = \left\langle \frac{\partial s(\boldsymbol{\theta})}{\partial \theta_1}, \frac{\partial s(\boldsymbol{\theta})}{\partial \theta_2}, \dots, \frac{\partial s(\boldsymbol{\theta})}{\partial \theta_P} \right\rangle^T$  é, então:

$$g(\boldsymbol{\theta}) = -2 \mathbf{v}_i^T [\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}_i)] + 2 \mathbf{v}_i^T \mathbf{v}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_i)$$

Igualando-se essa expressão a zero e rearranjando-se os valores, obtém-se:

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + [\mathbf{V}_i^T \mathbf{V}_i]^{-1} \mathbf{V}_i^T [\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}_i)] \quad (2.17)$$

Com valores iniciais para  $\boldsymbol{\theta}$  com  $i = 1$ , o processo continua até a convergência que ocorre quando  $\|\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}\|$  é menor do que alguma quantidade pequena pré-fixada.

Para Draper e Smith (1998), o método da linearização pode convergir muito lentamente; pode oscilar bastante, continuamente mudando-se de direção e, muitas vezes, de modo crescente, decrescendo-se a soma de quadrados. Entretanto, a solução pode estabilizar eventualmente e pode não convergir sob qualquer condição e, constantemente, divergir. Desde que tal modelo pode ser escrito da forma:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon},$$

o método de Gauss-Newton serve também para encontrar as estimativas em modelos de regressão linear em que  $\mathbf{X}$  é a matriz  $N \times P$  de variáveis regressoras, cuja primeira coluna pode ser um vetor unitário, se a função modelo incluir um termo constante. A matriz

jacobiana é:

$$\mathbf{V}_0 = \mathbf{X}$$

para a função modelo acima.

Para qualquer valor arbitrário utilizado como estimativa inicial, por exemplo,  $\boldsymbol{\theta}_0$ , o vetor de estimativas atualizado  $\boldsymbol{\theta}_1$  é:

$$\begin{aligned} \boldsymbol{\theta}_{i+1} &= \boldsymbol{\theta}_0 + [\mathbf{V}_0^T \mathbf{V}_0]^{-1} \mathbf{V}_0^T [\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}_0)] \\ &= \boldsymbol{\theta}_0 + [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T [\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_0] \\ &= \boldsymbol{\theta}_0 + [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y} - [\mathbf{X}^T \mathbf{X}]^{-1} [\mathbf{X}^T \mathbf{X}] \boldsymbol{\theta}_0 \\ &= \boldsymbol{\theta}_0 + [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y} - \boldsymbol{\theta}_0 \\ &= [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y} \end{aligned}$$

Segundo Draper e Smith (1998) se o processo iterativo é continuado é evidente que a estimativa fica imutável. E o modelo de regressão linear para o estimador de mínimos quadrados ficará do lado direito da igualdade. Desta forma, um modelo linear, converge para o estimador de mínimos quadrados em uma única iteração para qualquer valor inicial, segundo o método de Gauss-Newton. Existe vários métodos para a obtenção das estimativas de mínimos quadrados dos parâmetros de um modelo de regressão não-linear. Em princípio, tais métodos podem ser utilizados quando o método de Gauss-Newton apresentar problemas com relação à convergência segundo (BATES e WATTS, 1988). Dentre eles pode-se primar pelos métodos: Gradiente, Marquardt, Newton-Raphson e Steepest-descent.

#### 2.2.2.4 Valores iniciais

Segundo Souza (1998) o uso de uma função resposta apropriada e valores iniciais adequados está diretamente associado ao sucesso na convergência de um algoritmo para um método iterativo no processo de estimação não linear. O êxito na convergência de  $\hat{\boldsymbol{\theta}}$  não pode ser garantido por nenhum método iterativo (GALLANT,1987). O conhecimento prático permite ao pesquisador saber que a falta de convergência depende da distância do valor inicial à resposta correta e/ou do grau de parametrização da função resposta, relativos ao conjunto de dados utilizados. Quando não ocorrer a convergência, deve-se encontrar valores iniciais melhores (mais próximos de  $\boldsymbol{\theta}$ ) e/ou utilizar modelos com menor quantidade de parâmetros.

