

**Cr terios de informa o e sele o
de modelos lineares mistos**

Rodrigo Marques da Cruz

DISSERTA O APRESENTADA AO
INSTITUTO DE MATEM TICA E ESTAT STICA
DA UNIVERSIDADE DE S O PAULO
PARA OBTEN O DO T TULO DE
MESTRE EM CI NCIAS

Programa: Estat stica

Orientador: Prof. Dr. Julio da Motta Singer

S o Paulo, julho de 2020

Cr terios de informa o e sele o de modelos lineares mistos

Esta vers o da disserta o cont m as corre es e altera es sugeridas pela Comiss o Julgadora durante a defesa da vers o original do trabalho, realizada em 17/06/2020. Uma c pia da vers o original est  dispon vel no Instituto de Matem tica e Estat stica da Universidade de S o Paulo.

Comiss o Julgadora:

- Prof. Dr. Julio da Motta Singer (orientador) - IME-USP
- Prof. Dr. Juv ncio Santos Nobre - UFC
- Prof^a. Dr^a. Luzia Aparecida Trinca - UNESP

*The most merciful thing in the world, I think,
is the inability of the human mind to correlate all its contents..*

The Call of Cthulhu, H. P. Lovecraft

Agradecimentos

Em primeiro lugar, gostaria de agradecer a todos os professores responsáveis por minha formação até aqui. Em especial, à professora Elisabeti Kira, pela inspiradora conversa no corredor do IME-USP que resultou na minha transferência para o Bacharelado em Estatística, e ao professor Julio Singer, meu orientador, por todas as aulas, ensinamentos, correções e paciência. Sou grato também aos professores Luis Gustavo Esteves, Julia Pavan, Carmen André e Lucia Barroso por toda gentileza durante minha graduação no IME-USP. Por fim, gostaria de agradecer ao professor Vladimir Belitsky pelo grande apoio durante o começo da pós-graduação. Para mim, é uma honra e um privilégio poder compartilhar do conhecimento de pessoas tão brilhantes e inspiradoras!

Agradeço também ao apoio dos meus amigos Leandro, Rafael, André e Deivid, aos amigos da faculdade (Henrique, Veronica, Davi e Alexandre), aos amigos do trabalho (Daiane, Joyce, Paola, Fabio e Miguel), que me acompanharam em momentos de felicidade, alegria, estresse e ansiedade. Também não poderia deixar de agradecer à Beatriz Molina, minha psicóloga, pelo ótimo trabalho desenvolvido e por me ensinar a não tomar decisões em momentos de crise.

Ao Jorge, meu melhor amigo, só tenho palavras de agradecimento por todos esses anos de companheirismo. Você esteve comigo nos momentos felizes e nos momentos tristes, me ajudou a seguir adiante quando o cansaço era insuportável, mostrando que família vai muito além de laços sanguíneos. E ainda por cima me ajudou muito nas matérias de Computação!

Por fim, preciso agradecer ao apoio e confiança da minha família, em especial de minha mãe que, apesar da simplicidade e de algumas dificuldades, foi capaz de me prover cuidados e educação dignamente. Sem contar a excelente carga genética e o exemplo de pessoa batalhadora e inspiradora que sempre me fez seguir adiante!

Graças a todos vocês e meu empenho, persistência e resiliência, acordo todos os dias orgulhoso por ser estatístico.

Rodrigo Marques da Cruz

São Paulo, julho de 2020

Resumo

CRUZ, R. M. **Critérios de informação e seleção de modelos lineares mistos**. 2020. 44 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

O modelo linear misto é amplamente utilizado na análise de medidas repetidas e de dados longitudinais, e compreende duas abordagens. A primeira é apropriada quando estamos interessados em parâmetros populacionais (*efeitos fixos*) e a segunda é indicada quando os coeficientes individuais (*efeitos aleatórios*) são de interesse. A seleção de modelos mistos, quando realizada por meio de critérios de informação, leva em conta estas diferenças de abordagem: o Critério de Informação de Akaike (*AIC*) marginal baseia-se na log-verossimilhança marginal, e o Critério de Informação de Schwarz (*BIC*) é similar ao critério anterior, mas também inclui o número total de observações. O *AIC* condicional baseia-se na log-verossimilhança condicional aos efeitos aleatórios. Realizamos um estudo de simulação para observar o comportamento destes critérios frente a diversos cenários, concluindo que o *AIC* (marginal ou condicional) e o *BIC* apresentaram frequência de seleção do modelo correto elevada mesmo para tamanho amostral reduzido ($n = 10$), desde que a quantidade de observações por indivíduo seja elevada ($m \geq 13$). Além disso, quanto mais correlacionados os dados, são necessários tanto tamanho amostral quanto número de observações por indivíduo maiores ($n \geq 30$ e $m \geq 16$, respectivamente) para detecção do modelo correto. Consideramos dois exemplos em que ilustramos a aplicação do *AIC* marginal e do *AIC* condicional. Mostramos que a seleção de modelos por intermédio dos critérios *AIC* e *BIC* é compatível com a análise de resíduos associada.

Palavras-chave: *AIC*, *BIC*, dados longitudinais.

Abstract

CRUZ, R. M. **Information criteria and linear mixed model selection.** 2020. 44 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

Linear mixed models are widely used in repeated measures and longitudinal data analysis, and comprehend two distinct approaches. The first is adequate when the interest lies on populational-averaged parameters (*fixed effects*) and the second is adequate when the subject-specific coefficients (*random effects*) are of interest. The selection of mixed models, when accomplished via information criteria, takes these different approaches into account: the marginal Akaike Information Criterion (*AIC*) is based on the marginal log-likelihood, and the Schwarz Information Criterion (*BIC*) is similar to the aforementioned criterion, but also includes the total number of observations. The conditional *AIC* is based on the conditional log-likelihood on the random effects. We performed a simulation study to observe the behavior of these criteria in various scenarios, concluding that the *AIC* (marginal ou conditional) and the *BIC* presented a high frequency of selection of the correct model even for small sample size ($n = 10$), provided that the number of individual observations is large ($m \geq 13$). Besides that, the higher the correlation in the data, the higher the need for both the sample size and the number of individual observations to be large ($n \geq 30$ e $m \geq 16$, respectively) in order to detect the correct model. We considered two examples in which we illustrate the application of the marginal *AIC* and the conditional *AIC*. We showed that the model selection through the *AIC* and the *BIC* is compatible with the associated residual analysis.

