PAULO CÉSAR EMILIANO

CRITÉRIOS DE INFORMAÇÃO: COMO ELES SE COMPORTAM EM DIFERENTES MODELOS?

LAVRAS - MG

2013

PAULO CÉSAR EMILIANO

CRITÉRIOS DE INFORMAÇÃO: COMO ELES SE COMPORTAM EM DIFERENTES MODELOS?

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador
Dr. Mário Javier Ferrua Vivanco

Coorientador
Dr. Fortunato Silva de Menezes

LAVRAS - MG
2013

Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e Serviços da Biblioteca Universitária da UFLA

Emiliano, Paulo César.
Critérios de informação : como eles se comportam em diferentes modelos / Paulo César Emiliano. – Lavras : UFLA, 2013.
193 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2013.
Orientador: Mário Javier Ferrua Vivanco.
Bibliografia.

1. Seleção de modelos. 2. Critérios de informação. 3. Teste de médias. 4. Curvas de crescimento. 5. Séries temporais. I. Universidade Federal de Lavras. II. Título.

CDD – 519.5

PAULO CÉSAR EMILIANO

CRITÉRIOS DE INFORMAÇÃO: COMO ELES SE COMPORTAM EM DIFERENTES MODELOS?

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 30 de agosto de 2013.

Dr. Edwin Marcos Ortega	ESALQ
Dr. Fortunato Silva de Menezes	UFLA
Dr. Marcelo Angelo Cirillo	UFLA
Dra. Thelma Sáfadi	UFLA
Dr. Washington Santos Silva	IFMG

Dr. Mário Javier Ferrua Vivanco
Orientador

LAVRAS - MG
2013

Dedico à minha familia
meus pais,
Francisco e Alzira,
e aos meus irmãos,
Rosimeire, e
Washington.

AGRADECIMENTOS

Primeiramente a Deus, que deu-me forças em todos os momentos de minha vida, e a Nossa Senhora Aparecida, que sempre intercede por mim e da qual sou devoto.

Meus sinceros agradecimentos aos professores Mário (meu "pai" de Lavras) e Fortunato pela paciência com que me orientaram, disponibilidade em auxiliar-me a qualquer momento, pelas críticas e sugestões.

A todos os membros da banca, pelas críticas e sugestões, em especial Thelma e Marcelo Cirillo, que acompanharam todo o desenvolvimento desta tese.

Aos referees anônimos da RME e CSDA que muito contribuíram para o desenvolvimento deste trabalho.

Aos meus pais, Francisco e Alzira, pela confiança, compreensão, carinho, apoio e tudo que sou devo a eles. Aos meus irmãos Rosemeire e Washington, pelo carinho, compreensão e torcida em todos os momentos.

A Kamilla, minha princesinha com a qual caminho a partir de agora. Por todo apoio, incentivo, amizade, carinho! Te amo minha linda.

À Universidade Federal de Lavras, pela oportunidade de aprimoramento acadêmico.

Aos professores do DEX, da Matemática, Física e Estatística pelos seis anos de bom convívio com todos.

A todos os funcionários do DEX, Josi (da graduação), Maria, Miriam, Joice, Edila, Selminha e em especial a Josi (pós graduação), grande amiga e sempre nos ajudando, com seu grande coração, quando precisávamos.

A todos os colegas de mestrado e doutorado em Estatística, Ana Lúcia, Ana Paula, Augusto, Edcarlos, Eustáquio, Leandro, Moysés, Tania, pelo companheirismo, estudo, amizade e momentos de alegria.

A Francisca e Dalvinha, pelos ótimos momentos em que desfrutamos das guloseimas preparadas por ambas. Eustáquio que me desculpe se não agradeci na dissertação de mestrado, mas saiba que vocês três sempre me foram muito queridos e jamais serão esquecidos.

À Universidade Federal de Campina Grande - UFCG, em que fui docente quando ainda estudante do doutorado. Em particular a todos os amigos da Estatística da UAME-UFCG, Ana Cristina, Alexsandro, Amanda, Antônio José, Areli, Chico, Gilberto, Grayci-Mary, Iraponil, João Batista, Manoel, Michelli, Patrícia e Rosângela. Agradeço também a todos os funcionários e demais professores, em especial aos professores José Luiz (sábio em seus conselhos) e José de Arimatéia.

Aos amigos da UEPB, Tiago, Ana Patrícia, Silvio, Cléber e em especial Ricardo "Paraíba" pelas horas de almoço e apoio enquanto morei em Campina Grande.

À Universidade Federal de Viçosa - UFV, que também me propiciou a oportunidade de continuar meus estudos enquanto docente da Instituição. Em particular a todos os amigos da Estatística da UFV, Ana Carolina, Carlos Henrique, Cecon, Fabyano, Fernando, Gérson, José Ivo, Moysés, Nerilson, Peternelli, Policarpo e Sebastião.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e à Coordenação e Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro de ambas.

A todos, que direta ou indiretamente, contribuíram para a realização deste trabalho, muito obrigada!

"-What's the guy's name on first base?

-What's the guy's name on second base.

-I'm not askin' you who's on second.

-Who's on first.

-I don't know.

-He's on third. We're not talking about him."

Abbott and Costello - Who's On First?

"Explorers are we, intrepid and bold,

Out in the wild, amongst wonders untold.

Equipped wit our wits, a map, and a snack,

We're searching for fun and we're on the right track!"

The Indispensable Calvin and Hobbes - Bill Watterson

"If you have an apple and I have an apple and we exchange apples

then you and I still have one apple.

But if you have an idea and I have an idea and we exchange these ideas,

then each of us will have two ideas."

George Bernard Shaw

RESUMO

Um fenômeno pode ser explicado através de modelos, sendo estes destinados a ilustrar certos aspectos do problema, sem contudo, representar todos os detalhes havendo portanto perda de informação. Para não comprometer o entendimento do fenômeno em estudo, esta perda deve ser mínima. Não raro, mais de um modelo pode descrever um mesmo fenômeno, haja vista que não há uma única metodologia a ser seguida. Desse modo, ao se deparar com dois (ou mais modelos) é natural questionarmos: Dentre estes modelos qual é o mais adequado? Um bom modelo deve conseguir equilibrar a qualidade do ajuste e a complexidade, sendo esta, em geral, medida pelo número de parâmetros presentes neste; quanto mais parâmetros, mais complexo será, sendo pois, mais difícil explicá-lo. Assim, a seleção do melhor modelo torna-se, então, evidente. Diversas são as metodologias utilizadas para este fim, dentre eles os critérios de informação de Akaike (AIC), critério de informação de Akaike corrigido (AICc), e critério de informação bayesiano (BIC). Nesta perspectiva, o objetivo deste trabalho é utilizar e avaliar os critérios AIC, AICc e BIC em diferentes áreas: seleção de modelos normais, séries temporais e modelos de crescimento. Estudos de simulação em cada uma das três áreas foram desenvolvidas, bem como uma aplicação a dados reais foi realizada em cada uma das áreas.

Palavras-chave: Seleção de modelos. Critérios de informação. Séries Temporais. Curvas de crescimento. Teste de médias.

ABSTRACT

A phenomenon can be explained by models, which are intended to illustrate certain aspects of the problem, without, however, represent all details and consequently no loss of information. To avoid compromising the understanding of the phenomenon under study, this loss should be minimal. Often more than one model can describe the same phenomenon, given that there is no single methodology to be followed. Thus, when faced with two (or more models) is natural questioning: Among these models which is the most suitable? A good model should be able to balance the quality of fit and complexity, which is generally measured by the number of parameters present here, as more parameters, is more complex, and therefore more difficult to explain. Thus, the selection of the best model becomes then evident. There are several methods used for this purpose, including the Akaike information criterion (AIC), Akaike information criterion corrected (AICc) and Bayesian information criterion (BIC). In this perspective, the aim of this work is to use and evaluate the criteria AIC, AICc and BIC in different areas: selection of standard models, time series and growth models. Simulation studies in each of the three areas were developed, as an application to real data is performed in each of the areas.

Keywords: Model selection. Information criteria. Times series. Growth curves. Mean test.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

CAPÍTULO 1

Introdução geral

1 INTRODUÇÃO

Em geral, um fenômeno em estudo pode ser explicado através de modelos, sendo que estes são os principais instrumentos utilizados na estatística e constituem uma versão simplificada de algum problema ou situação da vida real, sendo destinados a ilustrar certos aspectos do problema, sem contudo, representar todos os detalhes.

Em situações práticas, o completo conhecimento do fenômeno não acontece. Geralmente os fenômenos observados são muito complexos, sendo portanto, impraticável descrever tudo aquilo que é observado com total exatidão. Neste sentido, devido às dificuldades na descrição exata dos fenômenos observados na forma de simbologias e fórmulas matemáticas, são utilizados os modelos estocásticos que possuem uma parte aleatória e outra sistemática. Entretanto, na representação de um fenômeno por um modelo probabilístico, há perda de informação. Para não comprometer o entendimento do fenômeno em estudo, esta perda deve ser mínima, caso contrário, nosso modelo não conseguirá explicar, de forma satisfatória, o entendimento do mesmo.

Não raro, mais de um modelo pode descrever um mesmo fenômeno, haja vista que não há uma única metodologia a ser seguida, já que cada pesquisador tem a liberdade de modelar o fenômeno seguindo aquela que julgar mais adequada.

Desse modo, ao se deparar com dois (ou mais modelos) é natural questionarmos: "Dentre estes modelos qual é o mais adequado?". Mazerolle (2004) afirma que o conceito de melhor modelo é controverso, e que um bom modelo deve conseguir equilibrar a qualidade do ajuste e a complexidade, sendo esta, em geral, medida pelo número de parâmetros presentes neste; quanto mais parâmetros, mais complexo será, sendo pois, mais difícil explicá-lo. torna-se, então, evidente.

Burnham e Anderson (2004), enfatizam a importância de selecionar modelos baseando-se em princípios científicos. Diversas são as abordagens utilizadas para este fim, tais como C_p de Mallows, critério de informação de Akaike (AIC) (AKAIKE, 1974), critério de informação de Akaike corrigido (AICc) (SUGIURA, 1978), critério de informação generalizado (GIC) (KONISHI; KITAGAWA, 2008), critério de informação bayesiano (BIC) (SCHWARZ, 1978), dentre outros.

Os critérios de informação são utilizados nas mais diversas áreas das ciências, sendo que os critérios AIC, AICc e BIC são os mais conhecidos e aplicados, sendo implementados na maioria dos softwares estatísticos.

Nos critérios AIC, AICc e BIC, de cada modelo é obtido um valor e aquele que apresentar a menor magnitude é considerado como o "melhor" modelo. Um questionamento recorrente que podemos fazer é: "Por que o critério com menor AIC (AICc ou BIC) é selecionado?".

Nesta perspectiva, o objetivo deste trabalho é utilizar e avaliar os critérios AIC, AICc e BIC em diferentes áreas: seleção de modelos normais, séries temporais e modelos de crescimento.

Para tanto, o trabalho está dividido em quatro capítulos e é organizado da seguinte maneira: No capítulo 1, apresentamos o referencial teórico com alguns conceitos básicos necessários para desenvolvimento dos capítulos 2 a 4. Consistirão em uma breve revisão sobre os critérios de informação, em que são apresenta-

dos alguns conceitos inerentes aos mesmos, tais como, sua utilização, importância e interpretação.

No capítulo 2, os critérios de informação serão utilizados em modelos normais. Nele haverá uma breve revisão acerca destes, em que apresentaremos as simulações referentes a esses tipos de modelos, além de que uma aplicação a dados reais será feita.

No capítulo 3, os critérios de informação serão utilizados em modelos de séries temporais. Haverá uma breve revisão acerca destes, sobre as quais apresentaremos as simulações referentes a esses tipos de modelos; uma aplicação a dados reais também será feita.

No capítulo 4, os critérios de informação serão utilizados em modelos de crescimento. Neste último capítulo haverá uma breve revisão acerca de tais modelos, a partir dos quais apresentaremos as simulações referentes a esses tipos de modelos; uma aplicação a dados reais será realizada.

2 REFERENCIAL TEÓRICO

Um modelo é uma versão simplificada de algum problema ou situação da vida real, destinado a ilustrar certos aspectos do mesmo sem levar em conta todos os detalhes. Se o modelo captura todos os aspectos inerentes ao fenômeno, temos um modelo determinístico; se, por outro lado, o modelo não captura todos os aspectos, temos um modelo estocástico. Antes da construção de modelos, é preciso que saibamos que não existem modelos verdadeiros, existem, apenas, modelos aproximados da realidade que causam perda de informações.

Faz-se necessário, então, minimizarmos esta perda. Box e Draper (1987) fizeram uma famosa afirmativa acerca disso: "Todos os modelos são errados, mas alguns são úteis" [1]. Deste modo, é necessário fazer a seleção do "melhor" modelo, dentre aqueles que foram ajustados, para explicar o fenômeno sob estudo.

De acordo com Mazerolle (2004), a seleção de modelos é a tarefa de escolher um modelo estocástico de um conjunto de modelos plausíveis. Em sua forma mais básica, esta é uma das tarefas fundamentais das pesquisas científicas nas mais diversas áreas, tais como Biologia, Agronomia, Engenharia, dentre outras. Mas, dos tantos modelos plausíveis que se poderia ajustar aos dados, como é possível escolher um bom modelo? Seria, pois, ingênuo esperarmos que os melhores resultados incluam todas as variáveis no mesmo. Isto violaria o princípio científico fundamentado na parcimônia, que requer que, dentre todos os modelos que expliquem bem os dados, o mais simples deva ser escolhido. Assim, é necessário encontrar um modelo mais simples que explique bem o fenômeno em estudo.

Para quantificar a informação perdida ao ajustarmos um modelo, existem diversas medidas propostas na literatura específica, como, por exemplo:

[1] "All models are wrong but some are useful". Tradução literal nossa.

1- A Estatística de χ^2 para aderência, dada por:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - g_i)^2}{f_i}.$$

2- A distância de Hellinger, dada por:

$$I_K(g; f) = \int_{-\infty}^{+\infty} \left\{ \sqrt{f(x)} - \sqrt{g(x)} \right\}^2 dx.$$

3- A informação generalizada, dada por:

$$I_\lambda(g; f) = \frac{1}{\lambda} \int_{-\infty}^{+\infty} \left\{ \left(\frac{g(x)}{f(x)} \right)^\lambda - 1 \right\} g(x)\, dx.$$

4- O critério deviance, dado por:

$$D(\theta) = -2 \left[\log L(\theta; x) - \log L\left(\widehat{\theta}; x \right) \right],$$

em que $\theta \in \Theta$, sendo que, Θ é o espaço paramétrico e $\widehat{\theta} \in \widehat{\Theta}$, sendo que, $\widehat{\Theta}$ é o espaço restrito.

5- A divergência, dada por:

$$D(g; f) = \int_{-\infty}^{+\infty} u(t(x)) g(x)\, dx = \int_{-\infty}^{+\infty} u\left(\frac{g(x)}{f(x)} \right) g(x)\, dx,$$

sendo que $t(x) = \dfrac{g(x)}{f(x)}$.

6- A $L^1 - norm$, dada por:

$$L_1(g; f) = \int_{-\infty}^{+\infty} |g(x) - f(x)| dx.$$

7- A $L^2 - norm$, dada por:

$$L_2(g; f) = \int_{-\infty}^{+\infty} \{g(x) - f(x)\}^2 dx.$$

8- A informação de Kullback-Leibler, dada por:

$$I(g; f) = E_g\left[\log\left(\frac{g(X)}{f(X)}\right)\right] = \int_{-\infty}^{+\infty} g(x) \log\left(\frac{g(x)}{f(x|\theta)}\right) dx, \tag{1}$$

sendo f, g, f_i e g_i funções densidades quaisquer, θ o parâmetro (ou vetor de parâmetros), $k \in \mathbb{N}^*$, $\lambda \in \mathbb{R}_+^*$, $x \in \mathbb{R}$ e $u(x) \geq 0$ uma função.

De acordo com Mazerolle (2004), Kullback e Leibler definiram a medida (1), posteriormente chamada Informação de Kullback-Leibler (K-L), para representar a informação perdida pela aproximação do modelo ajustado da realidade, e, segundo ele, muitos autores mostram que esta é uma medida natural para discriminar entre duas funções de probabilidade.

Akaike (1974) usou a informação de Kullback-Leibler e propriedades assintóticas dos estimadores de máxima verossimilhança para definir seu critério de informação.

A informação de Kullback-Leibler, dada pela equação (1) é usada em diversas áreas, tais como: (i) mecânica estatística não extensiva, que, segundo Tsallis (2009), estuda o comportamento estatístico de sistemas complexos através da função entropia $S(\{p\})$ e, (ii) em computação, no reconhecimento de padrões

(BISHOP, 2006), análise de imagens, dentre outros.

Fisicamente, a entropia $S(\{p\})$ quantifica a "desordem" de um sistema físico caracterizado por uma distribuição de probabilidade dos estados fisicamente acessíveis. Matematicamente, a entropia $S(\{p\})$ é dada por:

$$S(\{p\}) = -\sum_{i=1}^{W} p_i \ \log(p_i) \ , \text{com} \quad \sum_{i=1}^{W} p_i = 1, \tag{2}$$

em que W é o número de estados acessíveis. Quando o sistema está num estado bem definido (isto é, ordenado), a distribuição de probabilidade $(\{p\})$, associada ao conjunto de estados acessíveis, apresenta-se concentrada em certos valores próximos da unidade, o que resulta em um pequeno valor de $S(\{p\})$. Em particular, quando o estado acessível é único, a probabilidade desse estado (i_0) é unitário (isto é, $p_{i_0} = 1$) e a probabilidade de acessar qualquer outro estado $i \neq i_0$ é zero, que resulta num valor nulo (isto é, o valor mínimo de $S(\ p\ )$). Em outros termos, temos

$$S(p_{i_0} = 1, p_i = 0 \ \ \forall \ \ i \neq i_0) = 0.$$

Por outro lado, quando todos os possíveis estados têm uma probabilidade (igual) uniforme de acesso, como

$$p_i = \frac{1}{W}, \tag{3}$$

substituindo-se a equação (3) na equação (2), obtemos que o valor da entropia $S(\{p\})$ é máximo.

$$S(p_i) = -\sum_{i=1}^{W} \frac{1}{W} \ \log\left(\frac{1}{W}\right) = \log W.$$

A equação (1) pode também ser expressa como:

$$I\left(g;f\right)=\int_{-\infty}^{+\infty}g\left(x\right)\log\left[g\left(x\right)\right]dx-\int_{-\infty}^{+\infty}g\left(x\right)\log\left[f\left(x|\boldsymbol{\theta}\right)\right]dx.$$

Deste modo, para dois modelos $f_1(x|\boldsymbol{\theta})$ e $f_2(x|\boldsymbol{\theta})$, é possível obter de (1) que

$$I\left(f_1,g\right)=\int_{-\infty}^{+\infty}g\left(x\right)\log\left(g\left(x\right)\right)dx-\int_{-\infty}^{+\infty}g\left(x\right)\log\left(f_1\left(x|\boldsymbol{\theta}\right)\right)dx \tag{4}$$

e

$$I\left(f_2,g\right)=\int_{-\infty}^{+\infty}g\left(x\right)\log\left(g\left(x\right)\right)dx-\int_{-\infty}^{+\infty}g\left(x\right)\log\left(f_2\left(x|\boldsymbol{\theta}\right)\right)dx. \tag{5}$$

Note que nas equações (4) e (5), somente o segundo termo de cada é importante na avaliação dos modelos $f_1(x)$ e $f_2(x)$, pois o primeiro termo de ambas é o mesmo. Entretanto, os segundos termos dependem também da função densidade de probabilidade desconhecida g.

O segundo termo é a esperança matemática de uma função de variável aleatória X com função densidade de probabilidade $g(x)$, ou em outros termos,

$$E_g\left[\log f\left(X\right)\right]=\int g\left(x\right)\log\left(f\left(x|\boldsymbol{\theta}\right)\right)dx. \tag{6}$$

Conforme Akaike (1974), a K-L informação é apropriada para testar se um dado modelo é adequado. Porém, seu uso é limitado, pois depende da distribuição g, que é desconhecida. Se uma boa estimativa para a log verossimilhança esperada, dada pela equação (6), puder ser obtida através dos dados, esta poderá ser utilizada como um critério para comparar modelos.

Segundo Konishi e Kitagawa (2008), sob certas condições de regularidade, os estimadores de máxima verossimilhança (EMV) são assintoticamente eficien-

tes, pois a função de verossimilhança tende a ser mais sensível a pequenos desvios dos parâmetros do modelo de seus valores verdadeiros. Os estimadores de máxima verossimilhança, por suas propriedades assintóticas, se constituem como bons. Nesse sentido, utilizando estimadores de máxima verossimilhança, estima-se o vetor de parâmetros $\boldsymbol{\theta}$ em $f(x|\boldsymbol{\theta})$, obtendo $f(x|\widehat{\boldsymbol{\theta}})$, que será substituído na equação (6), a fim de encontrarmos:

$$E_g\left[\log f\left(X|\widehat{\boldsymbol{\theta}}\right)\right] = \int g\left(x\right) \log\left[f\left(x|\widehat{\boldsymbol{\theta}}\right)\right] dx. \tag{7}$$

O objetivo é encontrarmos um estimador de máxima verossimilhança para a equação (7). De acordo com Konishi e Kitagawa (2008), uma estimativa da função suporte esperada pode ser obtida pela substituição da função de distribuição desconhecida G, na equação (7), pela função de distribuição empírica $\widehat{G}$ baseada nos dados.

A fim de encontrar tal estimativa, são necessárias as seguintes definições:

Definição 2.1 *Seja $A = \{x_1, x_2, ..., x_n\}$ uma amostra aleatória de tamanho n de uma distribuição $G(x)$. A função de distribuição empírica $\widehat{G}$ é a função densidade acumulada que dá $\dfrac{1}{n}$ de probabilidade para cada x_i. Formalmente, temos*

$$\widehat{G}_n\left(x\right) = \frac{1}{n}\sum_{i=1}^{n} I\left(x_i \leq x\right),$$

em que

$$I\left(x_i \leq x\right) = \begin{cases} 1, \ se\ x_i \leq x \\ 0, \ se\ x_i > x. \end{cases}$$

Definição 2.2 *Um funcional estatístico $T(G)$ é qualquer função de G, em que G é uma distribuição e T uma função qualquer.*

Um estimador para $\theta = T(G)$ é dado por $\widehat{\theta}_n = \widehat{G}_n$. Se um funcional pode ser escrito na forma $T(G) = \int u(x) dG(x)$, Konishi e Kitagawa (2008) mostram que o estimador correspondente é dado por

$$\begin{aligned} T\left(\widehat{G}\right) &= \int u(x) d\widehat{G}(x) = \sum_{i=1}^{n} \widehat{g}(x_i)\, u(x_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} u(x_i), \end{aligned} \tag{8}$$

ou seja, é feita a substituição da função densidade de probabilidade acumulada G pela função de distribuição acumulada empírica $\widehat{G}$, e a função densidade $\widehat{g}_n = \frac{1}{n}$ para cada observação x_i.

Da equação (8), é possível estimar a função suporte esperada por:

$$\begin{aligned} E_{\widehat{G}}\left[\log f\left(x|\widehat{\boldsymbol{\theta}}\right)\right] &= \int \log f\left(x|\widehat{\boldsymbol{\theta}}\right) d\widehat{G}(x) \\ &= \sum_{i=1}^{n} \widehat{g}\left(x_i|\widehat{\boldsymbol{\theta}}\right) \log f(x_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} \log f\left(x_i|\widehat{\boldsymbol{\theta}}\right). \end{aligned}$$

2.1 Critérios de informação

Um modo de comparar n modelos, $g_1(x|\boldsymbol{\theta_1}), g_2(x|\boldsymbol{\theta_2}), \cdots, g_n(x|\boldsymbol{\theta_n})$, é simplesmente comparar as magnitudes da função suporte maximizada $L\left(\widehat{\boldsymbol{\theta}_i}\right)$. Tal método não fornece uma verdadeira comparação, haja vista que, não conhecendo o verdadeiro modelo $g(x)$, primeiramente o método da máxima verossimilhança estima os parâmetros $\boldsymbol{\theta_i}$ de cada modelo $g_i(x)$, $i = 1, 2, \cdots, n$ e posteriormente são utilizados os mesmos dados para estimar $E_G\left[\log\left(f(x|\widehat{\boldsymbol{\theta}})\right)\right]$; isto introduz um

viés em $L\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}\right)$, sendo que, a magnitude deste viés $(b\,(G))$ varia de acordo com a dimensão do vetor de parâmetros, sendo esta dada por Konishi e Kitagawa (2008):

$$b\,(G) = E_{G(\boldsymbol{x_n})}\left[\log f\left(\boldsymbol{x_n}|\widehat{\boldsymbol{\theta}}\,(x_n)\right) - nE_{G(Z)}\left[\log f\left(Z|\widehat{\boldsymbol{\theta}}\,(x_n)\right)\right]\right],$$

em que a esperança é tomada com relação à distribuição conjunta, e Z é a verdadeira distribuição de onde os dados são gerados, X_n são os dados e, $\widehat{\boldsymbol{\theta}}\,(x_n)$ é o vetor estimado por máxima verossimilhança como estimador de $\boldsymbol{\theta}$.

Desse modo, os critérios de informação são construídos para avaliar e corrigir o viés da função suporte. Segundo Konishi e Kitagawa (2008), um critério de informação (IC) tem a forma que se segue:

$$\begin{aligned} IC\left(\boldsymbol{X_n}, \widehat{G}\right) &= -2\sum_{i=1}^{n}\log f\left(X_i|\widehat{\boldsymbol{\theta}}\,(X_n)\right) + 2\,(\text{viés}) \\ &= -2\sum_{i=1}^{n}\log f\left(X_i|\widehat{\boldsymbol{\theta}}\,(X_n)\right) + 2\,(b\,(G))\,. \end{aligned} \tag{9}$$

Diversos critérios podem também serem utilizados, para a seleção de modelos. Em geral, esses critérios consideram a complexidade do modelo no critério de seleção, e, essencialmente, penalizam a verossimilhança utilizando o número de parâmetros do modelo, além de eventualmente, o tamanho da amostra. Esta penalização é feita subtraindo do valor da verossimilhança uma determinada quantidade, que depende do quão complexo é o modelo (quanto mais parâmetros, mais complexo).

Akaike (1974), propôs utilizar a informação de Kullback-Leibler para a seleção de modelos estabelecendo uma relação entre a máxima verossimilhança e a informação de Kullback-Leibler, desenvolvendo, então, um critério para estimar a informação de Kullback-Leibler, que posteriormente foi chamado de critério de

informação de Akaike (AIC).

Critérios de seleção de modelos como o critério de informação de Akaike (AIC), critério de informação de Akaike corrigido (AICc) e critério de informação bayesiano (BIC), são frequentemente utilizados para selecionar modelos em diversas áreas. Segundo esses critérios, o melhor modelo será aquele que apresentar menor valor de AIC, AICc ou BIC.

2.1.1 Critério de informação de Akaike - AIC

O critério de informação de Akaike (AIC), desenvolvido por Hirotugu Akaike sob o nome de "um critério de informação", em 1971 e proposto, em Akaike (1974), é uma medida relativa da qualidade de ajuste de um modelo estocástico estimado. Fundamentado no conceito de informação, ele oferece uma medida relativa das informações perdidas, quando um determinado modelo é utilizado para descrever a realidade. Akaike (1974) encontrou uma relação entre a esperança relativa da K-L informação e a função suporte maximizada, permitindo uma maior interação entre a prática e a teoria, em seleção de modelos e análises de conjuntos de dados complexos. Dessa forma, mostrou que o viés é dado assintoticamente por:

$$b\left(G\right) \quad = \quad tr\left\{ I\left(\boldsymbol{\theta}_0\right) J\left(\boldsymbol{\theta}_0\right)^{-1} \right\}, \tag{10}$$

sendo que $J\left(\boldsymbol{\theta_0}\right)$ é a matriz de informação de Fisher de g e $I\left(\boldsymbol{\theta}_0\right)$ é dado por

$$I\left(\boldsymbol{\theta_0}\right) = \int g\left(x\right) \frac{\partial f\left(x|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} \frac{\partial f\left(x|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta^T}} dx.$$

O AIC é um critério que avalia a qualidade do ajuste do modelo paramétrico, estimado pelo método da máxima verossimilhança; tendo fundamentos

ligados ao fato de que o viés (10) tende ao número de parâmetros a serem estimados no modelo, pois sob a suposição de que existe um $\boldsymbol{\theta}_0$ no espaço paramétrico Θ tal que $g(x) = f(x|\boldsymbol{\theta}_0)$, ocorre a igualdade das expressões $I(\boldsymbol{\theta}_0) = J(\boldsymbol{\theta}_0)$, e assim obtém-se na equação (10) que:

$$\begin{aligned} b(G) &= E_{G(\boldsymbol{x_n})}\left[\log f\left(\boldsymbol{X_n}|\widehat{\boldsymbol{\theta}}(X_n)\right) - nE_{G(Z)}\left[\log f\left(Z|\widehat{\boldsymbol{\theta}}(X_n)\right)\right]\right] \\ &= tr\left\{I(\boldsymbol{\theta_0})\, J(\boldsymbol{\theta_0})^{-1}\right\} = tr(I_p) = p, \end{aligned} \tag{11}$$

em que p é o número de parâmetros a serem estimados no modelo.

Substituindo o resultado obtido em (11) na expressão (9), Akaike (1974) definiu seu critério de informação da seguinte forma,

$$\begin{aligned} AIC &= -2\,(\text{função suporte maximizada}) + 2\,(\text{número de parâmetros}) \\ &= -2\log L\left(\widehat{\boldsymbol{\theta}}\right) + 2\,(p)\,. \end{aligned}$$

A derivação completa do resultado acima pode ser encontrada, por exemplo, em Emiliano (2009) e Konishi e Kitagawa (2008).

2.1.2 Critério de informação de Akaike corrigido - AICc

O critério de informação de Akaike (AIC) deriva da informação de Kullback-Leibler. Sugiura (1978) derivou uma variante de segunda ordem do AIC, chamado de critério de informação de Akaike corrigido (AICc), dado por:

$$AICc = -2\log L\left(\widehat{\boldsymbol{\theta}}\right) + 2p\left(\frac{n}{n-p-1}\right). \tag{12}$$

Esta expressão pode ser reescrita equivalentemente (SUGIURA, 1978) como:

$$AICc = AIC + \frac{2p\,(p+1)}{n-p-1}, \tag{13}$$

em que n é o tamanho amostral e p é o número de parâmetros do modelo.

De acordo com Sakamoto, Ishiguro e Kitagawa (1986) e Sugiura (1978), o critério de Akaike pode ter um desempenho ruim se existirem muitos parâmetros, em comparação com o tamanho da amostra. Assim, do exposto, podemos notar que o AICc é apenas uma correção, de segunda ordem, do viés do AIC. Segundo Burnham e Anderson (2004) a sua utilização é indicada quando a razão $\frac{n}{p}$ é pequena, (isto é, $\frac{n}{p} < 40$). Além disso, quando essa relação é suficientemente grande, ambos os critérios, AIC e AICc, apresentam resultados semelhantes.

2.1.3 Critério de informação bayesiano - BIC

O critério de informação bayesiano (BIC), também chamado de critério de Schwarz, foi proposto por Schwarz (1978), e consiste em um critério de avaliação de modelos definido em termos da probabilidade *a posteriori*, sendo assim chamado porque o autor deu um argumento bayesiano para prová-lo. A seguir, serão dados alguns conceitos que culminarão em um esboço da prova dada por Schwarz.

Teorema 1 *Seja $B_1, B_2, \ldots, B_n$ uma coleção de eventos mutuamente excludentes, satisfazendo $\Omega = \bigcup_{j=1}^{n} B_j$ e $P[B_j] > 0$, para $j = 1, 2, \ldots, n$, então para todo $A \in \mathcal{A}$, sendo $\mathcal{A}$ o espaço de eventos, tal que $P[A] > 0$, temos:*

$$P[B_k|A] = \frac{P[A|B_k]\,P[B_k]}{\sum_{j=1}^{n} P[A|B_j]\,P[B_j]}, \tag{14}$$

em que Ω é o espaço amostral.

Conforme Konishi e Kitagawa (2008), sejam $M_1, M_2, \ldots, M_k$, k modelos candidatos, cada um dos modelos M_i com uma distribuição de probabilidades $f_i\left(x|\theta_i\right)$ e uma *priori*, $\pi_i\left(\boldsymbol{\theta_i}\right)$ para o k_i-ésimo vetor θ_i. Se houverem n observações dadas $\boldsymbol{x_n} = \{x_1, x_2, \ldots, x_n\}$, então para o $i-$ésimo modelo M_i, a distribuição marginal de $\boldsymbol{x_n}$ é dada por:

$$p_i(x_n) = \int f_i\left(\boldsymbol{x_n}|\boldsymbol{\theta_i}\right) \pi_i\left(\boldsymbol{\theta_i}\right) d\boldsymbol{\theta_i}. \tag{15}$$

Essa quantidade pode ser considerada como a verossimilhança para o i-ésimo modelo e será referida como verossimilhança marginal dos dados.

Sendo $P(M_i)$ a distribuição *a priori* do i-ésimo modelo, da equação (14), a distribuição *a posteriori* do i-ésimo modelo é:

$$P\left(M_i|\boldsymbol{x_n}\right) = \frac{p_i\left(\boldsymbol{x_n}\right) P\left(M_i\right)}{\sum\limits_{j=1}^{n} p_j\left(\boldsymbol{x_n}\right) P\left(M_j\right)}. \tag{16}$$

A probabilidade a posteriori, $P\left(M_i|\boldsymbol{x_n}\right)$, indica a probabilidade dos dados serem gerados do i-ésimo modelo quando os dados $\boldsymbol{x_n}$ são observados. Se um modelo é selecionado de r modelos, o mais natural é adotar o modelo que tenha a maior probabilidade *a posteriori*. Esse princípio mostra que o modelo que maximiza o numerador $p_i\left(\boldsymbol{x_n}\right) P\left(M_i\right)$ deve ser selecionado, pois todos os modelos compartilham do mesmo denominador na equação (16).

Se as distribuições *a priori* $P(M_i)$ são iguais em todos os modelos, então o modelo que maximiza a probabilidade marginal dos dados $p_i(\boldsymbol{x_n})$ deve ser selecionado. Assim, se uma aproximação para a probabilidade marginal expressa em termos da integral na equação (15) puder ser obtida, a necessidade básica de en-

contrar a integral problema-por-problema desaparece; isto faz do BIC um critério satisfatório para seleção de modelos.

O BIC é definido como:

$$\begin{aligned} BIC &= -2\log p_i\left(\boldsymbol{x_n}\right) \\ &= -2\log \int f_i\left(\boldsymbol{x_n}|\boldsymbol{\theta_i}\right)\pi_i\left(\boldsymbol{\theta_i}\right)d\boldsymbol{\theta_i} \\ &\approx -2\log f_i\left(\boldsymbol{x_n}|\widehat{\boldsymbol{\theta_i}}\right) + k_i\log n, \end{aligned} \tag{17}$$

em que $\widehat{\boldsymbol{\theta_i}}$ é o estimador de máxima verossimilhança para o k_i-ésimo vetor paramétrico θ_i do modelo $f_i\left(\boldsymbol{x_n}|\boldsymbol{\theta_i}\right)$.

Consequentemente, dos r modelos avaliados ao utilizarmos o método da máxima verossimilhança, aquele que minimizar o valor do BIC será o melhor modelo para os dados.

Dessa maneira, sob a suposição de que todos os modelos tenham distribuição de probabilidades, *a priori* iguais, a probabilidade *a posteriori*, obtida usando a informação dos dados, poderá servir para comparar os modelos e ajudar na identificação do modelo que gerou os dados.

Sejam M_1 e M_2 dois modelos que tenhamos interesse em comparar. Para cada modelo existem as verossimilhanças marginais $p_i(\boldsymbol{x_n})$, as prioris $P(M_i)$ e as posterioris $P(M_i|\boldsymbol{x_n})$ com $i = \{1, 2\}$, sendo então, a razão *a posteriori* em favor do modelo M_1 *versus* o modelo M_2 é:

$$\frac{P\left(M_1|\boldsymbol{x_n}\right)}{P\left(M_2|\boldsymbol{x_n}\right)} = \frac{\frac{p_1(\boldsymbol{x_n})P(M_1)}{\sum_{j=1}^{n} p_j(\boldsymbol{x_n})P(M_j)}}{\frac{p_2(\boldsymbol{x_n})P(M_2)}{\sum_{j=1}^{n} p_j(\boldsymbol{x_n})P(M_j)}} = \frac{p_1\left(\boldsymbol{x_n}\right)P\left(M_1\right)}{p_2\left(\boldsymbol{x_n}\right)P\left(M_2\right)}.$$

A razão

$$\frac{p_1\left(\boldsymbol{x_n}\right)}{p_2\left(\boldsymbol{x_n}\right)} \tag{18}$$

é chamada de **fator de Bayes**.

O problema em encontrar o valor do BIC reside no fato de calcular o valor da integral na equação (15). Isso é feito utilizando a aproximação de Laplace para integrais. Neste sentido, temos:

i) **A aproximação de Laplace para integrais**

É necessária a aproximação de Laplace para a integral

$$\int \exp\left\{nq\left(\boldsymbol{\theta}\right)\right\} d\boldsymbol{\theta}, \tag{19}$$

em que $\boldsymbol{\theta}$ é um vetor de parâmetros p-dimensional e $q(\boldsymbol{\theta})$ é uma função real p-dimensional.

A grande vantagem da aproximação de Laplace é o fato de que quando o número n de observações é grande, o integrando concentra-se em uma vizinhança $\widehat{\boldsymbol{\theta}}$ de $q(\boldsymbol{\theta})$, e, consequentemente, o valor da integral depende somente do comportamento do integrando na vizinhança de $\widehat{\boldsymbol{\theta}}$, sendo

$$\begin{aligned} \left.\frac{\partial q\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= 0 \\ I_q\left(\widehat{\boldsymbol{\theta}}\right) = -\left.\frac{\partial^2 q\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &\neq 0. \end{aligned} \tag{20}$$

Se $q\left(\boldsymbol{\theta}\right)$ é uma função de valores reais avaliada em torno de $\widehat{\boldsymbol{\theta}}$, e $\boldsymbol{\theta}$ é um vetor de parâmetros, então a **aproximação de Laplace** para a integral é dada por:

$$\int \exp\left\{nq\left(\boldsymbol{\theta}\right)\right\} d\boldsymbol{\theta} \approx \frac{\left(2\pi\right)^{p/2}}{\left(n\right)^{p/2}\left|J_q\left(\widehat{\boldsymbol{\theta}}\right)\right|^{p/2}} \exp\left(nq\left(\widehat{\boldsymbol{\theta}}\right)\right), \tag{21}$$

em que $I_q\left(\widehat{\boldsymbol{\theta}}\right)$ é definido na equação (20).

Utilizando a aproximação de Laplace para determinarmos o valor da integral na equação (15), esta pode ser reescrita como

$$\begin{aligned} p(x_n) &= \int f_i\left(\boldsymbol{x_n}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta} \\ &= \int \exp\left\{\log f\left(\boldsymbol{x_n}|\boldsymbol{\theta}\right)\right\}\pi\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta} \\ &= \int \exp\left\{\ell\left(\boldsymbol{\theta}\right)\right\}\pi\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta}, \end{aligned} \tag{22}$$

em que $\ell\left(\boldsymbol{\theta}\right)$ é a função suporte, isto é, $\ell\left(\boldsymbol{\theta}\right) = \log f\left(x_n|\boldsymbol{\theta}\right)$.

Utilizando a equação (21) em (22) obtém-se, para n grande,

$$p(x_n) \approx \exp\left\{\ell\left(\widehat{\boldsymbol{\theta}}\right)\right\}\pi\left(\widehat{\boldsymbol{\theta}}\right)(2\pi)^{p/2}\, n^{-p/2}\left|J\left(\widehat{\boldsymbol{\theta}}\right)\right|^{-1/2}. \tag{23}$$

Ao tomar o logaritmo na equação (23) e multiplicar a expressão por -2 obtém-se:

$$-2\log p\left(x_n\right) = -2\ell\left(\widehat{\boldsymbol{\theta}}\right) + p\,\log n + \log\left|J\left(\widehat{\boldsymbol{\theta}}\right)\right| - p\,\log\left(2\pi\right) - 2\log\pi\left(\widehat{\boldsymbol{\theta}}\right).$$

Assim, o critério de informação bayesiano pode ser obtido da seguinte forma (ignorando os termos constantes na equação):

Definição 2.3 *Seja $f\left(\boldsymbol{x_n}|\widehat{\boldsymbol{\theta}}\right)$ um modelo estocástico estimado através do método da máxima verossimilhança. Então, o critério de informação bayesiano (BIC) é dado por:*

$$BIC = -2\log L\left(\widehat{\boldsymbol{\theta}}\right) + p\log\left(n\right), \tag{24}$$

em que $L\left(\widehat{\boldsymbol{\theta}}\right)$ é a verossimilhança do modelo ajustado, p é o número de parâmetros a serem estimados e n é o número de observações da amostra.