Bates e Watts (1988) sugerem que técnicas descritivas sejam utilizadas para a checagem do ajuste do modelo aos dados, bem como verificação de possíveis inconsistências. Para isso uma boa opção é plotar a resposta versus variável preditora, e observar o comportamento das variáveis, verificando-se os valores encontrados estão em harmonia com os dados sugeridos, traçando-se a curva ajustada com os dados e verificando-se não há nenhum resultado absurdo.

Em conformidade com Draper e Smith (1998), baseadas em qualquer informação disponível os valores iniciais podem ser aproximações preliminares, por exemplo, sugeridas por experiências anteriores. Quanto melhor os valores forem, menor será o número de interações necessárias para se encontrar a convergência. Para Bates e Watts (1988) e Draper e Smith (1998) vários princípios para determinar valores iniciais podem ser usados:

- i) interpretar o comportamento da função esperança e das derivadas da função esperança, em termos de parâmetros analiticamente ou graficamente;
- ii) usar linearidade condicional se possível;
- iii) reduzir dimensões substituindo-se valores para alguns parâmetros;
- iv) transformar a função esperança geometricamente ou analiticamente para alcançar comportamento mais simples, perfeitamente linear. Pode-se desconsiderar o mesmo apenas para a obtenção das estimativas iniciais, no caso em que o erro é aditivo.

## 2.3 Curvas de crescimentos

Dentre os diversos modelos não lineares, existem os modelos de curvas de crescimento. Entre estes modelos de crescimento, os cinco mais conhecidos são, Von Bertalanffy, Logístico e Gompertz, e ainda que existam variações quanto à interpretação e conteúdo dos mesmos, as funções não-lineares para as curvas de crescimento contém diversos parâmetros em comum, onde é possível agregar significado biológico a cada um deles (SILVA, 2008).

O ajuste de curva de crescimento peso-idade para animais tem um papel importante no planejamento da produção animal. No entanto, as curvas de crescimento ajustadas devem ser coerentes com as interpretações biológicas do crescimento do animal. A análise de curvas de crescimento consiste na análise de dados longitudinais por meio de ajustamento de um modelo matemático que descreve todo o período de vida do animal relacionando-se seu peso com sua idade. O ajuste destas curvas tem sido abordado em inúmeros

estudos e em diversas áreas de pesquisa. Pode-se definir os modelos de crescimento como funções não-lineares que relacionam os pesos de um animal, medido em diferentes idades (RAMOS, 2010)

Uma grande vantagem desses modelos é a simplicidade e facilidade na interpretação dos parâmetros, pois em muitas situações, são requeridos menos parâmetros nos modelos não-lineares do que nos lineares. Além disso, nos modelos não-lineares, os parâmetros fornecem um maior conhecimento sobre o fenômeno em estudo, o que gera um bom ajuste, com menos parâmetros.

Os parâmetros da curva de crescimento são utilizados para prever taxas de crescimento, necessidades alimentares, peso à maturidade e graus de maturidade, também servindo-se como critérios de seleção para programas de melhoramento animal. Nesse sentido, as curvas de crescimento refletem a relação entre a idade do animal e o seu impulso de crescimento e maturidade, sendo importantes para pesquisas e recomendações sobre eficiência de produção em programas de melhoramento, dessa maneira, para aumentar o lucro do produtor (MENDES, 2007).