Keywords: *AIC, BIC, longitudinal data.*

Sumário

1	Introdução	1
2	O modelo linear misto	3
3	Critério de Informação de Akaike	7
3.1	<i>AIC</i> marginal	10
3.2	<i>AIC</i> condicional	11
4	Critério de Informação de Schwarz	15
5	Simulação	19
5.1	Modelo misto com erros aleatórios independentes	20
5.2	Modelo misto com erros aleatórios autorregressivos	23
6	Exemplos	27
6.1	Dados de crescimento de bezerros	27
6.2	Dados de palatos de sapos	33
7	Conclusão	39
A	Tabelas	41
B	Figuras	45
C	Graus de liberdade efetivos	53
D	Códigos	55
	Referências	71

Capítulo 1

Introdução

O modelo linear misto é amplamente utilizado na análise de medidas repetidas e de dados longitudinais, situações que fogem do cenário clássico em que as observações são independentes entre si. Por exemplo, em um estudo analisado por [Singer et al. \(2017\)](#), um conjunto de bezerros recém-nascidos foi submetido a uma dieta experimental por um período de 26 semanas após o nascimento e, a cada 2 semanas, o peso de cada bezerro foi mensurado. Admitindo que os bezerros representam uma amostra aleatória avaliada em diversos instantes, o modelo misto é apropriado para, por exemplo, estudar o comportamento do valor esperado do peso (populacional ou de algum bezerro específico) ao longo do tempo. Em outras palavras, este tipo de modelo tem a capacidade de descrever o comportamento longitudinal tanto da resposta populacional (*population averaged*) quanto da específica (*subject-specific*).

Modelos mistos apresentam determinadas características, como estruturas de covariância mais complexas que permitem representar a possível dependência entre as observações da mesma unidade com um número menor de parâmetros e a possibilidade de previsão de efeitos aleatórios, que os tornam fundamentalmente diferentes de modelos lineares com apenas efeitos fixos. A resposta esperada é modelada por meio dos efeitos fixos (*populacionais*) e a estrutura de covariância é modelada por meio dos efeitos aleatórios (*condicionais* ou *específicos*), ou seja, esse tipo de modelo envolve a especificação tanto da estrutura de média quanto de covariância.

A estimação dos parâmetros é frequentemente concretizada por meio do método de máxima verossimilhança e de máxima verossimilhança restrita para os componentes de variância [[Demidenko \(2013\)](#)]. Após a estimação dos parâmetros, é necessário verificar se o modelo proposto é compatível com os dados, tanto no que diz respeito à resposta média quanto à estrutura de covariância. Dentre as principais técnicas diagnósticas disponíveis, destacam-se a análise de resíduos marginais, condicionais ou de efeitos aleatórios, além da análise de influência, por meio de pontos alavanca e distância de Cook. Mais detalhes sobre esse tópico são fornecidos em [Nobre \(2004\)](#), [Nobre & Singer \(2007\)](#), [Nobre & Singer \(2011\)](#) e [Singer et al. \(2017\)](#). Além disso, ferramentas descritivas, como gráficos de perfis médios e de variâncias amostrais, são instrumentos úteis para escolher estruturas para efeitos fixos e covariâncias [[Rocha & Singer \(2018\)](#)].

Além das técnicas de diagnóstico, a seleção de modelos mistos pode ser realizada por meio de critérios de informação, como o de Akaike (1973) e o de Schwarz (1978) — *AIC* e *BIC*, respectivamente. O *AIC marginal* (com base na distribuição marginal da variável resposta) é utilizado para seleção de modelos mistos em que o interesse está no valor esperado da distribuição marginal. Este critério depende tanto da verossimilhança da variável resposta quanto de um termo de penalização que, em geral, está relacionado ao número de parâmetros desconhecidos do modelo. Por outro lado, para a seleção de modelos mistos em que o interesse está relacionado à variação da resposta individual (modelos condicionais), Vaida & Blanchard (2005) sugerem a utilização do que denominaram *AIC condicional*, que se baseia na verossimilhança da variável resposta condicionada aos efeitos aleatórios e em um termo de penalização que depende do que foi definido por Hodges & Sargent (2001) como *graus de liberdade efetivos do modelo*.

Também é possível se utilizar o critério de informação bayesiano *BIC* para seleção de modelos mistos. Apesar de diferenças conceituais, como a atribuição de probabilidades a priori para os modelos, este critério tem expressão similar ao *AIC*, uma vez que envolve a verossimilhança da variável resposta e um termo de penalização que depende do número de parâmetros do modelo — além do logaritmo do número total de observações. Na prática, os pacotes computacionais informam ambos os critérios e é comum utilizá-los em conjunto para se decidir qual modelo selecionar.

O objetivo desta dissertação é fazer uma revisão dos critérios de informação *AIC* e *BIC*, utilizados como possíveis ferramentas de seleção de modelos lineares mistos (em conjunto com técnicas de diagnóstico), e ilustrar situações práticas em que tais critérios são indicados e situações em que é preferível utilizar outras técnicas, como análise de diagnóstico. A importância deste trabalho está relacionada não apenas à revisão teórica, mas também à maneira como os critérios *AIC* e *BIC* devem ser utilizados, na prática, para selecionar modelos mistos, uma vez que, para este tipo de modelo, alguns fatores influenciam a qualidade da seleção via *AIC* e *BIC* — dentre eles, o tamanho amostral e a quantidade de observações por indivíduo, além do grau de variabilidade e da correlação entre as observações.