2.2 Taxas de falso positivo (FP), falso negativo (FN) e verdadeiro positivo (TP)

2.2.1 Um pouco da história

Testes de hipóteses foram realizados desde os primórdios da civilização sem a formulação estatística necessária. Realizar um teste de hipóteses estatístico é, pois, uma prática relativamente recente. Segundo Lehmann (1993), a formulação e a filosofia dos testes de hipóteses, como a conhecemos hoje, foram, em grande parte, criada no período 1915 - 1933 por três estatísticos: R. A. Fisher (1890-1962), J. Neyman (1894-1981), e E. S. Pearson (1895-1980).

Ao realizarmos um teste de hipóteses, não teremos, na prática o teste "ideal", que, segundo Mood, Graybill e Boes (1974), é aquele que nunca rejeita uma hipótese H_0 se ela, de fato, for verdadeira; e sempre rejeita H_0, se ela de fato for falsa. Estaremos, então, sujeitos a cometer dois erros, conforme mostrado na Tabela 1, adaptada de Triola (2008).

Tabela 1 Erros tipo I e tipo II

		A hipótese nula é verdadeira	A hipótese nula é falsa
Decisão	Decidimos rejeitar a hipótese nula	Erro tipo I	Decisão correta
	Decidimos não rejeitar a hipótese nula	Decisão correta	Erro tipo II

Um teste ideal não apresentaria erros de nenhuma espécie, ou seja, seria um teste que possuiria probabilidades de erros tipo I e tipo II iguais a zero. Contudo, encontrar um teste com essa propriedade é inviável e uma possível solução seria encontrar testes que apresentem pequenas probabilidades de erros. De acordo com a abordagem proposta por Neyman-Pearson, uma maneira de encontrar um teste razoável é fixar a probabilidade máxima de erro tipo I, em que o decisor está

disposto a cometer e tentar encontrar um teste com a menor probabilidade de erro tipo II (MOOD; GRAYBILL; BOES, 1974).

No dia a dia, esta prática é utilizada na medicina, por exemplo. Uma das experiências mais rotineiras da prática médica é a solicitação de um teste diagnóstico. Os objetivos são vários, incluindo a triagem de paciente, o diagnóstico de doenças e o acompanhamento ou prognóstico da evolução do mesmo. Para chegar ao diagnóstico, o médico considera várias possibilidades, com níveis de certeza que variam de acordo com as informações disponíveis. Assim, podemos perceber que não existe teste perfeito aquele que com certeza absoluta determina a presença ou ausência da doença. Com base nos resultados obtidos nos exames, o médico diagnostica o paciente, podendo acertar ou não em seu diagnóstico; este é um exemplo típico da chamada análise ROC (VAZ, 2009).

Conforme Vaz (2009), a análise ROC teve origem na teoria da decisão estatística, desenvolvida para avaliar a detecção de sinais. Durante a Segunda Guerra Mundial, a teoria da detecção do sinal começou a ser utilizada pelos operadores de radares. Neste contexto, surgiu a curva ROC, que é uma ferramenta destinada a descrever o desempenho desses operadores (chamados de receiver operators). A curva ROC tinha como objetivo quantificar a habilidade dos operadores em distinguir um sinal de um ruído (REISER; FARAGGI, 1997). Essa habilidade ficou conhecida como *receiver operating characteristic* (ROC). Quando o radar detectava algo se aproximando, o operador decidia se o que foi captado era um avião inimigo (sinal), ou algum outro objeto irrelevante, como por exemplo, uma nuvem ou um bando de aves (ruído).

Durante a detecção do sinal, o problema envolvendo sinais eletromagnéticos na presença de ruídos foi tratado como um problema de teste de hipóteses. A hipótese nula (H_0) foi identificada como sendo o ruído, enquanto a hipótese

alternativa (H_1) estava associada ao ruído acrescido do sinal. O erro tipo I era um alarme qualquer (falso), enquanto o erro tipo II eram as falhas; ambos perigosos em relação à defesa.

Na teoria de detecção do sinal, cabe ao observador decidir, com base na aleatoriedade, qual dos estímulos é resultado somente do ruído ou do ruído acrescido do sinal. Assim, o problema da detecção do sinal pode ser visto da seguinte forma:

i) Existe uma ocorrência aleatória de dois acontecimentos: ruído e sinal;

ii) Cada acontecimento ocorre num intervalo de tempo bem definido;

iii) A evidência relativa a cada acontecimento (ou estímulo físico) varia de experiência para experiência, e tem um resultado que vem a ser a representação probabilística do acontecimento;

iv) Após cada observação, o operador deve tomar uma decisão do tipo: sim, o sinal encontra-se presente; ou não, o sinal encontra-se ausente, ou seja, tem-se evidências somente sobre o ruído.

Assim sendo, mesmo ciente dos erros que podem ser cometidos, e com base nos resultados disponíveis, uma decisão deve ser tomada e erros sempre podem ser cometidos.

2.2.2 Taxas TP, FP, FN e TN na seleção de modelos

Se tivermos dois modelos candidatos A e C, como geradores de n dados, digamos $X_1, X_2, \ldots, X_n$, podemos classificá-los como oriundos de qualquer um dos dois modelos e podemos incorrer em erros ou não. Suponhamos, sem perda de generalidade, que os dados verdadeiramente se originam de C (fato este, des-

conhecido na prática). Podemos então, ter quatro situações descritas a seguir e sumariadas na Figura (1), adaptada de Boone, Lindfors e Seibert (2002),

a) **Situação 1:** Classificar $X_1, X_2, \ldots, X_n$ como oriundos do modelo C, e de fato eles se originam de C. Neste caso, classificamos positivamente o modelo e temos um verdadeiro positivo (TP) - do inglês True Positive;

b) **Situação 2:** Classificar $X_1, X_2, \ldots, X_n$ como não oriundos do modelo C, embora de fato eles se originam de C. Neste caso, classificamos incorretamente o modelo como oriundo do modelo A e temos um falso negativo (FN) - do inglês False Negative;

c) **Situação 3:** Classificar $X_1, X_2, \ldots, X_n$ como oriundos do modelo C, embora de fato eles se originam de A. Neste caso, classificamos incorretamente o modelo e temos um falso positivo (FP) - do inglês False Positive;

d) **Situação 4:** Classificar $X_1, X_2, \ldots, X_n$ como não oriundos do modelo C, e de fato eles se originam de A. Neste caso, classificamos corretamente o modelo e temos um verdadeiro negativo (TN) - do inglês True Negative.

Conforme dito na seção 2.2.1, na prática médica, os chamados testes de diagnóstico são muito utilizados, e, por nos auxiliar no entendimento do assunto, voltaremos neles nesta seção, tratando-os sob o ponto de vista de Shapiro (1999). A precisão discriminatória de um teste de diagnóstico é comumente avaliada medindo o quão bem ele identifica corretamente os n_1, indivíduos conhecidos por serem sabidamente doentes $(D+)$ e n_0 indivíduos, sabidamente sadios $(D-)$. Se o teste de diagnóstico fornece um resultado dicotômico para cada sujeito, por exemplo, positivo $(T+)$ ou negativo $(T-)$, os dados podem ser resumidos em uma tabela 2×2 dos resultados do teste contra o estado verdadeiro da doença, conforme a Tabela 2.

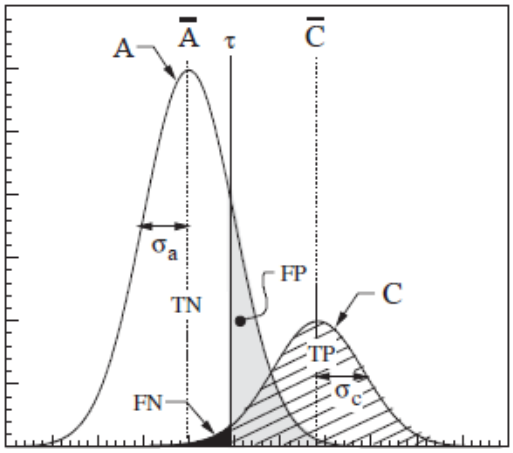

Figura 1 Possibilidades ao classificarmos dados quando dois modelos A e C competem

Tabela 2 Taxas TP, TN, FP e FN

		Estado da doença	
		Presente $(D+)$	Ausente $(D-)$
Resultado do teste	Positivo $(T+)$	TP	FP
	Negativo $(T-)$	FN	TN

Na literatura, a precisão discriminatória é normalmente medida pelas probabilidades condicionais de classificar corretamente pacientes doentes na Tabela 2. Assim temos:

a) $P(T+|D+) = \dfrac{TP}{TP+FN}$, é a chamada **sensitividade** ou **taxa de verdadeiro positivo** (TP), é a percentagem de indivíduos doentes, corretamente classificados como doentes pelo teste;

b) Para os indivíduos sadios $P(T-|D-) = \dfrac{TN}{FP+TN}$, chamada de **especificidade** ou **taxa de verdadeiro negativo** (TN), é a percentagem de indivíduos sadios corretamente classificados como sadios pelo teste;

c) $P(T+|D-) = \dfrac{FP}{FP+TN} = 1 -$ Especificidade, é a percentagem de indiví-

duos sadios erroneamente classificados como doentes pelo teste (**taxa de falso positivo**);

d) $P(T-|D+) = \dfrac{FN}{TP+FN} = 1-$Sensitividade, é a percentagem de indivíduos doentes erroneamente classificados como sadios (**taxa de falso negativo**).

Dessa forma, utilizando a notação de Mood, Graybill e Boes(1974), em que temos, em termos de testes de hipóteses, que

i) A taxa de verdadeiro positivo $TP(\%) = P(T+|D+) = \dfrac{TP}{TP+FN}$ nos dá um indicativo da confiança do teste $(1-\alpha)$;

ii) A taxa de verdadeiro negativo $TN(\%) = P(T-|D-) = \dfrac{TN}{FP+TN}$ nos dá um indicativo do poder do teste $(1-\beta)$;

iii) A taxa de falso positivo $FP(\%) = P(T+|D-) = \dfrac{FP}{FP+TN}$ nos dá um indicativo da probabilidade de cometermos o erro tipo II (β) do teste;

iv) A taxa de falso negativo $FN(\%) = P(T-|D+) = \dfrac{FN}{TP+FN}$ nos dá um indicativo da probabilidade de cometermos o erro tipo I (α) do teste.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control,** Notre Dame, v. 19, n. 6, p. 717-723, 1974.

BISHOP, C. M. **Pattern recognition and machine learning.** Singapore: Springer, 2006. 738 p.

BOONE J. M.; LINDFORS, K. K.; SEIBERT, J. A. Determining sensitivity of mammography from screening data, cancer incidence, and receiver-operating characteristic curve parameters. **Medical Decision Making,** Bethesda, v. 22, n. 3, p. 228-237, 2002.

BOX, G. E. P.; DRAPER, N. R. **Empirical model building and response surfaces**. New York: J. Wiley, 1987. 669 p.

BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: understanding aic and bic in model selection. **Sociological Methods and Research,** Beverly Hills, v. 33, n. 2, p. 261-304, 2004.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3rd ed. New York: J. Wiley, 1998. 706 p.

EMILIANO, P. C. **Fundamentos e aplicações dos critérios de informação:** Akaike e bayesiano. 2009. 92 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2009.

KONISHI, S.; KITAGAWA, G. **Information criteria and statistical modeling**. New York: Springer, 2008. 321 p.

LEHMANN, E. L. The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two? **Journal of the American Statistical Association**, New York, v. 88, n. 424, p. 1242-1249, 1993.

MAZEROLLE, M. J. **Mouvements et reproduction des amphibiens en tourbiéres perturbées**. 2004. 205 p. Thesis (Doctorat en Sciences Forestières) - Université Laval, Québec, 2004.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics.** Singapore: McGraw-Hill, 1974. 564 p.

REISER, B.; FARAGGI, D. Confidence intervals for generalized ROC criterion. **Biometrics**, Washington, v. 53, n. 2, p. 644-652, 1997.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. **Akaike information criterion statistics**. Tokyo: KTK Scientific, 1986. 320 p.

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, Hayward, v. 6, n. 2, p. 461-464, 1978.

SHAPIRO, D. E. The interpretation of diagnostic tests. **Statistical Methods in Medical Research**, Thousand Oaks, v. 8, n. 2, p. 113-134, 1999.

SUGIURA, N. Further analysts of the data by akaike's information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, Ontario, v. 7, n. 1, p. 13-26, 1978.

TRIOLA, M. F. **Introdução à estatística**. 10. ed. Rio de Janeiro: LTC, 2008. 696 p.

TSALLIS, C. **Introduction to nonextensive statistical mechanics approaching a complex world**. New York: Springer, 2009. 382 p.

VAZ, J. C. L. **Regiões de incerteza para a curva ROC em testes diagnósticos**. 2009. 151 p. Dissertação (Mestrado em Estatística) - Universidade Federal de São Carlos, São Carlos, 2009.

CAPÍTULO 2

Critérios de informação aplicados à seleção de modelos normais

RESUMO

Um dos problemas da inferência estatística consiste em testar hipóteses acerca de um ou mais parâmetros de uma população. Para o caso dos testes de médias, é importante sabermos se as variâncias são conhecidas ou desconhecidas, e, além disso, se são iguais ou diferentes. Sob a pressuposição de normalidade, para o caso do desconhecimento das variâncias populacionais, primeiramente procede-se a um teste para as variâncias, sendo o mais usual, o teste F. Posteriormente, procede-se o teste de igualdade de médias, e, em geral, utiliza-se para este fim o teste t de Student. Os critérios AIC, AICc e BIC podem ser utilizados com esta mesma finalidade. O que se faz é calcular o valor do AIC (ou AICc, ou BIC) e verificar qual deles apresenta o menor valor. No presente capítulo, procedeu-se os testes usuais (teste F e posteriormente teste t, aqui chamado teste "FT") e também utilizou-se os critérios de informação, sendo que o teste FT foi utilizado apenas como referência para o estudo dos critérios AIC, AICc e BIC, que foram os testes em que pretendemos avaliar seu desempenho neste momento. Foram verificadas as taxas de falso positivo (FP), falso negativo (FN) e verdadeiro positivo (TP) em estudos de simulação realizadas no software R. Em geral, o AIC e AICc apresentaram mesmo desempenho e o BIC apresentou melhor desempenho assintóticamente. Para amostras de tamanho 100 o AIC e AICc apresentaram desempenho similar e o BIC foi melhor na maioria das situações. Finalizando o capítulo, uma aplicação a dados reais de peso de suínos, aos 21 dias de nascimento, foi apresentada. Todos os critérios selecionaram o modelo que considera o peso dos suínos machos e fêmeas como iguais.

Palavras-chave: Testes de médias. Teste t. Teste F.

ABSTRACT

A problem of statistical inference is to test hypotheses about one or more parameters of a population. In the case of testing medium, it is important to know whether the variances are known or unknown, and furthermore, if they are the same or different. Under the normality assumption, for the case of lack of population variances, first proceed to a test for variances, the most usual, the F test Subsequently, the procedure is the test for equality of means, and, in general, is used for this purpose the Student t test. The criteria AIC, AICc and BIC can be used for the same purpose. What it does is calculate the value of AIC (or AICc, or BIC) and check which one has the lowest value. In this chapter, we proceeded to the usual tests (F test and t test later, here called test "FT") and also used the criteria of information, and the FT test was used as a reference for the study criteria AIC, AICc and BIC, which were tests that intend to evaluate their performance this moment. We studied the rates of false positive (FP), false negative (FN) and true positive (TP) in simulation studies using the software R. In general, the AIC and AICc showed the same performance and BIC performed better asymptotically. For samples of size 100 the AIC and AICc and BIC showed similar performance was better in most situations. Concluding chapter, an application to real data weight of pigs at 21 days of birth, was presented. All criteria selected the model that considers the weight of pigs males and females as equals.

Keywords: Mean test. t Test. F-test.

1 INTRODUÇÃO

Um dos problemas da inferência estatística consiste em testar hipóteses, isto é, determinar um procedimento estatístico com o objetivo de auxiliar o pesquisador a tomar uma decisão acerca de uma afirmação feita a respeito de um parâmetro da população, ou seja, verificar se os dados observados suportam ou não uma determinada hipótese.

Muita das vezes, o intuito do pesquisador é comparar duas médias de populações normalmente distribuídas. Neste caso, devemos considerar dois aspectos a saber: variâncias conhecidas e variâncias desconhecidas. Para a primeira circunstância pode-se utilizar o teste Z para testar as médias; entretanto o conhecimento das variâncias populacionais raramente ocorre. O segundo caso, mais realista, é o caso das variâncias desconhecidas, podendo elas serem diferentes ou iguais. O que é feito então, é um teste para as variâncias, e, posteriormente um teste para as médias.

Para descobrir a situação em que se encontram as variâncias populacionais, primeiramente procede-se a um teste para as variâncias, sendo o mais usual, o teste F. Posteriormente, procede-se o teste de igualdade de médias e utiliza-se para este fim o teste t de Student. Para tanto, os critérios AIC, AICc e BIC podem ser utilizados com essa mesma finalidade. O que se faz é calcular o valor do AIC (ou AICc, ou BIC) e verificar qual deles apresenta o menor valor.

No presente capítulo procederemos aos testes usuais (teste F e posteriormente teste t, aqui chamado teste "FT") e também utilizaremos os critérios de informação para este mesmo fim, sendo que, o teste FT será utilizado apenas como referência para o estudo dos critérios AIC, AICc e BIC, sobre os quais pretendemos avaliar seu desempenho aqui. Serão verificadas as taxas de falso positivo (FP),

falso negativo (FN) e verdadeiro positivo (TP) em estudos de simulação. Uma aplicação a dados reais será apresentada ao final do capítulo. O objetivo neste capítulo é, portanto, verificar o desempenho dos critérios de seleção de modelos na seleção de modelos normais, comparando-os com o teste "FT", comumente utilizado e, por fim, determinar, se possível, qual deles é o mais indicado a para este fim.

2 REFERENCIAL TEÓRICO

Os testes F e t de Student (que aqui constituem o chamado teste FT), serão utilizados como referência para o estudo dos critérios AIC, AICc e BIC (estudados nas seções 2.1.1, 2.1.2 e 2.1.3 do capítulo 1, respectivamente). A seguir temos uma discussão acerca dos mesmos e de outros conceitos necessários para este capítulo.

2.1 Teste t de Student

As seções 2.1.1 e 2.1.2 trazem um pouco da história do teste e dos conceitos envolvidos no mesmo.

2.1.1 História do teste t de Student

De 1894 a 1930, o University College, em Londres, foi o único lugar no Reino Unido em que havia o ensino avançado de Estatística, para onde afluíam estudantes do país e do exterior, em busca de pós-graduação. Foi nessa época que Karl Pearson, o pesquisador mais conceituado em matéria de estatística à época, foi procurado por William Sealy Gosset (1876 -1937), mais conhecido pelo pseudônimo de Student, para tirar suas dúvidas em relação ao tratamento das pequenas amostras (MEMÓRIA, 2004).

A estatística t foi introduzida em 1908 por Gosset, químico da cervejaria Guinness em Dublin, Irlanda. Gosset havia sido contratado devido à política inovadora de CLaude Guinness de recrutar os melhores graduados de Oxford e Cambridge para os cargos de bioquímico e estatístico da indústria Guinness (SAVAGE, 1976). Gosset desenvolveu o Teste t como um modo barato de monitorar a qualidade da cerveja tipo *stout*. Ele comunicou-se com Guinnes, que permitiu que

o artigo fosse publicado, mas foi forçado a usar um pseudônimo pelo seu empregador, que acreditava que o fato de usar estatística era um segredo industrial. Foram sugeridos "Pupil" e "Student", tendo Gosset optado pelo segundo e publicado o Teste t na revista acadêmica Biometrika em 1908 (BOX, 1987).

Segundo Memória (2004), o artigo de 1908 de Student é considerado o marco inicial do estudo das pequenas amostras. Nele, o pesquisador inicia o pensamento lapidar de que qualquer experimento pode ser considerado como um indivíduo de uma população de experimentos realizados sob as mesmas condições. Nesse sentido, uma série de experimentos é uma amostra extraída dessa população. Assim, esse trabalho surgiu da necessidade de se determinar dentro de quais limites a média da população μ estaria situada, conhecida a média da amostra $\bar{X}$ e seu erro-padrão. Segundo Raju (2005), o método usual era admitir que $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$, sendo $S = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}}$ obedecia a distribuição normal e utilizar as tábuas de probabilidades, mas Student intuiu que, para n pequeno, S estaria sujeito a um erro de amostragem maior, e assim, isso só seria válido para grandes amostras, pois o valor de S é praticamente equivalente a σ, sendo portanto $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \approx N(0,1)$.

A contribuição de Student não foi devidamente apreciada à época, pois, para Karl Pearson, as pequenas amostras não eram fidedignas, devendo ser evitadas. Foi Fisher quem reconheceu o mérito desse trabalho, ao qual emprestou seu gênio para desenvolvê-lo teoricamente. Gosset publicou ainda vários trabalhos, sempre com o pseudônimo de Student.

Em um de seus artigos de 1908 (STUDENT, 1908), Gosset criou a sua distribuição Z (que não é a distribuição normal padrão). Tendo levado em conta a distribuição de $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$ (que ele mostrou, por meio do cálculo dos momentos, ser do Tipo III das curvas de Pearson (em essência uma distribuição

de χ^2)) e ademais que $\bar{X}$ e S^2 eram independentemente distribuídas, Student derivou a distribuição de $Z = \frac{\bar{X}-\mu}{S} = \left(\bar{X}-\mu\right) \Big/ \sqrt{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \Big/ n}$, sendo $X_1, X_2, \ldots, X_n$ normalmente distribuídos, e foi por ele tabulada.

De acordo com Eisenhart (1979), Fisher se correspondeu com Gosset através de cartas a partir de 1912 e comunicou-lhe acerca de uma derivação matemática para sua distribuição Z, que foi publicada em 1923. Em cartas posteriores, Fisher achou mais conveniente trabalhar com a distribuição de $T = Z\sqrt{n-1}$, substituindo a convenção, antes utilizada, de se trabalhar em termos de somas de quadrados por quadrados médios divididos pelos "graus de liberdade". Em 1925, Fisher denominou-a formalmente de t e escreveu-a sob a forma que ela é hoje conhecida, $T = Z\sqrt{n-1}$ sendo

$$T = \frac{\left(\bar{X}-\mu\right)}{\sqrt{\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n(n-1)}}} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_\nu,$$

sendo $S^2 = \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n-1}$ e $\nu = n-1$ os graus de liberdade.

2.1.2 O teste t de Student

Conforme Mood, Graybill e Boes (1974), a função densidade de probabilidade t de Student é dada por,

$$f\left(x\right) = \frac{\Gamma\left((k+1)/2\right)}{\Gamma\left(k/2\right)} \frac{1}{\sqrt{k\pi}} \frac{1}{\left(1+x^2/k\right)^{(k+1)/2}},$$

em que $k > 0$, $-\infty < x < \infty$ e $\Gamma(\cdot)$ é a função gama, definida por

$$\Gamma\left(t\right) = \int_0^{+\infty} x^{t-1} e^{-x} dx, \tag{1}$$

para $t > 0$.

O teste t de Student ou somente teste t, é um teste de hipótese que usa conceitos estatísticos para rejeitar ou não uma hipótese nula quando a estatística de teste segue uma distribuição t de Student. O teste de Student, ou simplesmente teste t, é o método mais utilizado para se avaliarem as diferenças entre as médias de dois grupos.

Sejam $X_1, X_2, \ldots, X_{n_1}$ e $Y_1, Y_2, \ldots, Y_{n_2}$ dois conjuntos de variáveis aleatórias independentes, sendo que $X_i \sim N(\mu_1, \sigma_1^2)$ e $Y_j \sim N(\mu_2, \sigma_2^2)$. Segundo Snedecor e Cochran (1989), para testarmos a hipótese $H_0 : \mu_1 = \mu_2$ *versus* $H_1 : \mu_1 \neq \mu_2$ temos dois casos a considerar:

a) Variâncias populacionais desconhecidas, porém iguais $\left(\sigma_1^2 = \sigma_2^2 = \sigma^2\right)$.

Neste caso, utilizamos a estatística de teste $T = \frac{\bar{X} - \bar{Y}}{S_c\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_\nu$, sendo $\nu = n_1 + n_2 - 2$, $S_c = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$, $S_1^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2}{n_1 - 1}$ e $S_2^2 = \frac{\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_2 - 1}$ e rejeita-se H_0 se $P\left[|T| \geq |t|\right] \leq \alpha$, sendo α o nível de significância adotado.

b) Variâncias populacionais desconhecidas, e diferentes $\left(\sigma_1^2 \neq \sigma_2^2\right)$.

Segundo Sawilowsky (2002), o problema geral foi estabelecido inicialmente para amostras de mesmo tamanho por W. V. Behrens em 1929, e, posteriormente, tratado por R. A. Fisher em 1935, passando a ser conhecido por problema de Behrens-Fisher. Dudewicz et al. (2007) afirmam que diversos autores propuseram soluções para este problema, em que desejavam resolver alguma versão do mesmo, tais como Welch, Neyman e Bartlett, Scheffé, Hsu, Chapman, Dudewicz e Ahmed, dentre outros. Porém, para o caso tratado aqui, a versão mais utilizada é a estudada por Welch (WELCH, 1938, 1947), a qual, muita das vezes, é também

dita ser devido à Welch e Satterthwaite, por causa das contribuições deste último em seu artigo (SATTERTHWAITE, 1946).

Neste caso, utilizamos a estatística de teste $T_{WS} = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim t_\nu$, sendo que ν é dado pela fórmula de Welch-Satterthwaite, dada por

$$\nu = \left\lfloor \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2} \right\rfloor$$

e $\lfloor \cdot \rfloor$ representa arredondamento para baixo do valor obtido (SATTERTHWAITE, 1946). A este teste chamaremos de teste de Welch-Satterthwaite, conforme Trosset (2009), e rejeita-se H_0 se $P\left[|T_{WS}| \geq |t_{WS}|\right] \leq \alpha$, sendo α o nível de significância adotado.

Como as estatísticas de teste diferem em a) e b) acima, devemos saber em qual dos dois casos estamos: $\left(\sigma_1^2 = \sigma_2^2 = \sigma^2\right)$ ou $\left(\sigma_1^2 \neq \sigma_2^2\right)$. Uma saída, é fazer um teste para as variâncias (sobre o qual abordaremos na seção 2.2), e, posteriormente o teste de médias, T ou T_{WS}, dependendo da rejeição ou não de H_0 no teste F. A esta combinação do teste F com o posterior teste de médias, será referido aqui simplesmente como teste FT.

Autores como Markowski e Markowski (1990) e Moser e Stevens (1992), criticam esta saída alegando que o nível de significância e o poder do teste são severamente afetados, constituindo-se de uma combinação dos níveis de significância e do poder de cada teste utilizado.

Trosset (2009), reproduziu uma frase dita por seu professor Erich Lehmann (um grande estatístico do século XX) em um curso de graduação afirmando que, nunca deve-se utilizar o teste t de Student para duas médias. Segundo ele Lehmann afirmava que: “If you get just one thing out of this course, I’d like it to be that you

should never use Student's 2-sample t-test."[1] Desta forma, o autor sugere que ao invés do teste t do item a) acima, deve-se usar o teste de Welch-Satterthwaite, pois este equivale ao teste t, se de fato as variâncias forem iguais; caso sejam diferentes, o teste de Welch-Satterthwaite terá um desempenho melhor.

2.2 Teste F de Snedecor

Nas seções 2.2.1 e 2.2.2 encontra-se um pouco da história deste teste e dos conceitos envolvidos no desenvolvimento do mesmo.

2.2.1 História do teste F de Snedecor

Segundo Memória (2004), muitos consideram que o biólogo e estatístico Ronald Aylmer Fisher (1890 - 1962) contribuiu para o avanço da ciência Estatística mais do que qualquer outro indivíduo na história. Entre suas principais realizações está o desenvolvimento de uma teoria formal, para a análise de variância (ANOVA), que dependia de uma importante distribuição de probabilidade contínua, conhecida como distribuição F. A distribuição F foi descoberta por um matemático norte-americano chamado George W. Snedecor (1881-1974) e foi estendida e ampliada por Fisher em seu trabalho sobre a análise de variância.

Em seu artigo, Fisher (1922) estudou a distribuição dos coeficientes de regressão e se deparou com o problema de obter a distribuição da razão de duas χ^2 independentes. Naquele instante ele não dispunha ainda da distribuição F, mas de sua distribuição Z, (que não é a normal padrão). Não obstante, a distribuição da razão de duas qui-quadrado independentes aparecia quando Fisher utilizava-se da análise de variância, em que separava a variação total na variação dos resíduos e dos tratamentos.

[1] Se você aprender apenas uma coisa deste curso, gostaria que fosse que você nunca deve usar o teste t Student para duas amostras. Tradução literal nossa.

Fisher tabulou sua distribuição Z e, posteriormente George W. Snedecor tabulou razões de variâncias, percebendo que havia relação entre os valores por ele obtidos e os valores tabulados por Fisher, na verdade $e^{2Z} = F$. Snedecor renomeou a razão de duas variâncias para F, em homenagem a Fisher. Contudo, esta homenagem não o agradou, de acordo com Savage (1976).

2.2.2 O teste F de Snedecor

Conforme Mood, Graybill e Boes (1974), a função densidade de probabilidade F de Snedecor é dada por,

$$f(x) = \frac{\Gamma[(m+n)/2]}{\Gamma[m/2]\,\Gamma[n/2]} \left(\frac{m}{n}\right)^{m/2} \frac{x^{(m-2)/2}}{[1+(m/n)\,x]^{(m+n)/2}}$$

em que $m \in \mathbb{N}^*$, $n \in \mathbb{N}^*$, $0 < x < \infty$ e $\Gamma(\cdot)$ é a função gama, definida na equação (1).

Sejam $X_1, X_2, \ldots, X_{n_1}$ e $Y_1, Y_2, \ldots, Y_{n_2}$ dois conjuntos de variáveis aleatórias independentes, sendo que $X_i \sim N(\mu_1, \sigma_1^2)$ e $Y_j \sim N(\mu_2, \sigma_2^2)$. Segundo Snedecor e Cochran (1989), para testarmos a hipótese $H_0 : \sigma_1^2 = \sigma_2^2$ *versus* $H_1 : \sigma_1^2 > \sigma_2^2$ podemos utilizar a estatística de teste

$$F = \frac{S_X^2}{S_Y^2},$$

sendo $S_X^2 = \dfrac{1}{n_1 - 1} \displaystyle\sum_{i=1}^{n_1} \left(X_i - \overline{X}\right)^2$ e $S_Y^2 = \dfrac{1}{n_2 - 1} \displaystyle\sum_{j=1}^{n_2} \left(Y_j - \overline{Y}\right)^2$ e supondo $S_X^2 > S_Y^2$.

Sob H_0 verdadeira, temos que $F \sim F_{\nu_1,\nu_2}$, sendo $\nu_1 = n_1 - 1$ e $\nu_2 = n_2 - 1$ e rejeitaremos H_0 se $P[F \geq F_{\nu_1,\nu_2}] \leq \alpha$.

2.3 Teste de Shapiro-Wilk

Em todo o exposto até aqui, tanto na seção 2.1, quanto na seção 2.2, a pressuposição de normalidade é requerida. Segundo Stephens (1974), existem diversos testes para este fim, tais como: teste de Cramer-von Mises, teste Anderson-Darling, teste de Watson, teste Shapiro-Wilk, dentre outros.

Será utilizado aqui o teste proposto por Shapiro e Wilk (1965), para testar as hipóteses H_0 : "A amostra provém de uma população normalmente distribuída" *versus* H_1 : "A amostra não provém de uma população normalmente distribuída". Este teste utiliza a estatística,

$$W = \frac{\left(\sum_{i=1}^{n} a_i y_{(i)}\right)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},$$

em que as constantes $a_1, a_2, \cdots, a_n$ são calculadas como a solução de

$$(a_1, a_2, \cdots, a_n) = \frac{m^t V^{-1}}{(m^t V^{-1} V^{-1} m)^{1/2}},$$

sendo $m = (m_1, m_2, \cdots, m_n)^t$ o vetor dos valores esperados das estatísticas de ordem da amostra, V a matriz de covariância dessas estatísticas e $y_{(i)}$ é a $i-$ésima estatística de ordem.

A estatística de teste é exata se $n = 3$, e é aproximada se $n > 3$, podendo ser encontrada tabelada em Shapiro e Wilk (1965). Para este teste rejeita-se H_0 se $P[W \geq W_{tab}] < \alpha$.

3 METODOLOGIA

Para comparar variâncias de duas distribuições normais, usualmente utiliza-se o teste F para variâncias, seguido do teste t de Student se as variâncias forem iguais, ou ainda do teste de Welch-Satterthwaite, se as variâncias forem consideradas diferentes pelo teste F, referido aqui como teste FT. Muitos críticos desta metodologia, utilizam o teste de Welch-Satterthwaite diretamente, sem efetuar o teste F preliminarmente, referido aqui como teste T_{WS}. Comparamos a utilização do teste utilizado como referência, isto é, o teste FT com a metodologia de avaliação via critérios de informação AIC, AICc e BIC. Para isso, foram realizadas simulações Monte Carlo sob diferentes cenários das médias e variâncias das duas populações.

3.1 Simulação dos modelos normais

Considerando dois conjuntos de dados $\{X_1, X_2, \ldots, X_{n_1}\}$, em que $X_i \sim N\left(\mu_1, \sigma_1^2\right)$ e $\{Y_1, Y_2, \ldots, Y_{n_2}\}$, $Y_j \sim N\left(\mu_2, \sigma_2^2\right)$, para $1 \leq i \leq n_1$ e $1 \leq j \leq n_2$. Os critérios de informação AIC, AICc, e BIC podem ser utilizados para determinar de que tipo de modelo os dados foram gerados (os dados são gerados de distribuições normais). Eles podem ter: (i) mesma média e mesma variância (MIVI), (ii) mesma média e variâncias diferentes (MIVD), (iii) médias diferentes e mesma variância (MDVI), (iv) médias diferentes e variâncias diferentes (MDVD).

Para avaliar o desempenho dos critérios de informação e compará-los com os métodos usuais, empregamos simulação Monte Carlo utilizando o programa R (R DEVELOPMENT CORE TEAM, 2013). Duas etapas foram consideradas: na primeira, avaliou-se o desempenho em relação ao erro tipo I (FN) e, na segunda, em relação ao poder (TN), simulando-se amostras em cada uma das situações ci-

tadas, isto é,

$$\mu_1 = \mu_2 = \mu \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = \sigma^2 \quad \text{(MIVI)}, \tag{2}$$

$$\mu_1 = \mu_2 = \mu \quad \text{e} \quad \sigma_1^2 \neq \sigma_2^2 \quad \text{(MIVD)}, \tag{3}$$

$$\mu_1 \neq \mu_2 \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = \sigma^2 \quad \text{(MDVI)}, \tag{4}$$

$$\mu_1 \neq \mu_2 \quad \text{e} \quad \sigma_1^2 \neq \sigma_2^2 \quad \text{(MDVD)}, \tag{5}$$

sob a pressuposição de normalidade e independência.

As verossimilhanças para os casos descritos em (2)-(5) são dadas a seguir e sua derivação completa pode ser encontrada em Emiliano (2009).

Caso 1: $\mu_1 = \mu_2 = \mu$ e $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Para o caso descrito em (2), ou seja, $\mu_1 = \mu_2 = \mu$ e $\sigma_1^2 = \sigma_2^2 = \sigma^2$ existem dois parâmetros μ e σ^2 desconhecidos.

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2} - \frac{m}{2}\log\left(2\pi\sigma^2\right) - \frac{\sum_{i=n+1}^{n+m}(y_i-\mu)^2}{2\sigma^2},$$

$$L(\boldsymbol{\theta}) = -\frac{n+m}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n+m}(y_i-\mu)^2, \tag{6}$$

sendo $\boldsymbol{\theta} = \left(\mu, \sigma^2\right)$.

Maximizando (6) tem-se:

$$L\left(\widehat{\boldsymbol{\theta}}\right) = -\frac{n+m}{2}\left[\log\left(2\pi\widehat{\sigma}^2\right) + 1\right], \tag{7}$$

em que

$$\widehat{\mu} = \frac{1}{n+m}\sum_{i=1}^{n+m} y_i \tag{8}$$

e

$$\widehat{\sigma}_2^2 = \frac{1}{n+m}\sum_{i=1}^{n+m}(y_i - \widehat{\mu})^2. \tag{9}$$

Caso 2: $\mu_1 \neq \mu_2$ e $\sigma_1^2 \neq \sigma_2^2$

Se todos os parâmetros são desconhecidos tem-se então $\boldsymbol{\theta} = \left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right)$ e assim a função suporte é expressa como:

$$\begin{aligned} L(\boldsymbol{\theta}) &= L\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = -\frac{n}{2}\log\left(2\pi\sigma_1^2\right) - \frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(y_i - \mu_1)^2 \\ &- \frac{m}{2}\log\left(2\pi\sigma_2^2\right) - \frac{1}{2\sigma_2^2}\sum_{i=n+1}^{n+m}(y_i - \mu_2)^2 \end{aligned} \tag{10}$$

Logo,

$$L\left(\widehat{\boldsymbol{\theta}}\right) = -\frac{n}{2}\log\left(2\pi\widehat{\sigma_1^2}\right) - \frac{\sum_{i=1}^{n}(y_i - \widehat{\mu_1})^2}{2\widehat{\sigma_1^2}} - \frac{m}{2}\log\left(2\pi\sigma_2^2\right) - \frac{\sum_{i=n+1}^{m}(y_i - \widehat{\mu_2})^2}{2\widehat{\sigma_2^2}}, \tag{11}$$

e $\widehat{\mu_1}$, $\widehat{\mu_2}$, $\widehat{\sigma_1^2}$ e $\widehat{\sigma_2^2}$ são dados por respectivamente por (12), (13), (14) e (15).

$$\widehat{\mu_1} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{12}$$

$$\widehat{\mu_2} = \frac{1}{m}\sum_{i=n+1}^{n+m} y_i, \tag{13}$$

$$\widehat{\sigma_1^2} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\mu_1})^2, \tag{14}$$

$$\widehat{\sigma_2^2} = \frac{1}{m}\sum_{i=1}^{n}(y_i - \widehat{\mu_2})^2. \tag{15}$$

Caso 3: $\mu_1 \neq \mu_2$ e $\sigma_1^2 = \sigma_2^2 = \sigma^2$

No caso em que $\mu_1 \neq \mu_2$ $\sigma_1^2 = \sigma_2^2 = \sigma^2$, tem-se três parâmetros desconhecidos μ_1, μ_2 e σ^2, que devem ser estimados a fim de obter a estimativa da função suporte. Assim

$$L(\boldsymbol{\theta}) = -\frac{n+m}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu_1)^2 + \sum_{i=n+1}^{n+m}(y_i-\mu_2)^2\right], \quad (16)$$

em que $\boldsymbol{\theta} = \left(\mu_1, \mu_2, \sigma^2\right)$.