Todos os modelos citados anteriormente são derivados da curva de Richards (2.18) e a diferença entre eles consiste na variação do parâmetro de inflexão ( $m$ ). Os modelos von Bertalanffy (2.19) e Logístico (2.20) apresentam parâmetros de inflexão ( $m$ ) iguais  $2/3$  e  $2$  respectivamente, já no modelo Gompertz (2.21) esse parâmetro tende a  $1$  (THOLON e QUEIROZ, 2009).

i) Richards:

$$y_t = \alpha(1 - (\beta e^{-kt})^m) + \varepsilon \quad (2.18)$$

ii) Von Bertalanffy:

$$y_t = \alpha(1 - \beta e^{-kt})^3 + \varepsilon \quad (2.19)$$

iii) Logístico:

$$y_t = \alpha(1 + \beta e^{-kt})^{-m} + \varepsilon \quad (2.20)$$

iv) Gompertz:

$$y_t = \alpha e^{-\beta e^{-kt}} + \varepsilon \quad (2.21)$$

O parâmetro  $\alpha$ , definido como peso assintótico ou peso adulto, representa a estimativa de peso a maturidade, independente de flutuações de pesos devido a efeitos genéticos e ambientais. Um outro parâmetro,  $k$ , corresponde ao índice de maturidade ou a estimativa

de precocidade de maturidade e determina a eficiência do crescimento do animal. Quanto maior for o valor desse parâmetro, mais precoce é o animal e vice-versa. O parâmetro  $m$  é denominado parâmetro de inflexão. Este refere-se ao ponto em que o animal passa de uma fase de crescimento acelerado para uma fase de crescimento inibitório e indica o ponto a partir do qual o animal passa a crescer com menor eficiência. O  $\beta$  é denominado de parâmetro de integração ou interceptação com o eixo-y, não possui interpretação biológica e é utilizado apenas para adequar o valor inicial do peso vivo fazendo com que a curva passe pela origem quando  $y \neq 0$  e/ou  $t \neq 0$ , sendo  $t$  a expressão da idade e  $\varepsilon$  é o erro identicamente distribuído, normal com média zero e variância  $\sigma^2$  (SILVA et al., 2001).

### 2.3.1 Critérios para seleção de modelos

A frase atribuída ao estatístico George E. P. Box “ Todos os modelos são errados, mas alguns modelos são úteis” diz respeito à importância da escolha de um modelo, dentre todos os possíveis modelos, para se modelar um fenômeno. Usualmente em estatística, se faz uso da análise de resíduos após ajuste de um modelo aos dados, porém devido à natureza de alguns modelos esta estratégia nem sempre é a mais adequada, para modelos não-lineares com medidas correlacionadas no tempo se faz uso de diferentes critérios para avaliar a qualidade do ajuste de modelos não-lineares (SILVA, 2009).

Ao se fazer comparações entre os modelos é importante levar em conta a dificuldade computacional que está intrinsecamente relacionada à quantidade de parâmetros do modelo e ao bom ajuste do modelo. Vários autores dentre eles Oliveira (1995), Paz (2002) e Freitas (2005), utilizam como critérios para quantificar a qualidade de ajuste, a variância residual do ajuste do modelo a cada animal, desvios e correlação dos peso observados versus peso preditos nas várias fases da vida dos animais, erro de predição médio, o coeficiente de determinação dentre outros critérios. O coeficiente de determinação para modelos não-lineares pode ser concebido da seguinte forma (SCHABENBERGER, 2001):

$$R^2 = 1 - \frac{SQR}{SQT_c}$$

em que, SQR é a soma de quadrados do resíduo e  $SQT_c$  é a soma de quadrado total corrigida para o número de parâmetros do modelo. Deve-se tomar cuidado na avaliação da qualidade do ajuste ao utilizar o coeficiente de determinação, pois a soma dos resíduos não necessariamente é igual a zero e a soma dos quadrados dos resíduos mais a soma dos quadrados da regressão não é necessariamente igual a soma dos quadrados total no caso de modelos não-lineares, o que pode afetar o coeficiente de determinação.

Na validação do desempenho dos modelos, o Critério de informação de Akaike - Akaike's Information Criterion (AIC) tem sido bastante utilizado para diferentes estruturas de covariâncias. Akaike (1974) relacionou a discrepância, medida que existe entre o modelo verdadeiro e o modelo aproximado, com a máxima verossimilhança, que é o que possui melhor ajuste para os dados observados.