No Capítulo 2 descrevemos brevemente o modelo linear misto em sua formulação mais clássica, além de métodos de estimação de parâmetros e as principais ferramentas utilizadas na análise de diagnóstico, como resíduos marginais e condicionais. No Capítulo 3, fornecemos uma noção conceitual sobre o *AIC* e discutimos, com mais detalhes, sobre os critérios *AIC marginal* e *AIC condicional*. No Capítulo 4 apresentamos uma breve introdução sobre o *BIC* e sua formulação clássica. No Capítulo 5 conduzimos um estudo de simulação para observar o comportamento e a qualidade dos dois critérios na seleção de modelos com diferentes efeitos aleatórios e estruturas de covariância. No Capítulo 6 apresentamos duas aplicações dos critérios de informação na seleção de modelos mistos para dados de crescimento de bezerros [Singer et al. (2017)] e para dados de velocidade relativa de transporte mucociliar do palato de sapos [Rocha & Singer (2018)]. Por fim, no Capítulo 7, concluímos com uma discussão sobre as técnicas utilizadas. Resultados técnicos mais detalhados são apresentados no Apêndice.

Capítulo 2

O modelo linear misto

O modelo linear misto pode ser expresso como:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i, \quad (2.1)$$

$i = 1, 2, \dots, n$, em que \mathbf{y}_i ($m_i \times 1$) é o vetor de respostas observadas para a i -ésima unidade amostral, \mathbf{X}_i ($m_i \times p$) e \mathbf{Z}_i ($m_i \times s_i$) são matrizes de especificação (de posto completo) dos efeitos fixos e aleatórios, respectivamente, $\boldsymbol{\beta}$ ($p \times 1$) é o vetor de efeitos fixos, \mathbf{b}_i ($s_i \times 1$) é o vetor de efeitos aleatórios associado à i -ésima unidade amostral e \mathbf{e}_i ($m_i \times 1$) é o vetor de erros aleatórios, com $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}_{s_i}, \mathbf{G}_i)$ e $\mathbf{e}_i \sim \mathcal{N}(\mathbf{0}_{m_i}, \mathbf{R}_i)$ independentes. Em geral, o número de efeitos aleatórios por indivíduo (s_i) é constante.

Na forma compacta, o modelo (2.1) pode ser escrito como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}, \quad (2.2)$$

em que $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top$, $\mathbf{Z} = \oplus_{i=1}^n \mathbf{Z}_i$, $\mathbf{b} = (\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top)^\top$ e $\mathbf{e} = (\mathbf{e}_1^\top, \dots, \mathbf{e}_n^\top)^\top$, com $N = \sum_{i=1}^n m_i$ e $s = \sum_{i=1}^n s_i$, e o operador \oplus representa a soma direta de matrizes [Harville (2011)].

Seja $\boldsymbol{\tau} = (\boldsymbol{\gamma}^\top, \boldsymbol{\delta}^\top)^\top$ o vetor ($q \times 1$) com parâmetros associados a $\mathbf{G}(\boldsymbol{\gamma}) = \mathbf{G} = \oplus_{i=1}^n \mathbf{G}_i$ ($s \times s$) e $\mathbf{R}(\boldsymbol{\delta}) = \mathbf{R} = \oplus_{i=1}^n \mathbf{R}_i$ ($N \times N$), e $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top)^\top$ o conjunto dos $k = p + q$ parâmetros do modelo (2.2).

A suposição distribucional adotada, nomeadamente,

$$\begin{pmatrix} \mathbf{b} \\ \mathbf{e} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{0}_s \\ \mathbf{0}_N \end{pmatrix}, \begin{pmatrix} \mathbf{G} & \mathbf{0}_{s \times N} \\ \mathbf{0}_{N \times s} & \mathbf{R} \end{pmatrix} \right)$$

implica

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \quad (2.3)$$

em que $\mathbf{V} = \mathbf{V}(\boldsymbol{\tau}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R}$ é a matriz de covariâncias da variável resposta. Em geral, considera-se $\mathbf{R} = \sigma^2\mathbf{I}_N$, especificando o modelo de independência condicional homocedástico. O caso gaussiano é aquele mais utilizado na prática em virtude da relativa simplicidade de interpretação e da facilidade de implementação computacional. Para maiores detalhes, o leitor pode consultar [Diggle et al. \(2002\)](#), [Verbeke & Molenberghs \(2005\)](#), [Demidenko \(2013\)](#), entre outros.

O modelo misto permite a realização de inferências sob o enfoque populacional (*marginal*), por meio dos efeitos fixos, ou individual (*condicional*), por meio da previsão dos efeitos aleatórios. A abordagem adotada depende fundamentalmente do objetivo do estudo. No primeiro caso, o objetivo é modelar $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$, o valor esperado da variável resposta para a subpopulação com características definidas por \mathbf{X}_i e, no enfoque condicional, o objetivo é modelar $E(\mathbf{y}_i|\mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$, o valor esperado condicionado aos efeitos aleatórios, ou seja, a inferência diz respeito a unidades amostrais específicas.

A estimação dos parâmetros é frequentemente baseada no Estimador de Máxima Verossimilhança (*maximum likelihood estimator* ou *MLE*). É comum, ainda, o uso do *Estimador de Máxima Verossimilhança Restrita* (*restricted maximum likelihood estimator* ou *REML*) que reduz o viés de estimação de \mathbf{V} [[Patterson & Thompson \(1971\)](#)].