A função suporte estimada é dada por

$$L\left(\widehat{\boldsymbol{\theta}}\right) = -\frac{m+n}{2}\left(\log\left(2\pi\widehat{\sigma^2}\right) + 1\right) \quad (17)$$

Sendo os estimadores de μ_1, μ_2, e σ^2 dados respectivamente por:

$$\widehat{\mu_1} = \frac{\sum_{i=1}^{n} y_i}{n}, \quad (18)$$

$$\widehat{\mu_2} = \frac{\sum_{i=n+1}^{n+m} y_i}{m}, \quad (19)$$

$$\widehat{\sigma^2} = \frac{1}{(n+m)}\left[\sum_{i=1}^{n}(y_i-\widehat{\mu_1})^2 + \sum_{i=n+1}^{n+m}(y_i-\widehat{\mu_2})^2\right]. \quad (20)$$

Caso 4: $\mu_1 = \mu_2 = \mu$ e $\sigma_1^2 \neq \sigma_2^2$

Neste caso tem-se 3 parâmetros desconhecidos μ, σ_1^2, e σ_2^2, e $\boldsymbol{\theta} = \left(\mu, \sigma_1^2, \sigma_2^2\right)$. Assim sendo, tem-se:

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\log\left(2\pi\sigma_1^2\right) - \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma_1^2} - \frac{m}{2}\log\left(2\pi\sigma_2^2\right) - \frac{\sum_{i=n+1}^{n+m}(y_i-\mu)^2}{2\sigma_2^2}. \quad (21)$$

E assim

$$L\left(\widehat{\boldsymbol{\theta}}\right) = -\frac{(n+m)}{2}\left(\log 2\pi + 1\right) - \frac{n}{2}\log\widehat{\sigma_1^2} - \frac{m}{2}\log\widehat{\sigma_2^2}. \tag{22}$$

Sendo que

$$\widehat{\sigma_1^2} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\mu})^2, \tag{23}$$

$$\widehat{\sigma_2^2} = \frac{1}{m}\sum_{i=n+1}^{n+m}(y_i - \widehat{\mu})^2, \tag{24}$$

e o estimador de μ é encontrado resolvendo-se a equação

$$\widehat{\mu^3} + A\widehat{\mu^2} + B\widehat{\mu} + C = 0 \tag{25}$$

em que A, B e C, são constantes, conforme Emiliano(2009).

3.2 Avaliação da taxa TP *versus* tamanho amostral

Para verificar o comportamento dos critérios de informação AIC, AICc e BIC, ao aumentarmos os tamanhos amostrais, foram simuladas amostras de tamanhos $n_1 = n_2 = 100, 200, \ldots, 7900, 8000$, nas situações descritas em (2)-(5). Os parâmetros utilizados nestas simulações foram:

$$\mu_1 = \mu_2 = 10.0 \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = 1 \qquad \text{(MIVI)}, \tag{26}$$

$$\mu_1 = \mu_2 = 10.0 \quad \text{e} \quad \sigma_1^2 = 1.0, \sigma_2^2 = 0.25 \quad \text{(MIVD)}, \tag{27}$$

$$\mu_1 = 10.0, \mu_2 = 10.5 \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = 1 \qquad \text{(MDVI)}, \tag{28}$$

$$\mu_1 = 10.0, \mu_2 = 10.5 \quad \text{e} \quad \sigma_1^2 = 1.0, \sigma_2^2 = 0.25 \quad \text{(MDVD)}. \tag{29}$$

Para cada um dos casos (26) - (29) ajustou-se aos dados simulados a distribuição normal considerando-se os quatro casos possíveis e determinou-se o valor de AIC, AICc e BIC para cada um deles. Este processo foi repetido 100 vezes, e por meio dele pôde-se verificar a quantidade de vezes que cada um dos três critérios selecionou corretamente o verdadeiro modelo; por exemplo, das 100 vezes em que se simulou o modelo (26), em quantas vezes cada um dos critérios selecionou o modelo como tendo de fato mesma média e mesma variância. Este processo foi repetido para $n_1 = n_2 = 100, 200, \cdots, 7900, 8000$, para verificar como os critérios se comportam à medida em que o tamanho amostral aumenta.

Este mesmo processo, com as mesmas análises, foi aplicado para o caso

$$\mu_1 = \mu_2 = 10.0 \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = 1 \qquad \text{(MIVI)}, \tag{30}$$

$$\mu_1 = \mu_2 = 10.0 \quad \text{e} \quad \sigma_1^2 = 1.0, \sigma_2^2 = 0.81 \text{ (MIVD)}, \tag{31}$$

$$\mu_1 = 10.0, \mu_2 = 10.5 \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = 1 \qquad \text{(MDVI)}, \tag{32}$$

$$\mu_1 = 10.0, \mu_2 = 10.5 \quad \text{e} \quad \sigma_1^2 = 1.0, \sigma_2^2 = 0.81 \text{ (MDVD)}. \tag{33}$$

A razão pela qual se estabeleceu as duas situações acima, quais sejam, variâncias diferentes e próximas e variâncias diferentes e muito distantes, se dá pelo fato de que se espera que, para valores bem distintos das médias, todos os critérios se comportem bem, e, para o caso em que as médias sejam próximas, eles tenham menores taxas de acerto, dependendo do fato de as variâncias serem próximas ou bem distantes.

3.3 Avaliação das taxas TP, FN e FP

Para verificar o comportamento dos critérios de informação AIC, AICc e BIC, bem como dos testes FT e T_{WS}, para amostras de tamanho 100, foram

simulados dados de distribuições normais, sob as situações descritas em (2)-(5), considerando-se dois cenários:

a) Na primeira situação simulada, as variâncias populacionais, quando distintas, eram "bem distantes" para as duas populações. Os parâmetros utilizados nestas simulações foram:

$$\mu_1 = \mu_2 = 10.0 \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = 1 \qquad \text{(MIVI)}, \tag{34}$$

$$\mu_1 = \mu_2 = 10.0 \quad \text{e} \quad \sigma_1^2 = 1.0, \sigma_2^2 = 0.25 \text{ (MIVD)}, \tag{35}$$

$$\mu_1 = 10.0, \mu_2 = 10.5 \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = 1 \qquad \text{(MDVI)}, \tag{36}$$

$$\mu_1 = 10.0, \mu_2 = 10.5 \quad \text{e} \quad \sigma_1^2 = 1.0, \sigma_2^2 = 0.25 \text{ (MDVD)}. \tag{37}$$

b) Na segunda situação simulada, as variâncias populacionais, quando distintas, eram "próximas" para as duas populações. Os parâmetros utilizados nestas simulações foram:

$$\mu_1 = \mu_2 = 10.0 \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = 1 \qquad \text{(MIVI)}, \tag{38}$$

$$\mu_1 = \mu_2 = 10.0 \quad \text{e} \quad \sigma_1^2 = 1.0, \sigma_2^2 = 0.81 \text{ (MIVD)}, \tag{39}$$

$$\mu_1 = 10.0, \mu_2 = 10.5 \quad \text{e} \quad \sigma_1^2 = \sigma_2^2 = 1 \qquad \text{(MDVI)}, \tag{40}$$

$$\mu_1 = 10.0, \mu_2 = 10.5 \quad \text{e} \quad \sigma_1^2 = 1.0, \sigma_2^2 = 0.81 \text{ (MDVD)}. \tag{41}$$

Para cada par de conjunto de dados gerados $\{X_1, X_2, \ldots, X_{100}\}$ e $\{Y_1, Y_2, \ldots, Y_{100}\}$, foi ajustada a distribuição normal para os quatro modelos descritos pelas situações (2) - (5) e determinado o valor de AIC, AICc e BIC para cada um deles. Embora o intuito aqui não seja o de estudarmos o teste FT, mas apenas utilizá-lo como referência, aplicamos também o teste FT. Este processo foi repetido 100 vezes e por meio dele determinou-se a quantidade de vezes que cada um dos três critérios e o teste FT selecionaram corretamente o verdadeiro modelo; por exemplo, das 100 vezes em que simulou-se o modelo (38), em quantas vezes cada

um dos critérios selecionou o modelo como tendo de fato mesma média e mesma variância (TP). Além disso, pôde-se verificar, também, as taxas de FP e FN para os testes e os critérios.

4 APLICAÇÃO EM DADOS REAIS DE PESOS DE SUÍNOS

O conjunto de dados a ser utilizado aqui foram cedidos pelo Departamento de Zootecnia da Universidade Federal de Viçosa (UFV). A formação da população e a coleta dos dados fenotípicos foram realizadas na Granja de Melhoramento de Suínos do Departamento de Zootecnia da Universidade Federal de Viçosa (UFV), em Viçosa, Minas Gerais, Brasil, no período de novembro de 1998 a julho de 2001.

Foi construída uma população F_2 (345 suínos) proveniente do cruzamento de dois varões da raça naturalizada brasileira Piau com 18 fêmeas de linhagem desenvolvida na UFV pelo acasalamento de animais de raça comercial (Landrace $\times$ Large White $\times$ Pietrain). Indivíduos F_2 possuem informações fenotípicas (pesos) em 7 momentos distintos (0, 21, 42, 63, 77, 105, 150 dias), exceto para alguns animais que possuem informações perdidas em determinadas idades. Maiores detalhes sobre a constituição da população F_2 foram descritos por Peixoto et al. (2006). Dentre as diversas variáveis neste conjunto de dados, trabalharemos nesta seção com os pesos dos animais aos 21 dias após o nascimento de 100 animais machos e 100 animais fêmeas.

Os dados se encontram na Tabela 1 a seguir, e conforme apresentado na seção 5.2.3.1 os pesos seguem distribuição normal pelo teste de Shapiro-Wilk. O objetivo aqui é verificar se os pesos dos animais, machos e fêmeas, aos 21 dias de nascimento possuem mesma média e variância; médias iguais e variâncias diferentes; médias diferentes e variâncias iguais ou médias e variâncias diferentes.

Tabela 1 Peso de pesos dos animais aos 21 dias após o nascimento.

ID	M	F	ID	M	F	ID	M	F	ID	M	F
1	6,86	6,92	26	5,20	3,58	51	2,93	6,06	76	5,72	4,12
2	5,75	4,94	27	2,87	4,84	52	3,94	6,04	77	5,62	4,29
3	4,66	4,91	28	5,07	5,76	53	3,93	5,89	78	3,81	5,04
4	4,57	4,44	29	3,49	4,10	54	4,22	5,98	79	5,63	5,62
5	4,41	4,30	30	4,31	6,39	55	4,60	5,33	80	5,66	3,25
6	5,23	3,96	31	4,09	3,15	56	4,08	5,28	81	4,20	3,28
7	5,22	4,36	32	3,85	5,85	57	5,75	5,00	82	3,59	4,45
8	6,58	4,13	33	5,20	5,82	58	4,28	4,14	83	5,38	4,65
9	5,83	6,51	34	5,18	4,41	59	3,89	4,11	84	3,77	3,83
10	5,08	5,20	35	3,49	3,56	60	5,55	5,29	85	3,04	3,81
11	4,66	4,96	36	3,87	4,60	61	6,01	5,14	86	4,38	3,44
12	6,00	6,67	37	3,68	4,49	62	5,84	4,85	87	4,74	3,77
13	3,76	4,05	38	5,75	5,62	63	5,51	4,15	88	4,15	5,30
14	6,23	4,47	39	4,57	2,91	64	4,75	5,42	89	3,65	4,49
15	6,87	6,28	40	6,17	4,87	65	4,70	5,74	90	4,67	5,93
16	5,72	5,23	41	4,24	7,88	66	4,22	3,15	91	6,37	3,36
17	5,03	6,80	42	4,37	5,67	67	4,85	4,05	92	6,09	3,24
18	5,52	4,73	43	3,86	3,40	68	3,95	3,85	93	4,61	3,57
19	4,16	4,14	44	5,86	3,69	69	4,32	3,35	94	4,04	3,55
20	6,96	4,57	45	5,37	4,81	70	4,17	5,53	95	5,66	5,26
21	5,96	3,07	46	4,43	4,95	71	4,11	4,39	96	4,15	5,23
22	3,39	6,51	47	5,79	3,65	72	3,97	3,03	97	4,50	4,95
23	5,10	5,26	48	5,80	4,82	73	4,54	4,51	98	4,35	7,08
24	5,12	4,36	49	4,32	5,32	74	3,27	3,81	99	4,41	6,20
25	4,91	4,94	50	4,27	5,91	75	5,00	6,01	100	5,31	5,59

Nota: ID-Identificação do animal; M-Animal macho; F-Animal fêmea.

5 RESULTADOS E DISCUSSÃO

Os resultados referentes ao comportamento assintótico dos critérios de informação encontram-se na seção 5.1, enquanto que, para o tamanho amostral $N = 100$ temos os resultados apresentados na seção 5.2. As simulações referentes ao caso descrito em (38) encontra-se no apêndice A, sendo que, as simulações para os demais casos (39, 40 e 41) são similares.

5.1 Comportamento assintótico dos critérios AIC, AICc e BIC

Os resultados das simulações mostrando a taxa de acerto dos três critérios de informação, estudados sob as condições descritas nos casos (26)-(29) são discutidos na seção 5.1.1, enquanto que, na seção 5.1.2 encontram-se os resultados descritos nos casos (30)-(33).

5.1.1 Caso $\mu_1 = 10.0$; $\mu_2 = 10.5$; $\sigma^2 = 1$; $\sigma^2 = 0.25$

Analisando-se as Figuras 1 e 2 podemos notar que os critérios de Akaike e de Akaike corrigido possuem desempenho similar, apresentando praticamente as mesmas taxas de verdadeiro positivo (TP) para as quatro situações simuladas.

Ainda nas Figuras 1 e 2, podemos perceber que as taxas TP mantiveram-se estabilizadas desde o início e:

i) Para o caso descrito em (26), a taxa TP estabilizou-se em cerca de 70%;

ii) Para o caso descrito em (27), a taxa TP estabilizou-se em cerca de 90%;

iii) Para o caso descrito em (28), a taxa TP estabilizou-se em cerca de 85%;

iv) Para o caso descrito em (29), a taxa TP estabilizou-se em cerca de 99%.

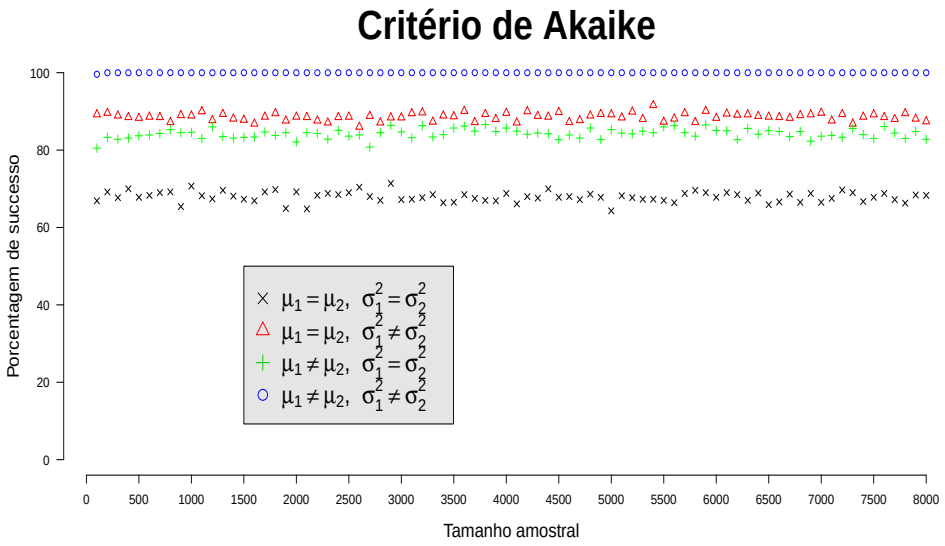

Figura 1 Porcentagem de sucessos (razão TP) para o critério de informação de Akaike AIC - caso 1

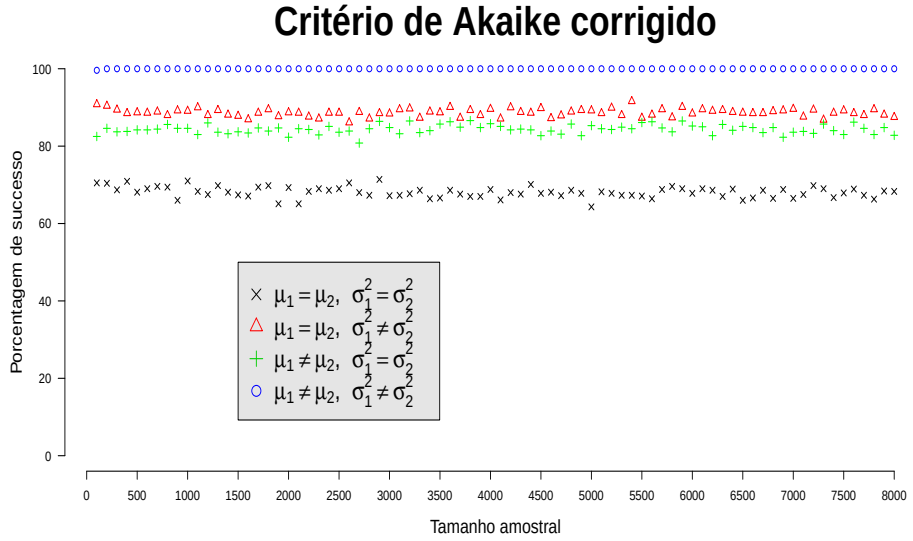

Figura 2 Porcentagem de sucessos (razão TP) para o critério de informação de Akaike corrigido AICc - caso 1

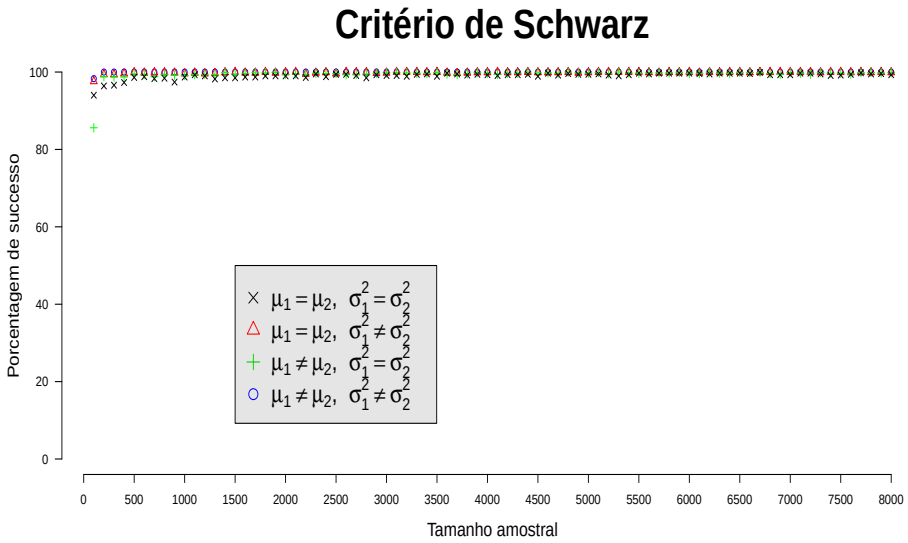

Figura 3 Porcentagem de sucessos (razão TP) para o critério de informação de Schwarz BIC - caso 1

Pela análise da Figura 3, vemos que o critério de Schwarz foi superior ao AIC e AICc, apresentando taxa TP superior à todas as taxas do AIC e AICc descritas pelas Figuras 1 e 2. Dessa forma, para o caso em que as variâncias são "bem distintas" o BIC mostrou-se superior ao AIC e ao AICc.

5.1.2 Caso $\mu_1 = 10.0$; $\mu_2 = 10.5$; $\sigma^2 = 1$; $\sigma^2 = 0.81$

Analisando as Figuras 4 e 5 notamos que os critérios de Akaike e de Akaike corrigido possuem desempenho similar, apresentando praticamente as mesmas taxas de verdadeiro positivo (TP) para as quatro situações simuladas. A taxas TP mantiveram-se estabilizadas desde o início somente para os casos em que as variâncias eram iguais;

i) Para o caso descrito em (30), a taxa TP estabilizou-se em cerca de 70%;

ii) Para o caso descrito em (32), a taxa TP estabilizou-se em cerca de 85%;

iii) Para o caso descrito em (33), a taxa TP estabilizou-se em cerca de 99%.

Para os casos em que as variâncias eram diferentes, a taxas TP estabilizaram-se somente para tamanhos amostrais superiores a 1200, o que, na prática, não se pode encontrar rotineiramente;

iv) Para o caso descrito em (31), a taxa TP estabilizou-se em cerca de 95%;

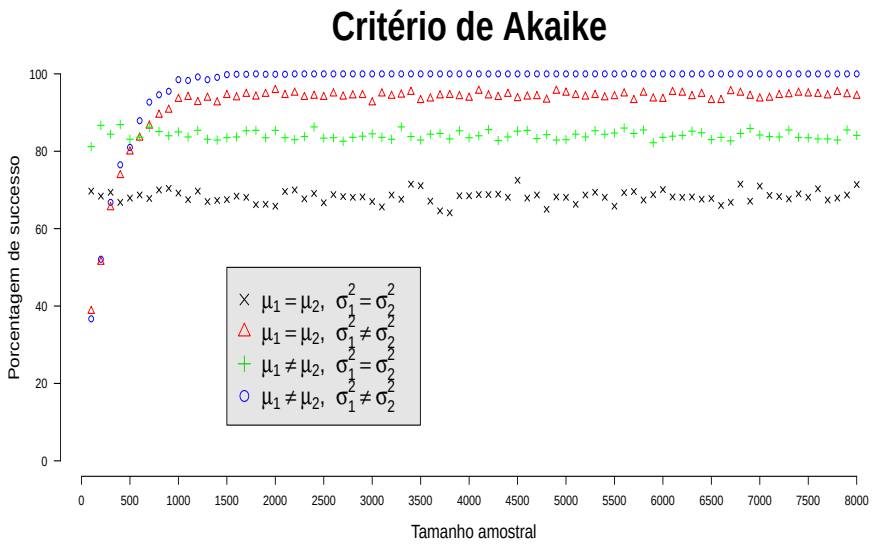

Figura 4 Porcentagem de sucessos (razão TP) para o critério de informação de Akaike AIC - caso 2

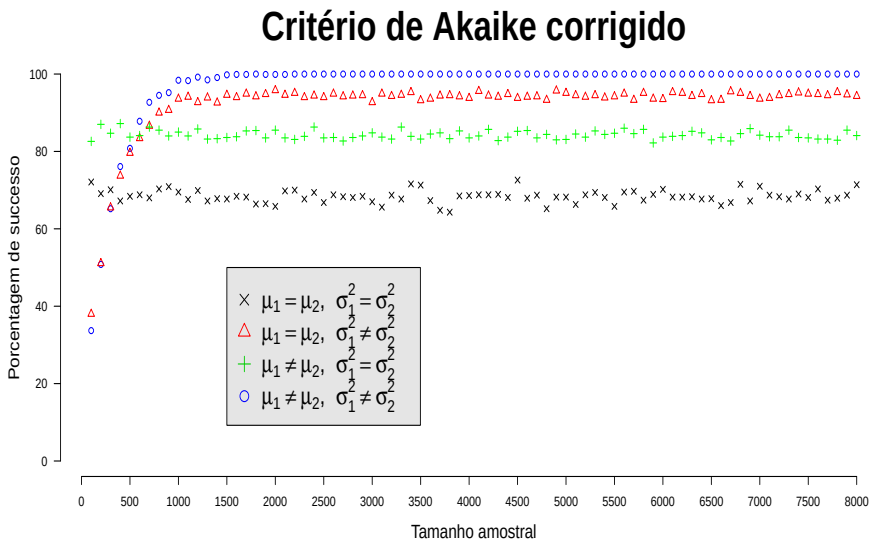

Figura 5 Porcentagem de sucessos (razão TP) para o critério de informação de Akaike corrigido AICc - caso 2

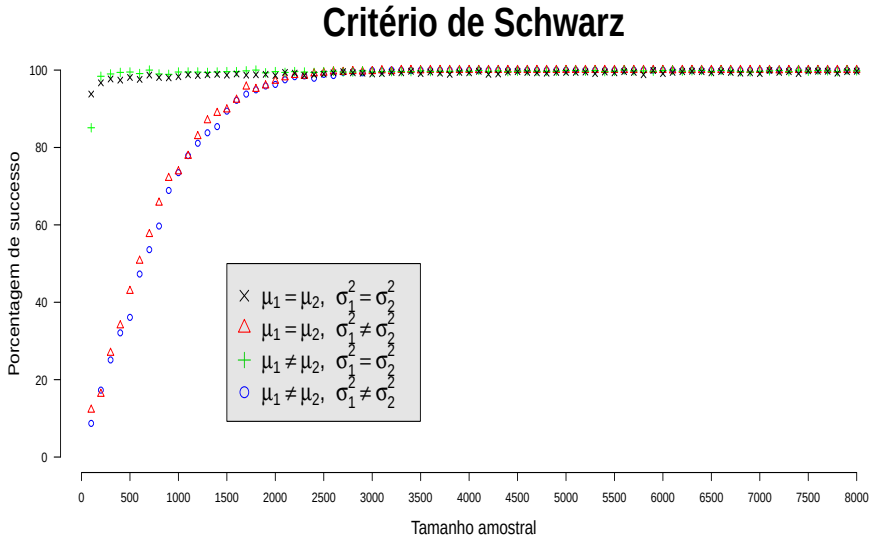

Figura 6 Porcentagem de sucessos (razão TP) para o critério de informação de Schwarz BIC - caso 2

Pela análise da Figura 6, vemos que o critério de Schwarz foi superior ao AIC e ao AICc (descritos pelas Figuras 4 e 5), apresentando taxa TP superior às do AIC e AICc, somente para os casos em que as variâncias eram iguais. Para os casos em que as variâncias eram diferentes, o BIC, assim como o AIC e AICc, não teve bom desempenho, não selecionando os modelos da forma correta como oriundos de populações com diferentes variâncias, − exceção feita a amostras de tamanho superior a 1200, para os quais o desempenho do BIC tornou-se melhor que o AIC e AICc. Todavia, amostras em que se deseja fazer um teste de médias e que tenham esta magnitude são raras.

Desta forma, para o caso em que as variâncias são distintas, porém "próximas", nenhum dos critérios é eficaz em selecionar corretamente os modelos se, de fato, as variâncias forem diferentes, − exceção feita a grandes amostras $(N > 1200)$, pois neste caso, o desempenho do BIC é superior aos demais critérios.

5.2 Taxas TP, FP e FN para o AIC, AICc, BIC e FT

Nesta etapa do estudo, foram simuladas amostras de tamanho $n_1 = n_2 = 100$ e verificadas as taxas TP, FP, e FN dos três critérios sob estudo e também o teste FT, que é frequentemente utilizado. O intuito é analisar, principalmente, a significância e o poder do teste. Assim como na seção 5.1, serão analisados os casos em que as variâncias são diferentes e "bem distintas" (seção 5.2.1) e o caso em que as variâncias serão diferentes, porém "próximas"(seção 5.2.2).

5.2.1 Caso $\mu_1 = 10.0$; $\mu_2 = 10.5$; $\sigma^2 = 1$; $\sigma^2 = 0.25$

Efetuamos as simulações de amostras de tamanho 100 dos modelos descritos em (34)-(37) verificando as taxas TP, FN e FP. O processo foi repetido

10 vezes para que uma estimativa do desvio padrão da taxa TP pudesse ser encontrada.

Analisando-se as Figuras 7 a 10, podemos observar que para as quatro situações simuladas, o desempenho do AIC e AICc foram muito similares, apresentando as mesmas taxas TP, FP e FN. Consequentemente os dois critérios apresentam as mesmas taxas de erro tipo I (α) e de erro tipo II (β), e, portanto, mesmo poder $1 - \beta = 1 - FP$.

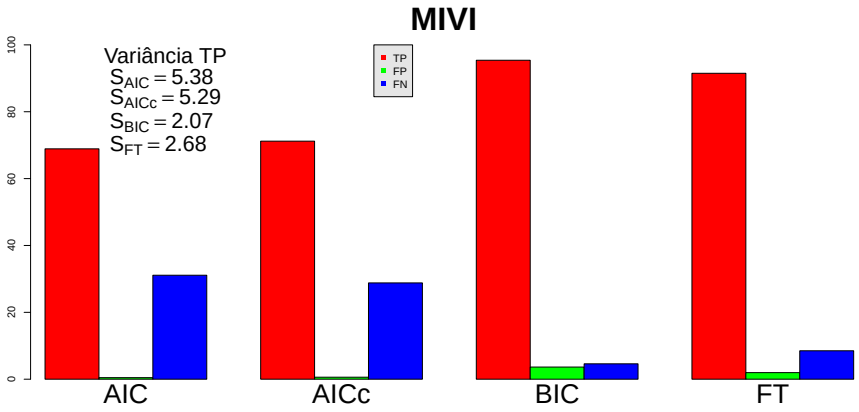

Figura 7 Taxas TP, FP, FN quando simulados modelos com mesma média e variância

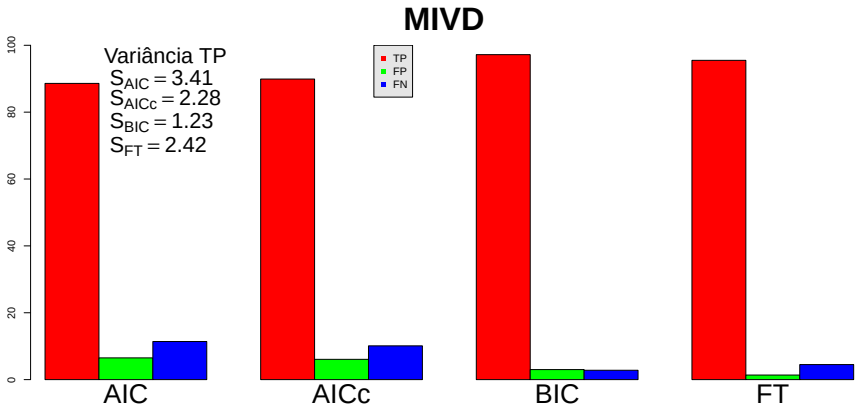

Figura 8 Taxas TP, FP, FN quando simulados modelos com mesma média e variância distintas e bem "diferentes"

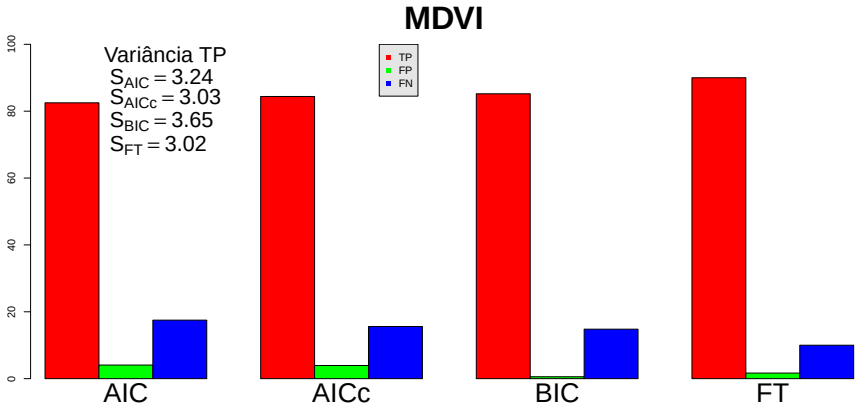

Figura 9 Taxas TP, FP, FN quando simulados modelos com médias diferentes e mesma variância

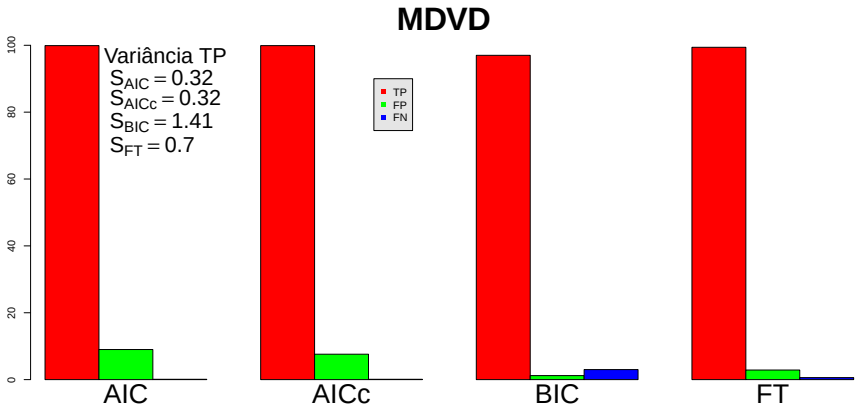

Figura 10 Taxas TP, FP, FN quando simulados modelos com média diferentes e variância distintas e bem "diferentes"

Analisando as Figuras 7-10, primeiramente podemos notar que o desempenho do AIC e AICc foram similares em todas as simulações realizadas, assim sendo, comparando os critérios BIC e AIC (ou AICc, haja vista que tem o mesmo desempenho), através das Figuras 7-10, percebemos que:

i) O BIC apresenta taxa TP superior nas situações (34) e (35);

ii) Para o caso (36), a taxa TP do BIC foi igual à taxa do AIC e AICc;

iii) Para o caso (37), a taxa TP foi inferior à taxa TP do AIC e AICc, sendo, entretanto, superior a 95%, ou seja, representando um bom desempenho do BIC;

iv) Quanto ao poder, o BIC apresentou poder superior ao AIC e AICc nas situações (35), (36) e (37), enquanto que, para o caso (34), o poder do AIC e AICc foram próximos a 99%, e, por sua vez o BIC apresentou poder de cerca de 97%, o que não prejudica o desempenho do BIC, pois o poder ainda assim é alto.

Comparando-se o critério BIC e o teste FT entendemos que:

i) O BIC apresenta taxa TP superior quando as médias são iguais nas situações (34) e (35), e TP inferior nas situações (36) e (37), em que as médias eram diferentes. Em ambas situações, quando um critério foi superior ao outro, a diferença não foi muito alta, como por exemplo, na situação (34), em que o BIC teve TP igual a 95% e o FT teve TP próximo a 91%;

ii) Quanto ao poder, o BIC apresentou poder inferior ao teste FT nas situações (34) e (35), enquanto que para os casos (36) e (37), o mesmo se apresentou superior ao teste FT. As diferenças entre o poder de um teste e do outro não foram grandes, como por exemplo, no caso (34), em que o BIC apresentou poder de cerca de 97%, e o teste FT apresentou poder de cerca de 98%. Desta forma, quanto ao poder podemos dizer que ambos apresentam bom desempenho.

Das discussões acima podemos observar que, para o caso simulado, em que as variâncias são diferentes e "bem distantes", o BIC e o teste FT apresentaram desempenho similar, sendo este desempenho bom e superior ao desempenho do AIC e AICc.

5.2.2 Caso $\mu_1 = 10.0$; $\mu_2 = 10.5$; $\sigma^2 = 1$; $\sigma^2 = 0.81$

Efetuamos as simulações de amostras de tamanho 100 dos modelos descritos em (38)-(41), verificando as taxas TP, FN e FP. O processo foi repetido 10 vezes para que uma estimativa do desvio padrão da taxa TP pudesse ser encontrada.

Analisando-se as Figuras 11 a 14 percebemos que, para as quatro situações simuladas, assim como os resultados obtidos quando as variâncias eram "bem dis-

tantes" (seção 5.2.1), o desempenho do AIC e AICc foram muito similares, apresentando as mesmas taxas TP, FP e FN. Consequentemente, os dois critérios apresentam as mesmas taxas de erro tipo I (α) e de erro tipo II (β), e, portanto, mesmo poder $1 - \beta = 1 - FP$.

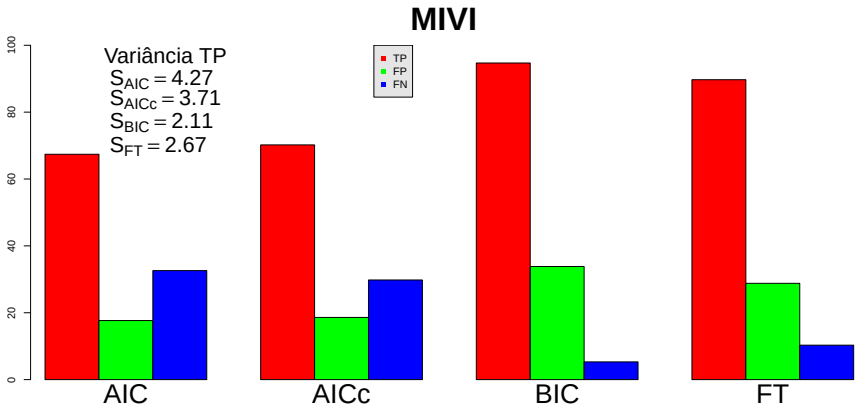

Figura 11 Taxas TP, FP, FN quando simulados modelos com mesma média e variância

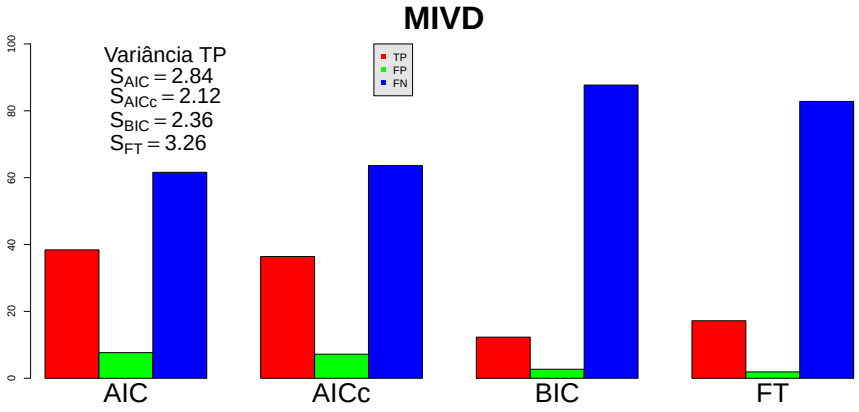

Figura 12 Taxas TP, FP, FN quando simulados modelos com mesma média e variância distintas e "próximas"

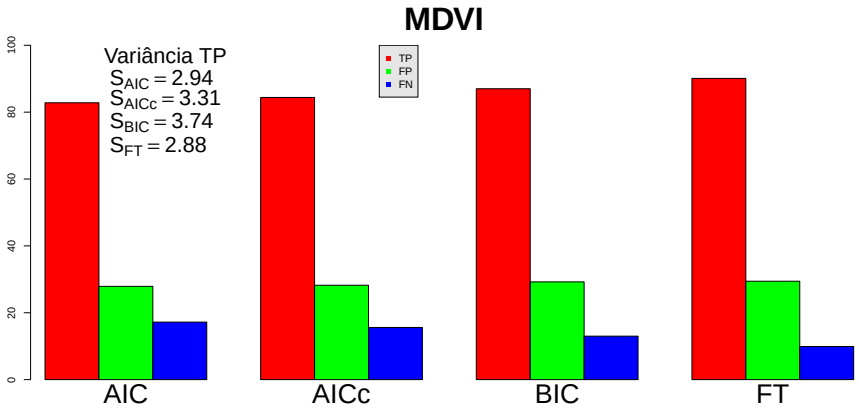

Figura 13 Taxas TP, FP, FN quando simulados modelos com médias diferentes e mesma variância

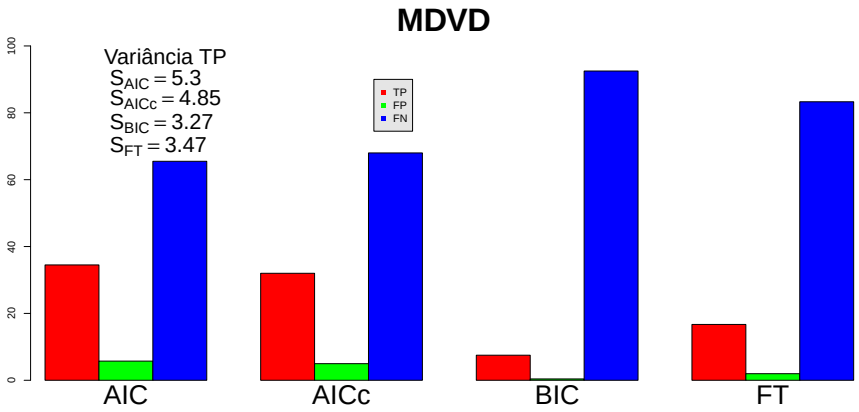

Figura 14 Taxas TP, FP, FN quando simulados modelos com média diferentes e variância distintas e "próximas"

Comparando-se os critérios BIC e AIC (ou AICc, haja vista, que possuem o mesmo desempenho), através das Figuras 11-14 notamos que:

i) O BIC apresenta taxa TP superior nas situações em que as variâncias eram iguais $-$ casos (38) e (40), e apresentou taxa TP inferior nas situações (39) e (41). Vale salientar que, para os casos em que as variâncias eram diferentes, nenhum dos critérios apresentou taxa TP alta, sendo cerca de 40% para o AIC e AICc e 10% para o BIC;

ii) Quanto ao poder, o BIC apresentou poder superior ao AIC e AICc nas situações (39), (40) e (41), enquanto que para o caso (38), o poder do AIC e AICc foram superiores ao BIC, que foi de cerca de 70%.

Para o teste FT temos que:

i) Assim como AIC, AICc e BIC, a taxa TP foi baixa nos casos em que as variâncias eram diferentes, e similar ao BIC quando as variâncias eram iguais;

ii) O poder do teste FT foi similar ao poder do BIC nas situações simuladas.

Das discussões acima podemos perceber que, para o caso simulado, em que as variâncias são diferentes e "próximas", AIC, AICc, BIC, além do teste FT, todos apresentaram desempenho ruim em selecionar o modelo corretamente.