O princípio do Critério de Informação de Akaike (AIC) é selecionar uma combinação de variáveis exploratórias a modelos para a função de correlação que minimize o valor do AIC (AKAIKE, 1974) . É importante observar que, em muitas situações clássicas, tais como regressão linear ou em modelos de séries temporais, o AIC não é uma condição consistente para a seleção de modelos. Ou seja, com o crescente aumento do tamanho das amostras, há uma alta probabilidade de que um modelo selecionado pelo AIC não corresponda ao verdadeiro modelo.

O critério de informação de Akaike é expresso por:

$$AIC = -2 \log L + 2(p + 1),$$

em que,  $L$  é o log de verossimilhança maximizado e  $p$  é o número de parâmetros. Segundo este critério, o melhor modelo é o que possui menor valor de AIC.

A comparação de modelos também pode ser feita a partir de medidas de adequabilidade, como o Bayesian Information Criterion (BIC) de Schwarz (SCHWARZ, 1978), os quais são aproximações do fator de Bayes. O BIC resolve este problema através da introdução de um termo de penalidade para o número de parâmetros do modelo dado por:

$$BIC = -2 \log L + (p + 1) \ln(n)$$

em que,  $n$  é o número de observações, ou equivalente ao tamanho da amostra,  $p$  é o número de parâmetros livres a serem estimados e  $\log L$  é o valor maximizado da função de verossimilhança para o modelo estimado. Menor valor do BIC indica o melhor ajuste do modelo.

### 3 Aplicação

Com o intuito de aplicar a teoria estudada, utilizou-se um conjunto de dados neste trabalho que foi extraído do site do 13<sup>th</sup> workshop *QTL-MAS* organizado pela universidade de Wageningen-UR na Holanda de 20 à 21 de Abril de 2009, a url está disponível em: [http : //www.qtlmas2009.wur.nl/UK/Dataset/](http://www.qtlmas2009.wur.nl/UK/Dataset/), acesso em: 20 de Setembro de 2011, sendo um conjunto de dados simulados de suínos com uma parte dos dados em mensurações fenotípicas e mais 3 partes de mensurações de cunho genético para a verificação de metodologias na área de genômica. O banco de dados fenotípico consiste de 2.025 indivíduos de duas gerações.

Todos os indivíduos têm informações de peso em cinco momentos distintos (0, 132, 265, 397, 560) dias correspondentes aos períodos de produção. Os primeiros 25 indivíduos são os pais, 20 do sexo feminino e 5 masculino. Os restantes 2.000 indivíduos são descendentes, 100 irmãos completos (FS) famílias, uma de cada combinação de um pai macho e fêmea. Cada família FS tem 20 filhos. Para a análise do crescimento dos animais as informações de parentesco não foram consideradas no modelo. Estes valores de rendimento podem ser vistos como representante de peso durante o crescimento de um animal.

Foram usados os três modelos não-lineares descritos anteriormente, o Von Bertalanffy (2.19), o Logístico (2.20) e o Gompertz (2.21) para estimar o crescimento dos animais em função da idade. Na Tabela 1 apresenta-se as derivadas em relação a cada parâmetro dos modelos estudados para que se possa utilizar no ajuste do modelo aos dados.

Tabela 1: Derivadas parciais dos modelos (2.19), (2.20) e (2.21) em função do peso assintótico ( $\alpha$ ), da constante de integração ( $\beta$ ), da taxa de maturação (K).

Modelo	$dy/d\alpha$	$dy/d\beta$	$dy/dK$
Von Bertalanffy	$(1 - \beta e^{-kt})^3$	$-3\alpha e^{-kt}(1 - \beta e^{-kt})^2$	$3\alpha \beta t e^{-kt}(1 - \beta e^{-kt})^2$
Logístico	$(1 + \beta e^{-kt})^{-m}$	—	$\alpha m t e^{-kt}(1 + e^{-kt})^{-m-1}$
Gompertz	$e^{-\beta e(-kt)}$	$-\alpha e^{-kt} e^{(-\beta e(-kt))}$	$\alpha \beta t e^{-kt} e^{-\beta e(-kt)}$