A função log-verossimilhança de \mathbf{y} , aqui denotada por *log-verossimilhança marginal de \mathbf{y}* , é

$$\mathcal{L}(\boldsymbol{\theta}) = \log f(\mathbf{y}|\boldsymbol{\theta}) = -\frac{1}{2} \left[N \log 2\pi + \log |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right]. \quad (2.4)$$

Do mesmo modo, a *log-verossimilhança condicional de \mathbf{y} , dado \mathbf{b}* , é

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{b}) = \log f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{b}) = -\frac{1}{2} \left[N \log 2\pi + \log |\mathbf{R}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})^\top \mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) \right]. \quad (2.5)$$

De fato, a quantidade (2.5) não é uma verossimilhança legítima, dado que \mathbf{b} é uma variável latente.

Considerando $\boldsymbol{\tau}$ conhecido, o *MLE* de $\boldsymbol{\beta}$, denotado por $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau})$, é

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}) = \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}. \quad (2.6)$$

O *MLE* de $\boldsymbol{\tau}$ é o vetor $\hat{\boldsymbol{\tau}}$ que maximiza a função de verossimilhança perfilada

$$-\frac{1}{2} \left[N \log 2\pi + \log |\mathbf{V}| + [\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\tau})]^\top \mathbf{V}^{-1}[\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\tau})] \right]. \quad (2.7)$$

Por meio de $\hat{\boldsymbol{\tau}}$ e $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\tau}})$, obtém-se $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\tau}}^\top \right)^\top$, o *MLE* de $\boldsymbol{\theta}$. Entretanto, como $\hat{\boldsymbol{\tau}}$ é enviesado para $\boldsymbol{\tau}$ ([Harville \(1977\)](#)), é dada preferência ao *REML* $\hat{\boldsymbol{\tau}}_R$ de $\boldsymbol{\tau}$. Nesse cenário, $\hat{\boldsymbol{\tau}}_R$ é o vetor que maximiza a função

$$-\frac{1}{2} \left[N \log 2\pi + \log |\mathbf{V}| + \log |\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}| + [\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\tau})]^\top \mathbf{V}^{-1}[\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\tau})] \right], \quad (2.8)$$

que é semelhante a (2.7), com a inclusão do termo de correção de viés $-\frac{1}{2} \log |\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}|$. Por meio de $\hat{\boldsymbol{\tau}}_R$ e $\hat{\boldsymbol{\beta}}_R = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\tau}}_R)$, obtém-se $\hat{\boldsymbol{\theta}}_R = \left(\hat{\boldsymbol{\beta}}_R^\top, \hat{\boldsymbol{\tau}}_R^\top \right)^\top$, o REML de $\boldsymbol{\theta}$.

Os efeitos aleatórios são previstos a partir da distribuição conjunta de \mathbf{y} e \mathbf{b} , por meio das chamadas *Equações de Henderson* [Henderson (1950)]:

$$\begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{y} \end{pmatrix}. \quad (2.9)$$

Supondo \mathbf{G} e \mathbf{R} conhecidas, o *melhor estimador linear não enviesado* (*best linear unbiased estimator* ou BLUE) de $\boldsymbol{\beta}$ e o *melhor preditor linear não enviesado* (*best linear unbiased predictor* ou BLUP) de \mathbf{b} são, respectivamente,

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}, \quad (2.10)$$

$$\hat{\mathbf{b}} = \mathbf{G} \mathbf{Z}^\top \mathbf{V}^{-1} \left(\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right) = \mathbf{G} \mathbf{Z}^\top \mathbf{W} \mathbf{y}, \quad (2.11)$$

em que $\mathbf{W} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{V}^{-1}$. Consequentemente,

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \text{Var}(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} \right)^{-1},$$

$$E(\hat{\mathbf{b}}) = \mathbf{0}, \quad \text{Var}(\hat{\mathbf{b}}) = \mathbf{G} \mathbf{Z}^\top \mathbf{W} \mathbf{Z} \mathbf{G}.$$

Na prática, como geralmente são desconhecidas, \mathbf{G} e \mathbf{R} são substituídas por $\hat{\mathbf{G}}$ e $\hat{\mathbf{R}}$ — obtidas preferencialmente por máxima verossimilhança restrita — e, nesse caso, (2.10) e (2.11) são denominados *BLUE e BLUP empíricos* (*EBLUE e EBLUP*, respectivamente). Para mais detalhes sobre o processo de estimação de parâmetros no contexto de modelos mistos, o leitor pode consultar Robinson (1991), Diggle et al. (2002), Demidenko (2013), entre outros.

No caso de modelos lineares mistos, a análise de resíduos é levemente diferente e envolve três tipos de resíduos:

- Resíduos marginais, $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}$, que predizem $\boldsymbol{\varepsilon} = \mathbf{y} - E(\mathbf{y}) = \mathbf{y} - \mathbf{X} \boldsymbol{\beta}$ (erros marginais);
- Resíduos condicionais, $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}} - \mathbf{Z} \hat{\mathbf{b}}$, que predizem $\mathbf{e} = \mathbf{y} - E(\mathbf{y}|\mathbf{b}) = \mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}$ (erros condicionais);
- Resíduos de efeitos aleatórios, $\hat{\mathbf{Zb}} = \mathbf{Z} \hat{\mathbf{b}}$, que predizem $\mathbf{Zb} = E(\mathbf{y}|\mathbf{b}) - E(\mathbf{y})$.

Os resíduos marginais são indicados para verificação de linearidade dos efeitos fixos, presença de observações atípicas e validação da matriz de covariâncias \mathbf{V} . Os resíduos condicionais são utilizados para verificação da hipótese de homocedasticidade e de normalidade dos erros condicionais \mathbf{e} , também, são úteis para detecção de observações atípicas. Os resíduos de efeitos aleatórios, além de ser indicados para detecção de unidades amostrais atípicas, são utilizados para verificação da hipótese de normalidade dos

efeitos aleatórios. Mais detalhes sobre ferramentas de diagnóstico para modelos lineares mistos podem ser encontrados em [Hilden-Minton \(1995\)](#), [Tan et al. \(2001\)](#), [Fung et al. \(2002\)](#), [Nobre \(2004\)](#), [Nobre & Singer \(2007\)](#), [Nobre & Singer \(2011\)](#) e [Singer et al. \(2017\)](#).