5.2.3 Seleção do modelo para os dados reais

Aplicamos aqui o teste FT e os critérios AIC, AICc e BIC na seleção do melhor modelo para pesos de suínos aos 21 dias.

5.2.3.1 Aplicação do teste FT

Para os dados descritos na seção 4, considerando-se as variáveis X_1 : "Peso médio das fêmeas aos vinte e um dias de nascimento" e X_2 : "Peso mé-

dio dos machos aos vinte e um dias de nascimento" tem-se que

Tabela 2 Resumo dos dados dos pesos de suínos

Fêmeas	Machos
$\overline{X}_1 = 4,786$	$\overline{X}_2 = 4,792$
$\widehat{\sigma_1} = 0,9227$	$\widehat{\sigma_2} = 1,0532$

Além disso, pelo o teste de Shapiro-Wilk (Tabela 3), tem-se que as variáveis X_1 e X_2 são normalmente distribuídas.

Tabela 3 Teste de Shapiro-Wilk para os pesos de suínos

	Estatística W	Valor p
Fêmeas	$W_1 = 0,9791$	0,1134
Machos	$W_2 = 0,9818$	0,1818

Aplicando o teste F para verificar a hipótese de igualdade de variâncias, não há indícios, de que as variâncias sejam diferentes pelo teste F ao nível de 5% de significância. Aplicando-se então o teste t de Student temos que não há indícios de que as médias sejam diferentes ao nível de 5% de significância. Podemos concluir, assim, pelo teste FT que os pesos dos suínos machos e fêmeas possuem médias e variâncias estatísticamente iguais.

5.2.3.2 Seleção do melhor modelo via AIC, AICc e BIC

Aplicando-se os critérios de informação para selecionar o melhor modelo obtivemos os resultados da Tabela 4 abaixo, em que se pode perceber que há concordância entre os três critérios de informação, selecionando o melhor modelo como sendo aquele que apresenta mesma média e variâncias homogêneas.

Notamos, assim que, de acordo com o teste FT na seção 5.2.3.1 e com os testes AIC, AICc e BIC, sumariados na Tabela 4, as variáveis X_1 e X_2 possuem mesma média e mesma variância.

Tabela 4 Critérios de informação para os pesos de suínos

	AIC	AICc	BIC
MIVI	565,5825	565,7062	572,1792
MIVD	567,8367	568,2577	581,0299
MDVI	567,5807	567,8307	577,4756
MDVD	565,8375	566,0875	575,7324

6 CONCLUSÃO

Foi possível verificarmos que nenhum dos critérios de informação AIC, AICc e BIC, bem como do teste F combinado ao teste T (ou T_{WS}, dependendo da igualdade ou não das variâncias) foram eficazes na seleção de modelos enquanto oriundos de variáveis com médias e variâncias iguais e ou distintas, a menos que estas difiram enormemente ou as amostras sejam muito grandes.

REFERÊNCIAS

BOX, J. F. Guinness, Gosset, Fisher, and small samples. **Statistical Science,** Hayward, v. 2, n. 1, p. 45-52, 1987.

DUDEWICZ, E. J. et al. Exact solutions to the Behrens-Fisher problem: asymptotically optimal and finite sample efficient choice among. **Journal of Statistical Planning and Inference,** Atlanta, v. 137, n. 5, p. 1584-1605, 2007.

EINSENHART, C. On the transition from student's z to student's t. **The American Statistician,** Washington, v. 33, n. 1, p. 6-10, 1979.

FISHER, R. A. On the mathematical foundations of theoretical statistics. **Philosophical Transactions of the Royal Society A**, v. 222, p. 309-368, 1922.

MARKOWSKI, C. A.; MARKOWSKI, E. P. Conditions for the effectiveness of a preliminary test of variance. **The American Statistician,** Washington, v. 44, n. 4, p. 322-326, 1990.

MEMÓRIA, J. M. P. **Breve história da estatística**. Brasília: EMBRAPA Informação Tecnológica, 2004. 111 p.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics.** Singapore: McGraw-Hill, 1974. 564 p.

MOSER, B. K.; STEVENS, G. R. Homogeneity of variance in the two-sample means test. **The American Statistician,** Washington, v. 46, n. 1, p. 19-21, 1992.

PEIXOTO, J. O. et al. Associations of leptin gene polymorphisms with production traits in pigs. **Journal of Animal Breeding and Genetics,** Weinheim, v. 123, n. 6, p. 378-383, 2006.

R DEVELOPMENT CORE TEAM. **R:** a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <http://www.r-project.org>. Acesso em: 11 maio 2013.

RAJU, T. N. K. William Sealy Gosset and William A. Silverman: two "students" of science. **Pediatrics,** Elk Grove Village, v. 116, n. 3, p. 732-735, 2005.

SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. **Biometrics Bulletin**, Washington, v. 2, n. 6, p. 110-114 , 1946.

SAVAGE, L. J. On rereading R. A. Fisher. **The Annals of Statistics,** Hayward, v. 4, n. 3, p. 441-500, 1976.

SAWILOWSKY, S. S. Fermat, Schubert, Einstein, and Behrens-Fisher: the probable difference between two means when $\sigma_1^2 \neq \sigma_2^2$. **Journal of Modern Applied Statistical Methods,** Hayward, v. 1, n. 2, p. 461-472, 2002.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality: complete samples. **Biometrika**, Oxford, v. 52, n. 3, p. 591-611, 1965.

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical methods.** 8^{th} ed. Ames: Iowa State University, 1989. 503 p.

STEPHENS, M. A. EDF statistics for goodness of fit and some comparisons. **Journal of the American Statistical Association,** Washington, v. 69, n. 347, p. 730-737, 1974.

STUDENT. The probable error of a mean. **Biometrika,** New York, v. 6, n. 1, p. 1-25, 1908.

TROSSET, M. W. **An introduction to statistical inference and its applications with R.** London: Chapman & Hall, 2009. 496 p.

WELCH, B. L. The generalization of student's problem when several different population variances are involved. **Biometrika,** New York, v. 34, n. 1, p. 28-35, 1947.

——· The significance of the difference between two means when the population variances are unequal. **Biometrika**, New York, v. 29, n. 4, p. 350-362, 1938.

APÊNDICE

Página


```
u=100
matrizaic<-matrix(rep(0,4*u),u,4)
matrizbic<-matrix(rep(0,4*u),u,4)
menoraic<-c(rep(0,u))
menoraicc<-c(rep(0,u))
menorbic<-c(rep(0,u))
contaic<-0;contaicc<-0;contbic<-0;contaic1<-0;
contaic2<-0;contaic3<-0;contaic4<-0;contaicc1<-0;
contaicc2<-0;contaicc3<-0;contaicc4<-0;contbic1<-0;
contbic2<-0;contbic3<-0;contbic4<-0
FtMIVI<-0;FtMIVD<-0;FtMDVI<-0;FtMDVD<-0
for(k in 1:u){
#######################################################
q=100                    #numero de dados da 1ª amostra
g=100                    #numero de dados da 2ª amostra
mediaamar=10             #media da 1ª amostra
mediaverm=10             #media da 2ª amostra
varamar=1                #DP 1ª amostra
varverm=1                #DP 2ª amostra
Amarelo=rnorm(q,mediaamar,varamar)
Vermelho=rnorm(g,mediaverm,varverm)
juntos=c(Amarelo,Vermelho)
n=length(Amarelo)
m=length(Vermelho)
#######################################################
#teste Ft
#######################################################
var.test(Amarelo, Vermelho, alternative = c("two.sided"))
pvalF=var.test(Amarelo, Vermelho, alternative = c("two.sided"))[3]
if (pvalF>0.05){
pvtEV=t.test(Amarelo, Vermelho,var.equal = FALSE)[3]
{if (pvalF>0.05 & pvtEV>0.05){FtMIVI<-FtMIVI+1}}
{if (pvalF>0.05 & pvtEV<0.05){FtMDVI<-FtMDVI+1}}
}
if (pvalF<0.05){
pvtEV=t.test(Amarelo, Vermelho,var.equal = FALSE)[3]
{if (pvalF<0.05 & pvtEV>0.05){FtMIVD<-FtMIVD+1}}
{if (pvalF<0.05 & pvtEV<0.05){FtMDVD<-FtMDVD+1}}
}
#####################################################################
#CASO 1 - mi1, mi2 iguais; sigma1 e sigma2 iguais
```

```
#####################################################################
k1=2
miestJ=mean(juntos)
D=juntos-mean(juntos)
s=D^2
sigmaest1=(1/(n+m))*sum(s)
vero1=-((n+m)/2)*(log(2*pi*sigmaest1)+1)
AIC1=(n+m)*(log(2*pi*sigmaest1,exp(1))+1)+2*k1                    #AIC
BIC1=(n+m)*(log(2*pi*sigmaest1,exp(1))+1)+k1*log(n+m,exp(1))      #BIC
#####################################################################
#CASO 2 - mi1, mi2 diferentes; sigma1 e sigma2 diferentes
#####################################################################
k2=4
miest2A=mean(Amarelo)
miest2V=mean(Vermelho)
D1=Amarelo-mean(Amarelo)
s1=D1^2
sigmaest2A=mean(s1)
D2=Vermelho-mean(Vermelho)
s2=D2^2
sigmaest2V=mean(s2)
vero2=(-(n/2)*(log(2*pi*sigmaest2A,exp(1)))
      -(m/2)*(log(2*pi*sigmaest2V,exp(1)))-(m+n)/2)
AIC2=((n+m)*(log(2*pi,exp(1))+1)+n*log(sigmaest2A,exp(1))         #AIC
+m*log(sigmaest2V,exp(1))+2*k2)
BIC2=((n+m)*(log(2*pi,exp(1))+1)+n*log(sigmaest2A,exp(1))         #BIC
     +m*log(sigmaest2V,exp(1))+k2*log(n+m,exp(1)))
#####################################################################
#CASO 3 - mi1 e mi2 diferentes, sigma1 e sigma2 iguais
#####################################################################
k3=3
miest3A=mean(Amarelo)
miest3V=mean(Vermelho)
D3=Amarelo-mean(Amarelo)
s3=D3^2
D4=Vermelho-mean(Vermelho)
s4=D4^2
sigmaest3=(1/(m+n))*(sum(s3)+sum(s4))
vero3=-(((n+m)/2)*(log(2*pi*sigmaest3,exp(1))+1))
AIC3=(n+m)*(log(2*pi*sigmaest3,exp(1))+1)+2*k3
BIC3=(n+m)*(log(2*pi*sigmaest3,exp(1))+1)+k3*log(n+m,exp(1))
#####################################################################
#CASO 4 - mi1 e mi2 iguais, sigma1 e sigma2 diferentes
#####################################################################
k4=3
medA=mean(Amarelo)
medV=mean(Vermelho)
v=m/(n+m)
```

```
w=n/(n+m)
aux=Amarelo^2
aux1=Vermelho^2
sq1=sum(aux)/n
sq2=sum(aux1)/m
A=-(medA*(1+v)+(1+w)*medV)
B=(2*medA*medV+v*sq1+w*sq2)
C=-(w*medA*sq2+v*medV*sq1)
media<-c(rep(mean(juntos),3))
vetor<-polyroot(c(C,B,A,1))
vetor2<-c(10^25,10^25,10^25)
for(i in 1:3)
    {if(abs(Im(vetor[i]))<10^{-10})
    {vetor2[i]<-vetor[i]}
}
indice<-0
dif<-abs(vetor2-media)
menor<-min(dif)
for(j in 1:3)
    {if (abs(vetor2[j]-media[j])-menor<10^{-10})
    indice<-j
    }
maple<-vetor2[indice]
#####################################################################
miest4=maple
sigmaest4A=(1/n)*sum((Amarelo-miest4)^2)
sigmaest4V=(1/m)*sum((Vermelho-mean(Vermelho))^2)
L= - ((n +m)/2)*(log( 2*pi   ,exp(1))+ 1)-(n/2)*log(sigmaest4A,exp(
    1))-(m/2)*log(sigmaest4V)
AIC4=((n+m)*(log(2*pi,exp(1))+1)+n*log(sigmaest4A,exp(1))+
    m*log(sigmaest4V,exp(1))+2*k4)                          #AIC
BIC4=((n+m)*(log(2*pi,exp(1))+1)+n*log(sigmaest4A,exp(1))+
    m*log(sigmaest4V,exp(1))+k4*log(n+m,exp(1)))            #BIC
###############FIM DA LOG VEROSSIMILHANÇA PARA MODELOS NORMAIS####
AICc1=AIC1+(2*k1*(k1+1))/(n-k1-1)
AICc2=AIC2+(2*k2*(k2+1))/(n-k2-1)
AICc3=AIC3+(2*k3*(k3+1))/(n-k3-1)
AICc4=AIC4+(2*k4*(k4+1))/(n-k4-1)
menoraic[k]<-min(AIC1, AIC2, AIC3, Re(AIC4))
    {if (menoraic[k]==Re(AIC4))
    contaic4<-contaic4+1
    }
    {if (menoraic[k]==AIC3)
    contaic3<-contaic3+1
    }
    {if (menoraic[k]==AIC2)
    contaic2<-contaic2+1
    }
```

```
    {if (menoraic[k]==AIC1)
    contaic1<-contaic1+1
    }
menoraicc[k]<-min(AICc1, AICc2, AICc3, Re(AICc4))
 {if (menoraicc[k]==Re(AICc4))
 contaicc4<-contaicc4+1}
     {if(menoraicc[k]==AICc3)
    contaicc3<-contaicc3+1}
 {if (menoraicc[k]==AICc2)
    contaicc2<-contaicc2+1}
 {if (menoraicc[k]==AICc1)
    contaicc1<-contaicc1+1}
menorbic[k]<-min(BIC1, BIC2, BIC3, Re(BIC4))
    {if (menorbic[k]==Re(BIC4))
    contbic4<-contbic4+1
    }
    {if (menorbic[k]==BIC3)
    contbic3<-contbic3+1
    }
    {if (menorbic[k]==BIC2)
    contbic2<-contbic2+1
    }
    {if (menorbic[k]==BIC1)
    contbic1<-contbic1+1
    }
}
dados=c(contaic1,contaic2,contaic3,contaic4,contaicc1,contaicc2,
contaicc3,contaicc4,contbic1,contbic2,contbic3,contbic4,FtMIVI,
    FtMDVD,FtMDVI,FtMIVD)
paraAR2AICc=matrix(dados,4,4)
row=c("mivi","mdvd","mdvi","mivd")
col=c("AIC","AICc","BIC","Ft")
rownames(paraAR2AICc)<-row
colnames(paraAR2AICc)=col
paraAR2AICc
```

CAPÍTULO 3

Critérios de informação em modelos de séries temporais

RESUMO

A Análise de Séries Temporais é um ramo da Ciência Estatística dedicado ao tratamento analítico de uma série de instâncias ou observações dependentes. Neste capítulo foi feita uma breve descrição teórica dos modelos de Box e Jenkins; além disso, procedeu-se simulações no programa R, acerca dos modelos AR(1), AR(2), MA(1), MA(2), ARMA$(1,1)$, ARMA$(1,2)$, ARMA$(2,1)$ e ARMA$(2,2)$ para verificar o desempenho assintótico dos critérios AIC, AICc e BIC na seleção de modelos. O BIC apresentou desempenho superior (taxa de acerto TP) aos critérios AIC e AICc à medida que o tamanho amostral aumentou. Para amostras de tamanho 100, os critérios apresentaram-se bons em determinadas situações e insatisfatórios em outras, não havendo sobressaimento em relação aos demais. Para os dados reais, oriundos da série de dados de consumo de energia da região Sudeste, os critérios AIC e AICc selecionaram o modelo ARIMA$(1,1,1)$ enquanto que o BIC selecionou o modelo ARIMA$(0,1,1)$ como o melhor modelo que se ajusta aos dados.

Palavras-chave: Arima. Series temporais. Modelos de Box e Jenkins.

ABSTRACT

The Time Series Analysis is a branch of science devoted to statistical analytic treatment of a number of instances or dependent observations. In this chapter, a brief description of the theoretical models of Box and Jenkins, also proceeded simulations in R program, about the models AR(1), AR(2), MA(1), MA(2), ARMA$(1,1)$, ARMA$(1,2)$, ARMA$(2,1)$ and ARMA$(2,2)$ to check the asymptotic performance of AIC, AICc and BIC in model selection. The BIC outperformed (hit rate TP) to AIC and AICc as the sample size increased. For samples of size 100, the criteria had to be good in certain situations and unsatisfactory in others, there is no sobressaimento in relation to others. For the actual data, derived from the data series of energy consumption in the Southeast, the AIC and AICc selected model ARIMA$(1,1,1)$ while the BIC selected the ARIMA$(0,1,1)$ as the model that best fits the data.

Keywords: Arima. Time series. Box-Jenkins Models.

1 INTRODUÇÃO

A Análise de Séries Temporais é um ramo da Ciência Estatística dedicado ao tratamento analítico de uma série de instâncias ou observações dependentes. Segundo Box, Jenkins e Reinsel (1994), uma série temporal pode ser definida como qualquer conjunto de observações tomadas sequencialmente no tempo, podendo ser discretas ou contínuas. No que tange ao conjunto índice, as séries são ditas contínuas se $T = \{t : t_1 < t < t_2\}$, ou seja, as observações são coletadas de forma ininterrupta, como, por exemplo, o registro de um eletrocardiograma. Por outro lado, o conjunto índice das séries temporais discretas é denotado por $T = \{t_1, t_2, ..., t_N\}$, representando que a coleta das instâncias ocorre em instantes pontuais, podendo ser igualmente espaçados ou não. Contudo, uma série temporal discreta pode ser obtida através da amostragem de uma série contínua, limitada por instantes específicos (MORETTIN; TOLOI, 2006). Muitos são os exemplos que podem ser dados por estas observações, tais como: valores mensais da temperatura média de determinada localidade, valores da precipitação, número mensal de ocorrência de determinado tipo de crime e outros.

Quando os dados são tomados sequencialmente no tempo, espera-se que exista uma correlação entre as observações no instante t e em um tempo subsequente $t+h$. Baseados neste fato, as metodologias estatísticas clássicas não podem ser aplicadas, pois, nestas, algumas pressuposições devem ser satisfeitas tais como a independência dos dados. Sendo assim, uma melhor abordagem a ser utilizada consiste nas técnicas de séries temporais.

Com base em propósitos determinados, o objetivo da análise de séries temporais é construir modelos para as séries em estudo. Pode-se estar interessado em apenas descrever o comportamento da série como, investigar o mecanismo gera-

dor desta, procurar periodicidades ou fazer previsões. No entanto, objetiva-se, que os modelos construídos sejam o mais simples possível e com o menor número de parâmetros.

Assim como ocorre nos procedimentos desenvolvidos para observações independentes, as séries temporais podem ser univariadas ou multivariadas. As séries temporais, consideradas no presente trabalho, representam apenas uma característica de interesse Y_t, séries univariadas; em tempo discreto $T = \{1, 2, ..., N\}$.

Geralmente, as séries temporais possuem variações ou componentes específicos, que caracterizam os seguintes comportamentos:

a) Tendência:

Para Morettin e Toloi (2006), a tendência pode ser entendida como aumento, ou diminuição, gradual das observações ao longo do tempo. Ela representa o efeito de longo prazo ao redor da média, ou seja, o movimento dominante em uma série temporal. Tal movimento pode ser de crescimento/decrescimento linear ou não-linear. Na Figura 1, adaptada de Cryer e Chan (2008), o autor mostra o exemplo de um passeio aleatório com tendência linear de crescimento.

b) Sazonalidade:

É um componente que Morettin e Toloi (2006) caracterizam como difícil de definir, do ponto de vista conceitual e estatístico, mas que pode ser definido como a variação sazonal, refere-se a movimentos similares ou repetição de um padrão que ocorre na série em anos sucessivos. Como exemplos podemos citar que, em média, é de praxe que as vendas de produtos de praia sejam menores no inverno do que no verão. Outro exemplo, discutido por Cryer e Chan (2008) e representado na Figura 2, refere-se aos níveis de dióxido de carbono (CO_2), que são mais altos durante os meses de inverno e menores no verão, para todos os anos.

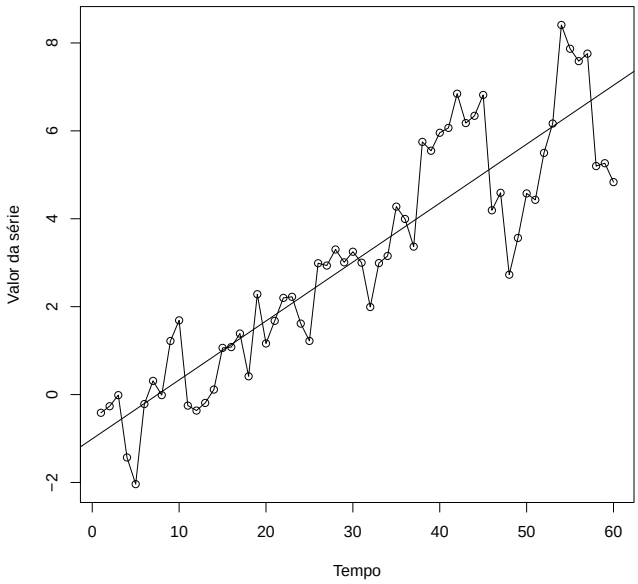

Figura 1 Passeio aleatório com tendência linear

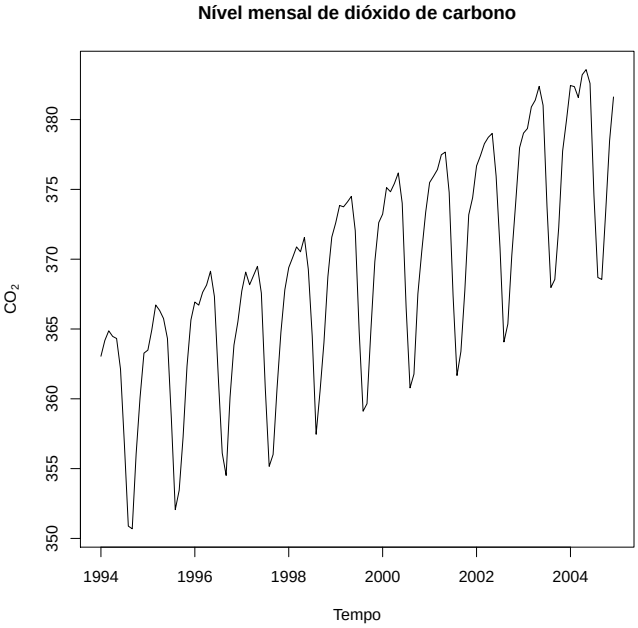

Figura 2 Nível mensal de dióxido de carbono CO_2 em Northwest - Canadá

c) Variações aleatórias:

Movimentos de grande instabilidade que correspondem às flutuações de uma série temporal, devido ao acaso ou fatores aleatórios. Tal movimento é também denominado como "erro aleatório" e representa as variações não capturadas pelos demais componentes.

Segundo Morretin e Toloi (2006), para modelar uma série temporal observada, uma possibilidade é escrevê-la na forma

$$Y_t = f(t) + a_t,$$

em que $f(t)$ é chamada sinal e a_t ruído. De acordo com as hipóteses feitas acerca de $f(t)$ e a_t podemos ter diversas classes de modelos. Na literatura há sugestões sobre a forma de decomposição da série em componentes constituintes, permitindo melhor compreensão sobre a série ao isolar as componentes não-observáveis. A maneira clássica de decomposição fundamenta-se em um modelo aditivo ou multiplicativo envolvendo os componentes da série (tendência, sazonalidade, ruído). A decomposição aditiva pode ser representada através da expressão

$$Y_t = m_t + S_t + \varepsilon_t,$$

e caso o modelo seja multiplicativo, pode-se representar a decomposição como

$$Y_t = m_t S_t \varepsilon_t,$$

em que, m_t e S_t representam os componentes de tendência e sazonalidade (incluindo a variação cíclica), respectivamente; e ε_t retrata o erro aleatório.

Geralmente, utiliza-se a decomposição multiplicativa quando a amplitude da variação da série temporal cresce no decorrer do tempo. Ou seja, há indicativo de que as flutuações sazonais são proporcionais ao nível médio da série.

Conforme apresentado por Morettin e Toloi (2006), os principais objetivos

ao estudar séries temporais podem ser sintetizados a partir dos seguintes itens:

1) Descrição:

Descrever o padrão de comportamento da série através de gráficos, medidas sumárias, verificação de tendência, ciclos e variações sazonais, entre outras ferramentas descritivas. Geralmente este objetivo é parte integrante dos estudos.

2) Modelagem:

Construção de modelos que expliquem o mecanismo gerador da série temporal; uso da variação de uma série para explicar a variação de outra série; investigar a relação de uma série com outras de interesse. A construção desses modelos envolve a estimação de parâmetros e a avaliação da qualidade estatística do ajuste. Pode-se citar, por exemplo, a série que modela o consumo mensal de energia dos estados da região Sul do Brasil, de 1940 a 1990. Se o intuito for, apenas determinar o mecanismo gerador do consumo, temos um exemplo em que estamos interessados apenas na modelagem do fenômeno.

3) Previsão:

Estimação de valores futuros da série temporal, considerando os valores em instantes anteriores. As previsões podem ser a curto, médio ou a longo prazo. Em muitos estudos de séries temporais, especialmente econômicos, a previsão é crucial, uma vez que possibilita conhecer a provável evolução da série, e, consequentemente, tomar decisões apropriadas. Como exemplo, se pode citar um investidor da bolsa de valores que, com base em cotações anteriores de uma determinada companhia decide se deve, ou não, investir na mesma e, para isto, se baseia nas previsões das ações desta empresa.

4) Controle de processos:

Uso específico para monitorar os valores da série que representam a qualidade de um processo, o que preconiza a detecção de mudanças, estruturais ou periodicidades na série. Estas alterações estruturais podem indicar, no âmbito do controle estatístico de qualidade, que o modelo adotado para retratar o funcionamento do processo já não é válido. Pode-se citar, por exemplo, a série que modela o número de pessoas infectadas pela dengue anualmente; se houver uma mudança brusca no número de casos, pode ser indício de um surto (se houve um aumento brusco), ou então, as políticas adotadas surtiram efeito (se o número de casos reduzir drasticamente). Em ambos os casos, há uma mudança no padrão da série. Nesta perspectiva, vale salientarmos que, em uma investigação pode-se explorar mais de um dos objetivos descritos acima.

Segundo Brillinger (2001), os conceitos de séries temporais são utilizados com diferentes enfoques, por exemplo: modelos probabilísticos, métodos não-paramétricos, estudos longitudinais e medidas repetidas.

As abordagens utilizadas em séries temporais são analisadas de acordo com dois domínios: temporal e de frequências. Segundo Morettin e Toloi (2006), em ambos, sendo a análise feita no domínio temporal ou no domínio de frequências, o propósito é construir modelos para as séries. Quando a análise é realizada no domínio do tempo, os modelos obtidos são paramétricos, enquanto que, no domínio de frequências os modelos construídos são não-paramétricos (MORETTIN; TOLOI, 2006). Neste sentido consideraremos o primeiro enfoque.

A classe dos modelos paramétricos caracteriza-se pela sua quantidade finita de parâmetros. Os modelos dessa classe, comumente usados, são os autorregressivos e de médias móveis integrado, ARIMA(p, d, q), sendo p a ordem da parte autorregressiva do modelo, d o número de diferenças para tornar a série estacionária e q a ordem da parte de médias móveis. Além dos modelos ARIMA(p, d, q),

suas variações também são utilizadas. São elas, os modelos de regressão, modelos estruturais e modelos não-lineares.

Dentre a classe dos modelos não-paramétricos, os mais conhecidos são a função de autocovariância/autocorrelação e a sua transformada de Fourier, também conhecida como espectro de frequência da série temporal, e Wavelets (TSAY, 2000). Comumente, do ponto de vista clássico, quando os modelos não-paramétricos são aplicados, costuma-se denominar a investigação como análise espectral.

Para a seleção dos melhor modelo que se ajusta aos dados, os critérios de AIC, AICc e BIC podem ser utilizados para este fim. Nosso objetivo, no presente capítulo, é verificar via simulações, o desempenho dos três critérios; em um primeiro momento os critérios serão analisados assintóticamente e, posteriormente, para amostras de tamanho $N = 100$, estudaremos as taxas de verdadeiro positivo (TP), falso positivo (FP) e falso negativo (FN) para modelos de séries temporais comumente utilizados. Esperamos que, via simulações, possamos determinar qual critério deve ser utilizado quando estivermos comparando modelos de séries temporais.

2 REFERENCIAL TEÓRICO

2.1 Breve história sobre a análise de séries temporais

Conforme apresentado em Tsay (2000), as aplicações contribuíram fundamentalmente para o desenvolvimento das metodologias de séries temporais. Há bastante tempo, são realizadas coletas e acompanhamento de séries temporais. Uma das séries mais antigas, conhecida na literatura através de uma das publicações do Graunt (1662), intitulada por *Natural and Political Observations upon the Bills of Mortality*, retrata a distribuição das idades de morte no século XVII em Londres e lança as bases para a demografia (ESQUIVEL, 2012). Outra aplicação que motivou os estudos da análise de séries temporais, segundo os historiadores, refere-se à clássica série do número médio anual de manchas solares de Wolfer para o período de 1749 a 1924 (BRILLINGER, 2001; MORETTIN; TOLOI, 2006).

No que tange ao desenvolvimento metodológico, a análise espectral pode ser considerada como técnica pioneira, pois produziu a primeira aplicação, datada em 1664, através do experimento de Newton, que envolvia a passagem da luz solar decomposta em seus componentes por um prisma (BRILLINGER, 2001).

Com o passar do tempo, houveram outros desenvolvimentos técnicos na metodologia de séries temporais. Podemos citar o artigo de T. N. Thiele de 1880 que contém contribuições relacionadas à formulação e à análise de um modelo para uma série temporal constituída de uma soma de componentes da regressão, movimento Browniano e um ruído branco, − a partir do qual Kalman desenvolveu um procedimento recursivo para estimar a componente da regressão e predizer o movimento Browniano, procedimento conhecido a partir de 1961 por filtro de Kalman (KALMAN; BUCY, 1961)

Todavia, a explosão no uso da técnica ocorreu na década de 80 como evidenciam os diversos artigos publicados nas revistas específicas de Estatística, cujo título possui o termo "filtro de Kalman". E a partir daí, foram desenvolvidas diversas formas de aplicabilidade da técnica, como por exemplo, tratamento de valores faltantes ou *missings*; no desenvolvimento de novos métodos para a extração de sinal; para alisamento e ajuste sazonal; entre outros (TSAY, 2000).

Vale ressaltarmos que as abordagens utilizadas em séries temporais (domínio temporal e domínio de frequências) se desenvolveram de forma independente a partir das contribuições dos seus defensores. Segundo Tsay (2000), no início da história ocorriam muitos debates e críticas entre as duas abordagens; contudo, a forte separação entre as escolas está diminuindo de acordo com o mesmo, e, atualmente, a escolha entre qual abordagem usar depende, principalmente, do objetivo da análise e da experiência do analista, podendo utilizar técnicas de ambas as abordagens, conforme percebemos.

De acordo com a literatura especializada, o século XX proporcionou muitos avanços importantes na análise de séries temporais, principalmente na área da previsão. A partir de 1920 a previsão deixou de ser realizada através de extrapolações simples ao redor do valor médio e passou a levar em consideração os valores passados da série temporal. O desenvolvimento na área iniciou-se com Yule (1927), ao ajustar um modelo linear aos dados de manchas solares, chamando-o de autorregressivo (AR) (KLEIN, 1997).

Com a evolução dos estudos na área da previsão de séries temporais e com a chegada do computador nos anos 50, houve uma popularização dos modelos de previsão, já que se tratava de uma época em que foi criada uma técnica simples, eficiente e razoavelmente precisa para tratar as causas de flutuações e realizar previsões a partir do padrão básico identificado na série. Tal técnica ficou conhecida

como alisamento ou suavização exponencial. Sendo Brown o precursor da técnica, ele trabalhou em prol do desenvolvimento da metodologia de alisamento exponencial e publicou dois importantes livros da época, o *Statistical Forecasting for Inventory Control* de 1959 e o *Smoothing, Forecasting and Prediction of Discrete Time Series*, em 1963 (GARDNER, 2006).

Em paralelo aos trabalhos de Brown, Charles Holt, com o apoio do *Office of Naval Research* (ONR), desenvolvia técnicas de suavização exponencial para tratar séries com tendências e sazonalidades. Inicialmente, o seu trabalho foi documentado no memorando do próprio ONR, em 1957, e somente em 2004 publicado no *International Journal of Forecasting* (HOLT, 2004a, 2004b). Apesar do registro do trabalho em 1957, a técnica desenvolvida por Holt apenas atingiu ampla divulgação a partir do trabalho de Winters (1960), que testou os métodos de Holt a partir de dados empíricos. Desde então, a técnica ficou conhecida como suavização exponencial de Holt-Winters (ESQUIVEL, 2012).

Na década de 70 houve um marco importante para a análise de séries temporais; uma nova classe de modelos vinha à tona. Esta classe ficou conhecida como metodologia de Box e Jenkins a partir da publicação, em 1970, do clássico livro *Time Series Analysis, Forecasting and Control*. Em linhas gerais, a metodologia de Box e Jenkins ou ARIMA trabalha com modelos probabilísticos lineares e considera um processo sistemático na modelagem das séries. Tal processo inicia-se na especificação e na identificação do modelo mais adequado aos dados, passando pela estimação dos parâmetros, verificação da adequação do modelo, e a previsão dos valores futuros (MORETTIN; TOLOI, 2006).

Com o passar das décadas e com a evolução dos estudos, os pesquisadores da época notaram a necessidade de considerar a não-linearidade na análise de séries temporais, haja vista que muitas das aplicações não apresentavam a lineari-

dade suposta nos modelos criados até o momento (TSAY, 2000).

A partir da década de 80, com a melhora no acesso aos microprocessadores, os pesquisadores vêm utilizando técnicas alternativas para a detecção de padrões e tratamento de séries não-lineares, principalmente, de uma forma automatizada.

2.2 Definições básicas

Antes de apresentar os modelos clássicos para análise de séries temporais, faz-se necessário abordarmos algumas definições relacionadas aos processos estocásticos dentro do contexto de séries temporais.

2.2.1 Processos estocásticos

Os modelos utilizados na representação dos fenômenos investigados na forma de séries temporais são ditos processos estocásticos ou aleatórios, ou seja, fenômenos controlados por leis probabilísticas.

Um processo estocástico é geralmente interpretado como uma coleção de variáveis aleatórias dependentes de outras variáveis, como o tempo. Em outras palavras, um processo estocástico é o conjunto de todas as possíveis trajetórias de um fenômeno ou uma coleção de realizações do processo físico.

A ideia de um processo estocástico e uma de suas realizações é análoga à utilizada na descrição de população e de amostra. Na investigação de muitos problemas práticos coleta-se informações de uma amostra do fenômeno visando conhecê-lo na população. De forma semelhante ocorre nos estudos de séries temporais, em que se observa uma realização do processo aleatório estudado, $-$ denominada série temporal $-$, com intuito de conhecer o processo estocástico associado. Por exemplo, a temperatura do ar, em um dado local e intervalo de tempo,

é considerada como uma realização ou trajetória do processo estocástico gerador das temperaturas (populacional) no local do experimento (GUJARATI, 2006; MORETTIN; TOLOI, 2006).

Considere um processo estocástico $\{Y(t), t \in T\}$, em que T indexa a evolução do parâmetro físico ou conjunto índice, em particular o tempo. Usualmente, descreve-se um processo estocástico através de suas funções média, autocovariância e variância. A função média de $Y(t)$ pode ser denotada como

$$E[Y(t)] = \mu_t,$$

a autocovariância entre $Y(t)$ e $Y(s)$ por

$$\gamma_{t,s} = Cov[Y(t), Y(s)] = E[(Y(t) - \mu_t)(Y(s) - \mu_s)],$$

sendo t e s dois instantes distintos. Ou ainda,

$$\gamma(h) = Cov[Y(t), Y(t+h)] = E[(Y(t) - \mu_t)(Y(t+h) - \mu_{t+h})],$$

representando a autocovariância entre $Y(t)$ e $Y(t+h)$, sendo h a defasagem ou *lag*, isto é, a distância entre dois instantes de tempo. E caso h seja zero, tem-se a variância de $Y(t)$ dada por

$$\gamma(0) = Var[Y(t)] = \sigma_t^2.$$

Existem alguns processos estocásticos com comportamentos específicos, caracterizando os diferentes tipos de processo. Como exemplo, pode-se citar os processos estocásticos conhecidos na literatura e que recebem atenção em muitos estudos de séries temporais, a saber: processos estacionários; processos Gaussi-

anos ou normais; ruído branco; passeio aleatório; processos Markovianos; movimento browniano; entre outros.

Para o propósito do presente trabalho as ideias sobre os processos estacionários, ruído branco e passeio aleatório serão apresentadas e utilizadas.

Os processos estocásticos podem apresentar estacionariedade "fraca", também denominada como "ampla" ou ainda de "segunda ordem"; ou estacionariedade "forte" (estrita). Contudo, nas situações práticas, restringe-se a análise ao caracterizar os processos através dos seus dois primeiros momentos, o que possibilita investigar os "processos estacionários de segunda ordem" (MORETTIN; TOLOI, 2006).

Um processo estocástico é dito fracamente estacionário se e somente se

1. $E\left[Y\left(t\right)\right] = \mu_t = \mu$, constante, para todo $t \in T$;
2. $Var\left[Y\left(t\right)\right] = \sigma_t^2 = \sigma^2$, constante, para todo $t \in T$;
3. $\gamma\left(h\right) = Cov\left[Y\left(t\right), Y\left(t+h\right)\right]$, é uma função de h.

Por convenção, um processo fracamente estacionário é conhecido simplesmente como *processo estacionário* (GUJARATI, 2006; MORETTIN; TOLOI, 2006).

Caso uma série não seja estacionária, esta pode transformar-se em tal ao ser diferenciada uma quantidade d finita de vezes. Ao realizar esse procedimento, a série também estará livre de tendência, ou seja, a componente de tendência será removida da série temporal.

Quando uma série temporal não é estacionária, mas torna-se estacionária após d diferenças, costumamos chamá-la como série integrada de ordem d e denotá-la como $I(d)$. Assim, por convenção, o processo $I(d = 0)$ representa uma série estacionária (GUJARATI, 2006).

A diferenciação na série ocorre ao aplicar o conhecido operador diferença dado por

$$\Delta^d Y_t = \Delta\left[\Delta^{d-1} Y_t\right],$$

em que $\Delta Y_t = Y_t - Y_{t-1}$ é a primeira diferença da série temporal e denota a variação dos dados entre dois tempos consecutivos; e d representa o número de diferenças. Geralmente, no máximo duas diferenças são suficientes para tornar a série temporal estacionária.

Todavia, para algumas séries temporais as diferenças não serão suficientes para atingir a estacionariedade. Costuma-se aplicar uma transformação não linear nos dados antes de realizar as diferenças, de modo que a série temporal passe a ser estacionária. De uma forma geral, aplica-se uma transformação na série temporal com o intuito de alcançar a estabilidade da variância e ou simetria nos dados, ou ainda, em séries sazonais, tornar o efeito sazonal aditivo. A escolha da transformação deve ser realizada de acordo com o comportamento da série temporal. Tipicamente, utiliza-se a conhecida transformação de Box-Cox; que pode ser vista em Morettin e Toloi (2006).

Um importante processo estocástico refere-se ao ruído branco. Um processo $\{\varepsilon(t), t \in T\}$ é dito ser **ruído branco** se

1. $E[\varepsilon(t)] = 0$, constante, para todo $t \in T$;
2. $Var[\varepsilon(t)] = \sigma_\varepsilon^2$, para todo $t \in T$;
3. $Cov[\varepsilon(t), \varepsilon(s)] = 0$, para todo $t \neq s$.

Essas especificações representam uma coleção de variáveis aleatórias não correlacionadas, centradas em zero e de variância finita e constante. Usualmente utiliza-se a notação $\varepsilon(t) \sim RB\left(0, \sigma_\varepsilon^2\right)$ para indicar o particular tipo de processo estocástico.

Um processo estocástico é dito ser um passeio aleatório se a primeira diferença deste processo resulta em um ruído branco. Uma definição mais formal consiste em considerar uma sequência de variáveis aleatórias $\{\varepsilon_t, t \geq 1\}$, identicamente distribuídas com média μ_ε e variância σ_ε^2. Assim, um processo Y_t é chamado de **passeio aleatório** se

$$Y_t = Y_{t-1} + \varepsilon_t.$$

Na prática existem séries temporais que se comportam como um passeio aleatório, como a difusão gasosa ou líquida, o caminho percorrido por uma molécula e os preços diários de ações.