Os modelos foram ajustados às séries de peso por idade de cada animal, individual-

mente, de acordo com o método de Gauss-Newton para modelos não-lineares. Foi utilizado o procedimento “nls” do Software livre R para obtenção das estimativas individuais dos parâmetros “e dos parâmetros médios a partir da totalidade dos dados disponíveis”. Em cada modelo, foram calculados o desvio médio entre o peso predito e o observado, e a correlação entre os dois. O primeiro desses critérios serve como indicativo da presença de vícios em fases distintas da vida do animal. A correlação entre o peso predito e o observado mostra como as variações dos pesos observados são representadas por variações nos pesos preditos a cada idade. Os pesos preditos utilizados no cálculo dos desvios e estimação das correlações foram obtidos em cada idade em que o animal foi pesado.

A comparação entre modelos quanto ao seu ajustamento aos dados, quando não existe um modelo verdadeiro, pressupõe o teste da hipótese nula de que todos os modelos são igualmente bons para este fim, contra a hipótese alternativa de que um ou mais deles são melhores que os outros. Desta forma, foram estabelecidos alguns critérios para comparação dos parâmetros, embora a escolha de um modelo como o melhor para descrever os dados fosse subjetiva, já que alguns desses critérios podem ser mais importantes que outros, dependendo da aplicação que terão os resultados.

Na Figura 1, apresenta-se as curvas observadas em função do tempo para cada animal com mensuração válida. Por meio desta figura percebe-se que o crescimento é não-linear, indicando-se a necessidade de aplicação de modelos não lineares para descrição dos dados.

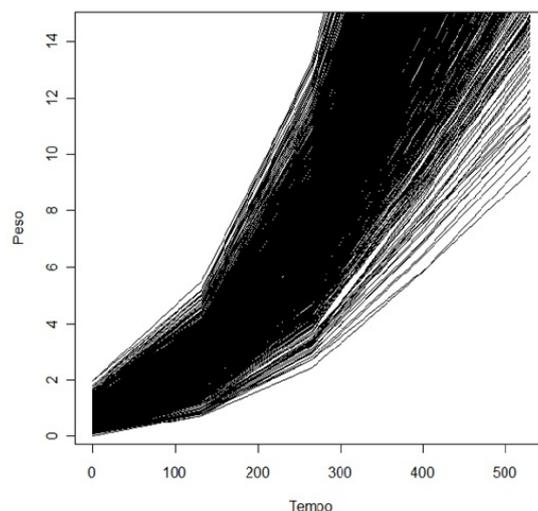


Figura 1: Relação entre o peso dos animais ao longo dos dias.

Em função do comportamento dos dados foi possível ajustar os modelos não-lineares e obter as estimativas dos parâmetros considerando-se todos os dados, para cada um dos modelos (2.19), (2.20) e (2.21) que relaciona peso com a idade do animal. A Figura 2 esboça o ajuste do modelo Logístico.

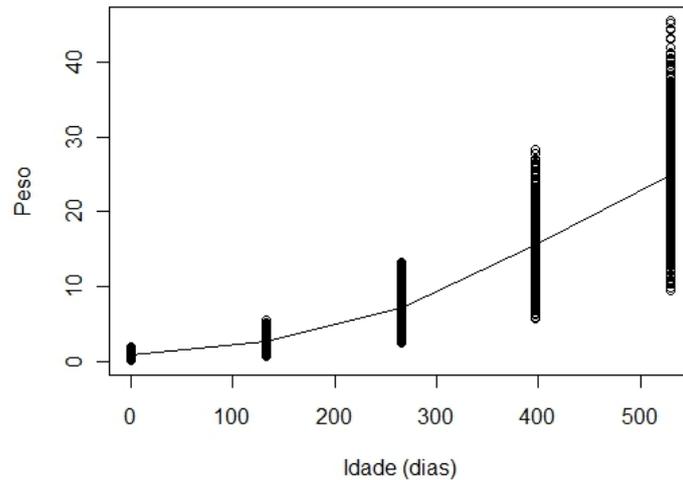


Figura 2: Representação gráfica do ajuste do modelo Logístico.