Capítulo 3

Critério de Informação de Akaike

Os critérios de informação, juntamente com técnicas de diagnóstico, são úteis para a seleção de modelos. O Critério de Informação de Akaike, mais conhecido como *AIC*, é derivado da informação ou distância de Kullback-Leibler, que mede a discrepância entre dois modelos [Kullback & Leibler (1951), Akaike (1973)].

Seja $f(\cdot)$ a densidade de probabilidade geradora dos dados e $\mathcal{G} = \{g_{\theta_j}(\cdot) = g_j(\cdot|\theta_j), \theta_j \in \Theta_j, j = 1, 2, \dots, m\}$ uma família de m modelos com parâmetros desconhecidos θ_j candidatos à aproximação de f . Para $g_{\theta} \in \mathcal{G}$, $\theta \in \{\theta_j, j = 1, \dots, m\}$, a Informação de Kullback-Leibler entre f e g_{θ} é definida como

$$\mathcal{I}(f, g_{\theta}) = E_f [\log f(x)] - E_f [\log g(x|\theta)] = \int \log \left(\frac{f(x)}{g(x|\theta)} \right) f(x) dx, \quad (3.1)$$

e, segundo Burnham & Anderson (2002), pode ser interpretada intuitivamente como a distância entre os dois modelos ou como a informação perdida ao se utilizar g_{θ} para aproximar f .

Em (3.1), $E_f [\log f(x)]$ é constante, pois depende apenas de f , ou seja,

$$-E_f [\log g(x|\theta)] = \mathcal{I}(f, g_{\theta}) - E_f [\log f(x)], \quad (3.2)$$

e o termo $\mathcal{I}(f, g_{\theta}) - E_f [\log f(x)]$ é uma medida de distância relativa entre f e g_{θ} . Se $\mathcal{I}(f, g_{\theta_i}) < \mathcal{I}(f, g_{\theta_j})$ ou, equivalentemente, $-E_f [\log g(x|\theta_i)] < -E_f [\log g(x|\theta_j)]$, então g_{θ_i} é mais adequado que g_{θ_j} . Consequentemente, o termo $E_f [\log g(x|\theta)]$ torna-se a quantidade de interesse e foi utilizada por Akaike (1973) como base para a criação do *AIC*. Entretanto, diferentemente de (3.1), $-E_f [\log g(x|\theta)]$ é uma medida de distância que não possui um zero absoluto, explicando porque é referida como uma medida de distância *relativa* entre f e g_{θ} [Burnham & Anderson (2002)].

No contexto de seleção de modelos, o objetivo é escolher, dentre um conjunto de modelos candidatos, aquele que minimiza (3.2). Contudo, na prática, (3.2) não pode ser utilizada diretamente, pois requer o conhecimento prévio de f e g_{θ} , que são desconhecidas. Isso indica que, na verdade, a seleção de modelos com base na Informação de Kullback-Leibler busca minimizar o valor esperado de um estimador de (3.2).

Com base nisso, no desenvolvimento do *AIC*, Akaike (1973) utilizou uma versão modificada de (3.2) denominada Informação de Akaike. Denotando por $\mathbf{y} = (y_1, \dots, y_n)^\top$ e $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^\top$ amostras aleatórias independentes com densidade desconhecida f , a Informação de Akaike é

$$AI = E_{\mathbf{y}} \left[E_{\mathbf{y}^*} \left(\log g(\mathbf{y}^* | \hat{\boldsymbol{\theta}}(\mathbf{y})) \right) \right] = \int \left[\int \log g(\mathbf{y}^* | \hat{\boldsymbol{\theta}}(\mathbf{y})) f(\mathbf{y}^*) d\mathbf{y}^* \right] f(\mathbf{y}) d\mathbf{y}, \quad (3.3)$$

em que $g(\mathbf{y}^* | \hat{\boldsymbol{\theta}}(\mathbf{y}))$ é um modelo candidato para f avaliado em $\hat{\boldsymbol{\theta}}(\mathbf{y})$, o MLE de $\boldsymbol{\theta}$ obtido por meio de \mathbf{y} . Em teoria, (3.3) depende de duas réplicas independentes providas do mesmo processo gerador f .

Akaike (1973) mostrou que (3.3) pode ser estimado por meio de $n^{-1} \log g(\mathbf{y} | \hat{\boldsymbol{\theta}}) = n^{-1} \sum_{i=1}^n \log g(y_i | \hat{\boldsymbol{\theta}})$, mas que este estimador é enviesado, uma vez que utiliza o mesmo conjunto de dados para estimar $\boldsymbol{\theta}$ e (3.3), isto é, por depender de apenas uma das amostras. Sob certas condições de regularidade, que envolvem principalmente a suposição de independência nos dados e a forma do espaço paramétrico de $\boldsymbol{\theta}$ [Akaike (1973), Burnham & Anderson (2002) e Greven & Kneib (2010)], o viés é aproximadamente igual a k/n , em que k é a quantidade de parâmetros estimáveis do modelo [em geral, $k = \dim(\boldsymbol{\theta})$]. Logo, um estimador assintoticamente não enviesado para (3.3) é

$$\frac{1}{n} \sum_{i=1}^n \log g(y_i | \hat{\boldsymbol{\theta}}) - \frac{k}{n}. \quad (3.4)$$

Multiplicando (3.4) por $-2n$, Akaike definiu o critério

$$AIC = -2 \sum_{i=1}^n \log g(y_i | \hat{\boldsymbol{\theta}}) + 2k. \quad (3.5)$$

Com base no ajuste dos modelos candidatos à aproximação da verdadeira distribuição dos dados, o modelo a ser selecionado é aquele que produz o menor valor de (3.5). É importante apontar que, no desenvolvimento do *AIC*, não se supõe que o verdadeiro modelo pertença ao conjunto de modelos candidatos, isto é, o *AIC* serve para selecionar o modelo que mais bem aproxima o verdadeiro processo gerador dos dados dentre os candidatos.