2.2.2 Função de autocovariância e autocorrelação

Uma importante ferramenta para analisar a estrutura de uma série temporal consiste na quantificação da dependência entre as instâncias da série. Essa quantificação é realizada através da autocovariância ou autocorrelação amostral, de forma análoga ao coeficiente de correlação amostral ou covariância entre duas diferentes variáveis. A autocorrelação nada mais é do que a correlação defasada em um número h de unidades de tempo, de uma série consigo mesma, e que pode ser estimada por

$$r_h = \frac{\sum_{t=1}^{n-h} \left[Y(t) - \overline{Y}(t)\right] \left[Y(t+h) - \overline{Y}(t)\right]}{\sum_{t=1}^{n} \left[Y(t) - \overline{Y}(t)\right]^2},$$

ou $r_h = \dfrac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)}$, em que $\widehat{\gamma}(h)$ é a covariância amostral na defasagem h ($h = 0, 1, 2, \cdots$); $\widehat{\gamma}(0)$ é a variância amostral, ao considerar séries estacionárias; e $\overline{Y}(t)$ é a média da série temporal. Para a autocorrelação temos que ela é adimensional e varia de -1 a 1.

Uma forma de representar as autocorrelações é através do gráfico denominado correlograma ou função de autocorrelação amostral. Em tal gráfico esboça-se os primeiros valores de r_h contra os valores não-negativos das defasagens h considerados na análise.

Em linhas gerais, podemos interpretar o correlograma buscando a identificação de padrões que retratem características da série temporal. Por exemplo, se uma série temporal é puramente aleatória, isto é, um ruído branco, os seus valores defasados não são correlacionados e espera-se que r_h seja zero; teoricamente, os coeficientes de autocorrelação amostral se distribuem, aproximadamente, como uma normal com média zero e variância igual ao inverso do tamanho amostral $\left(\frac{1}{N}\right)$, sob a suposição de estacionariedade fraca. Já se o correlograma apresenta decaimento muito lento das autocorrelações amostrais há uma indicação de não estacionariedade na série (ESQUIVEL, 2012; GUJARATI, 2006).

O correlograma pode ser utilizado, também, na identificação do modelo paramétrico mais adequado para ajustar aos dados investigados, especialmente o ARIMA, como estimativa da função de autocorrelação (fac). A identificação do modelo também pode ser realizada através das funções estimadas de autocovariância (facv) e autocorrelação parcial (facp), assim como são apresentadas na seção 2.3.

A ideia da função de autocorrelação parcial é semelhante à ideia do coeficiente de regressão parcial, já que ambos controlam a influência dos demais fatores considerados na correlação; variáveis, no caso da regressão, e observações no contexto de séries temporais. A autocorrelação parcial quantifica a correlação entre observações que estejam em h períodos afastados (isto é, entre as observações Y_t e Y_{t-h}), após ocorrer o controle das correlações nas defasagens anteriores a h (GUJARATI, 2006; MORETTIN; TOLOI, 2006).

2.2.3 Testes estatísticos para avaliar características em séries temporais

Antes de estimar os componentes ou iniciar o processo de modelagem de uma série, recomenda-se analisar as características da série temporal. A seguir, serão apresentados alguns testes úteis para este fim. Caso a série temporal tenha mais de um componente não aleatório, é conveniente testar a existência de um deles (tendência ou sazonalidade) após a eliminação do outro componente (MORETTIN; TOLOI, 2006).

2.2.3.1 Testes para a detecção de estacionariedade

i) Teste com base no correlograma

Assim como mencionado na seção anterior (2.2.2), se o correlograma apresentar autocorrelações amostrais que diminuam gradualmente, teremos um indicativo de não estacionariedade na série temporal. Por outro lado, se as autocorrelações forem todas nulas, a partir da primeira defasagem, encontraremos o padrão das autocorrelações de um processo estocástico puramente aleatório.

Entre essas duas situações extremas, na prática, pode-se encontrar correlogramas com perfis não tão facilmente identificáveis. Por esse motivo, costuma-se também avaliar a significância estatística das autocorrelações estimadas (r_h). Essa avaliação pode ser realizada através do erro padrão do estimador do coeficiente de autocorrelação $\left(EP(r_h) = \frac{1}{\sqrt{N}}\right)$, assim como, através do intervalo de confiança para o parâmetro ρ, isto é, o coeficiente de autocorrelação populacional. O intervalo de confiança pode, então, ser incluído no correlograma. Ao considerar 95% de confiança, por exemplo, o intervalo de confiança para qualquer ρ_h será dado por

$$IC\left(\rho_h; 95\%\right) = \left[0 \pm 1,96\frac{1}{\sqrt{N}}\right] = \left[-1,96\frac{1}{\sqrt{N}}; 1,96\frac{1}{\sqrt{N}}\right].$$

Desta forma, há 95% de chance de que cada ρ_h esteja dentro dos limites de confiança. E se um particular r_h estiver incluso no $IC\left(\rho_h; 95\%\right)$, a hipótese de que $\rho_h = 0$ não será rejeitada (ESQUIVEL, 2012).

Pode-se também testar se todos os coeficientes de autocorrelação são simultaneamente iguais a zero. Para testar a hipótese conjunta, costuma-se utilizar a estatística Q de Box e Pierce ou o teste alternativo de Ljung-Box, que se mostra mais poderoso em pequenas amostras do que a estatística Q de Box e Pierce (BROCKWELL; DAVIS, 2002; GUJARATI, 2006).

ii) Testes para raiz unitária

A estacionariedade em uma série pode, ainda, ser avaliada através dos testes de raiz unitária. Estes testes verificam se o polinômio (autorregressivo e ou de médias móveis) analisado possui uma raiz sobre o círculo unitário.

Para efeito de ilustração, seja um modelo que considera a regressão do valor de Y_t sobre o seu valor no instante imediatamente anterior $(t-1)$, conhecido como autorregressivo de ordem 1 $(AR\left(1\right))$,

$$Y_t = \phi Y_{t-1} + \varepsilon, \text{ sendo } \varepsilon \sim RB\left(0, \sigma_\varepsilon^2\right),$$

em que, o peso ϕ, em valor absoluto, é inferior a um $(|\phi| < 1)$.

Se a hipótese nula de existência de uma raiz unitária $(\phi = 1)$ não for rejeitada, temos uma situação de não estacionariedade, conhecida como passeio aleatório (GUJARATI, 2006).

Na literatura especializada, o teste de raiz unitária mais conhecido é o clássico Dickey - Fuller (abreviadamente DF), proposto no trabalho de Dickey e Fuller (1979). O teste de Dickey - Fuller pode ser encontrado, por exemplo, no apêndice B de Morettin e Toloi (2006).

Houveram melhorias no teste de Dickey-Fuller (como o DF aumentado e testes para mais de uma raiz unitária) e propostas de testes específicos que serviriam para complementar a investigação de estacionariedade. É o caso, por exemplo, do teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS), proposto por Kwiatkowski et al. (1992), sendo esta a melhoria mais conhecida.

2.2.3.2 Testes para a detecção de tendência

A presença do componente de tendência pode ser avaliada através de testes estatísticos, além, é claro, da inspeção gráfica da série temporal. Na literatura sobre séries temporais existem algumas propostas de testes que investigam a presença do componente de tendência. Esquivel (2012), cita os testes de estabilidade de Mann-Kendall; o teste de Wald-Wolfowitz; o teste de Cox-Stuart e o teste de Spearman para séries temporais.

i) Teste de sequências (Wald-Wolfowitz)

Considere as N observações Y_t, $t = 1, 2, \ldots, N$, de uma série temporal, e seja m a mediana destes valores. Atribuímos a cada valor Y_t o símbolo A, se ele for maior ou igual a m, e B, se ele for menor que m. Teremos, então, $N =$ (n_1 pontos $A + n_2$ pontos B). A estatística de teste é

$T_1 =$ Número total de sequências (isto é, grupos de símbolos iguais).

Rejeitamos a hipótese nula H_0 se há poucas sequências, ou seja, se T_1 for pequeno. Para um dado α, rejeitamos H_0 se $T_1 < w_\alpha$, em que w_α é o α-quantil da distribuição de T_1, que é tabelado e conforme Moretttin e Toloi (2006), pode ser encontrado em Conover (1998).

Para n_1 ou n_2 maior que 20, podemos utilizar a aproximação normal, isto é, $T_1 \sim N(\mu, \sigma^2)$, em que

$$\mu = \frac{2n_1n_2}{N} + 1,$$

$$\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)}}.$$

2.2.3.3 Testes para a detecção de sazonalidade

Os testes utilizados para avaliar a presença de sazonalidade são específicos para o caso determinístico, isto é, as séries que possuem sazonalidade determinística, ou seja, que podem ser previstas perfeitamente a partir de observações anteriores.

A sazonalidade determinística pode ser testada usando testes usuais, ao considerarmos as suposições dos modelos lineares generalizados, como pode ser visto em Draper e Smith (1998); ou fazendo uso dos testes não-paramétricos, como, por exemplo, o teste de Friedman (adequado para várias amostras dependentes/emparelhadas); e teste de Kruskall-Wallis (apropriado para várias amostras independentes) (ESQUIVEL, 2012).

i) Teste de Kruskall-Wallis

Ao aplicar o teste de Kruskall-Wallis em uma série temporal mensal, cada mês da série é suposto ser uma das k amostras de uma população dentro de cada um dos p anos. Considere Y_{ij} $(i = 1, \cdots, n_j; j = 1, \cdots, k)$ as observações da série. Sejam R_{ij} os postos obtidos ao ordenar todas as N observações $\left(N = \sum_{j=1}^{k} n_j\right)$ e R_j a soma dos postos associados à $j-$ésima amostra $\left(R_j = \sum_{i=1}^{n_j} R_{ij}\right)$.

A estatística do teste de Kruskal-Wallis será expressa por

$$T_{KW} = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{(R_j)^2}{n_j} - 3(N+1).$$

Sob H_0, a distribuição assintótica de T_{KW} é uma qui-quadrado com $k-1$ graus de liberdade. A hipótese nula de ausência de sazonalidade é rejeitada se o valor calculado a partir de T_{KW} for maior ou igual ao valor crítico da distribuição da estatística T_{KW}.

2.3 Modelos de Box e Jenkins

Na literatura sobre séries temporais, encontramos disponíveis duas estratégias clássicas de modelagem. Uma das estratégias é considerada a mais simples e se refere aos modelos de suavização exponencial, e a outra abordagem caracteriza a conhecida metodologia de Box-Jenkins.

A metodologia de Box e Jenkins considera os modelos autorregressivos integrados de médias móveis (ARIMA, sigla oriunda do termo "AutoRegressive Integrate Moving Average") e suas extensões. Tal metodologia fundamenta-se na construção de modelos ajustados aos dados à luz das suas propriedades probabilísticas. Isto é, o método possibilita realizar a descrição e previsão em séries temporais estacionárias, podendo ter o seu uso estendido para séries não estacionárias homogêneas.

A construção dos modelos na metodologia de Box e Jenkins segue um ciclo iterativo que possibilita identificar o processo estocástico gerador dos dados, assim como os seus parâmetros. O ciclo consiste de quatro estágios, que são:

1. A especificação dos modelos a serem considerados na análise, de acordo com as características existentes na série temporal. Ou seja, uma classe de

modelos é proposta, como, por exemplo, a classe dos modelos autorregressivos (AR);

2. A identificação de um particular modelo e seus parâmetros;

3. A estimação dos parâmetros do modelo identificado na etapa anterior com base nos dados. As variações dos métodos de máxima verossimilhança e dos mínimos quadrados são os métodos de estimação usuais;

4. A checagem ou diagnóstico do modelo ajustado. Estágio em que verifica se o modelo retrata satisfatoriamente a dinâmica dos dados. Usualmente, investiga-se as estimativas resíduos, com o intuito de avaliar se estes são adequados para o objetivo principal da análise. Espera-se que os resíduos sejam ruídos brancos, e, para este fim, pode-se utilizar os testes para autocorrelação residual, semelhantes aos apresentados na seção 2.2.3; sendo que agora analisa-se a autocorrelação entre os resíduos ao invés da autocorrelação entre as observações da série temporal.

Caso o modelo selecionado seja adequado à realidade dos dados, este será usado no processo de previsão da série temporal. Por outro lado, se o ajuste realizado não for satisfatório, reinicia-se o ciclo a partir do segundo estágio ou fase, que é considerada a fase crítica da metodologia. Isso porque, há a possibilidade de que diferentes analistas identifiquem modelos distintos ao investigar a mesma série temporal (GUJARATI, 2006; MORETTIN; TOLOI, 2006).

De uma forma geral, os analistas utilizam modelos com poucos parâmetros, mas com uma representação satisfatória da estrutura dos dados; prevalecendo, na modelagem, o princípio da parcimônia, isto é, modelos mais simples costumam ser melhores.

Na fase de identificação, a escolha do modelo é realizada através das autocorrelações e autocorrelações parciais estimadas, as quais devem representar as funções teóricas de autocorrelação e autocorrelação parcial, fac e facp, respectivamente. Além dos métodos alternativos que utilizam, por exemplo, funções que penalizam o ajuste de modelos com muitos parâmetros ou não parcimoniosos, tais como: Critério de informação de Akaike (AIC) (AKAIKE, 1974); Critério de informação de Akaike corrigido(AICc) (SUGIURA, 1978); Critério de Schwarz (BIC) (SCHWARZ, 1978); entre outros. Para o presente trabalho, utilizamos os critérios AIC, AICc e BIC, pelo fato de que esses são comumente aplicados na seleção de modelos paramétricos.

2.3.1 Operadores

Conforme Morettin e Toloi (2006), alguns operadores são amplamente utilizados no desenvolvimento dos modelos de Box e Jenkins. São eles:

i) B operador de retardo, defasagem ou translação para o passado, o qual ocasiona uma defasagem de um período de tempo m para trás a cada vez que é utilizado. É definido por:

$$BY_t = Y_{t-1}; B^2Y_t = Y_{t-2}; \cdots ; B^mY_t = Y_{t-m}.$$

ii) F operador de translação para o futuro, responsável por uma defasagem de um período de tempo m para frente a cada vez que é utilizado. Definido por:

$$FY_t = Y_{t+1}; F^2Y_t = Y_{t+2}; \cdots ; F^mY_t = Y_{t+m}.$$

iii) Δ operador de diferença. Definido por:

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - B^1 Y_t = (1 - B)Y_t,$$

logo,

$$\Delta = 1 - B.$$

2.3.2 Modelos autorregressivos - AR(p)

Considere $\{Y_t\}$ uma série temporal com N observações. O modelo autorregressivo, de ordem p, AR(p), é expresso como

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t, \tag{1}$$

em que ε_t representa o erro aleatório que não pode ser explicado pelo modelo, sobre o qual supõe-se ser um ruído branco $\left(\varepsilon_t \sim RB\left(0, \sigma^2\right)\right)$; $\phi_1, \phi_2, \cdots, \phi_p$ são parâmetros autorregressivos do modelo AR; e c é uma constante, que não é a média.

Ao ser aplicado o operador defasagem em (1), e evidenciando Y_t, resultará na seguinte expressão resumida do modelo AR:

$$\phi(B)Y_t = c + \varepsilon_t,$$

em que $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ é o polinômio operador de defasagens ou operador autorregressivo de ordem p.

Segundo Esquivel (2012), a esperança da $j-$ésima autocovariância de um

processo AR(p) é dada, por

$$E\left[Y_t\right] = \frac{c}{1 - \phi_1 - \phi_2 - \cdots - \phi_p},$$

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \cdots + \phi_p\gamma_{j-p}, \quad \gamma_j = \gamma_{-j}, \quad j > 0.$$

Ao dividir γ_j por $\gamma_0 = Var\left[Y_t\right]$, resulta na função de autocorrelação (fac),

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \cdots + \phi_p\rho_{j-p}, \quad j > 0,$$

em que:

$$Var\left[Y_t\right] = \frac{\sigma_\varepsilon^2}{1 - \phi_1\rho_{j-1} - \phi_2\rho_{j-2} - \cdots - \phi_p\rho_{j-p}}.$$

A função autocorrelação (fac) teórica de um processo autorregressivo decai segundo exponenciais e/ou senóides amortecidas, e é infinita em extensão. E as autocorrelações parciais de um AR(p) têm as p primeiras autocorrelações estatisticamente diferentes de zero. Ou seja, a função autocorrelação parcial (facp) de um processo autorregressivo apresenta um corte na ordem do modelo correspondente.

Um processo autorregressivo de ordem p é dito estacionário se todas as raízes do polinômio $\phi(B)$ estiverem fora do círculo unitário, que pode ser denotado por $a^2 + b^2 = 1$. A estacionariedade é determinada pelos valores dos parâmetros autorregressivos. Por exemplo, um AR(1) será estacionário se $-1 < \phi_1 < 1$ (ou $|\phi_1| < 1$); e para o modelo AR(2), a região de estacionariedade corresponde à $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$ e $\phi_2 - \phi_1 < 1$.

2.3.3 Modelos de médias móveis - MA(q)

Considere $\{Y_t\}$ uma série temporal. Esta série é dita ser um modelo de média móvel de ordem q, MA(q), se puder ser expressa como:

$$Y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}, \tag{2}$$

em que $\left(\varepsilon_t \sim RB\left(0, \sigma^2\right)\right)$; $\theta_1, \theta_2, \cdots, \theta_q$ são os parâmetros de médias móveis do modelo MA; e μ é a média do processo.

A equação (2) pode ser reescrita na forma

$$Y_t = \mu + \theta(B)\varepsilon_t,$$

em que $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ é o polinômio operador de médias móveis de ordem q.

Para um processo MA(q) tem-se que

$$E\left[Y_t\right] = \mu;$$

$$\gamma_j = \begin{cases} 1 + \left(\theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)\sigma_\varepsilon^2, j = 0, \\ \left(-\theta_j + \theta_1\theta_{j+1} + \theta_2\theta_{j+2} + \cdots + \theta_{q-j}\theta_q\right)\sigma_\varepsilon^2, j = 1, 2, \cdots, q, \\ 0, j > q, \end{cases}$$

em que $E\left[Y_t\right]$ representa a média do processo; γ_j a função de autocovariância (facv), e γ_0 é a variância de Y_t. Novamente, ao dividir γ_j por γ_0, temos a função de autocorrelação (fac) e esta pode ser representada por

$$\rho_j = \begin{cases} 1, j = 0, \\ \dfrac{\left(-\theta_j + \theta_1\theta_{j+1} + \theta_2\theta_{j+2} + \cdots + \theta_{q-j}\theta_q\right)}{1 + \left(\theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)}, j = 1, 2, \cdots, q, \\ 0, j > q. \end{cases}$$

A ordem de um modelo de médias móveis é indicada através da função de autocorrelação (fac), ao invés da função autocorrelação parcial (facp), como nos modelos autorregressivos. De acordo com a função de autocorrelação do MA, pode-se verificar que até o lag q as autocorrelações são não nulas, indicando que a fac de um MA(q) é finita. Já a facp de um MA(q) comporta-se de forma semelhante à fac de um AR(p).

Os processos de médias móveis são sempre estacionários, isto é, os modelos MA não possuem restrições nos seus parâmetros. Porém, esses modelos não têm unicidade. Para que a unicidade possa ser garantida, as restrições com relação à invertibilidade do processo devem ser consideradas.

Assim,condição de invertibilidade para um modelo MA(q) é de que todas as raízes da equação característica $\theta(B) = 0$ tenham módulo maior que um, ou seja, estejam fora do círculo unitário. Para exemplificar, considere um MA(2); este processo será invertível se as seguintes condições forem satisfeitas

$$\begin{cases} -1 < \theta_2 < 1, \\ \theta_1 + \theta_2 < 1, \\ \theta_2 - \theta_1 < 1, \end{cases}$$

o que equivale às condições de estacionariedade para um AR(2).

2.3.4 Modelo autorregressivos e de médias móveis - ARMA(p, q)

Um processo de médias móveis puro é pouco intuitivo para representar o comportamento de uma particular série temporal, enquanto que a aplicação dos modelos autorregressivos é bastante natural em muitas áreas do conhecimento. Na prática, considera-se os termos autorregressivos e de médias móveis simultaneamente, buscando a melhoria no ajuste ao considerar poucos parâmetros. Ao se

combinar o modelo autorregressivo (AR) e o modelo médias móveis (MA), obtém-se o modelo autorregressivo e de médias móveis (ARMA).

Considerando a série temporal $\{Y_t\}$, com $t = 1, 2, \ldots, N$ o modelo ARMA de ordem (p, q), sendo p a ordem da parte autorregressiva e q a ordem da parte médias móveis, podemos transcrevê-lo da seguinte forma

$$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}, \qquad (3)$$

em que $\varepsilon_t \sim RB\left(0, \sigma_\varepsilon^2\right)$.

A expressão de um ARMA(p, q) também pode ser reescrita, considerando os operadores autorregressivo e de médias móveis usuais, o que resulta na forma compacta dada por:

$$\phi(B) Y_t = c + \theta(B) \varepsilon_t.$$

As condições de estacionariedade em um processo ARMA(p, q) são as mesmas de um processo AR(p) e as restrições para que haja invertibilidade permanecem idênticas a de um MA(q).

De igual modo, a função autocovariância (facv) de um processo ARMA(p, q) é a mesma de um AR(p) para defasagens ou "lags" maiores que q, sendo que esta costuma ser representada por:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p}, \quad j > q; \qquad (4)$$

e ao dividir γ_j em (4) por $\gamma_0 = Var\left[Y_t\right]$, resulta na função de autocorrelação (fac) dos modelos ARMA, qual seja,

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \cdots + \phi_p \rho_{j-p}, \quad j > q.$$

Para “lags” menores ou iguais a q, podemos deduzir que as autocorrelações são afetadas pelos parâmetros de médias móveis. É possível verificarmos que, se o processo ARMA for formado por mais parâmetros autorregressivos do que de médias móveis $(q < p)$, a fac comporta-se como exponenciais e/ou senóides amortecidas após o lag $q - p$; todavia, se $q = p$, as autocorrelações iniciais $\rho_0, \rho_1, \cdots, \rho_{q-p}$ não seguirão este padrão (MORETTIN; TOLOI, 2006).

Segundo Morettin e Toloi (2006), um modelo frequentemente usado é o ARMA$(1, 1)$, no qual

$$\begin{aligned} p &= q = 1, \\ \phi(B) &= 1 - \phi_1 B, \\ \theta(B) &= 1 - \theta_1 B, \end{aligned}$$

podendo assim ser escrito como:

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}.$$

Para um processo ARMA$(1, 1)$, a condição de estacionariedade é a mesma que para um processo AR(1), e a condição de invertibilidade é a mesma que para um processo MA(1), isto é, as condições são $-1 < \phi < 1$, e $-1 < \theta < 1$.

3 METODOLOGIA

Conforme visto na seção 2.3, na fase de identificação, os critérios de informação podem ser utilizados para a escolha do melhor modelo que se ajusta aos dados. Comparou-se aqui o desempenho dos critérios de informação Akaike (AIC); Akaike corrigido (AICc) e de Schwarz (BIC), quando utilizados com esta finalidade em modelos de séries temporais. Para isso, foram realizadas simulações Monte Carlo sob diferentes cenários, simulando diversos modelos e analisando o comportamento dos três critérios sob estudo.

Foram feitas simulações para oito modelos de séries temporais, a saber:

$$\begin{aligned} &AR(1),\ \ AR(2), ARMA(1,1), ARMA(1,2), \\ &MA(1),\ \ MA(2), ARMA(2,1), ARMA(2,2), \end{aligned} \tag{5}$$

para os modelos descritos acima os parâmetros adotados nas simulações foram:

$$\phi_1 = 0.8, \qquad \phi_2 = -0.75, \qquad \theta_1 = 0.4, \qquad \theta_2 = -0.6.$$

Estudamos o comportamento assintótico da taxa TP na seção 3.1, enquanto que na seção 3.2, em que utilizou-se amostras de tamanho 100, determinamos, também, as taxas FP e FN, além do desvio padrão para a taxa TP, quando o processo foi repetido 10 vezes.

3.1 Avaliação assintótica da taxa TP

Na primeira fase, os dados foram gerados levando-se em conta diferentes tamanhos amostrais, quais sejam, $N = 100, 150, \cdots, 1000, 1100, \cdots 5000, 5500, \cdots, 30000$, em cada um dos oito modelos de séries temporais indicados em (5). Foram simulados N dados de um dos modelos estudados (AR(1), por exemplo) e ajustados os oito modelos a este conjunto de dados, determinando-se os

valores de AIC, AICc e BIC para cada um dos oito modelos. Buscamos verificar, então, qual dos modelos apresentava menor valor de AIC, AICc e BIC. Se o critério (AIC, por exemplo) selecionou corretamente o modelo teríamos um acerto, e, caso contrário, teríamos que o modelo não selecionou corretamente o conjunto de dados como oriundos de um determinado modelo (AR(1), por exemplo). Este processo foi repetido 1000 vezes e determinou-se a quantidade de vezes que o critério de informação (AIC, por exemplo) selecionou corretamente o modelo. Tal procedimento nos fornece a taxa TP (conforme a seção 2.2.2 do capítulo 2), sobre a qual fizemos um gráfico representando o tamanho amostral *versus* a taxa de verdadeiro positivo (TP) - em porcentagem - para cada um dos oito modelos simulados, conforme (5).

3.2 Avaliação das taxas TP, FP e FN em amostras de tamanho 100

Na segunda fase, os dados foram gerados com tamanho amostral $N = 100$ para um particular modelo de séries temporais, por exemplo AR(1), e, em seguida, procedeu-se o ajuste dos oito modelos de séries temporais estudados em (5). Calculamos então, os valores de AIC, AICc e BIC para cada um dos oito modelos simulados, buscando verificar qual apresentou menor valor para cada um dos critérios. O processo foi repetido 1000 vezes e foram determinadas as taxas TP, FP e FN para cada um dos critérios. Este procedimento foi repetido 10 vezes, obtendo-se então, as médias das taxas TP, FP e FN. Determinou-se também, o desvio padrão obtido para a taxa TP.

4 APLICAÇÃO EM DADOS REAIS DE CONSUMO DE ENERGIA

O conjunto de dados a ser utilizado aqui se encontra na Tabela 1 e foi obtido pelo sítio eletrônico do Instituto de Pesquisa Econômica Aplicada - IPEA (2013), que se refere ao consumo de energia elétrica na Região Sudeste (SE). Os dados são medidos em Gigawatt-hora (GWh) e possuem frequência mensal, de janeiro de 1979 até maio de 2013. Entretanto apenas as últimas 100 observações foram utilizadas aqui, isto é, de fevereiro de 2005 a maio de 2013.

Tabela 1 Consumo mensal de energia da região Sudeste em Gigawatt-hora, de fevereiro de 2005 a maio de 2013.

Data	GWh	Data	GWh	Data	GWh	Data	GWh
2005, 02	14528	2007, 03	17593	2009, 04	17282	2011, 05	18837
2005, 03	14852	2007, 04	17664	2009, 05	16781	2011, 06	18677
2005, 04	15550	2007, 05	17337	2009, 06	16620	2011, 07	18616
2005, 05	14987	2007, 06	16896	2009, 07	16936	2011, 08	19155
2005, 06	15164	2007, 07	16622	2009, 08	17469	2011, 09	19699
2005, 07	14815	2007, 08	17013	2009, 09	17796	2011, 10	19244
2005, 08	14904	2007, 09	17586	2009, 10	18111	2011, 11	19309
2005, 09	15285	2007, 10	17489	2009, 11	18388	2011, 12	19298
2005, 10	15245	2007, 11	17983	2009, 12	18492	2012, 01	18926
2005, 11	15468	2007, 12	17622	2010, 01	17876	2012, 02	19334
2005, 12	15242	2008, 01	17552	2010, 02	18358	2012, 03	20226
2006, 01	17289	2008, 02	17445	2010, 03	18501	2012, 04	20304
2006, 02	16809	2008, 03	17566	2010, 04	18824	2012, 05	19251
2006, 03	17006	2008, 04	17576	2010, 05	18239	2012, 06	19344
2006, 04	16122	2008, 05	17862	2010, 06	18211	2012, 07	18747
2006, 05	16260	2008, 06	17415	2010, 07	19103	2012, 08	19388
2006, 06	15473	2008, 07	17747	2010, 08	18653	2012, 09	19896
2006, 07	15209	2008, 08	18282	2010, 09	18977	2012, 10	19852
2006, 08	15789	2008, 09	18157	2010, 10	18816	2012, 11	20508
2006, 09	15860	2008, 10	18526	2010, 11	18969	2012, 12	19525
2006, 10	15890	2008, 11	18439	2010, 12	19580	2013, 01	20062
2006, 11	16723	2008, 12	16842	2011, 01	19180	2013, 02	19766
2006, 12	16690	2009, 01	16346	2011, 02	19192	2013, 03	19905
2007, 01	16528	2009, 02	16418	2011, 03	19276	2013, 04	20036
2007, 02	16512	2009, 03	17161	2011, 04	19317	2013, 05	19894

5 RESULTADOS E DISCUSSÃO

Nas seções 5.1 e 5.1.1 discutiremos acerca dos resultados obtidos a partir das simulações realizadas. Na seção 5.2, a discussão versará acerca dos dados reais. As simulações referentes ao modelo MA(1) encontra-se no apêndice A, sendo que, as simulações para os demais casos (descritos em (5)) são similares.

5.1 Avaliação assintótica da taxa TP

O comportamento dos critérios de informação AIC, AICc e BIC, à medida que o tamanho das amostras cresce, é sumariado nas Figuras 3 a 10. Pela análise destas figuras poderemos perceber que assintoticamente (para $N > 2000$) o desempenho (taxa TP) do AIC e do AICc foi inferior ao desempenho do BIC, em todas as situações simuladas. O desempenho dos critérios AIC e AICc foi similar para todos as situações, e, assintoticamente estabilizaram-se em torno de 85% para todos os modelos. Por outro lado, a taxa TP do BIC estabilizou-se em torno de 96% para todos os modelos simulados, o que confirma a superioridade da taxa de acertos do BIC em relação ao AIC e AICc assintoticamente.

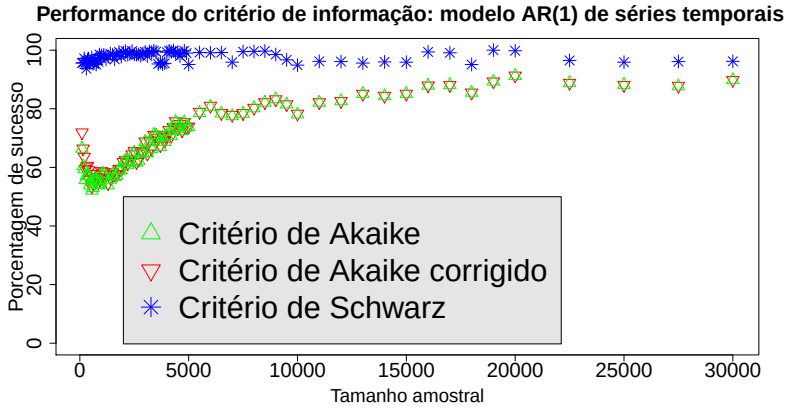

Figura 3 Porcentagem de sucesso (TP) *versus* tamanho amostral para o modelo AR(1)

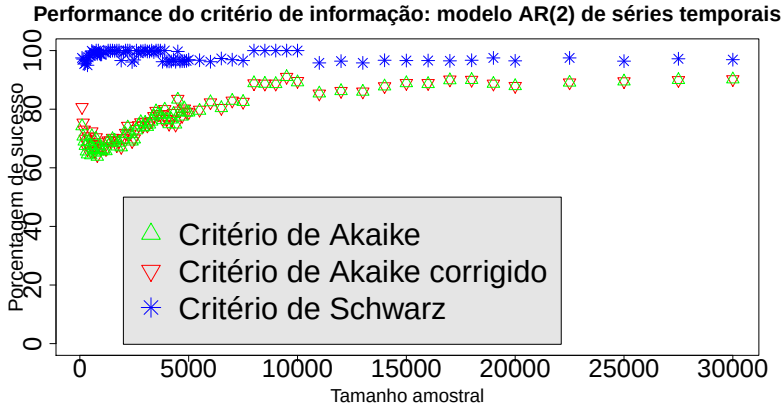

Figura 4 Porcentagem de sucesso (TP) *versus* tamanho amostral para o modelo AR(2)

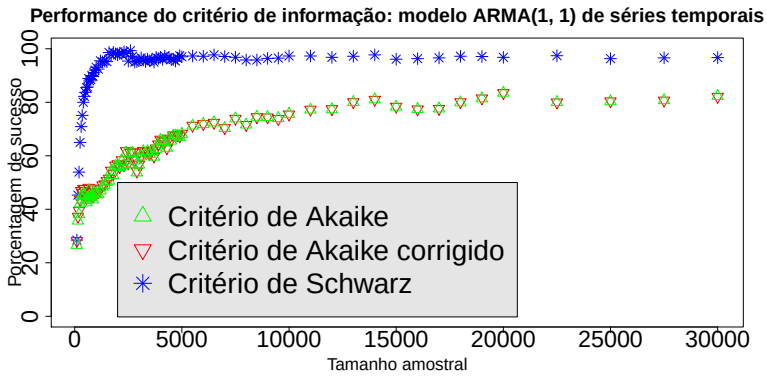

Figura 5 Porcentagem de sucesso (TP) *versus* tamanho amostral para o modelo ARMA$(1, 1)$

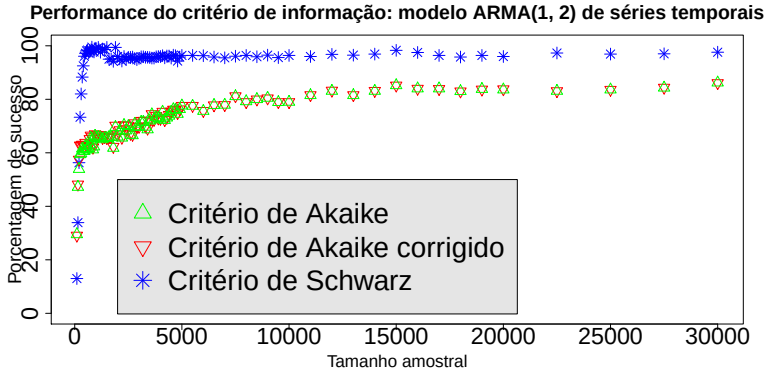

Figura 6 Porcentagem de sucesso (TP) *versus* tamanho amostral para o modelo ARMA$(1, 2)$

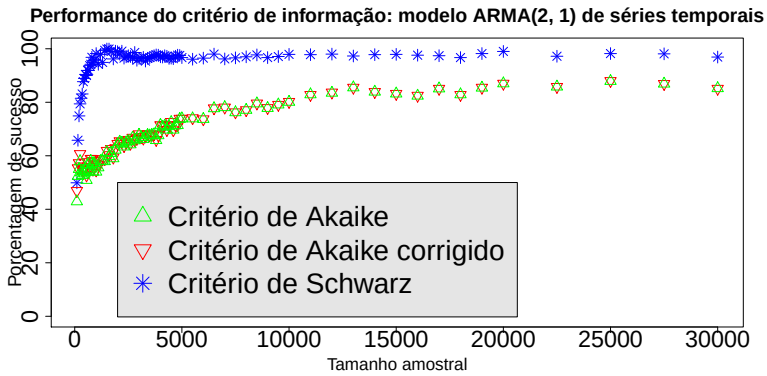

Figura 7 Porcentagem de sucesso (TP) *versus* tamanho amostral para o modelo ARMA$(2, 1)$

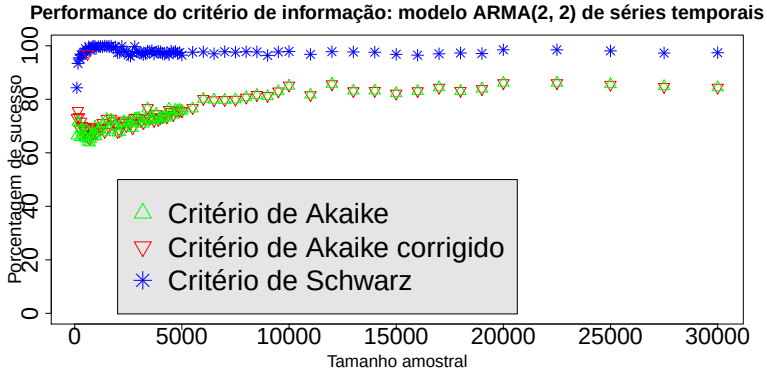

Figura 8 Porcentagem de sucesso (TP) *versus* tamanho amostral para o modelo ARMA$(2, 2)$

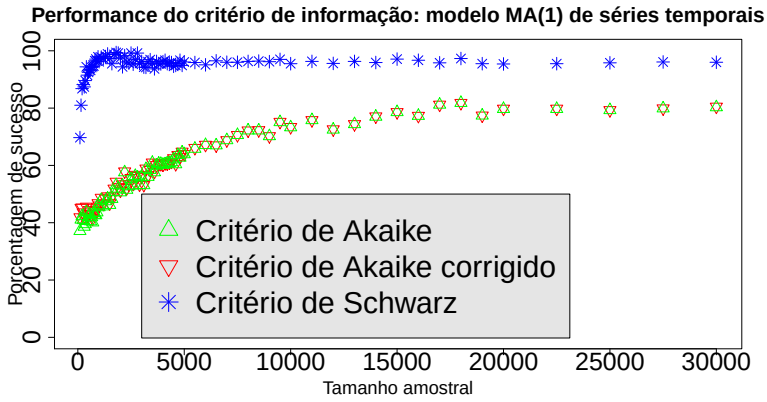

Figura 9 Porcentagem de sucesso (TP) *versus* tamanho amostral para o modelo MA(1)

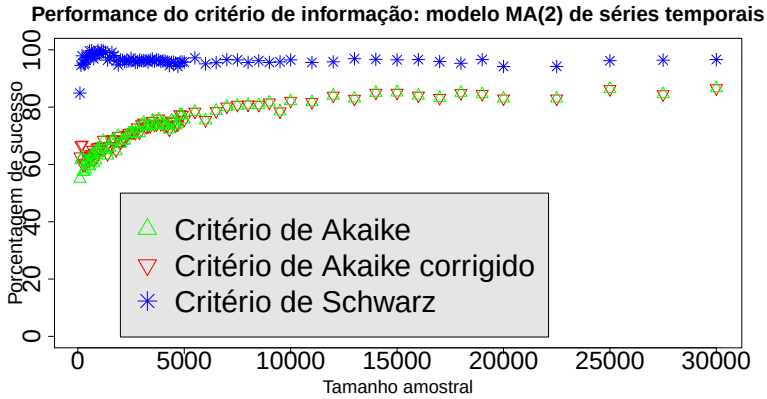

Figura 10 Porcentagem de sucesso (TP) *versus* tamanho amostral para o modelo MA(2)

5.1.1 Taxas TP, FP, e FN na seleção de modelos de séries temporais

Os resultados para amostras de tamanhos $N = 100$ estão resumidos nas Figuras 11 a 18. Pela análise destas figuras, podemos perceber que a taxa de acerto (TP) do AIC e AICc é inferior à taxa de acerto do BIC em seis dos oito casos estudados. São eles: AR(1), AR(2), ARMA$(2,1)$, ARMA$(2,2)$, MA(1) e MA(2). Em outras palavras, para o tamanho amostral $N = 100$, em seis dos oito modelos simulados, a taxa de acerto do BIC foi superior ao AIC e AICc. Mesmo

quando a taxa de acerto dos critérios, foi baixa, a maior delas refere-se ao BIC em torno de 50% (Figura 5).

Assim como ocorreu na seção 5.2 do capítulo 2, para a simulação de modelos normais em que em determinadas simulações as taxas TP mostraram-se baixas, em alguns dos casos simulados, nenhum dos critérios de informação foi capaz de selecionar satisfatoriamente o modelo correto (ou seja, o modelo do qual os dados foram gerados). É o caso dos modelos ARMA$(1,1)$, $ARMA(1,2)$ e, ARMA$(2,1)$ (ver Figuras 13, 14 e 15). Nestes casos, as razões de TP são todas abaixo de 45%. No último caso, ARMA$(2,1)$, a taxa TP do BIC é cerca de 50% e maior que a taxa FN para este critério (que é inferior a 50%). Isto permite justificar que o modelo ARMA$(2,1)$ é selecionado corretamente pelo BIC na maioria dos casos; porém o mesmo não ocorre para os critérios da AIC ou AICc, uma vez que a sua taxa TP para esse modelo particular são menores do que 45% e a taxa FN é alta, superando a taxa TP.