As estimativas dos parâmetros com respectivos erros-padrão, teste t e intervalos de confiança para cada parâmetro dos 3 modelos avaliados se encontram na Tabela 2. Por meio desta tabela, é possível perceber que o modelo Logístico foi o que apresentou menor amplitude no intervalo para os parâmetros estimados. Todos os parâmetros diferiram de zero pelo teste t, indicando-se que os mesmos são significativos nos modelos. As estimativas dos parâmetros foram distintas apresentando-se maior discrepância no modelo de Von Bertalanffy em relação aos outros dois modelos. Os limites inferiores e superiores dos intervalos de confiança a 95% de cada estimativa de parâmetro possuem estreita amplitude de estimação e modelo Logístico foi o que apresentou a amplitude mais estreita dentre os modelos ajustados (Tabela 2).

A qualidade de ajustamento foi, inicialmente, avaliada por dois critérios: Soma de quadrado do resíduo (SQR) e o coeficiente de determinação ( $R^2$ ). As comparações entre as médias dos modelos, nesses dois critérios, foram feitas considerando-se, para o cálculo das médias, apenas as observações em que a convergência foi atingida em todos os modelos. Outros dois critérios utilizados na comparação da qualidade de ajustamento foram obtidos em cada classe de idade foram a Informação de Akaike (AIC) e a Informação de Bayes (BIC). Na Tabela 3 encontram-se os valores das estatística AIC, BIC,  $R^2$  e SQR. Assim

Tabela 2: Estimativas de peso assintótico ( $\alpha$ ), maturidade do animal ao nascimento ( $\beta$ ), taxa de maturação ( $k$ ), o erro padrão (E. Pad), Parâmetro (Par.), Estimativas (Estim.) e seus respectivos intervalos de confiança (Min. IC e Max. IC)

Modelo	Par.	Estim.	E. Pad.	t	p	Min. IC	Max. IC
Von Bert.	$\alpha$	202,9	38,86	5,2208	< .0001	143,0808	322,6858
	$\beta$	0,8825	0,0035	247,5057	< .0001	0,8756	0,8904
	$K$	0,001073	0,0001	9,2434	< .0001	0,0008	0,0013
Logístico	$\alpha$	34,61	0,7139	48,4807	< .0001	33,3145	36,1141
	$\beta$	38,99	1,4850	26,2576	< .0001	36,2985	42,0931
	$K$	0,008797	0,0002	47,7542	< .0001	0,0084	0,0092
Gompertz	$\alpha$	67,60	4,3910	15,3944	< .0001	59,9085	77,8216
	$\beta$	4,894	0,0537	90,9852	< .0001	4,7921	5,0123
	$K$	0,003034	0,0001	23,0769	< .0001	0,0028	0,0033

o melhor modelo ajustado é aquele que apresentar menores valores para AIC, BIC e SQR e maior valor para  $R^2$ .

Tabela 3: Estatísticas da qualidade de informação de ajuste, valores da soma de quadrado dos resíduos (SQR) e coeficiente de determinação ( $R^2$ ), AIC (Informação de Akaike) e BIC (Informação de Bayes).

	Von Bertalanffy	Logístico	Gompertz
SQR <sub>res</sub>	59661.19	59218.41	59372.77
$R^2$	0.8741	0.875	0.8747
AIC	26593.60	26556.36	26569.37
BIC	26619.67	26582.43	26595.44

Os coeficientes de determinação obtidos foram altos para os modelos em que houve convergência, ficando-se em torno de 87% o que indica que todos os modelos apresentaram desempenho semelhantes. Porém, foi observado na Tabela 3 que o melhor ajuste foi do modelo Logístico, o qual apresentou os menores valores para AIC, BIC, SQR e o maior valor para  $R^2$ .