Uma vez que (3.2) é uma medida de distância relativa, isto é, sem zero absoluto, temos que a ordenação e a diferença entre os valores de *AIC* dos modelos candidatos é mais importante do que seu valor propriamente dito. Por esse motivo, Burnham & Anderson (2002) recomendam a utilização de um indicador que denominaram *diferenças de AIC*, denotado por:

$$\Delta AIC_i = AIC_i - AIC_{min} \quad (3.6)$$

em que AIC_i é o valor do *AIC* obtido para o modelo candidato i e AIC_{min} é o valor mínimo do *AIC* dentre esses modelos. Portanto, as diferenças de *AIC* constituem um método direto de identificação do modelo com *AIC* mínimo e de discriminação entre os modelos, sem afetar a ordenação e a diferença entre

os mesmos. Segundo [Burnham & Anderson \(2002\)](#), quanto maior Δ_i , menos plausível é que o modelo correspondente seja o melhor. Com base nisso, esses autores propuseram uma regra empírica de escolha do melhor modelo, apresentada na Tabela 3.1.

Tabela 3.1: Regra empírica para seleção do melhor modelo via AIC (sugerida por [Burnham & Anderson \(2002\)](#))

Δ_i	Nível de suporte empírico do modelo i
0 - 2	Substancial
4 - 7	Consideravelmente baixo
> 10	Essencialmente nenhum

Considere, por exemplo, o cenário em que os modelos candidatos g_1 , g_2 e g_3 apresentaram valores de AIC iguais a 3, 10 e 14, respectivamente. Como o AIC de g_1 é o mínimo, as diferenças de AIC para este conjunto de modelos são iguais a $\Delta AIC_{g_1} = 0$, $\Delta AIC_{g_2} = 7$ e $\Delta AIC_{g_3} = 11$. Nesse caso, g_1 é escolhido como o melhor modelo, pois as diferenças de AIC de g_2 e g_3 são relativamente altas, levando em conta a regra empírica definida na Tabela 3.1.

Por outro lado, considere o caso em que os modelos g_1 , g_2 e g_3 apresentaram valores de AIC iguais a 3, 4 e 14, respectivamente. As diferenças de AIC, nesse caso, são iguais a $\Delta AIC_{g_1} = 0$, $\Delta AIC_{g_2} = 1$ e $\Delta AIC_{g_3} = 11$ e, tanto g_1 como g_2 podem ser considerados como os melhores modelos. Nesse caso, para decidir entre g_1 e g_2 , podemos escolher o modelo mais parcimonioso (com o menor número de parâmetros) ou utilizar outras técnicas, como análise de diagnóstico.

Entre as limitações do AIC, destaca-se o baixo desempenho quando o tamanho amostral é pequeno em relação ao número de parâmetros do modelo. Com base nisso, [Sugiura \(1978\)](#) e [Hurvich & Tsai \(1989\)](#) propuseram a utilização de um AIC corrigido para amostras finitas, dado por

$$AIC_c = -2 \sum_{i=1}^n \log g(y_i | \hat{\theta}) + 2 \left(\frac{n}{n-k-1} \right) k \quad (3.7)$$

em que n é o tamanho amostral. Quando $n \rightarrow +\infty$, $AIC_c \rightarrow AIC$, ou seja, ambos os estimadores são assintoticamente equivalentes. Portanto, quando o tamanho amostral é pequeno, é aconselhável utilizar (3.7) em vez de (3.5).

Para seleção de modelos mistos, duas versões do AIC costumam ser empregadas de acordo com o objetivo do estudo: estimar parâmetros populacionais ou prever efeitos aleatórios. No primeiro caso, o AIC *marginal* é recomendado e, no segundo, o critério denominado AIC *condicional*, com versões assintóticas e para amostras finitas, conforme indicado nas seções seguintes.

3.1 AIC marginal

O *AIC marginal* é um critério de informação derivado de (3.3) utilizado na seleção de modelos mistos quando o objetivo é comparar modelos cujo foco é o valor esperado (populacional) da variável resposta.

Exemplo 3.1. Suponha um conjunto de n indivíduos cuja pressão sanguínea foi observada ao longo de m instantes de tempo. Denotamos por y_{ij} a pressão sanguínea do i -ésimo indivíduo no j -ésimo instante de observação. Considere o conjunto de modelos

$$\begin{aligned}\mathcal{M}_1: & y_{ij} = \beta_0 + \beta_1 t_{ij} + e_{ij} \\ \mathcal{M}_2: & y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + e_{ij} \\ \mathcal{M}_3: & y_{ij} = (\beta_0 + b_{0i}) + \beta_1 t_{ij} + e_{ij} \\ \mathcal{M}_4: & y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + e_{ij},\end{aligned}$$

em que β_0 é o valor esperado da resposta no instante $t_{ij} = 0$, β_1 é a variação esperada da resposta por unidade de tempo, $i = 1, \dots, n, j = 1, \dots, m$, $e_{ij} \sim \mathcal{N}(0, \sigma^2)$, $b_{0i} \sim \mathcal{N}(0, \tau_0^2)$, $b_{1i} \sim \mathcal{N}(0, \tau_1^2)$ e $\text{Corr}(b_{0i}, b_{1i}) = \rho_{01}$, com $(b_{0i}, b_{1i})^\top$ independente de e_{ij} . O objetivo é determinar, entre os quatro modelos apresentados, qual é o mais adequado para modelar a média populacional $E(y_{ij})$. Neste caso, os quatro modelos candidatos possuem a mesma média para a variável resposta, $E(y_{ij}) = \beta_0 + \beta_1 t_{ij}$, porém estruturas de covariância distintas, induzidas pela presença dos efeitos aleatórios. Após o ajuste dos modelos, calculamos o *AIC marginal* para $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ e \mathcal{M}_4 , e selecionamos o modelo com ΔAIC mínimo. No Capítulo 6 apresentamos um exemplo detalhado com aplicação desta técnica.