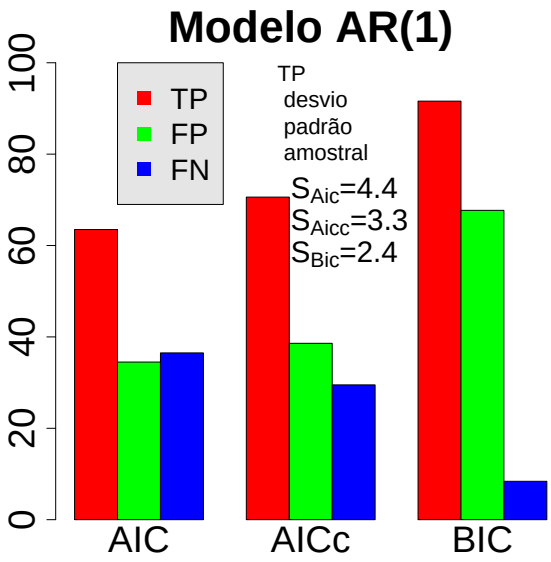

Figura 11 Taxas TP, FP e FN para o modelo AR(1)

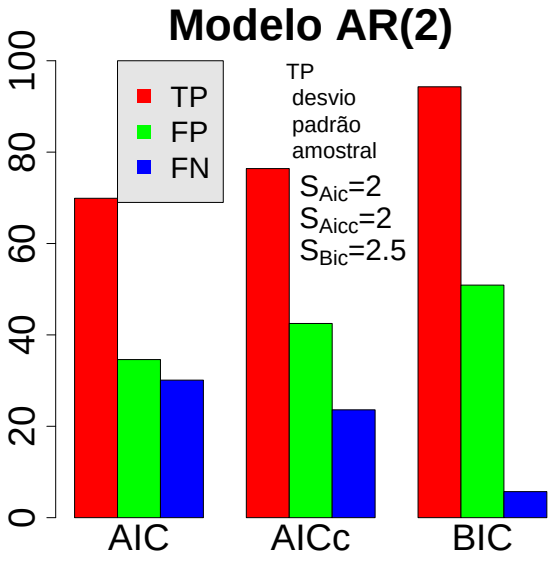

Figura 12 Taxas TP, FP e FN para o modelo AR(2)

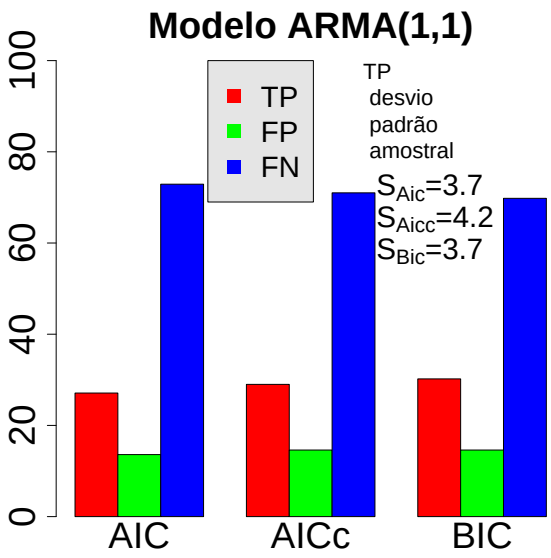

Figura 13 Taxas TP, FP e FN para o modelo ARMA$(1, 1)$

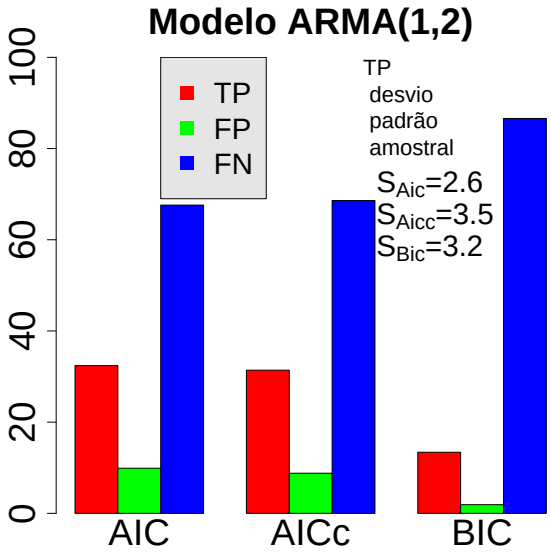

Figura 14 Taxas TP, FP e FN para o modelo ARMA$(1, 2)$

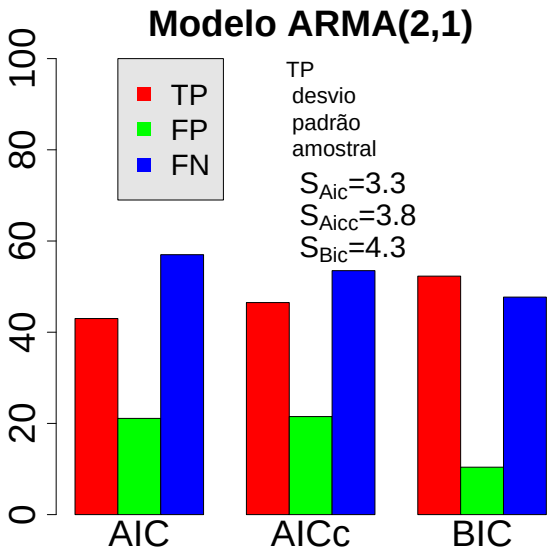

Figura 15 Taxas TP, FP e FN para o modelo ARMA$(2, 1)$

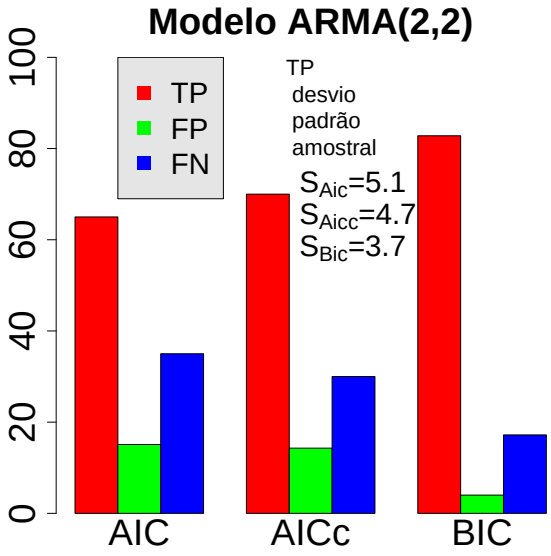

Figura 16 Taxas TP, FP e FN para o modelo ARMA$(2, 2)$

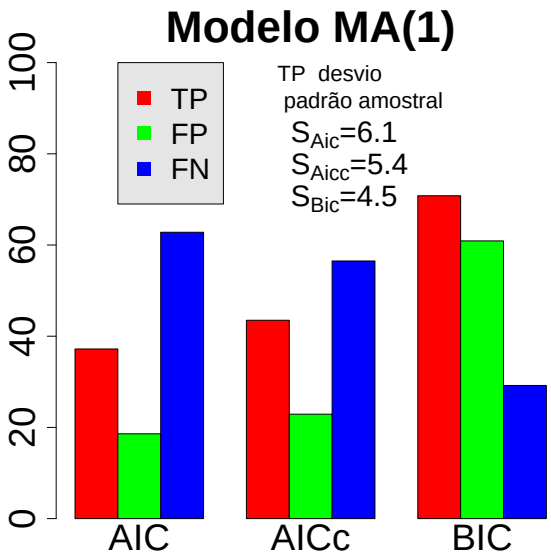

Figura 17 Taxas TP, FP e FN para o modelo MA(1)

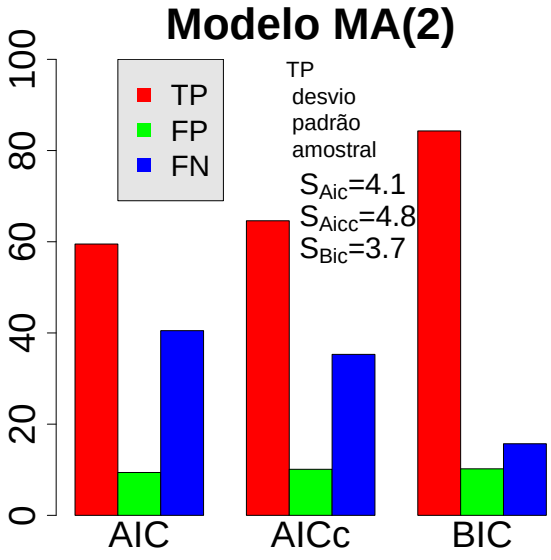

Figura 18 Taxas TP, FP e FN para o modelo MA(2)

5.2 Aplicação aos dados de consumo de energia elétrica na região Sudeste

O gráfico da série encontra-se na Figura 19, e, pela análise visual, a série aparenta ter tendência. De fato, pelo teste Wald-Wolfowits obtemos um valor p inferior a 2.2×10^{-16}, o que demonstra a existência de tendência nesta série, como entendemos.

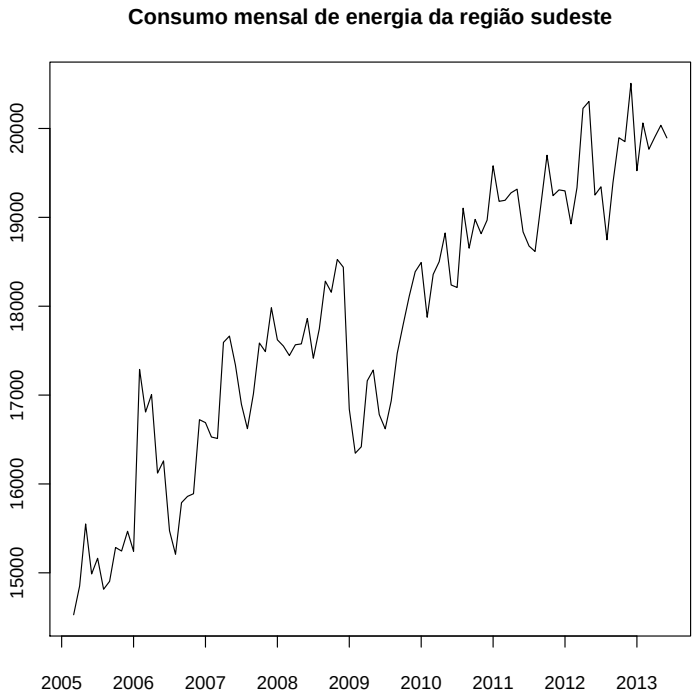

Figura 19 Série do consumo de energia elétrica da região Sudeste de fevereiro de 2005 a maio de 2013

Aplicando o operador diferença na série, esta tornou-se livre de tendência pelo teste de Wald-Wolfowits, que resultou em um valor p igual a 0.0536. O gráfico da série após a primeira diferença encontra-se na Figura 20. Quanto à sazonalidade, ao aplicar o teste de Kruskall-Wallis através da função `kruskal.test` do software R, foi obtido valor p igual a 0.9739, o que mostra que a série diferen-

ciada não apresenta sazonalidade, ao nível de 5% de significância. A estacionariedade foi verificada através do teste de Dickey-Fuller, pela função `adf.test` do software R, obteve-se valor p 0.01, e assim sendo, rejeitamos a hipótese nula de existência de raiz unitária, desta forma a série diferenciada é estacionária, ao nível de 5% de significância.

Nas Figuras 21 e 22 encontram-se os gráficos das funções de autocorrelação e autocorrelação parcial, respectivamente. Pela análise visual destas figuras, notamos que, a princípio, há o indicativo do modelo ARIMA$(1, 1, 1)$.

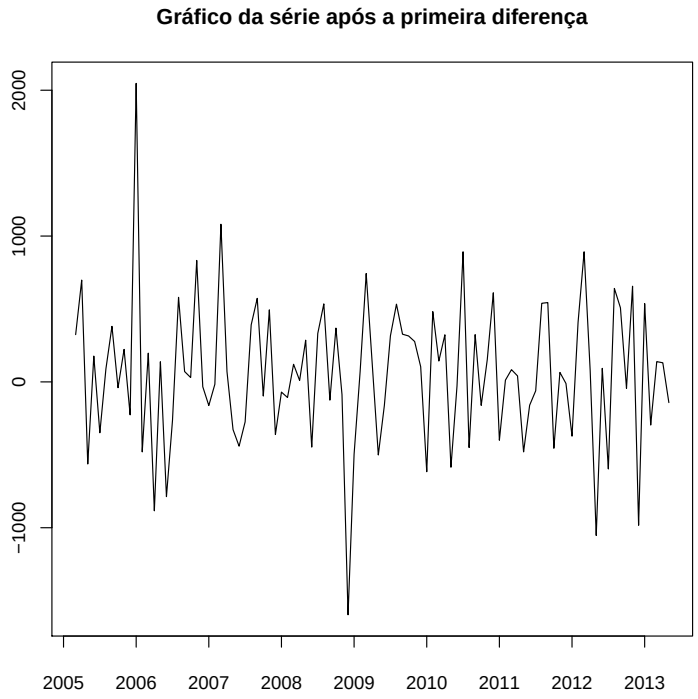

Figura 20 Série do consumo de energia elétrica após a primeira diferença

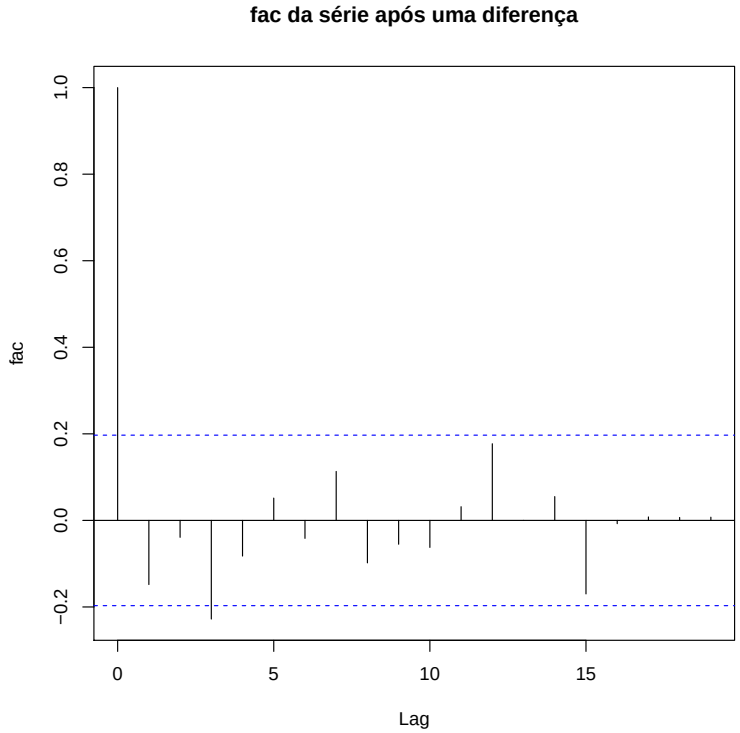

Figura 21 fac da série do consumo de energia elétrica após a primeira diferença

Na Tabela 2 são apresentados os valores do AIC, AICc e BIC para oito modelos ajustados.

Analisando a Tabela 2, notamos que os critérios AIC e AICc indicaram o modelo ARIMA$(1, 1, 1)$ como sendo o melhor, enquanto que o BIC indicou que o modelo ARIMA$(0, 1, 1)$. Pelo teste de Box-Pearce não se pode rejeitar a hipótese de que os resíduos são ruídos brancos, ao nível de 5% de significância, pois obteve-se valores p 0.2393 e 0.559, respectivamente para os modelos ARIMA$(0, 1, 1)$ e ARIMA$(1, 1, 1)$.

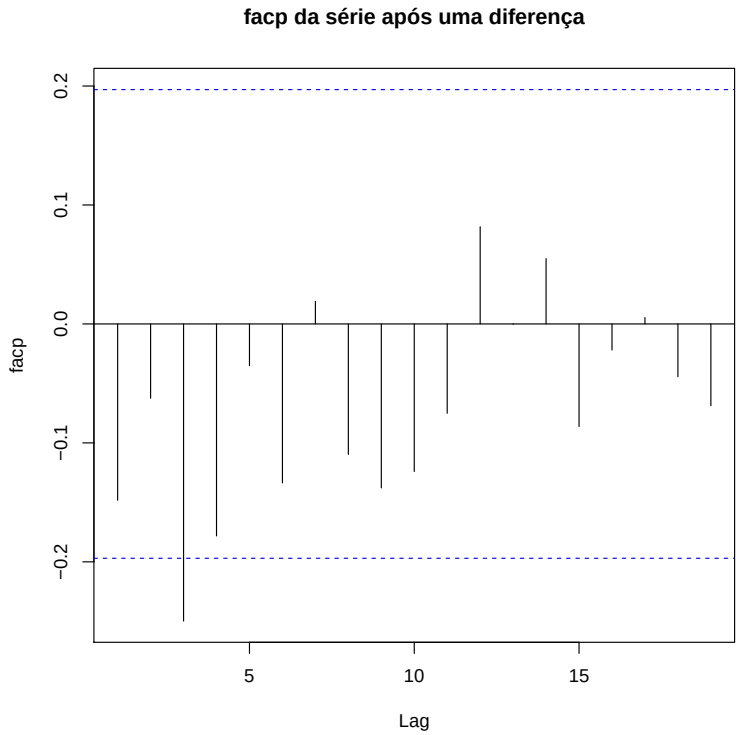

Figura 22 facp da série do consumo de energia elétrica após a primeira diferença

Tabela 2 Valores de AIC, AICc e BIC para os modelos ajustados.

Modelo	AIC	AICc	BIC
ARIMA$(1,1,0)$	$1514,03$	$1514,15$	$1519,24$
ARIMA$(2,1,0)$	$1515,78$	$1516,03$	$1523,60$
ARIMA$(0,1,1)$	$1513,69$	$1513,81$	$1518,90$
ARIMA$(0,1,2)$	$1514,02$	$1514,27$	$1521,84$
ARIMA$(1,1,1)$	$1511,50$	$1511,69$	$1519,32$
ARIMA$(1,1,2)$	$1513,38$	$1513,80$	$1523,80$
ARIMA$(2,1,1)$	$1513,31$	$1513,73$	$1523,73$
ARIMA$(2,1,2)$	$1515,43$	$1516,06$	$1528,45$

6 CONCLUSÃO

As principais conclusões das análises mostradas nas seções anteriores deste capítulo são dissertadas nessa seção. No estudo de simulações, pudemos constatar que, assintóticamente, o critério BIC se mostrou superior ao AIC e AICc em todas as situações simuladas, apresentando taxa TP próxima de 96% para todos os modelos simulados, enquanto que os demais critérios estabilizaram-se em torno de 85%. Os critérios AIC e AICc apresentaram desempenho assintótico similar, estabilizando-se em torno dos mesmos valores e para os mesmos tamanhos amostrais.

Nas simulações de tamanho amostral $N = 100$, embora o BIC tenha apresentado taxa TP superior aos demais critérios, em seis das oito situações simuladas houve situações em que ele não teve bom desempenho, apresentando baixas taxas de TP e altas taxas de FP. Novamente aqui, o desempenho do AIC e AICc não diferiram muito, apresentando comportamento similar em todos os modelos simulados.

De forma geral, não foi possível indicar um dos critérios como sendo melhor que os demais na seleção de modelos de séries temporais, a menos que as amostras tenham tamanho superior a 1200; neste caso o BIC mostrou-se superior aos demais critérios quanto a taxa TP.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control,** Notre Dame, v. 19, n. 6, p. 717-723, 1974.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis:** forecasting and control. 3rd ed. Englewood Cliffs: Prentice Hall, 1994. 592 p.

BRILLINGER, D. R. **Time series:** data analysis and theory: classics in applied mathematics. Philadelphia: SIAM, 2001. 540 p.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to time series and forecasting**. 2nd ed. New York: Springer-Verlag, 2002. 434 p.

CONOVER, W. J. **Practical nonparametric statistics.** 3rd ed. New York: J. Wiley, 1998. 592 p.

CRYER, J. D.; CHAN, K. S. **Time series analysis with applications in R.** 2nd ed. New York: Springer, 2008. 491 p.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association,** New York, v. 74, n. 366, p. 427-431, 1979.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3rd ed. New York: J. Wiley, 1998. 706 p.

ESQUIVEL, R. M. **Análise espectral singular:** modelagem de séries temporais através de estudos comparativos usando diferentes estratégias de previsão. 2012. 160 p. Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) - Serviço Nacional de Aprendizagem Industrial, Salvador, 2012.

GARDNER, E. S. Exponential smoothing: the state of the art, part II. **International Journal of Forecasting,** Amsterdam, v. 22, n. 4, p. 637-666, 2006.

GUJARATI, D. N. **Econometria básica.** 4. ed. São Paulo: Pearson Makron Books, 2006. 812 p.

HOLT, C. C. Author's retrospective on "Forecasting seasonals and trends by exponentially weighted moving averages". **International Journal of Forecasting**, Amsterdam, v. 20, n. 1, p. 11-13, 2004a.

——· Forecasting seasonals and trends by exponentially weighted moving averages. **International Journal of Forecasting**, Amsterdam, v. 20, n. 1, p. 5-10, 2004b.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Periodicidade**. Disponível em: <http://www.ipeadata.gov.br//>. Acesso em: 28 jul. 2013.

KALMAN, R. E.; BUCY, R. New results in linear filtering and prediction theory. **Journal of Basic Engineering,** New York, v. 83, p. 95-108, Mar. 1961.

KLEIN, J. L. **Statistical visions in time:** a history of time series analysis, 1662-1938. Cambridge: Cambridge University, 1997. 368 p.

KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? **Journal of Econometrics**, Amsterdam, v. 54, n. 1, p. 159-178, 1992.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: E. Blücher, 2006. 538 p.

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, Hayward, v. 6, n. 2, p. 461-464, 1978.

SUGIURA, N. Further analysts of the data by akaike's information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, Ontario, v. 7, n. 1, p. 13-26, 1978.

TSAY, R. S. Time series and forecasting: brief history and future research. **Journal of the American Statistical Association,** New York, v. 95, n. 450, p. 638-643, 2000.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. **Management Science,** Baltimore, v. 6, n. 3, p. 324-342, 1960.

YULE, G. U. On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. **Philosophical Transactions of the Royal Society of London,** New York, v. 226, p. 267-298, Feb. 1927.

APÊNDICE

Página


```
library(nnet)
l<-35;c<-10;n<-100;f<-1;MM1<--0.8;MM2<-0;MAR1<-0;MAR2<-0
escolha1<-c(rep(0,n)); escolha2<-c(rep(0,n)); escolha3<-c(rep(0,n))
contaic<-0;contbic<-0;contaiccor<-0;amostra<-100
##########################################
box<-c(rep(0,l))
vetorpar<-c(rep(1,l)); vetorpar2<-c(rep(1,l))
erropadrao<-matrix(c(rep(0,l*c)),l,c)
diferenca<-matrix(c(rep(0,l*c)),l,c)
guarda<-matrix(c(rep(0,l*c)),l,c)
akaike<-c(rep(0,l)); akaikecor<-c(rep(0,l))
#####################################################################
for(k in f:n){
vetorpar<-c(rep(1,l))
    #PHI_1=0.8; PHI_2=-0.75;
    THETA_1=0.4; THETA_2=-0.6; media=315
    erro<-rnorm(amostra,0,6)
    Y<-c(rep(0,amostra))
    Y[1]<-313+rnorm(1,0,1)
    Y[2]<-323+rnorm(1,0,1)
    for (i in 3:amostra)
    Y[i]<-315+MM1*(Y[i-1]-315)+MM2*(Y[i-2]-315)+
    erro[i]+MAR1*erro[i-1]+MAR2*erro[i-2]
    ajuste1<-arima(Y,order=c(1,0,0), include.mean=TRUE)
    ajuste1$code<-0
    box[1]<-Box.test(residuals(ajuste1),lag=36)$p.value
    guarda[1,1]<-ajuste1$coef[1]
    erropadrao[1,1]<-sqrt(ajuste1$var.coef[1,1])
    akaike[1]<-ajuste1$aic
    vetorpar2[1]<-AIC(ajuste1,k = log(amostra))
 para=3
    akaikecor[1]<-ajuste1$aic+2*para*(para+1)/(amostra-para-1)
    #Modelo AR(2)
    ajuste2<-arima(Y,order=c(2,0,0), include.mean=TRUE)
    ajuste2$code<-0
    box[2]<-Box.test(residuals(ajuste2),lag=36)$p.value
    guarda[2,1:2]<-ajuste2$coef[1:2]
    erropadrao[2,1:2]<-sqrt(diag(ajuste2$var.coef)[1:2])
    akaike[2]<-ajuste2$aic
    vetorpar2[2]<-AIC(ajuste2,k = log(amostra))
 para=3
```

```
    akaikecor[2]<-ajuste2$aic+2*para*(para+1)/(amostra-para-1)
    #Modelo MA(1)
    ajuste3<-arima(Y,order=c(0,0,1), include.mean=TRUE)
    ajuste3$code<-0
    box[3]<-Box.test(residuals(ajuste3),lag=36)$p.value
    guarda[3,1]<-ajuste3$coef[1]
    erropadrao[3,1]<-sqrt(ajuste3$var.coef[1])
    akaike[3]<-ajuste3$aic
    vetorpar2[3]<-AIC(ajuste3,k = log(amostra))
para=3
    akaikecor[3]<-ajuste3$aic+2*para*(para+1)/(amostra-para-1)
    #Modelo MA(2)
    ajuste4<-arima(Y,order=c(0,0,2), include.mean=TRUE)
    ajuste4$code<-0
    box[4]<-Box.test(residuals(ajuste4),lag=36)$p.value
    guarda[4,1:2]<-ajuste4$coef[1:2]
    erropadrao[4,1:2]<-sqrt(diag(ajuste4$var.coef)[1:2])
    akaike[4]<-ajuste4$aic
    vetorpar2[4]<-AIC(ajuste4,k = log(amostra))
para=4
    akaikecor[4]<-ajuste4$aic+2*para*(para+1)/(amostra-para-1)
    #Modelo ARMA(1,1)
    ajuste5<-arima(Y,order=c(1,0,1), include.mean=TRUE)
    ajuste5$code<-0
    box[5]<-Box.test(residuals(ajuste5),lag=36)$p.value
    guarda[5,1:2]<-ajuste5$coef[1:2]
    erropadrao[5,1:2]<-sqrt(diag(ajuste5$var.coef)[1:2])
    akaike[5]<-ajuste5$aic
    vetorpar2[5]<-AIC(ajuste5,k = log(amostra))
para=4
    akaikecor[5]<-ajuste5$aic+2*para*(para+1)/(amostra-para-1)
    #Modelo ARMA(1,2)
    ajuste6<-arima(Y,order=c(1,0,2), include.mean=TRUE)
    box[6]<-Box.test(residuals(ajuste6),lag=36)$p.value
    guarda[6,1:3]<-ajuste6$coef[1:3]
    erropadrao[6,1:3]<-sqrt(diag(ajuste6$var.coef)[1:3])
    akaike[6]<-ajuste6$aic
    vetorpar2[6]<-AIC(ajuste6,k = log(amostra))
para=5
    akaikecor[6]<-ajuste6$aic+2*para*(para+1)/(amostra-para-1)
    #Modelo ARMA(2,1)
    ajuste7<-arima(Y,order=c(2,0,1), include.mean=TRUE)
    ajuste7$code<-0
    box[7]<-Box.test(residuals(ajuste7),lag=36)$p.value
    guarda[7,1:3]<-ajuste7$coef[1:3]
    erropadrao[7,1:3]<-sqrt(diag(ajuste7$var.coef)[1:3])
    akaike[7]<-ajuste7$aic
    vetorpar2[7]<-AIC(ajuste7,k = log(amostra))
```

```
 para=5
    akaikecor[7]<-ajuste7$aic+2*para*(para+1)/(amostra-para-1)
    #Modelo ARMA(2,2)
    ajuste8<-arima(Y, order=c(2,0,2), include.mean=TRUE)
    ajuste8$code<-0
    box[8]<-Box.test(residuals(ajuste8),lag=36)$p.value
    guarda[8,1:4]<-ajuste8$coef[1:4]
    erropadrao[8,1:4]<-sqrt(diag(ajuste8$var.coef)[1:4])
    akaike[8]<-ajuste8$aic
    vetorpar2[8]<-AIC(ajuste8,k = log(amostra))
 para=6
    akaikecor[8]<-ajuste8$aic+2*para*(para+1)/(amostra-para-1)
    for(i in 1:l){
        for(j in 1:c){
            if(guarda[i,j]=="NaN" | erropadrao[i,j]=="NaN"){
                diferenca[i,j]<--1
            }
                else{diferenca[i,j]<-abs(guarda[i,j])-
                2*erropadrao[i,j]
                    if(diferenca[i,j]<0){
                    vetorpar[i]<-10^25
                    vetorpar2[i]<-10^25
   akaikecor[i]<-10^25}
                  }
        }
    }
    for(i in 1:l){
        if(vetorpar[i]==1){
        if(box[i]<0.05)
        {vetorpar[i]<-10^25
        vetorpar2[i]<-10^25
 akaikecor[i]<-10^25}else{vetorpar[i]<-akaike[i]}
        }
    }
    escolha1[k]<-which.is.max(-vetorpar)
    escolha2[k]<-which.is.max(-vetorpar2)
 escolha3[k]<-which.is.max(-akaikecor)
    if(escolha1[k]==1){
        contaic<-contaic+1}
    if(escolha3[k]==1){
        contaiccor<-contaiccor+1}
    if(escolha2[k]==1){
        contbic<-contbic+1}
        f<-k+1
}
contaic;contbic;contaiccor
```

CAPÍTULO 4

Critérios de informação aplicados a dados de crescimento

RESUMO

As curvas de crescimento são de fundamental importância em diversas áreas, em especial no crescimento bovino. Em geral, os parâmetros dos modelos lineares não são de fácil interpretação, o mesmo não ocorre com os modelos não lineares, que resumem uma grande quantidade de informações em poucos parâmetros que são biologicamente interpretáveis. Uma das formas de se selecionar o melhor modelo para curvas, de crescimento é através dos critérios de informação, em especial o AIC, AICc e o BIC. Objetivou-se aqui avaliar o desempenho destes critérios na seleção de modelos de crescimento de bovinos Hereford. Foram simuladas diversas situações e ajustados diversos modelos. Os critérios de AIC e AICc, em geral, apresentaram-se melhores que o BIC, com maiores taxas de verdadeiro positivo e menor taxa de falso positivo. Na aplicação a dados reais, todos os critérios selecionaram o modelo de von Bertalanffy como o que melhor se ajusta aos dados, enquanto que os modelos de Michaelis-Menten e Brody superestimaram o peso assintótico dos animais.

Palavras-chave: Curva de crescimento. Modelos não-lineares. Heterocedasticidade. Autocorrelação.

ABSTRACT

The growth curves are of fundamental importance in many areas, especially in growing cattle. In general, the parameters of linear models are not easy to interpret, the same does not occur with non-linear models, which summarize a lot of information in a few parameters that are biologically interpretable. One of the ways to select the best model for curves, growth is through the information criteria, in particular the AIC, AICc and BIC. The objective here is to evaluate the performance of these criteria in the selection of growth models of Hereford cattle. Many situations were simulated and adjusted various models. Criteria AIC and AICc in general showed up better than the BIC, with higher rates of true positive and lower false positive rate. In the application to real data, all the selected criteria von Bertalanffy model as the one that best fits the data, while models of Michaelis-Menten and Brody overestimated the asymptotic weight of the animals.

Keywords: Growth curves. Non-linear models. Heteroscedasticity. Autocorrelation.

1 INTRODUÇÃO

O rebanho de gado de corte brasileiro é constituído basicamente por duas origens raciais, as europeias e as asiáticas. A maioria das raças europeias é criada na região Sul do Brasil, onde o Rio Grande do Sul detém 10-12% do rebanho efetivo nacional. Dentre as raças europeias, destaca-se a raça Hereford, originária da Inglaterra, do Condado de mesmo nome. Suas principais vantagens são: adaptação aos mais diversos ambientes e sistemas de produção, graças a docilidade e rusticidade; índice de fertilidade dos mais altos da espécie, quando favorecidos com manejo e alimentação adequados; excepcional ganho de peso a pasto, sendo comum novilhos de 450Kg aos 18-24 meses; preponderância nos cruzamentos com outras raças, especialmente as zebuínas; altamente lucrativa para criadores, invernadores e frigoríficos, graças ao insuperável índice de rendimento de carcaça, entre as raças europeias (ASSOCIAÇÃO BRASILEIRA DE HEREFORD E BRAFORD, 2013).

Na pecuária de corte, os criadores estão cada vez mais conscientes da importância do crescimento dos bovinos, para gerenciar e avaliar a rentabilidade da atividade econômica. Em um sistema de produção de carne, crescimento é uma função primordial, pois apresenta relação direta com a quantidade e qualidade do produto final - carne.

A representação gráfica do peso ou massa corporal em relação à idade resulta na curva de crescimento (GOTTSCHALL, 1999). A análise de dados de crescimento é, em geral, muito complexa em modelos lineares. Neste contexto, a utilização de modelos não-lineares é de grande utilidade, uma vez que sintetizam um grande número de medidas, em apenas alguns parâmetros interpretáveis biologicamente (BROWN; FITZHUGH JUNIOR; CARTWRIGHT, 1976).

Algumas informações que auxiliam no manejo da propriedade, são obtidas

através do emprego da curva de crescimento, quais sejam: maior precocidade; redução no tempo para atingir a maturidade sexual, antecipando desta forma a idade de entrar em reprodução e diminuição no percentual de gordura da carcaça em função do aumento da precocidade (TEDESCHI, 1996). Outro emprego da curva de crescimento é no melhoramento genético, o qual torna possível obter previsões do tamanho das matrizes e touros que serão mantidos no plantel de reprodução, sendo que em rebanhos especializados, isto representa um alto valor na produção.

No estudo da curva de crescimento pode ocorrer heterogeneidade das variâncias dos pesos corporais, pois à medida que a idade aumenta, a variância dos pesos corporais também aumenta. Outro fato que pode ocorrer, em se tratando de curvas de crescimento peso-idade, é a autocorrelação dos resíduos. Quando não levados em consideração podem levar a obtenção de estimativas viesadas (PASTERNAK; SHALEV, 1994).

No presente capítulo, este trabalho teve como objetivo comparar o desempenho dos critérios de informação de Akaike(AIC), Akaike corrigido (AICc) e de Schwarz (BIC), na seleção de modelos de crescimento comumente utilizados para descrever as relações peso-idade dos animais. O intuito aqui é determinar, via simulação, se os critérios de informação possuem bom desempenho ao selecionar modelos de crescimento. Tal desempenho será verificado através das taxas TP, FP e FN.

2 REFERENCIAL TEÓRICO

2.1 Curvas de crescimento

De acordo com Fitzhugh Junior (1976), o termo curva de crescimento, usualmente evoca a imagem de uma curva sigmóide, descrevendo o tempo de vida em uma sequência de medidas de tamanho, frequentemente peso corporal. O crescimento dos animais tem uma forte relação com a quantidade e a qualidade da carne. Assim, torna-se de fundamental importância o conhecimento do processo de ganho de massa corporal do animal, pois esse conhecimento possibilita que se faça um controle da produção de carne, otimizando os lucros dessa atividade. Este ganho pode ser influenciado pela alimentação, por condições climáticas, pelo estado sanitário e pelas características genéticas associadas aos animais (GOTTSCHALL, 1999).

De acordo com Tedeschi et al. (2000), o processo de crescimento e desenvolvimento dos animais é assunto de bastante interesse para os pesquisadores, pois o seu domínio permite que o manejo nutricional possa ser conduzido eficientemente, além de permitir que programas de seleção sejam elaborados para as características de crescimento inerentes a cada raça. Fitzhugh Junior (1976), considera que o ajuste das funções peso-idade fornece informações descritivas da espécie, devendo observar alguns requisitos para que a descrição possa ser bem feita. Entre estes requisitos, os parâmetros devem ser biologicamente interpretáveis e o ajuste dos dados deve ser adequado, sendo que, a dificuldade computacional também deve ser considerada. Há casos onde as funções podem ser ajustadas matematicamente, mas não apresentam uma interpretação biológica.

Kshirsagar e Smith (1995) e Mazzini (2001), consideram que a premissa básica em modelos de curvas de crescimento é que existe uma relação funcional

entre um efeito de tratamento e sua aplicação no tempo, e esta relação pode ser modelada. As funções utilizadas para ajustar peso em relação à idade podem ser lineares ou não-lineares (TEDESCHI, 1996).

Segundo Braccini Neto et al. (1996), os modelos de regressão linear proporcionam um bom ajuste, porém, seus parâmetros não possuem significado biológico. Desta forma, a relação entre o peso do animal e o tempo geralmente é descrita por meio de modelos de regressão não-linear, os quais condensam informações de peso-idade de todo o período de vida de um indivíduo em um conjunto de parâmetros biologicamente interpretáveis (FITZHUGH JUNIOR, 1976). Dentre tais parâmetros, geralmente se destacam o peso à maturidade, ou peso adulto e taxa de maturidade, que representa a velocidade de crescimento, de forma que quanto mais alto for o seu valor, mais precoce é o animal (SILVA et al., 2004). Segundo Silva (2010), as estimativas destes parâmetros constituem características fenotípicas alternativas para programas de seleção que visam a obtenção de animais mais precoces, com maior peso e melhor qualidade de carcaça.

Dentre as diversas funções não-lineares utilizadas para ajustar as relações peso-idade pode-se citar Brody, Gompertz, Logística, von Bertalanffy e Michaelis-Menten, dentre outras.

Segundo Fitzhugh Junior (1976), os seguintes requisitos devem ser atendidos para que um modelo de regressão não-linear descreva adequadamente a relação peso-idade: interpretação biológica dos parâmetros, “alta qualidade” de ajuste e facilidade de convergência. Assim, nos estudos de curvas de crescimento é imprescindível que se tenha bom conhecimento sobre a teoria de modelos de regressão não-linear, a fim de facilitar a compreensão do fenômeno abordado.

2.2 Modelos de regressão não-linear

Avaliar a possível relação entre uma variável dependente e uma ou mais variáveis independentes é uma das tarefas mais comuns em análises estatísticas. Pode-se atingir este objetivo através dos bem conhecidos modelos de regressão, os quais se dividem em duas classes distintas: os lineares e os não-lineares. Existem muitas diferenças entre estas classes de modelos. No caso linear, a partir de um conjunto de observações, busca-se o modelo que melhor explica a relação, se existir alguma, entre as variáveis inerentes a um dado fenômeno. Os modelos não-lineares, na maioria das vezes, são baseados em considerações teóricas inerentes ao fenômeno que se tem interesse em estudar (MAZUCHELI; ACHCAR, 2002).

Segundo Souza (1998), modelos de regressão não-linear com resposta univariada y_i são da forma

$$y_i = f\left(\boldsymbol{x_i}, \boldsymbol{\theta}\right) + \varepsilon_i, \tag{1}$$

sendo que $i = 1, \cdots, n$; y_i representa a observação da variável dependente; $f\left(\boldsymbol{x_i}, \boldsymbol{\theta}\right)$ é a função esperança ou função resposta conhecida; $\boldsymbol{x_i}$ é um vetor k dimensional formado por observações em variáveis exógenas; $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_p)'$ é um vetor de parâmetros p dimensional desconhecido e ε_i representa o efeito do erro aleatório não observável.

Segundo Draper e Smith (1998), dizer que um modelo é linear ou não-linear, é referir-se à linearidade ou não-linearidade, em relação aos parâmetros. Desta forma, os modelos podem ser classificados em lineares, linearizáveis e não-lineares.

Os modelos são lineares em relação aos parâmetros quando:

$$\frac{\partial f_i(X;\boldsymbol{\theta})}{\partial \theta_j} = h(X),$$

para $i = 1, 2, \cdots, n$ e $j = 1, 2, \cdots, p$.

Os modelos linearizáveis são aqueles que podem ser transformados em lineares através de anamorfose. Os modelos de regressão são não-lineares em relação aos parâmetros quando:

$$\frac{\partial f_i(X;\boldsymbol{\theta})}{\partial \theta_j} = h(X;\boldsymbol{\theta}),$$

pelo menos para algum i e j.

Para Ratkowsky (1993), um modelo de regressão não-linear é aquele no qual os parâmetros apresentam não-linearidade, por exemplo: $Y_t = X_t^{\theta} + \varepsilon_t$, em que θ é o parâmetro a ser estimado. De forma semelhante aos modelos lineares, o processo de estimação do parâmetro θ, pode ser obtido através da minimização da soma de quadrados dos resíduos, obtendo-se, o sistema de equações normais não-linear, o qual não apresenta uma solução explícita para $\hat{\theta}$, sendo obtida por processos iterativos.