Após o ajuste dos modelos e a escolha do melhor modelo, procedeu-se com a análise gráfica referentes as Figuras 3 e 4. Nestas figuras apresenta-se os resíduos versus idade(dias) e valores observados vs valores preditos, para o modelo Logístico, por este ter sido o melhor modelo dentre os modelos ajustados.

Na Figura 3, percebe-se uma heterogeneidade de variâncias ao longo do tempo, o que sugere que apesar do modelo Logístico ter sido o melhor modelo, é plausível se pensar em estudar modelos que comportem em sua estrutura heterogeneidade de variâncias. Pelo comportamento gráfico da Figura 4, valores preditos versus valores observados, pode-se estudar outros modelos mais complexos, pois devido a variabilidade não capturada entre

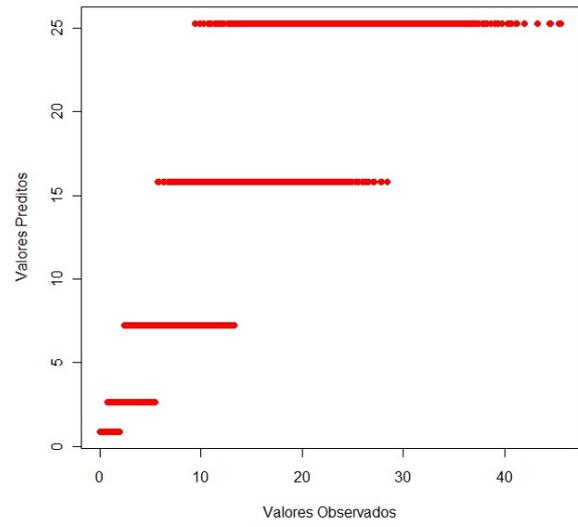
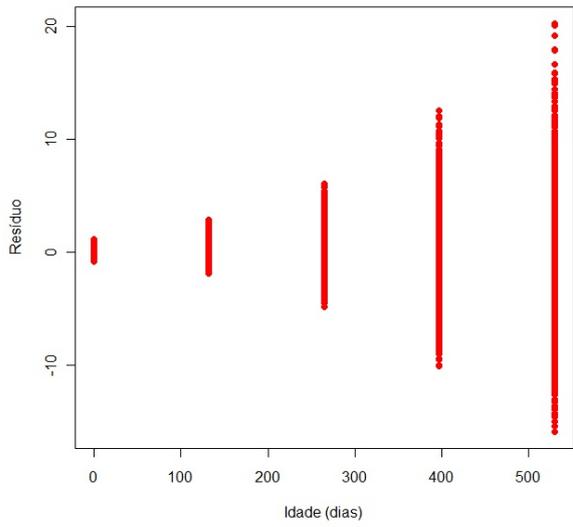


Figura 3: Representação gráfica dos resíduos versus dias

Figura 4: Representação gráfica dos resíduos versus dias.

os tempos, o comportamento não é linear entre as observações e as predições.

## 4 Conclusão

Por meio deste trabalho verificou-se a importância dos ajustes de modelos não-lineares, representados através das curvas de crescimentos, pois além de simples e fácil interpretação, tem se mostrado muito útil para caracterizar a espécie em estudo. Os modelos ajustados, o Logístico, o Von Bertalanffy e o Gompertz se mostraram apropriados para descrever a curva de crescimento para os dados simulados de suínos. Após a análise dos dados, através dos critérios e dos métodos estatísticos o modelo que melhor descreveu o crescimento dos animais foi o Logístico, o mesmo obteve ajuste superior a 80% utilizando a metodologia de aproximação linear, sendo o mais indicado para representar a curva de crescimento médio, logo à importância da modelação matemática do crescimento é incontestável.