Considere o modelo misto (2.2), em que $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$. Lembremos que \mathbf{V} depende da matriz de covariâncias dos efeitos aleatórios (\mathbf{G}) e da matriz de covariâncias dos erros aleatórios (\mathbf{R}).

Normalmente, o *AIC marginal* é calculado por meio do *MLE* ($\hat{\boldsymbol{\theta}}$) ou do *REML* ($\hat{\boldsymbol{\theta}}_R$) de $\boldsymbol{\theta}$. Ambas as opções são usualmente fornecidas por pacotes computacionais (R, SAS etc). Segundo Vaida & Blanchard (2005), o *AIC marginal* com base em $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\tau}}^\top)^\top$ é definido como

$$mAIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}) + 2(p + q) \quad (3.8)$$

em que $\mathcal{L}(\hat{\boldsymbol{\theta}})$ é a função (2.4) avaliada em $\hat{\boldsymbol{\theta}}$ e $p + q = \dim(\boldsymbol{\theta})$ é o número de parâmetros no modelo (2.2). O *AIC marginal corrigido* correspondente a (3.7) é definido como

$$mAICc = -2\mathcal{L}(\hat{\boldsymbol{\theta}}) + 2 \left(\frac{N}{N - p - q - 1} \right) (p + q) \quad (3.9)$$

em que $N = \sum_{i=1}^n m_i$ é o total de observações das n unidades experimentais.

O *AIC* marginal com base em $\hat{\boldsymbol{\theta}}_R = \left(\hat{\boldsymbol{\beta}}_R^\top, \hat{\boldsymbol{\tau}}_R^\top\right)^\top$ é

$$m\text{AIC}_R = -2\mathcal{L}(\hat{\boldsymbol{\theta}}_R) + 2q \quad (3.10)$$

em que $\mathcal{L}(\hat{\boldsymbol{\theta}})$ é a função (2.4) avaliada em $\hat{\boldsymbol{\theta}}_R$ e $q = \dim(\boldsymbol{\tau})$ é o número de parâmetros não redundantes da matriz de covariâncias \mathbf{V} . Nesse caso, o *AIC* marginal corrigido é dado por:

$$m\text{AIC}_{cR} = -2\mathcal{L}(\hat{\boldsymbol{\theta}}_R) + 2 \left(\frac{N-p}{N-p-q-1} \right) q. \quad (3.11)$$

O *AIC* marginal é um estimador assintoticamente enviesado de (3.3) [Grevén & Kneib (2010)] e as razões para existência de viés, mesmo assintótico, estão relacionadas às condições de regularidade assumidas em seu desenvolvimento, que não são necessariamente verdadeiras no modelo linear misto. Por exemplo, o espaço paramétrico Θ não é uma transformação de \mathbb{R}^k , dadas as restrições nos parâmetros da matriz de covariâncias, e as observações não são independentes devido às correlações induzidas pelos efeitos aleatórios.

Essas peculiaridades dão abertura ao desenvolvimento de critérios de informação que buscam minimizar o viés. Müller et al. (2013) discutem com mais detalhes alguns pontos, como a falta de um termo simples de correção de viés que torne o *AIC* marginal não enviesado e o fato de o termo de penalização dar o mesmo peso para os parâmetros de $\boldsymbol{\beta}$ e $\boldsymbol{\tau}$. Shang & Canavaugh (2008), por exemplo, propõem a utilização de *bootstrapping* para estimar um termo de correção de viés apropriado.

3.2 AIC condicional

O *AIC* condicional se baseia no modelo condicional $\mathbf{y}|\mathbf{b} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{R})$ e é utilizado de maneira similar ao *AIC* marginal. A principal diferença é que o *AIC* condicional é indicado para a seleção de modelos mistos cujo interesse está relacionado à variação da resposta individual. Portanto, a comparação direta de valores de *AIC* marginal e *AIC* condicional não é recomendada, uma vez que estes critérios são utilizados em contextos distintos.

Exemplo 3.2. Em relação ao Exemplo 3.1., suponha que o objetivo seja prever o valor esperado individual da pressão sanguínea, isto é, condicionada a indivíduos específicos. Nesse caso, considere os modelos:

$$\mathcal{M}_1: \quad y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + e_{ij}$$

$$\mathcal{M}_2: \quad y_{ij} = (\beta_0 + b_{0i}) + \beta_1 t_{ij} + e_{ij}$$

$$\mathcal{M}_3: \quad y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + e_{ij}$$

em que β_0 e β_1 têm a mesma interpretação do Exemplo 3.1, $e_{ij} \sim \mathcal{N}(0, \sigma^2)$, $b_{0i} \sim \mathcal{N}(0, \tau_0^2)$, $b_{1i} \sim \mathcal{N}(0, \tau_1^2)$

e $\text{Corr}(b_{0i}, b_{1i}) = \rho_{01}$, com $(b_{0i}, b_{1i})^\top$ independente de e_{ij} . Neste caso, sob a ótica condicional, em \mathcal{M}_1 supomos que a inclinação (β_1) é comum a todos os indivíduos, em \mathcal{M}_2 adotamos interceptos aleatórios e em \mathcal{M}_3 modelamos tanto o intercepto como a inclinação por meio de efeitos aleatórios. Após o ajuste dos modelos, calculamos o AIC condicional (e, conseqüentemente, o $\Delta cAIC$) para os três modelos mistos e selecionamos o modelo com $\Delta cAIC$ mínimo. No Capítulo 6 apresentamos um exemplo mais detalhado com aplicação desta técnica.