Souza (1998), considera o modelo de regressão não-linear na forma

$$Y = f(\theta^0) + \varepsilon,$$

em que Y tem componentes y_t; $f(\theta^0)$ tem componentes $f(x_t, \theta^0)$ e ε tem componentes ε_t, sendo $F(\theta)$ a matriz jacobiana de $f(\theta)$ e $F = F(\theta^0)$.

O estimador $(\widehat{\theta})$ de mínimos quadrados de θ^0 é obtido mediante a pesquisa do mínimo (em Θ) da soma de quadrados residuais,

$$SSE\,(\theta) = \sum_{t=1}^{n} [(y_t - f(x_t, \theta)]^2 = [(Y - f(\theta)]'[Y - f(\theta)].$$

Conforme Souza (1998), de modo análogo ao modelo linear, como es-

timador da variância σ^2, toma-se $\widehat{\sigma}^2 = \dfrac{SSE(\hat{\theta})}{n-p}$, uma outra alternativa, seria $\widehat{\sigma^2} = \dfrac{SSE(\widehat{\theta})}{n}$. Neste caso, com erros normais, o par $(\widehat{\theta}', \widehat{\sigma}^2)$ define o estimador de máxima verossimilhança de (θ^0, σ^2). O estimador de mínimos quadrados $\widehat{\theta}$ satisfaz a equação $\dfrac{\partial SSE(\theta)}{\partial(\theta)} = 0$, portanto deve-se ter $-2F'(\widehat{\theta})[Y - f(\widehat{\theta})] = 0$.

Resulta que o vetor residual $\hat{e} = Y - f(\hat{\theta})$ satisfaz $F'(\hat{\theta})\hat{e} = 0$ e é, portanto, ortogonal às colunas da matriz jacobiana $F(\theta)$ calculada em $\theta = \hat{\theta}$. Em regressão linear $F = F(\hat{\theta}) = X$, a identificação entre X no caso linear e F no caso não-linear vale em geral, isto é, as expressões utilizadas no estudo da inferência do modelo linear e com erros normais tem uma relação para o caso não-linear, obtida por intermédio da substituição da matriz X por F. A razão disto é que a teoria de inferência estatística que se desenvolve para o modelo de regressão não-linear baseia-se essencialmente na aproximação linear por série de Taylor,

$$f(\theta) \approx f\left(\theta^0\right) + F\left(\theta - \theta^0\right).$$

Ainda de acordo com Souza (1998), o método de mínimos quadrados generalizados representa um conjunto de ideias extremamente importante para o estudo de modelos lineares e não-lineares. A abordagem se faz necessária, particularmente, na presença de heterogeneidade de variâncias.

Quando a matriz de variância dos resíduos é da forma $\sigma^2\Omega$ com $\Omega \neq I$, o estimador dos mínimos quadrados de β não é o mais eficiente. Se Ω é positiva definida, existe uma matriz P não-singular tal que $\Omega^{-1} = P'P$.

Na maioria das aplicações que exigem o uso de mínimos quadrados generalizados, a matriz Ω não será conhecida. Nestas circunstâncias tipicamente $\Omega = \Omega(\gamma)$, em que γ tem dimensão finita e um estimador consistente de γ estará disponível. Seja γ este estimador $\hat{\Omega} = \Omega(\gamma)$. Neste caso utiliza-se $\hat{\beta}_E =$

$(X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y$. Sob certas condições de regularidade adicionais, que devem ser verificadas em cada caso, $\hat{\beta}_E$ será mais eficiente do que o estimador de mínimos quadrados ordinários, consistente e normal.

Segundo Mazzini (2001), a caracterização da regressão em função das suposições do vetor de erros ε, distingue-se da seguinte maneira:

a) modelos ordinários: aqueles cuja estrutura dos erros não viola nenhuma das pressuposições. Pode ser escrito de forma mais eficiente como: $\boldsymbol{\varepsilon} \sim N(\boldsymbol{\phi}; I\sigma^2)$;

b) modelos ponderados: são aqueles cuja estrutura dos erros viola a pressuposição de homogeneidade de variâncias. Nesse caso diz-se que os erros são heterocedásticos. Escreve-se: $\boldsymbol{\varepsilon} \sim N(\phi; D\sigma^2)$ em que D é uma matriz diagonal, positiva definida, que pondera a variância σ^2;

c) modelos generalizados: são aqueles cuja estrutura dos erros viola a pressuposição de independência dos erros e possivelmente a de homogeneidade de variâncias. Diz-se que os erros são correlacionados (e possivelmente heterocedásticos). Escreve-se: $\boldsymbol{\varepsilon} \sim N(\boldsymbol{\phi}; W\sigma^2)$ em que W é uma matriz simétrica, positiva definida, que representa as variâncias e covariâncias dos erros.

A heterocedasticidade, foi citada por Elias (1998), que observou um aumento das variâncias ao longo do tempo para as raças Gir, Guzerá e Nelore, entretanto o modelo básico de regressão não-linear deve atender às seguintes suposições: erros com média zero; erros não correlacionados; homogeneidade de variâncias e erros com distribuição normal. No caso de dados longitudinais, as suposições de erros não correlacionados e homogeneidade de variâncias não são realísticas. Resultados similares foram obtidos por Mazzini (2001) e Mazzini et al. (2005), para crescimento de novilhos da raça Hereford.

Vários métodos iterativos são propostos na literatura para obtenção das estimativas de mínimos quadrados dos parâmetros de um modelo de regressão não-

linear. Os mais utilizados são: o método de Gauss-Newton, o método "Steepest-Descent" ou método do Gradiente e o método de Marquardt, os quais fazem uso das derivadas parciais da função esperança $f(x_i;\theta)$ com relação a cada parâmetro. Outro método, bastante similar ao de Gauss-Newton, exceto pelo fato de não exigir a especificação das derivadas parciais da função esperança, é chamado método de DUD (doesn't use derivates) (MAZUCHELI; ACHCAR, 2002). Um outro método iterativo bastante utilizado é o de Gauss-Newton modificado ou método de linearização, o qual usa o resultado de mínimos quadrados lineares em uma sucessão de passos, convertendo o problema de uma regressão não-linear para uma série de regressões lineares. Para isso faz uma expansão em série de Taylor até 1º grau e depois minimiza a soma de quadrados residual (MAZZINI, 2001).

2.2.1 Autocorrelação

Os modelos básicos de regressão, em geral, assumem que os erros são independentes. No caso de estudo de curva de crescimento, em que as medidas são tomadas em sequência, no mesmo animal, a hipótese de independência dos erros é frequentemente não apropriada (NETER; WASSERMAN; KUTNER, 1985).

A característica geral da autocorrelação dos resíduos é a de existir uma variação sistemática dos valores em observações sucessivas. Quando isso ocorre diz-se que os resíduos são serialmente correlacionados ou autocorrelacionados (MORETTIN; TOLOI, 2006). Segundo Hoffman e Vieira (1998) e Mazzini (2001), a autocorrelação dos resíduos surge, geralmente, quando se trabalha com séries cronológicas de dados, admitindo que o erro da observação relativa a um período está correlacionado com o erro da observação anterior.

Para Neter, Wasserman e Kutner (1985), nos casos de modelos de regressão com erros autocorrelacionados positivamente, o uso do método dos quadra-

dos mínimos ordinário tem importantes consequências, tais como: os estimadores dos coeficientes obtidos pelo método dos quadrados mínimos ordinário são não tendenciosos, mas levam a uma subestimativa da variância, podendo ser completamente ineficientes; o quadrado médio do resíduo pode subestimar a variância dos erros; em consequência o desvio padrão calculado de acordo com o método dos mínimos quadrados ordinário pode subestimar o verdadeiro desvio padrão do coeficiente de regressão estimado, invalidando os intervalos de confiança e testes usando as distribuições t e F.

Conforme Souza (1998), o procedimento estatístico adequado ao modelo não-linear subordinado à estrutura do AR(1) envolve a determinação de um estimador consistente $\hat{\Omega}$ de Ω e a busca de um estimador de θ mais eficiente do que o de mínimos quadrados ordinários. Neste contexto, procura-se pelo mínimo da soma de quadrados residual "ponderada"

$$[Y - f(\theta)]' \hat{\Omega}^{-1} [Y - f(\theta)] .$$

O valor de θ que minimiza esta soma de quadrados é o estimador de mínimos quadrados generalizados $\hat{\theta}_G$. Este vetor é determinado pela fatoração $\hat{\Omega}^{-1} = P'P$ e da regressão não-linear $PY = Pf(\theta) + v$ em que a matriz P é determinada a partir dos resíduos de mínimos quadrados ordinários não-lineares e da solução do sistema de equações de Yule-Walker. Após as primeiras l observações, a transformação P induz o modelo $(PY)_t = [Pf(\theta)]_t + v_t$ em que

$$(PY)_t = y_t + \hat{\phi}_1 y_{t-1} + \hat{\phi}_2 y_{t-2} + \cdots + \hat{\phi}_l y_{t-l},$$

e

$$\{P[f(\theta)]\}_t = f(x_t, \theta) + \hat{\phi}_1 f(x_{t-1}, \theta) + \hat{\phi}_2 f(x_{t-2}, \theta) + \cdots + \hat{\phi}_l f(x_{t-l}, \theta).$$

Em estudo de crescimento de vacas leiteiras, Kroll (1990), comparou os modelos polinomial, Gompertz, Logístico e Mitscherlich com estrutura de erros independentes e autorregressivos de primeira ordem, concluindo que, o processo autorregressivo diminui significativamente a autocorrelação nos modelos.

Segundo Mazzini (2001), em trabalho objetivando avaliar a qualidade e as características do ajuste da função logística monofásica e difásica, com estrutura de erros independentes e autorregressivos de primeira e segunda ordens, AR(1) e AR(2), em dados simulados e reais de vacas leiteiras, Medeiros et al. (2000) concluiu que a introdução da estrutura de autocorrelação nos erros melhorou o ajuste, minimizando o problema das medidas repetidas. Mendes et al. (2009), também utilizaram os modelos Logístico Monofásico e Difásico para descrever o crescimento de vacas, e obtiveram melhor qualidade de ajuste quando assumiram estruturas de erros autorregressivos.

Segundo Elias (1998), embora a ponderação pelo inverso da variância melhore a qualidade das estimativas, não atende a todas as especificações, em particular a condição de independência dos erros, o que implica na necessidade de utilização de procedimentos alternativos.

Para maiores detalhes de autocorrelação, vide seção 2.2.2 do capítulo 3.

2.2.2 Estimação dos parâmetros de modelos não-lineares

Para ajustar os modelos de regressão não-linear aos dados de crescimento, em geral, utiliza-se o método dos quadrados mínimos ordinários cujas soluções podem ser obtidas por meio de processos iterativos.

2.2.2.1 Método dos mínimos quadrados ordinários

Considerando o modelo (1) na forma matricial, tem-se:

$$\mathbf{y} = \mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta^0}\right) + \boldsymbol{\varepsilon},$$

em que:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, f\left(\mathbf{x}, \boldsymbol{\theta^0}\right) = \begin{bmatrix} f\left(\mathbf{x}, \theta_{\mathbf{1}}^{\mathbf{0}}\right) \\ f\left(\mathbf{x}, \theta_{\mathbf{2}}^{\mathbf{0}}\right) \\ \vdots \\ f\left(\mathbf{x}, \theta_{\mathbf{n}}^{\mathbf{0}}\right) \end{bmatrix} \text{ e } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

A soma dos quadrados dos erros aleatórios (SQE) deverá ser minimizada por θ, portanto a função de mínimos quadrados para um modelo não-linear é dada por:

$$SQE\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{n} \left[y_i - f\left(x_i, \boldsymbol{\theta}\right)\right]^2,$$

a qual pode ser representada matricialmente por:

$$SQE\left(\boldsymbol{\theta}\right) = \left[y - f\left(\boldsymbol{\theta}\right)\right]' \left[y - f\left(\boldsymbol{\theta}\right)\right].$$

Segundo Souza (1998), em modelos não-lineares não se pode fazer afirmações gerais acerca das propriedades dos estimadores de quadrados mínimos, tais como não tendenciosidade e variância mínima, exceto para grandes amostras, os chamados resultados assintóticos. Para uma melhor compreensão do processo de obtenção destes estimadores, considere a seguinte notação de diferenciação matricial:

$$f(\boldsymbol{\theta}) = \begin{bmatrix} f_1(\theta) \\ f_2(\theta) \\ \vdots \\ f_n(\theta) \end{bmatrix}, \text{ e } F(\theta) = \frac{\partial f(\theta)}{\partial \theta'} = \begin{bmatrix} \frac{\partial f_1(\theta)}{\partial \theta_1} & \frac{\partial f_1(\theta)}{\partial \theta_2} & \cdots & \frac{\partial f_1(\theta)}{\partial \theta_p} \\ \frac{\partial f_2(\theta)}{\partial \theta_1} & \frac{\partial f_2(\theta)}{\partial \theta_2} & \cdots & \frac{\partial f_2(\theta)}{\partial \theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n(\theta)}{\partial \theta_1} & \frac{\partial f_n(\theta)}{\partial \theta_2} & \cdots & \frac{\partial f_n(\theta)}{\partial \theta_p} \end{bmatrix},$$

em que:

$f(\theta)$ é uma função vetor coluna $n \times 1$ de um argumento p dimensional θ e $F(\theta)$ é a matriz Jacobiana de $f(\theta)$. Dessa forma, o estimador de mínimos quadrados, $\hat{\theta}$, satisfaz a equação $\frac{\partial SQE(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}} = \emptyset$, a qual representa a minimização de interesse. Sendo, $\frac{\partial SQE(\theta)}{\partial \theta'} = \frac{\partial}{\partial \theta'}[y - f(\theta)]'[y - f(\theta)] = -2[y - f(\theta)]'F(\theta)$, tem-se: $F'\left(\hat{\theta}\right)\left[y - f\left(\hat{\theta}\right)\right] = \emptyset$.

Portanto, o sistema de equações normais (SEN) é dado por:

$$\begin{bmatrix} \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_1} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_1} & \cdots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_1} \\ \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_2} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_2} & \cdots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_p} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_p} & \cdots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_p} \end{bmatrix} \times \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} f_1\left(\hat{\theta}\right) \\ f_2\left(\hat{\theta}\right) \\ \vdots \\ f_n\left(\hat{\theta}\right) \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2}$$

2.2.2.2 Método dos mínimos quadrados ponderados

De acordo com Hoffman e Vieira (1998) e Mazzini (2001), em presença de heterogeneidade de variâncias, o método dos quadrados mínimos ponderado é mais adequado, por fornecer estimadores não tendenciosos e de mínima variância.

Seja o modelo linear

$$Y = X\beta + u,$$

supondo-se que $\varepsilon \sim N(0; V\sigma^2)$, em que V é uma matriz diagonal, positiva defi-

nida, que representa as variâncias associadas a cada u_i, com $E(u) = 0$ e

$$E(u'u) = V\sigma^2 = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_n \end{bmatrix} \sigma^2.$$

De acordo com Mazzini (2001), o fato de serem nulos os elementos fora da diagonal principal da matriz V, significa que é válida a pressuposição de independência das várias observações, isto é $E(u_j u_h) = 0$ para $j \neq h$.

Defina a matriz diagonal Λ, cujos elementos são dados por: $\lambda_j = \frac{1}{\sqrt{V_j}}$, ou seja,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix},$$

desta forma, tem-se que:

$$\Lambda\Lambda = V^{-1} \text{ e } V = \Lambda^{-1}\Lambda^{-1},$$

pré-multiplicando cada um dos termos de $Y = X\beta + u$ por Λ, obtém-se o modelo:

$$\Lambda Y = \Lambda X\beta + \Lambda u.$$

No modelo $Y = X\beta + u$, tem-se o vetor de erros $\varepsilon = \Lambda u$, uma vez que $E(u) = 0$, tem-se que $E(\varepsilon) = 0$.

$$E(\varepsilon\varepsilon') = E(\Lambda u u'\Lambda) = \Lambda V \Lambda \sigma^2,$$

de acordo com

$$V = \Lambda^{-1}\Lambda^{-1},$$

$$E(\varepsilon\varepsilon') = \Lambda\Lambda^{-1}\Lambda^{-1}\Lambda\sigma^2 = I\sigma^2.$$

O modelo $\Lambda Y = \Lambda X\beta + \Lambda u$ é homocedástico. O método dos quadrados mínimos ordinário fornece o SEN dado por

$$X'V^{-1}X\hat{\beta} = X'V^{-1}Y,$$

a solução do SEN leva ao estimador

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y.$$

Como o método dos quadrados mínimos para os modelos lineares é similar aos modelos não-lineares, pode-se por analogia, escrever o SEN (não-linear) ponderado da seguinte forma:

$$(X'V^{-1}X)f(\hat{\theta}) = X'V^{-1}Y. \tag{3}$$

2.2.2.3 Método dos mínimos quadrados generalizados

Hoffman e Vieira (1998), consideram que em presença de heterogeneidade de variâncias e autocorrelação dos resíduos, o método dos quadrados mínimos generalizados é mais eficiente do que o método dos quadrados mínimos ponderados e ordinários.

Seja o modelo linear

$$Y = X\beta + u$$

supondo-se que $\varepsilon \sim N(0; W\sigma^2)$, em que W é uma matriz simétrica, positiva de-

finida, que representa as variâncias e covariâncias dos erros. Admitindo-se que os erros são autocorrelacionados na forma de um processo autorregressivo estacionário de primeira ordem,

$$u_t = \rho_v u_{t-1} + \varepsilon_t,$$

em que $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$, $E(\varepsilon_t \varepsilon_{t-h}) = 0$ se $h \neq 0$, $-1 < \rho_v < 1$ para $t = 1, 2, \cdots, n$; e $v = 1, 2, \cdots, n$.

Nestas condições $\sigma_u^2 = \dfrac{\sigma_\varepsilon^2}{1-\rho_v^2}$ e $Cov[u_i, u_j] = \dfrac{\sigma_\varepsilon^2}{1-\rho_v^2}\rho_v^h = \rho_v^h \sigma_u^2$.

De maneira análoga, ao método dos quadrados mínimos ponderados, encontra-se $\hat{\beta}$

$$\hat{\beta} = (X'W^{-1}X)^{-1}X'W^{-1}Y,$$

em que

$$W = \frac{\sigma_\epsilon^2}{1-\rho_v^2}\begin{bmatrix} 1 & \rho_v & \rho_v^2 & \cdots & \rho_v^{n-1} \\ \rho_v & 1 & \rho_v & \cdots & \rho_v^{n-2} \\ \rho_v^2 & \rho_v & 1 & \cdots & \rho_v^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_v^{n-1} & \rho_v^{n-2} & \rho_v^{n-3} & \cdots & 1 \end{bmatrix}. \tag{4}$$

Substituindo-se a equação 4 no SEN (não-linear), dado pela equação3, tem-se

$$(X'W^{-1}X)^{-1} f(\hat{\theta}) = X'W^{-1}Y.$$

Segunda a notação de Morettin e Toloi (2006), se os erros forem autocorrelacionados na forma de um processo autorregressivo estacionário de segunda ordem tem-se,

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t,$$

em que u_t é estacionário se:

$$\phi_1 + \phi_2 < 1,$$
$$\phi_2 - \phi_1 < 1,$$
$$-1 < \phi_2 < 1,$$

sendo que ϕ_1 e ϕ_2 são os parâmetros de autocorrelação.

Logo, temos que:

$$\sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_1\rho_1 - \phi_2\rho_2},$$

enquanto as funções de autocorrelação são dadas por:

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2},\ j > 0,$$

sendo

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \quad e \quad \rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2.$$

2.2.3 Processo iterativo

Segundo Silveira (2010), para o SEN não-linear, (equação 3), não existe uma solução explícita, sendo assim a solução para o sistema deve ser obtida por meio de processos iterativos. Um dos métodos iterativos é a linearização da função não-linear, chamado método de Gauss-Newton, o qual se resume ao seguinte procedimento.

2.2.3.1 Método de Gauss-Newton

De acordo com Gallant (1987), o método de Gauss-Newton na forma original, consiste no desenvolvimento em série de Taylor até o termo de primeira

ordem da função $f(X_i;\theta)$, em torno do ponto θ^0.

Considere o modelo não-linear $y_i = f(x_i,\theta) + \varepsilon_i$, e $\hat{\theta}_0$ um valor tal que $F'\left(\hat{\theta}_0\right)\left[Y - f\left(\hat{\theta}_0\right)\right] \approx 0$. Aproximando $f\left(\hat{\theta}\right)$ pelo ponto $\hat{\theta}_0$ por uma série de Taylor de primeira ordem, tem-se:

$$f\left(\hat{\theta}\right) \approx f\left(\hat{\theta}_0\right) + F\left(\hat{\theta}_0\right)\left(\hat{\theta} - \hat{\theta}_0\right) \tag{5}$$

$$F'\left(\hat{\theta}\right)\left[Y - f\left(\hat{\theta}\right)\right] \approx 0 \tag{6}$$

Aplicando (5) em (6): $F'\left(\hat{\theta}\right)\left[Y - f\left(\hat{\theta}_0\right) - F\left(\hat{\theta}_0\right)\left(\hat{\theta} - \hat{\theta}_0\right)\right] \approx 0$, e multiplicando à esquerda, ambos os lados da igualdade, por $\left[F'\left(\hat{\theta}\right)\right]^{-1}$, obtém-se:

$$Y - f\left(\hat{\theta}_0\right) - F\left(\hat{\theta}_0\right)\hat{\theta} + F\left(\hat{\theta}_0\right)\hat{\theta}_0 \approx 0.$$

Logo, $F\left(\hat{\theta}_0\right)\hat{\theta} \approx F\left(\hat{\theta}_0\right)\hat{\theta}_0 + \left[Y - f\left(\hat{\theta}_0\right)\right]$. Multiplicando novamente à esquerda, ambos os lados da igualdade, por $\left[F\left(\hat{\theta}\right)\right]^{-1}$, verifica-se que:

$$\hat{\theta} \approx \hat{\theta}_0 + \left[F\left(\hat{\theta}_0\right)\right]^{-1}\left[Y - f\left(\hat{\theta}_0\right)\right].$$

Fazendo $\hat{\theta} = \hat{\theta}_{k+1}$ e $\hat{\theta}_0 = \hat{\theta}_k$, tem-se para a k-ésima iteração, a expressão (7), a qual representa o processo iterativo conhecido como Gauss-Newton:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \left[F\left(\hat{\theta}_k\right)\right]^{-1}\left[Y - f\left(\hat{\theta}_k\right)\right]. \tag{7}$$

Este processo iterativo prossegue até que algum critério adotado para convergência seja atingido.

2.2.4 Funções não-lineares para descrever modelos de crescimento

Segundo Denise e Brinks (1985), os modelos mais utilizados para descrever o crescimento de animais são modelos biológicos como Brody (BRODY, 1945), von Bertalanffy (BERTALANFFY, 1957), Logístico (NELDER, 1961), Michaelis-Menten (MICHAELIS; MENTEN, 1913) e Gompertz (LAIRD; HOWARD, 1967). Estes modelos apresentam dois parâmetros de grande importância para a avaliação do crescimento, o peso assintótico (A) e a taxa de maturidade (k). Tais modelos também contêm um parâmetro que se caracteriza como constante matemática (B), o qual não é interpretável na prática. O peso assintótico, ou peso adulto, que representa a estimativa de peso à maturidade, independente de flutuações de pesos em razão de efeitos genéticos e ambientais. A taxa de maturidade corresponde a uma estimativa de precocidade, determinando assim a eficiência do crescimento. No crescimento de bovinos, por exemplo, as curvas de crescimento são importantes para a obtenção de animais com maior produção, menores custos e menor tempo para atingir determinado peso (LAIRD; HOWARD, 1967). Os modelos são apresentados na Tabela 1.

Tabela 1 Expressões matemáticas dos modelos.

Modelo	Equação
Brody	$Y = A\left(1 - Bexp(-kt)\right) + \epsilon$
von Bertalanffy	$Y = A\left(1 - Bexp(-kt)\right)^{3} + \epsilon$
Logístico	$Y = A\left(1 + Bexp(-kt)\right)^{-1} + \epsilon$
Gompertz	$Y = Aexp(-Bexp(-kt)) + \epsilon$
Michaelis-Menten	$Y = At/\left(B + t\right) + \epsilon$

Em todas as funções Y representa o peso corporal, e t representa a idade, em meses, a partir do nascimento. Segundo Silveira (2010), de forma geral, estes modelos têm por objetivo descrever uma trajetória assintótica da variável depen-

dente peso, em função da variável independente tempo. Geralmente, a diferença entre tais modelos é dada pela definição do ponto de inflexão da curva, que confere uma forma sigmóide a mesma, porém para alguns modelos este ponto pode não existir.

2.2.4.1 Função de Brody

Segundo Duarte (1975) e Mazzini (2001), a função de Brody ou Monomolecular foi estudada inicialmente por Robertson (em 1908); Brody (em 1945) e von Bertalanffy (em 1957), para descrever o crescimento de bovinos.

A função de Brody é escrita da seguinte forma:

$$Y_t = A(1 - Be^{-Kt}),$$

em que

A = valor assintótico, sendo $A > 0$;

B = constante de integração, sendo $B \in \mathbb{R}$;

K = taxa de maturidade, sendo $K > 0$.

Logo, seus parâmetros são $\theta_1 = A$; $\theta_2 = B$; $\theta_3 = K$, e sua variável independente é t.

As derivadas parciais para esta função são:

$$\frac{\partial f}{\partial A} = 1 - Be^{-Kt}, \qquad \frac{\partial f}{\partial B} = -Ae^{-Kt}, \qquad \frac{\partial f}{\partial K} = ABte^{-Kt}.$$

Logo, a matriz $F(\theta^0) = X$, é uma matriz $n \times p$ de derivadas parciais, em que:

n = número de pesagens;

p = número de parâmetros da função; $p = 3$ (A, B e K);

$$X = \begin{bmatrix} 1 - B_0 e^{-Kt_1} & -A_0 e^{-K_0 t_1} & A_0 B_0 t_1 e^{-K_0 t_1} \\ 1 - B_0 e^{-Kt_2} & -A_0 e^{-K_0 t_2} & A_0 B_0 t_2 e^{-K_0 t_2} \\ \vdots & \vdots & \vdots \\ 1 - B_0 e^{-Kt_n} & -A_0 e^{-K_0 t_n} & A_0 B_0 t_n e^{-K_0 t_n} \end{bmatrix}_{n \times 3}.$$

Neste caso a matriz $X'X$ é uma matriz (3×3) simétrica, a qual pode ser escrita da seguinte forma:

$$a_{11} = \sum_{i=1}^{n} \left(1 - B_0 e^{-K_0 t_i}\right)^2,$$

$$a_{12} = a_{21} = A_0 \sum_{i=1}^{n} \left[\left(B_0 e^{-K_0 t_i} - 1\right) e^{-K_0 t_i}\right],$$

$$a_{13} = a_{31} = A_0 B_0 \sum_{i=1}^{n} \left[\left(1 - B e^{-K_0 t_i}\right) t_i e^{-K_0 t_i}\right],$$

$$a_{22} = A_0^2 \sum_{i=1}^{n} e^{-2K_0 t_i},$$

$$a_{23} = a_{32} = -A_0^2 B_0 \sum_{i=1}^{n} t_i e^{-2K_0 t_i},$$

$$a_{33} = A_0^2 B_0^2 \sum_{i=1}^{n} t_i^2 e^{-2K_0 t_i}.$$

2.2.4.2 Função de Gompertz

De acordo com Duarte (1975) e Mazzini (2001), a função de Gompertz foi estudada por Gompertz (em 1825), Winsor (em 1932) e Laird (em 1965) para descrever a taxa de mortalidade numa população e seu emprego, para descrever modelos sigmoidais de crescimento, foi sugerido por Wright (em 1926).

A função de Gompertz é escrita da seguinte forma:

$$Y_t = A \exp(-B \exp(-Kt)),$$

em que,

A = valor assintótico, sendo $A > 0$;

B = constante de integração, sendo $B \in \mathbb{R}$;

K = taxa de maturidade, sendo $K > 0$.

Logo, seus parâmetros são $\theta_1 = A$; $\theta_2 = B$; $\theta_3 = K$, e sua variável independente é t.

As derivadas parciais para esta função são:

$$\frac{\partial f}{\partial A} = e^{-Be^{-Kt}}, \qquad \frac{\partial f}{\partial B} = -Ae^{-Be^{-Kt}}e^{-Kt}, \qquad \frac{\partial f}{\partial K} = ABte^{-Be^{-Kt}}e^{-Kt}.$$

Logo, a matriz $F(\theta^0) = X$, é uma matriz $n \times p$ de derivadas parciais, em que:

n = número de pesagens;

p = número de parâmetros da função; $p = 3$ (A, B e K);

$$X = \begin{bmatrix} e^{-B_0e^{-K_0t_1}} & -A_0e^{-B_0e^{-K_0t_1}}e^{-K_0t_1} & A_0B_0t_1e^{-B_0e^{-K_0t_1}}e^{-K_0t_1} \\ e^{-B_0e^{-K_0t_2}} & -A_0e^{-B_0e^{-K_0t_2}}e^{-K_0t_2} & A_0B_0t_2e^{-B_0e^{-K_0t_2}}e^{-K_0t_2} \\ \vdots & \vdots & \vdots \\ e^{-B_0e^{-K_0t_n}} & -A_0e^{-B_0e^{-K_0t_n}}e^{-K_0t_n} & A_0B_0t_ne^{-B_0e^{-K_0t_n}}e^{-K_0t_n} \end{bmatrix}_{n\times 3}.$$

Neste caso a matriz $X'X$ é uma matriz (3×3) simétrica, a qual pode ser escrita da seguinte forma:

$$a_{11} = \sum_{i=1}^{n} e^{-2B_0e^{-K_0t_i}},$$

$$a_{12} = a_{21} = -A_0 \sum_{i=1}^{n} e^{-2B_0e^{-K_0t_i}}e^{-K_0t_i},$$

$$a_{13} = a_{31} = A_0B_0 \sum_{i=1}^{n} t_ie^{-2B_0e^{-K_0t_i}}e^{-K_0t_i},$$

$$a_{22} = A_0^2 \sum_{i=1}^{n} e^{-2B_0e^{-K_0t_i}}e^{-2K_0t_i},$$

$$a_{23} = a_{32} = -A_0^2 B_0 \sum_{i=1}^{n} t_i e^{-2B_0 e^{-K_0 t_i}} e^{-2K_0 t_i},$$
$$a_{33} = A_0^2 B_0^2 \sum_{i=1}^{n} t_i^2 e^{-2B_0 e^{-K_0 t_i}} e^{-2K_0 t_i}.$$

2.2.4.3 Função logística

Conforme Hoffmann e Vieira (1998) e Mazzini (2001), a função logística foi indicada para o estudo descritivo do crescimento de populações humanas por Verhulst (1845), que a denominou de "curva logística". Muitos anos mais tarde, Pearl e Reed (1920), sem conhecerem a contribuição de Verhulst, obtiveram a mesma curva, a qual utilizaram para descrever o crescimento da população dos EUA, de 1870 a 1910, com base em dados censitários.

Sua função é escrita da seguinte forma:

$$Y_t = A(1 + B\exp(-Kt))^{-1},$$

em que,

A = valor assintótico, sendo $A > 0$.

B = constante de integração, sendo $B \in \mathbb{R}$;

K = taxa de maturidade, sendo $K > 0$.

Logo, seus parâmetros são $\theta_1 = A$; $\theta_2 = B$; $\theta_3 = K$, e sua variável independente é t.

As derivadas parciais para esta função são:

$$\frac{\partial f}{\partial A} = \frac{1}{(1 + Be^{-Kt})}, \qquad \frac{\partial f}{\partial B} = \frac{-Ae^{-Kt}}{(1 + Be^{-Kt})^2}, \qquad \frac{\partial f}{\partial K} = \frac{ABte^{-Kt}}{(1 + Be^{-Kt})^2}.$$

Logo, a matriz $F(\theta^0) = X$, é uma matriz $n \times p$ de derivadas parciais, em que:

n = número de pesagens;

p = número de parâmetros da função; $p = 3$ (A, B e K);

$$X = \begin{bmatrix} \frac{1}{(1+B_0e^{-K_0t_1})^2} & \frac{-A_0e^{-K_0t_1}}{(1+B_0e^{-K_0t_1})^2} & \frac{A_0B_0t_1e^{-K_0t_1}}{(1+B_0e^{-K_0t_1})^2} \\ \frac{1}{(1+B_0e^{-K_0t_2})^2} & \frac{-A_0e^{-K_0t_2}}{(1+B_0e^{-K_0t_2})^2} & \frac{A_0B_0t_2e^{-K_0t_2}}{(1+B_0e^{-K_0t_2})^2} \\ \vdots & \vdots & \vdots \\ \frac{1}{(1+B_0e^{-K_0t_n})^2} & \frac{-A_0e^{-K_0t_n}}{(1+B_0e^{-K_0t_n})^2} & \frac{A_0B_0t_ne^{-K_0t_n}}{(1+B_0e^{-K_0t_n})^2} \end{bmatrix}_{n\times 3}.$$

Neste caso a matriz $X'X$ é uma matriz (3×3) simétrica, a qual pode ser escrita da seguinte forma:

$$a_{11} = \sum_{i=1}^{n} (1 + B_0e^{-K_0t_i})^{-2},$$

$$a_{12} = a_{21} = -A_0 \sum_{i=1}^{n} e^{-K_0t_i}(1 + B_0e^{-K_0t_i})^{-3},$$

$$a_{13} = a_{31} = A_0B_0 \sum_{i=1}^{n} t_ie^{-K_0t_i}(1 + B_0e^{-K_0t_i})^{-3},$$

$$a_{22} = A_0^2 \sum_{i=1}^{n} e^{-2K_0t_i}(1 + B_0e^{-K_0t_i})^{-4},$$

$$a_{23} = a_{32} = -A_0^2B_0 \sum_{i=1}^{n} t_ie^{-2K_0t_i}(1 + B_0e^{-K_0t_i})^{-4},$$

$$a_{33} = A_0^2B_0^2 \sum_{i=1}^{n} t_i^2e^{-2K_0t_i}(1 + B_0e^{-K_0t_i})^{-4}.$$

2.2.4.4 Função de von Bertalanffy

De acordo com Elias (1998) e Mazzini (2001), a forma da função de von Bertalanffy é similar à de Gompertz, e foi desenvolvida com base na suposição de que o crescimento de um organismo é a diferença entre taxas de anabolismo e catabolismo de seus tecidos.

Segundo Duarte (1975) e Mazzini (2001), a relação entre tamanho corporal e taxa metabólica capacitou von Bertalanffy a modificar a forma geral de sua equação, tornando-a apropriada para algumas espécies ou tipos de medidas. De acordo

com o mesmo autor, o modelo de von Bertalanffy tem, talvez, o mais rigoroso suporte nas teorias biológicas, o que lhe capacita resistir melhor a interpretação de seus parâmetros.

Sua função é escrita da seguinte forma:

$$Y_t = A(1 - B\exp(-Kt))^3,$$

em que:

A = valor assintótico, sendo $A > 0$;

B = constante de integração, sendo $B \in \mathbb{R}$;

K = taxa de maturidade, sendo $K > 0$.

Logo, seus parâmetros são $\theta_1 = A$; $\theta_2 = B$; $\theta_3 = K$, e sua variável independente é t.

As derivadas parciais para esta função são:

$$\frac{\partial f}{\partial A} = (1 - Be^{-Kt})^3,$$

$$\frac{\partial f}{\partial B} = -3A(1 - Be^{-Kt})^2 e^{-Kt},$$

$$\frac{\partial f}{\partial K} = 3ABt(1 - Be^{-Kt})^2 e^{-Kt}.$$

Logo, a matriz $F(\theta^0) = X$, é uma matriz $n \times p$ de derivadas parciais, em que:

n = número de pesagens;

p = número de parâmetros da função; $p = 3$ (A, B e K);

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & X_{n3} \end{bmatrix}_{n \times 3}$$

sendo, para $1 \leq i \leq n$,

$X_{i1} = (1 - B_0 e^{-K_0 t_i})^3$,

$X_{i2} = -3A_0(1 - B_0 e^{-K_0 t_i})^2 e^{-K_0 t_i}$,

$X_{i3} = 3A_0 B_0 t_i (1 - B_0 e^{-K_0 t_i})^2 e^{-K_0 t_i}$.

Neste caso a matriz $X'X$ é uma matriz (3x3) simétrica, a qual pode ser escrita da seguinte forma:

$a_{11} = \sum_{i=1}^{n} (1 - B_0 e^{-K_0 t_i})^6$,

$a_{12} = a_{21} = -3A_0 \sum_{i=1}^{n} e^{-K_0 t_i}(1 - B_0 e^{-K_0 t_i})^5$,

$a_{13} = a_{31} = 3A_0 B_0 \sum_{i=1}^{n} t_i e^{-K_0 t_i}(1 - B_0 e^{-K_0 t_i})^5$,

$a_{22} = 9A_0^2 \sum_{i=1}^{n} e^{-2K_0 t_i}(1 - B_0 e^{-K_0 t_i})^4$,

$a_{23} = a_{32} = -9A_0^2 B_0 \sum_{i=1}^{n} t_i e^{-2K_0 t_i}(1 - B_0 e^{-K_0 t_i})^4$,

$a_{33} = 9A_0^2 B_0^2 \sum_{i=1}^{n} t_i^2 e^{-2K_0 t_i}(1 - B_0 e^{-K_0 t_i})^4$.

2.2.4.5 Função de Michaelis-Menten

A função de Michaelis-Menten, introduzida em 1913 tem sua função escrita da seguinte forma:

$$Y_t = \frac{At}{B + t},$$

em que,

A = valor assintótico, sendo $A > 0$;

B = constante de integração, sendo $B \in \mathbb{R}$.

Logo, seus parâmetros são $\theta_1 = A$; $\theta_2 = B$; e sua variável independente é t.

As derivadas parciais para esta função são:

$$\frac{\partial f}{\partial A} = \frac{t}{B+t},$$

$$\frac{\partial f}{\partial B} = -\frac{At}{(B+t)^2}.$$

Logo, a matriz $F(\theta^0) = X$, é uma matriz $n \times p$ de derivadas parciais, em que:

n = número de pesagens;

p = número de parâmetros da função; $p = 2$ (A e B);

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ \vdots & \vdots \\ X_{n1} & X_{n2} \end{bmatrix}_{n \times 2},$$

sendo, para $1 \leq i \leq n$, $X_{i1} = \dfrac{t_i}{B+t_i}$ e $X_{i2} = -\dfrac{At_i}{(B+t_i)^2}$.

Neste caso a matriz $X'X$ é uma matriz (2×2) simétrica, a qual pode ser escrita da seguinte forma:

$$a_{11} = \sum_{i=1}^{n} \left(\frac{t_i}{B+t_i}\right)^2,$$

$$a_{12} = a_{21} = -A \sum_{i=1}^{n} \frac{t_i^2}{(B+t_i)^3},$$

$$a_{22} = A^2 \sum_{i=1}^{n} \left(\frac{t_i}{(B + t_i)^2} \right)^2.$$

2.3 Comparação entre os modelos

Segundo Silveira et al. (2011), diversos avaliadores podem ser utilizados para avaliar a qualidade de ajuste, tais como, coeficiente de determinação ajustado (R^2_{aj}), critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), erro quadrático médio de predição (MEP) e coeficiente de determinação de predição (R^2_p). O interesse aqui concentra-se nos critérios AIC, $AICc$ e BIC.

2.3.1 Critério de informação de Akaike - AIC

Permite utilizar o princípio da parcimônia na escolha do melhor modelo, ou seja, de acordo com este critério nem sempre o modelo mais parametrizado é melhor (BURNHAM; ANDERSON, 2004). Menores valores de AIC refletem um melhor ajuste (AKAIKE, 1974). Sua expressão é dada por:

$$AIC = -2 \log L \left(\widehat{\boldsymbol{\theta}} \right) + 2 \left(p \right),$$

em que: p é o número de parâmetros e $\log L \left(\widehat{\boldsymbol{\theta}} \right)$ é o valor do logaritmo da função de verossimilhança considerando as estimativas dos parâmetros.

Para maiores detalhes acerca do critério de informação de Akaike (AIC), vide seção 2.1.1 do capítulo 1.