## 5 Referências Bibliográficas

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Boston, v.19, n.16, p.716-723, Dec. 1974
- BATES, D. M.; WATTS, D. G. Relative curvature measures of nonlinearity (With discussion). **Journal of the Royal Statistical Society**, Ser. B, v.42, n. 1, p. 1-25, 1980.
- BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its applications**. New York: Wiley series in probability e mathematical statistics, 1988. 365p.
- CURRIE, D. J. Estimating Michaelis-Menten parameters: bias, variance e o experimental design. **Biometrics**, Oxford, v.38 n.4, p.907-919, 1982.
- DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3. ed. New York: J. Wiley, 1998. 706p.
- FREITAS, A.R. Curvas de crescimento na produção animal. **Revista Brasileira de Zootecnia**, Viçosa, v. 34, n. 3, p.786-795, 2005.
- GALLANT, A. R. **Nonlinear Statistical model**. New York: Wiley series in probability e methematical statistics, 1987. 624p.
- GUJARATI, D. **Econometria básica**. 4. ed. São Paulo: Campus, 2006. 806p.
- MEMÓRIA, J.M.P. **Breve história da estatística**. Brasília, DF: Embrapa Informação Tecnológica: Embrapa-Secretaria de Gestão e Estatística, 2004. 111p. (Embrapa-Secretaria de Gestão e Estatística. Texto para discussão, 21).
- MENDES,P. N. **Curvas de crescimento difásicas de fêmeas hereford com erros auto regressivos e heterogeneidade de variâncias**. 2007. 98p. Dissertação (Mestrado em Agronomia)-Universidade Federal de Lavras, Lavras, MG.
- OLIVEIRA, H.N. de; ALENCAR, M.M.; LIMA, R. Eficiência produtiva de vacas da raça Nelore. **Revista Brasileira de Zootecnia**, Viçosa, v.24, n.3, p.445-453, 1995.

OLIVEIRA, H.N.; LÔBO, R.B.; PEREIRA, C.S. Comparação de modelos não-lineares para descrever o crescimento de fêmeas da raça Guzerá. **Pesquisa Agropecuária Brasileira**, Brasília, v.35, n.9, p.1843-1851, set. 2000.

PAZ, C.C.P. **Associação entre polimorfismos genéticos e parâmetros da curva de crescimento em bovinos de corte**. 2002. 106p. Tese (Doutorado em Ciência Animal e Pastagens) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2002.

RAMOS, S. B. **Ajustes de curvas de crescimento e estimativas da variabilidade genética de peso corporal de avestruzes (*Struthio camelus*)**. 2010. 63p. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, 2010.

RATKOWSKY, D. A. **Nonlinear regression modelling**. a unified practical approach, New York: Marcel Dekker, 1983. 276p.

SCHABENBERGER, O.; PIERCE, F.J. **Contemporary Statistical Models for the Plant and Soil Sciences**. Boca Raton: CRC Press, 2002. 738 p.

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**. Hayward, v.6, p.461-464, Mar.1978.

SILVA, A.M.S. **Modelos hierárquico bayesiano aplicado na avaliação genética de curva de crescimento de bovinos de corte**. 2008, 47p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)-Universidade Federal de Lavras, Lavras, MG.

SILVA, F.L. **Curvas de crescimento e produtividade de vacas Nelore e cruzadas, de diferentes tipos biológicos, em sistema de produção intensiva**. 2009 89p. Dissertação (Mestrado em Ciência Animal e Pastagens)-Piracicaba, SP.

SILVA, F. F., AQUINO, L. H., OLIVEIRA, A. I. G. **Influência de fatores genéticos e ambientais sobre as estimativas dos parâmetros das funções de crescimento em gado nelore**. **Ciênc. agrotec.**, Lavras, v.25, n.5, p. 1195-1205, set./out., 2001.

SOUZA, E. M. **Modelagem não linear da extração de zinco em um solo tratado com lodo de esgoto**. 2008, 95p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)-Universidade Federal de Lavras, Lavras, MG.

SOUZA, G. da S. **Introdução aos modelos de regressão linear e não-linear**. Brasília: Embrapa-SPI/Embrapa-SEA, 1998. 489 p.

THOLON P.; QUEIROZ, S.A. Modelos matemáticos utilizados para descrever curvas

de crescimento em aves aplicados ao melhoramento animal. **Ciência Rural**, v.39, n.7, p.2261-2269, out 2009.