2.3.2 Critério de informação de Akaike corrigido - AICc

Sugiura (1978), derivou uma variante de segunda ordem do AIC, chamado de critério de informação de Akaike corrigido (AICc), dado por

$$AICc = -2\log L\left(\widehat{\boldsymbol{\theta}}\right) + 2p\left(\frac{n}{n-p-1}\right),$$

em que n é o tamanho amostral e p é o número de parâmetros do modelo.

Para maiores detalhes acerca do critério de informação de Akaike (AIC), vide seção 2.1.2 do capítulo 1.

2.3.3 Critério de informação bayesiano - BIC

Assim como o AIC, também leva em conta o grau de parametrização do modelo, e da mesma forma, quanto menor for o valor de BIC (SCHWARZ, 1978), melhor será o ajuste do modelo. Sua expressão é dada por:

$$BIC = -2\log L\left(\widehat{\boldsymbol{\theta}}\right) + p\log\left(n\right),$$

em que, n é o número de observações utilizadas para ajustar a curva.

Para maiores detalhes acerca do critério de informação de Schwarz (BIC), vide seção 2.1.3 do capítulo 1.

3 METODOLOGIA

Com o intuito de estudar o crescimento de animais (em especial de bovinos), pode-se utilizar os modelos não lineares de Gompertz, Logístico, von Bertalanffy, Michaelis-Menten e Brody, dentre outros, para modelar este fenômeno. Na etapa da seleção do melhor modelo, os critérios AIC, AICc e BIC podem ser utilizados, a fim de selecionar o melhor modelo que se ajusta aos dados sob estudo. Comparou-se aqui a utilização destes critérios de informação e para isso, foram realizadas simulações Monte Carlo para os cinco modelos e estudou-se os critérios em diferentes cenários simulados.

3.1 Simulação dos modelos de crescimento

Dados de peso e idade foram simulados a partir dos modelos Brody, Gompertz, Logístico, Michaelis-Menten e von Bertalanffy, tendo como referência o peso de bovinos da raça Hereford, ajustado a um conjunto de dados, conforme relatado por Mazzini et al. (2005), e reproduzido na Tabela 2. Considerou-se os ajustes médios para os animais, cujas estimativas foram obtidas através do método de mínimos quadrados generalizados, com erros auto-regressivos de primeira ordem. As rotinas de simulação foram implementadas no software R (R DEVELOPMENT CORE TEAM, 2013) e os modelos ajustados utilizando a biblioteca $nlme$ através da função $gnls()$.

Para cada um dos cinco modelos estudados, foram simulados pesos de 50 animais ao longo de 721 dias, em pesagens a cada 60 dias, isto é, os tempos adotados foram $1, 61, 121, \ldots, 721$, que perfizeram um total de treze diferentes pesos para cada animal. Em posse dos dados simulados de um dos modelos estudados (Brody, por exemplo), os cinco modelos foram ajustados aos dados e determinou-

Tabela 2 Parâmetros utilizados nas simulações dos modelos, adotados conforme relatado por Mazzini et al. (2005).

Modelo	α	β	k
Brody	1069	0,9512	0,0014
von Bertalanffy	862	0,6408	0,0029
Logístico	703	16,6503	0,0089
Gompertz	786	3,0610	0,0042
Michaeles-Menten	600	350	Ausente

se os valores de AIC, AICc e BIC para cada um deles. Caso o menor valor do critério de seleção (AIC, por exemplo), refere-se ao verdadeiro modelo simulado, temos um acerto (TP), caso contrário o critério não selecionou corretamente o modelo (FN); o modelo que foi selecionado incorretamente (Logístico, por exemplo) terá sua taxa de falso positivo (FP) incrementada. Este processo foi repetido 100 vezes, para cada um dos modelos, e determinou-se as taxas TP, FP e FN, além do desvio padrão da taxa de verdadeiro positivo (TP).

4 APLICAÇÃO EM DADOS REAIS DE CONSUMO DE CRESCIMENTO BOVINO

Os dados utilizados neste trabalho, provém de animais da raça Hereford, machos não castrados, nascidos nos anos de 1992, 1993 e 1994, na Agropecuária Recreio de propriedade do Sr. Roberto Silveira Collares, situada no município de Bagé, Estado do Rio Grande do Sul. Os mesmos foram cedidos pelo Centro de Pesquisas Pecuária Sul, pertencente a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), situada no mesmo município.

Um dos principais objetivos da Agropecuária é a criação de touros, os quais são vendidos, com aproximadamente 2 anos de idade, época em que estão aptos a entrarem em reprodução.

O manejo alimentar da propriedade é basicamente pasto nativo, pastagem cultivada e suplementação à campo estratégica (do desmame até 1° ano).

As pesagens consideradas no estudo, foram coletadas de 160 animais, desde o nascimento até aproximadamente 790 dias de idade, obtendo-se no grupo 47 animais com 16 pesagens; 30 animais com 17 pesagens; 24 animais com 18 pesagens; 51 animais com 19 pesagens e 8 animais com 20 pesagens. Foram consideradas aqui somente as 16 primeiras pesagens dos animais, não trabalhando-se então com diferentes números de pesagens. Os dados utilizados encontram-se na Tabela 3.

Tabela 3 Valores médios dos pesos observados e respectivas variâncias.

Idade (em dias)	Peso (em Kg)	Variância
0	35	13,95
168	189,97	3105,72
214	213,76	3037,99
243	243,64	2383,03
275	281,12	2657,6
311	324,39	2704,35
344	363,9	3372,86
377	404,22	3238,55
414	443,92	3468,11
472	489,51	4722,59
508	518,54	6041,73
542	541,34	5853,09
594	559,21	5788,24
624	583,81	6578,37
660	611,76	6542,95
692	639,24	8667,74

5 RESULTADOS E DISCUSSÃO

Na seção 5.1 discute-se acerca dos resultados obtidos a partir das simulações realizadas, enquanto que na seção 5.2 a discussão versa acerca dos dados reais. As simulações referentes ao modelo Gompertz encontra-se no apêndice A, sendo que, as simulações para os demais modelos descritos na Tabela 1 são similares.

5.1 Estudo de simulações para modelos de crescimento

Analisando as Figuras (1)-(5), nota-se que, o BIC não apresentou o melhor desempenho (maior TP e menor FP) em nenhumas das situações. Para o modelo Michaelis-Menten (Figura 4), o BIC apresentou TP superior aos outros critérios, entretanto a taxa FP foi também superior. Para o modelo Brody (Figura 1), a situação foi oposta. No modelo Gompertz (Figura 2) o desempenho dos três critérios foi similar e bom, com TP em torno de 76% e taxa FP baixa. Para o modelo Logístico (Figura 3), tanto o AIC quanto o AICc apresentaram TP superior ao BIC e a taxa FP foi similar para os três critérios. No modelo de von Bertallanffy (Figura 5), as taxas TP e FP foram baixas para todos os critérios.

Em geral, para os modelos estudados, o desempenho do AIC e AICc foi tão bom ou superior ao desempenho do BIC (ver Figuras 1, 2 e 3). Para o caso do modelo de von Bertalanffy (veja a Figura 5), os critérios de informação AIC e AICc mostram-se tão ruins quanto o BIC, isto é, nenhum dos critérios (AIC, AICc e BIC) são capazes de selecionar o modelo corretamente. Somente para o caso do modelo de Michaelis-Menten (Figura 4) o BIC apresentou TP superior aos outros critérios, entretanto a taxa de falso positivo (FP) foi superior e, em termos de teste de hipóteses, este modelo apresentará poder inferior ao AIC e AICc.

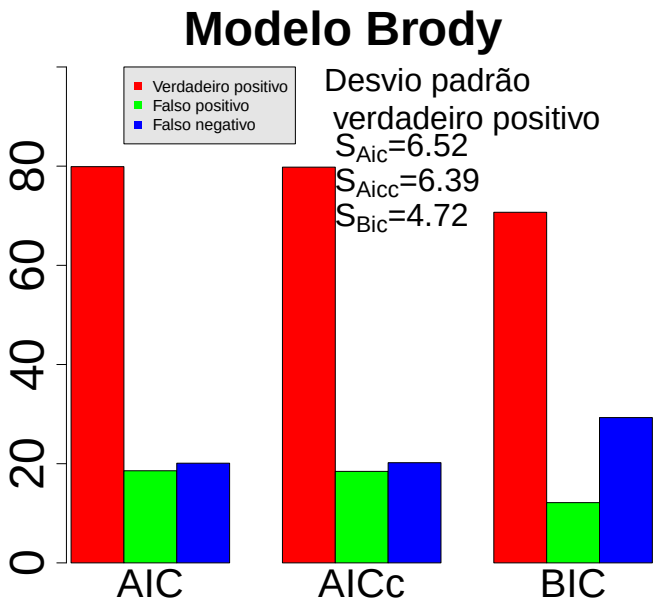

Figura 1 Resultados da simulação para o modelo Brody

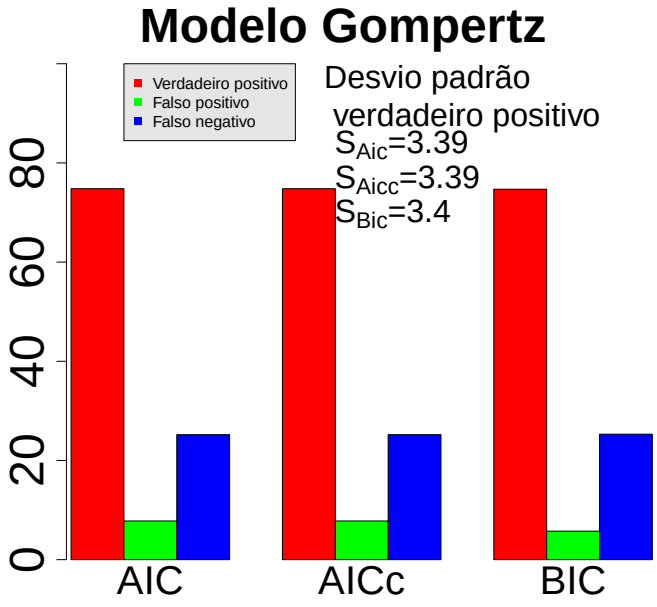

Figura 2 Resultados da simulação para o modelo Gompertz

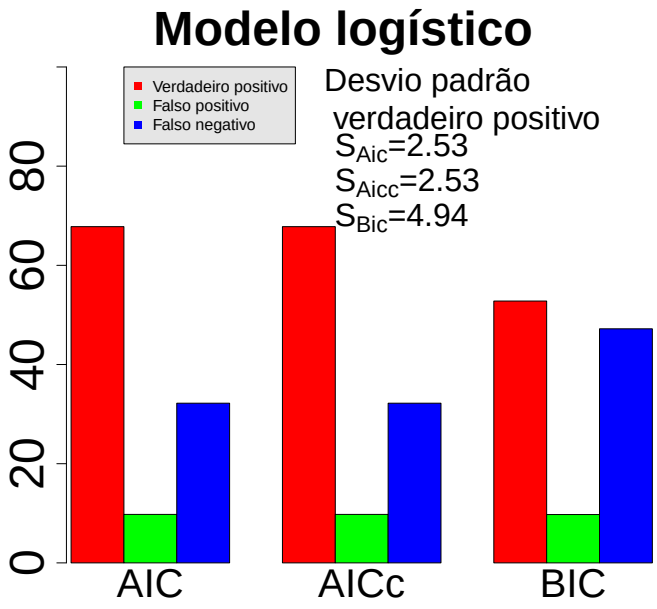

Figura 3 Resultados da simulação para o modelo Logístico

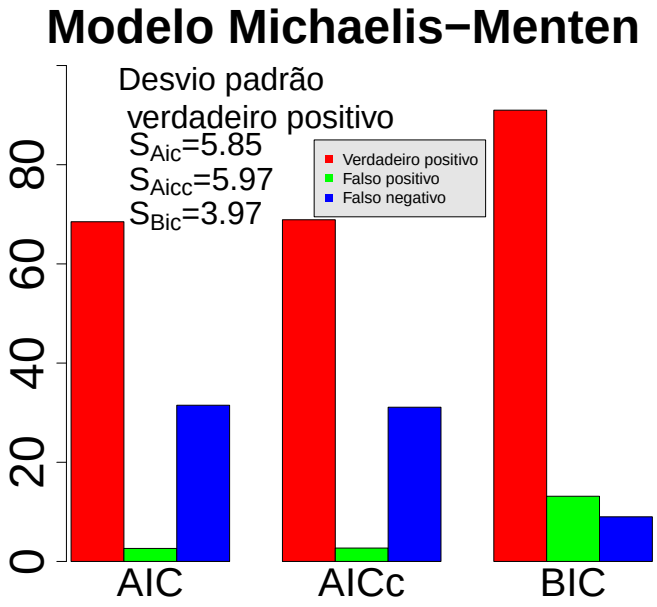

Figura 4 Resultados da simulação para o modelo Michaelis-Menten

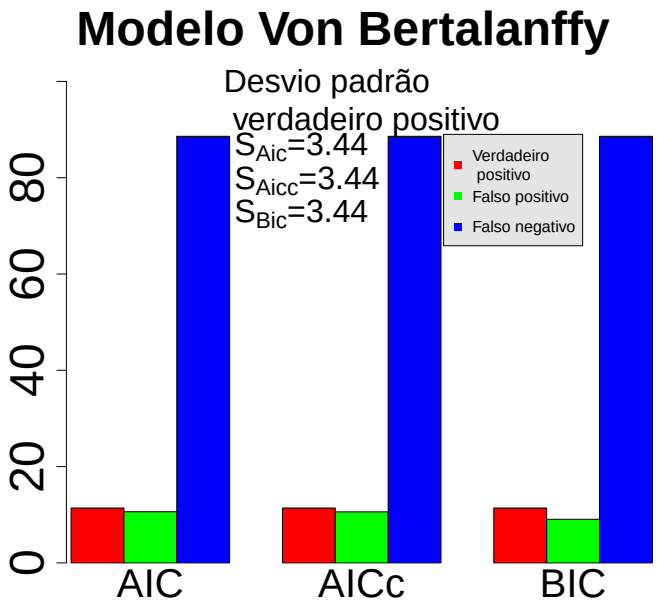

Figura 5 Resultados da simulação para o modelo von Bertalanffy

5.2 Aplicação aos dados de crescimento de novilhos machos da raça Hereford

Na Figura 6, constam os valores médios dos pesos observados e suas variâncias em função da idade, em dias, dos 160 animais em estudo.

Pode-se observar que à medida que a idade aumentou, houve um incremento nas variâncias dos pesos corporais. Estes resultados estão de acordo com Pasternak e Shalev (1994), os quais relataram que a variância dos pesos corporais aumenta com a idade, ocorrendo a heterocedasticidade, o que os autores denominam "distúrbios de regressão". Este resultado justifica a ponderação pelo inverso da variância que foi efetuado aqui.

Pela análise da Tabela 4, pode-se observar que, sem a ponderação e a estrutura de correlação dos erros, os três critérios selecionaram o mesmo modelo (von

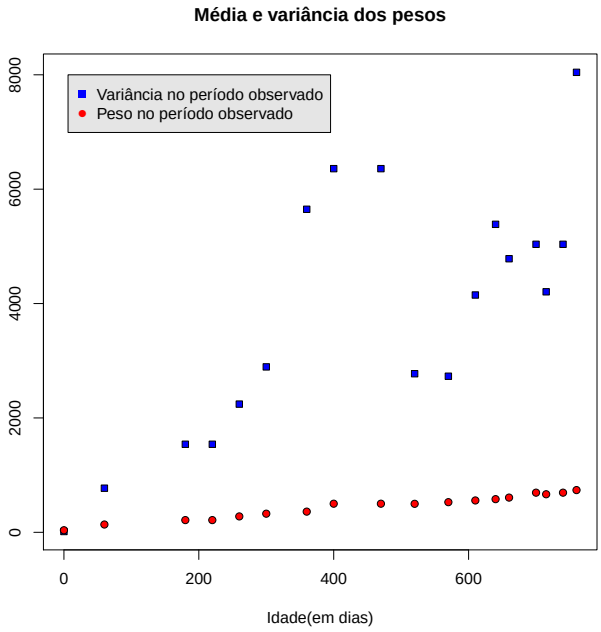

Figura 6 Média e variância dos pesos observados, dentro de cada período, dos machos Hereford

Bertalanffy), entretanto tal resultado não apresenta as pressuposições necessárias à validade do ajuste dos modelos de regressão não-linear. Quando ponderou-se pelo inverso da variância, as conclusões do AIC e AICc BIC também são as mesmas, apontando o modelo de von Bertalanffy como o melhor. Este último resultado entretanto, não leva em consideração a correlação, e desta forma nossos modelos podem ser invalidados. Considerando-se a correlação e a ponderação pelo inverso da variância, novamente os três critérios levam a mesma conclusão de que o melhor modelo é o de von Bertalanffy.

Estes resultados condizem com aqueles apresentados por Mazzini (2001) e Mazzini et al. (2005), em que os modelos selecionados foram os de Gompertz e de von Bertalanffy, sendo que nestes trabalhos os critérios de seleção foram o erro

de predição médio (EPM), a porcentagem de convergência e o quadrado médio do resíduo (QMR). Deve-se notar também, que na Tabela 5 os modelos Gompertz, Logístico e de von Bertalanffy apresentaram taxas mais realistas para o parâmetro A, enquanto que os modelos Michaelis-Menten e Brody convergiram para valores muito acima daqueles esperados para a raça Hereford.

Tabela 4 Critérios de informação para os modelos ajustados pelos métodos dos quadrados mínimos ordinários (QMO), quadrados mínimos ponderados (QMP) e quadrados mínimos ponderados com erros autoregressivos de primeira ordem (QMPG-AR1).

Método	Modelo	AIC	AICc	BIC
QMO	Brody	$147,82$	$153,28$	$151,16$
	Gompertz	$140,02$	$145,47$	$143,35$
	Michaelis-Menten	$147,99$	$151,32$	$150,49$
	Logístico	$150,42$	$155,88$	$153,76$
	von Bertalanffy	$138,45$	$143,90$	$141,78$
QMP	Brody	$149,21$	$154,67$	$153,38$
	Gompertz	$140,22$	$145,67$	$144,38$
	Michaelis-Menten	$147,57$	$150,91$	$150,91$
	Logístico	$151,71$	$157,17$	$155,88$
	von Bertalanffy	$138,93$	$144,39$	$143,10$
QMPG-AR1	Brody	$138,24$	$143,70$	$147,24$
	Gompertz	$137,96$	$143,42$	$146,96$
	Michaelis-Menten	$138,53$	$141,86$	$146,69$
	Logístico	$148,06$	$153,52$	$157,06$
	von Bertalanffy	$134,45$	$139,91$	$143,45$

Os erros padrões para cada uma das estimativas encontram-se na Tabela 6.

O modelo de von Bertalanffy, selecionado pelos três critérios, e ajustado aos dados de crescimento de novilhos Hereford encontra-se na Figura 7.

Tabela 5 Estimativas dos parâmetros A, B e K, para os modelos ajustados.

Método	Modelo	Parâmetros		
		A	B	K
QMO	Brody	$2197,93$	$0,9915$	$0,00049$
	Gompertz	$785,61$	$2,8487$	$0,00377$
	Michaelis-Menten	$3129,71$	$2648,0540$	
	Logístico	$692,05$	$8,8570$	$0,00653$
	von Bertalanffy	$867,47$	$0,6587$	$0,00278$
QMP	Brody	$1771,33$	$0,9951$	$0,00065$
	Gompertz	$769,51$	$2,9013$	$0,00392$
	Michaelis-Menten	$2888,91$	$2400,9308$	
	Logístico	$707,29$	$8,2126$	$0,00616$
	von Bertalanffy	$955,16$	$0,6641$	$0,00249$
QMPG-AR1	Brody	$2463,76$	$0,9798$	$0,00037$
	Gompertz	$772,58$	$2,9010$	$0,00391$
	Michaelis-Menten	$3857,50$	$3403,6322$	
	Logístico	$742,20$	$6,7984$	$0,00556$
	von Bertalanffy	$871,33$	$0,6584$	$0,00277$

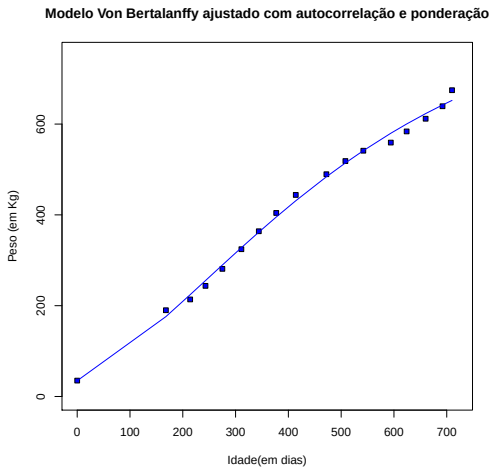

Figura 7 Ajuste da curva de crescimento para o modelo de von Bertalanffy ponderado e autoregressivo

Tabela 6 Erros padrões para as estimativas dos parâmetros A, B e K, para os modelos ajustados.

Método	Modelo	Parâmetros		
		A	B	K
QMO	Brody	$756,3136$	$0,0050$	$0,00021$
	Gompertz	$31,4747$	$0,1364$	$0,00029$
	Michaelis-Menten	$599,3401$	$610,6376$	
	Logístico	$22,2587$	$0,9963$	$0,00049$
	Von Bertalanffy	$44,8319$	$0,0221$	$0,00024$
QMP	Brody	$517,7133$	$0,0092$	$0,00023$
	Gompertz	$27,3214$	$0,1064$	$0,00024$
	Michaelis-Menten	$461,9862$	$476,8020$	
	Logístico	$24,3366$	$0,9538$	$0,00050$
	Von Bertalanffy	$41,0799$	$0,0155$	$0,00020$
QMPG-AR1	Brody	$4904,2163$	$0,0082$	$0,00022$
	Gompertz	$31,5816$	$0,1326$	$0,00029$
	Michaelis-Menten	$1229,2895$	$1276,7966$	
	Logístico	$57,5605$	$2,0511$	$0,00108$
	Von Bertalanffy	$51,6114$	$0,0219$	$0,00026$

6 CONCLUSÃO

Nas condições do presente estudo, e de acordo com os resultados obtidos, pode-se concluir que:

Para os modelos de crescimento estudados, o AIC e o AICc apresentaram desempenho similar em todos os modelos simulados, sendo que este desempenho foi igual ou superior ao desempenho do BIC, exceto para o modelo de Michaelis-Menten.

Deve-se notar também que, para a aplicação em dados reais, todos os modelos indicaram o mesmo modelo (von Bertalanffy), como sendo o que melhor se ajusta aos dados.

As funções de Brody e Michaelis-Menten não apresentam um bom ajuste aos dados.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control,** Notre Dame, v. 19, n. 6, p. 717-723, 1974.

ASSOCIAÇÃO BRASILEIRA DE HEREFORD E BRAFORD. **Hereford**: carne de qualidade tipo exportação. Disponível em: <http://www.hereford.com.br/?bW9kdWxvPTEmbWVudT01NiZhcnF1aXZvP-WNvbnRldWRvLnBocA==>. Acesso em: 19 jul. 2013.

BERTALANFFY, L. V. Quantitative laws in metabolism and growth. **The Quarterly Review of Biology**, New York, v. 32, n. 3, p. 217-231, 1957.

BRACCINI NETO, J. et al. Análise de curvas de crescimento de aves de postura. **Revista Brasileira de Zootecnia**, Brasília, v. 25, n. 6, p. 1062-1073, 1996.

BRODY, S. **Bioenergetics and growth**. New York: Reinhold. 1945. 1023 p.

BROWN, J. E.; FITZHUGH JUNIOR, H. A.; CARTWRIGHT, T. C. A comparison of nonlinear models for describing weight-age relationships in cattle. **Journal of Animal Science**, Champaing, v. 42, n. 4, p. 810-818, 1976.

BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: understanding aic and bic in model selection. **Sociological Methods and Research,** Beverly Hills, v. 33, n. 2, p. 261-304, 2004.

DENISE, R. S. K.; BRINKS, J. S. Genetic and environmental aspects of the growth curve parameters in beef cows. **Journal of Animal Science**, Champaing, v. 61, n. 6, p. 1431-1440, 1985.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3rd ed. New York: J. Wiley, 1998. 706 p.

DUARTE, F. A. M. **Estudo da curva de crescimento de animais da raça "Nelore" (Bos taurus indicus) através de cinco modelos estocásticos**. 1975. 284 p. Tese (Livre Docência em Genética e Matemática Aplicada à Biologia) - Universidade de São Paulo, Ribeirão Preto, 1975.

ELIAS, A. M. **Análise de curvas de crescimento de vacas das raças Nelore, Guzerá e Gir**. 1998. 128 p. Dissertação (Mestrado em Ciência Animal e Pastagens) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 1998.

FITZHUGH JUNIOR, H. A. Analysis of growth curves and strategies for altering their shape. **Journal of Animal Science**, Champaing, v. 42, n. 4, p. 1036-1051, 1976.

GALLANT, A. R. **Nonlinear statistical models**. New York: J. Wiley, 1987. 610 p.

GOTTSCHALL, C. S. Impacto nutricional na produção de carne-curva de crescimento. In: LOBATO, J. F. P.; BARCELLOS, J. O. J.; KESSLER, A. M. (Ed.). **Produção de bovinos de corte**. Porto Alegre: EDIPUCRS, 1999. p. 169-192.

HOFFMANN, R.; VIEIRA, S. **Análise de regressão:** uma introdução à econometria. 3. ed. São Paulo: HUCITEC, 1998. 379 p.

KROLL, L. B. **Estudo do crescimento de vacas leiteiras através de modelos com autocorrelação nos erros**. 1990. 96 p. Tese (Doutorado em Energia na Agricultura) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Botucatu, 1990.

KSHIRSAGAR, A. M.; SMITH, W. B. **Growth curves**. New York: M. Dekker, 1995. 359 p.

LAIRD, A. K.; HOWARD, A. Growth curves in inbred mice. **Nature**, New York, v. 213, n. 5078, p. 786-788, 1967.

MAZUCHELI, J.; ACHCAR, J. Algumas considerações em regressão não linear. **Acta Scientiarum**, Maringá, v. 24, n. 6, p. 1761-1770, 2002. Disponível em:
<http://www.periodicos.uem.br/ojs/index.php/ActaSciTechnol/article/view/25-51/1574>. Acesso em: 19 jul. 2013.

MAZZINI, A. R. A. **Estudo da curva de crescimento de machos hereford**. 2001. 47 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2001.

MAZZINI, A. R. A. et al. Growth curve for hereford males: heterocedasticity and autoregressives residuals. **Ciência Rural**, Santa Maria, v. 35, n. 2, p. 422-427, mar./abr. 2005.

MEDEIROS, H. A. et al. Avaliação da qualidade do ajuste da função logística monofásica com estrutura de erros independentes e autorregressivos através de simulação. **Ciência e Agrotecnologia**, Lavras, v. 24, n. 4, p. 973-985, jul./ago. 2000.

MENDES, P. N. et al. Modelo logístico difásico no estudo do crescimento de fêmeas da raça Hereford. **Ciência Rural**, Santa Maria, v. 38, n. 7, p. 1984-1990, 2008.

MICHAELIS, L.; MENTEN, M. L. Die Kinetik der Invertinwirkung. **Biochemische Zeitschrift**, Berlin, v. 49, n. 1, p. 333-369, 1913.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: E. Blücher, 2006. 538 p.

NELDER, J. A. The fitting of a generalization of the logistic curve. **Biometrics**, Washington, v. 17, n. 1, p. 89-94, 1961.

NETER, J.; WASSERMAN, W.; KUTNER, M. H. **Applied linear statistical models:** regression, analysis of variance, and experimental designs. 2nd ed. Homewood: R. D. Irwin, 1985. 1127 p.

PASTERNAK, H.; SHALEV, B. A. The effect of a feature of regression disturbance on the efficiency of fitting growth curves. **Growth, Development & Aging**, Bar Harbor, v. 58, n. 1, p. 33-39, 1994.

PEARL, R.; REED, L. J. On the rate of growth of the population of the United States since 1790 and its mathematical representation. **Proceedings of the National Academy of Sciences**, Washington, v. 6, n. 6, p. 275-88, 1920.

R DEVELOPMENT CORE TEAM. **R:** a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <http://www.r-project.org>. Acesso em: 11 maio 2013.

RATKOWSKY, D. A. Principles of nonlinear regression modeling. **Journal of Industrial Microbiology**, Berlin, v. 12, n. 3/5, p. 195-199, 1993.

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, Hayward, v. 6, n. 2, p. 461-464, 1978.

SILVA, N. A. M. **Seleção de modelos de regressão não lineares e aplicação do algoritmo SAEM na avaliação genética de curvas de crescimento bovinos de nelore**. 2010. 58 p. Tese (Doutorado em Zootecnia) - Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

SILVA, N. A. M. et al. Curvas de crescimento e influência de fatores não genéticos sobre as taxas de crescimento de bovinos da raça Nelore. **Ciência e Agrotecnologia**, Lavras, v. 28, n. 3, p. 647-654, maio/jun. 2004.

SILVEIRA, F. G. **Classificação multivariada de modelos de crescimento para grupos genéticos de ovinos de corte.** 2010. 59 p. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa, MG, 2010.

SILVEIRA, F. G. et al. Análise de agrupamento na seleção de modelos de regressão não-lineares para curvas de crescimento de ovinos cruzados. **Ciência Rural**, Santa Maria, v. 41, n. 4, p. 692-698, 2011.

SOUZA, G. S. **Introdução aos modelos de regressão linear e não-linear**. Brasília: EMBRAPA, 1998. 489 p.

SUGIURA, N. Further analysts of the data by akaike's information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, Ontario, v. 7, n. 1, p. 13-26, 1978.

TEDESCHI, L. O. **Determinação dos parâmetros da curva de crescimento de animais da Raça Guzerá e seus cruzamentos alimentados a pasto, com e sem suplementação.** 1996. 140 p. Dissertação (Mestrado em Ciência Animal e Pastagens) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 1996.

TEDESCHI, L. O. et al. Estudo da curva de crescimento de animais da raça Guzerá e seus cruzamentos alimentados a pasto, com e sem suplementação: 1., análise e seleção das funções não-lineares. **Revista Brasileira de Zootecnia**, Brasília, v. 29, n. 2, p. 630-637, mar./abr. 2000.

VERHULST, P. F. Recherches mathématiques sur la loi d'accroissement de la population. **Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles**, Bruxelles, v. 18, n. 1, p. 1-45, 1845.

APÊNDICE

Página


```
require(MASS); require(nlme); require(lattice)
require(VGAM); require(qpcR)
u=2
num=50
menoraic<-c(rep(0,u)); menoraicc<-c(rep(0,u));
menorbic<-c(rep(0,u))
contaic<-0;contaicc<-0;contbic<-0
contaic1<-0#MicaMen
contaic2<-0#Gompertz
contaic3<-0#Logistico
contaic4<-0#VonBerta
contaic5<-0#Brody
contaicc1<-0;contaicc2<-0;contaicc3<-0;contaicc4<-0;
contaicc5<-0;
contbic1<-0;contbic2<-0;contbic3<-0;contbic4<-0;contbic5<-0
#Número de parâmetros
paraBrody=3+1+1 #1devido a phi de AR(1) e outro a sigma^2
paraVonBerta=3+1+1
paraLogistico=3+1+1
paraGompertz=3+1+1
paraMicaMen=2+1+1
#
Brody<-matrix(rep(0,num*u),u,num)
VonBerta<-matrix(rep(0,num*u),u,num)
Gompertz<-matrix(rep(0,num*u),u,num)
MicaMen<-matrix(rep(0,num*u),u,num)
Logistico<-matrix(rep(0,num*u),u,num)
matrizaic<-matrix(rep(0,num*u),u,num)
matrizbic<-matrix(rep(0,num*u),u,num)
#
AICBrody<-matrix(rep(0,num*u),u,num)
AICVonBerta<-matrix(rep(0,num*u),u,num)
AICLogistico<-matrix(rep(0,num*u),u,num)
AICGompertz<-matrix(rep(0,num*u),u,num)
AICMicaMen<-matrix(rep(0,num*u),u,num)
menorAIC<-matrix(rep(0,num*u),u,num)
#
AICBrodyc<-matrix(rep(0,num*u),u,num)
AICVonBertac<-matrix(rep(0,num*u),u,num)
AICLogisticoc<-matrix(rep(0,num*u),u,num)
```

```
AICGompertzc<-matrix(rep(0,num*u),u,num)
AICMicaMenc<-matrix(rep(0,num*u),u,num)
menorAICc<-matrix(rep(0,num*u),u,num)
#
BICBrody<-matrix(rep(0,num*u),u,num)
BICVonBerta<-matrix(rep(0,num*u),u,num)
BICLogistico<-matrix(rep(0,num*u),u,num)
BICGompertz<-matrix(rep(0,num*u),u,num)
BICMicaMen<-matrix(rep(0,num*u),u,num)
menorBIC<-matrix(rep(0,num*u),u,num)
for(g in 1:u){
##########################################################
animal <- 1:num
tempo <- seq(1,(2*365),by=60) # distâncias uniformes = lag
da <- data.frame(y=rep(0,(length(tempo))*(length(animal))),
 animal=animal, tempo=rep(tempo, each=length(animal)))
da1 <- da[order(da[ ,2]), ]
rownames(da1)<-NULL
a <- 786+rnorm(1,0,30);
b<-3.06+rnorm(1,0,0.07);
k=0.0042;
n=(length(tempo))*(length(animal))
e=matrix(0,length(animal),length(tempo))
for(i in 1:length(animal))
{
 e[i,]=arima.sim(n = length(tempo), list(ar = c(-0.5897)),
    sd = sqrt(100))
}
e1=matrix(t(e),n,1,byrow=T)
########
#da1$y <- a*(1-b*exp(-da1$tempo*k))+e1####Brody
#da1$y<-a*(1-b*exp((-tempo*k)))^3+e1###Von Bertalanffy
#da1$y<-a*(1+b*exp((-tempo*k)))^(-1)+e1###Logistico
da1$y<-a*(exp(-b*exp((-tempo*k))))+e1###Gompertz
#da1$y<-a*tempo/(b+tempo)###MicaMem
da1$y
da1=data.frame(da1)
for(u in 1:n){
 if (da1$y[u]<0) {
da1$y[u]<-10+rnorm(1,0,3)
    }
}
##
for(h in 1:length(animal)){
#Brody#
Brody[g,h]<- suppressWarnings(gnls(y~a1*(1-b1*exp(-tempo*k1)),
data=da1[da1[,2]==h,],
            start=list(a1=1400, b1=1,k1=0.005),
```

```
            corr=corAR1(form=~1|animal),
            control=list(returnObject=TRUE))$logLik)
#Von Bertalanffy#
VonBerta[g,h]<-suppressWarnings(gnls(y~al*(1-bl*exp(
-tempo*kl))^3,data=da1[da1[,2]==h,],start=list(al=1000,
 bl=1,kl=0.005),              corr=corAR1(form=~1|animal),
  control=list(returnObject=TRUE))$logLik)
#Logistico#
Logistico[g,h]<- suppressWarnings(gnls(y~al*(1+bl*exp((
-tempo*kl)))^(-1), data=da1[da1[,2]==h,],start=list(
al=350, bl=6,kl=0.005), corr=corAR1(form=~1|animal),
control=list(returnObject=TRUE))$logLik)
#Gompertz#
Gompertz[g,h]<-suppressWarnings(gnls(y~al*(exp(-bl*exp((
-tempo*kl)))), data=da1[da1[,2]==h,],start=list(al=600,
 bl=2,kl=0.005),corr=corAR1(form=~1|animal),
  control=list(returnObject=TRUE))$logLik)
#Michaelis-Menten#
fit = vglm( da1[da1[,2]==h,]$y~ 1, micmen,  da1[da1[,2]==h,],
    trace=F, crit="c", form2 = ~ tempo - 1)
att=fit@coefficients[1]
att=as.numeric(att)
btt=fit@coefficients[2]
btt=as.numeric(btt)
MicaMen[g,h] <- suppressWarnings(gnls(y~al*tempo/(
bl+tempo), data= da1[da1[,2]==h,], start=list(al=att,
 bl=btt), corr=corAR1( form=~1|animal), control=list(
 returnObject = TRUE))$logLik)
#AIC#
AICBrody[g,h]=-2*Brody[g,h]+2*paraBrody
AICVonBerta[g,h]=-2*VonBerta[g,h]+2*paraVonBerta
AICLogistico[g,h]=-2*Logistico[g,h]+2*paraLogistico
AICGompertz[g,h]=-2*Gompertz[g,h]+2*paraGompertz
AICMicaMen[g,h]=-2*MicaMen[g,h]+2*paraMicaMen
#AICc#
AICBrodyc[g,h]=AICBrody[g,h]+2*paraBrody*(paraBrody+1)/(
    n-paraBrody-1)
AICVonBertac[g,h]=AICVonBerta[g,h]+2*paraVonBerta*(
paraVonBerta+1)/(n-paraVonBerta-1)
AICLogisticoc[g,h]=AICLogistico[g,h]+2*paraLogistico*(
    paraLogistico+1)/(n-paraLogistico-1)
AICGompertzc[g,h]=AICGompertz[g,h]+2*paraGompertz*(
paraGompertz+1)/(n-paraGompertz-1)
AICMicaMenc[g,h]=AICMicaMen[g,h]+2*paraMicaMen*(
paraMicaMen+1)/(n-paraMicaMen-1)
#BIC#
BICBrody[g,h]=-2*Brody[g,h]+paraBrody*log(n)
BICVonBerta[g,h]=-2*VonBerta[g,h]+paraVonBerta*log(n)
```

```
BICLogistico[g,h]=-2*Logistico[g,h]+paraLogistico*log(n)
BICGompertz[g,h]=-2*Gompertz[g,h]+paraGompertz*log(n)
BICMicaMen[g,h]=-2*MicaMen[g,h]+paraMicaMen*log(n)
################################################
menorAIC[g,h]<-min(AICBrody[g,h], AICVonBerta[g,h],
    AICLogistico[g,h],AICGompertz[g,h],AICMicaMen[g,h])
menorAICc[g,h]<-min(AICBrodyc[g,h], AICVonBertac[g,h],
    AICLogisticoc[g,h],AICGompertzc[g,h],AICMicaMenc[g,h])
menorBIC[g,h]<-min(BICBrody[g,h], BICVonBerta[g,h],
    BICLogistico[g,h],BICGompertz[g,h],BICMicaMen[g,h])
#AIC
    {if (menorAIC[g,h]==AICBrody[g,h])
    contaic5<-contaic5+1
    }
    {if (menorAIC[g,h]==AICVonBerta[g,h])
    contaic4<-contaic4+1
    }
    {if (menorAIC[g,h]==AICLogistico[g,h])
    contaic3<-contaic3+1
    }
    {if (menorAIC[g,h]==AICGompertz[g,h])
    contaic2<-contaic2+1
    }
    {if (menorAIC[g,h]==AICMicaMen[g,h])
    contaic1<-contaic1+1
    }
#AICc
    {if (menorAICc[g,h]==AICBrodyc[g,h])
    contaicc5<-contaicc5+1
    }
    {if (menorAICc[g,h]==AICVonBertac[g,h])
    contaicc4<-contaicc4+1
    }
    {if (menorAICc[g,h]==AICLogisticoc[g,h])
    contaicc3<-contaicc3+1
    }
    {if (menorAICc[g,h]==AICGompertzc[g,h])
    contaicc2<-contaicc2+1
    }
    {if (menorAICc[g,h]==AICMicaMenc[g,h])
    contaicc1<-contaicc1+1
    }
#BIC
    {if (menorBIC[g,h]==BICBrody[g,h])
    contbic5<-contbic5+1
    }
    {if (menorBIC[g,h]==BICVonBerta[g,h])
    contbic4<-contbic4+1
```

```
    }
    {if (menorBIC[g,h]==BICLogistico[g,h])
    contbic3<-contbic3+1
    }
    {if (menorBIC[g,h]==BICGompertz[g,h])
    contbic2<-contbic2+1
    }
    {if (menorBIC[g,h]==BICMicaMen[g,h])
    contbic1<-contbic1+1
    }
}
}
x=matrix(c(contaic1,contaic2,contaic3,contaic4,contaic5,
 contaicc1,contaicc2,contaicc3,contaicc4,contaicc5,
 contbic1,contbic2,contbic3,contbic4,contbic5),15,1)
row=c("AIC MMen","AIC Gomp","AIC Logi","AIC Von B","AIC Brod",
 "AICc MMen","AICc Gomp","AICc Logi","AICc Von B","AICc Brod",
 "BIC MMen","BIC Gomp","BIC Logi","BIC Von B","BIC Brod")
col=c("Gompertz")
rownames(x)<-row
colnames(x)=col
x
```