Analogamente ao desenvolvimento do AIC por meio de (3.3), Vaida & Blanchard (2005) basearam o AIC condicional em

$$-2E_{f(\mathbf{y}, \mathbf{b})} \left\{ E_{f(\mathbf{y}^* | \mathbf{b})} \left[\log g(\mathbf{y}^* | \hat{\boldsymbol{\theta}}(\mathbf{y}), \hat{\mathbf{b}}(\mathbf{y})) \right] \right\} = -2 \int \int \left[\int \log g(\mathbf{y}^* | \hat{\boldsymbol{\theta}}(\mathbf{y}), \hat{\mathbf{b}}(\mathbf{y})) f(\mathbf{y}^* | \mathbf{b}) d\mathbf{y}^* \right] f(\mathbf{y}, \mathbf{b}) d\mathbf{y} d\mathbf{b}, \quad (3.12)$$

em que \mathbf{y} e \mathbf{y}^* correspondem a amostras condicionalmente independentes da variável resposta com verdadeira distribuição $f(\cdot | \mathbf{b})$. Além disso, $f(\mathbf{y}, \mathbf{b})$ é a distribuição conjunta de \mathbf{y} e \mathbf{b} e $\log g(\mathbf{y}^* | \hat{\boldsymbol{\theta}}(\mathbf{y}), \hat{\mathbf{b}}(\mathbf{y}))$ é a função log-verossimilhança condicional de \mathbf{y}^* , avaliada em $\hat{\boldsymbol{\theta}}$ e $\hat{\mathbf{b}}$.

A expressão para o AIC condicional é similar a (3.8), com duas diferenças importantes: a função log-verossimilhança é condicional aos efeitos aleatórios e o termo de correção de viés está relacionado com os *graus de liberdade efetivos* do modelo (ρ) [Hodges & Sargent (2001)]. Formalmente, $\rho = \text{tr}\{\mathbf{H}\}$, em que \mathbf{H} é a matriz de projeção tal que $\hat{\mathbf{y}}_c = \mathbf{H}\mathbf{y}$. No modelo condicional, $\hat{\mathbf{y}}_c = E(\mathbf{y} | \hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}) = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{b}}$ e, conseqüentemente,

$$\rho = \text{tr} \left\{ \left(\begin{array}{cc} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{array} \right)^{-1} \left(\begin{array}{cc} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} \end{array} \right) \right\}. \quad (3.13)$$

Em termos de interpretação, ρ está relacionado ao número de parâmetros da estrutura de média do modelo condicional, que engloba tanto os efeitos fixos quanto os efeitos aleatórios. Para mais detalhes sobre a obtenção de (3.13), consulte o Apêndice C.

Suponha que, por exemplo, $\mathbf{R} = \sigma^2 \mathbf{I}_N$, $\mathbf{G}_0 = \sigma^{-2} \mathbf{G}$ e $\mathbf{V} = \sigma^2 (\mathbf{Z} \mathbf{G}_0 \mathbf{Z}^\top + \mathbf{I}_N)$. Vaida & Blanchard (2005) mostram que, se σ^2 e \mathbf{G}_0 são conhecidos,

$$cAIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}} | \hat{\mathbf{b}}) + 2\rho \quad (3.14)$$

é não enviesado para (3.12). Nesse caso, $\mathcal{L}(\boldsymbol{\theta} | \mathbf{b})$ é a função (2.5) avaliada no MLE $\hat{\boldsymbol{\theta}}$ (ou $REML$ $\hat{\boldsymbol{\theta}}_R$) e no $BLUP$ $\hat{\mathbf{b}}$ (2.11).

Nas condições anteriores, supondo σ^2 desconhecido e \mathbf{G}_0 conhecido, o AIC condicional com base no

MLE $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\sigma})^\top$ e no EBLUP $\hat{\mathbf{b}}$ é

$$cAIC_c = -2\mathcal{L}(\hat{\boldsymbol{\theta}}|\hat{\mathbf{b}}) + 2 \left[\frac{N(N-p-1)}{(N-p)(N-p-2)} (\rho+1) + \frac{N(p+1)}{(N-p)(N-p-2)} \right]. \quad (3.15)$$

O AIC condicional com base no REML $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\beta}}_R^\top, \hat{\sigma}_R)^\top$ é

$$cAIC_{cR} = -2\mathcal{L}(\hat{\boldsymbol{\theta}}_R|\hat{\mathbf{b}}) + 2 \left[\frac{N-p-1}{N-p-2} (\rho+1) + \frac{p+1}{N-p-2} \right]. \quad (3.16)$$

Quando $n \rightarrow +\infty$,

$$cAIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}|\hat{\mathbf{b}}) + 2(\rho+1), \quad (3.17)$$

ou seja, o termo de correção de viés $2(\rho+1)$ inclui o número de graus de liberdade efetivos do modelo — relacionados aos efeitos fixos e aos efeitos aleatórios — e um parâmetro da matriz \mathbf{R} (σ^2). [Vaida & Blanchard \(2005\)](#) provam que (3.14), (3.15) e (3.16) são não enviesados.

Na prática, como \mathbf{G} e \mathbf{R} geralmente são desconhecidas, ρ é substituído pelo estimador $\hat{\rho}$, obtido por meio da estimação de \mathbf{G} e \mathbf{R} (via máxima verossimilhança restrita). Nesse contexto, critérios diferentes com base no AIC condicional são propostos. [Vaida & Blanchard \(2005\)](#) propõem simplesmente substituir ρ por $\hat{\rho}$ em (3.17), gerando o critério

$$cAIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}|\hat{\mathbf{b}}) + 2(\hat{\rho}+1), \quad (3.18)$$

alegando que a correção nos graus de liberdade devida aos parâmetros desconhecidos em \mathbf{G} é desprezável assintoticamente e que a log-verossimilhança condicional (2.5) não depende de \mathbf{G} . Para \mathbf{R} não necessariamente igual a $\sigma^2\mathbf{I}_N$, [Overholser & Xu \(2008\)](#) provam que o AIC condicional

$$cAIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}|\hat{\mathbf{b}}) + 2(\rho+q_R), \quad (3.19)$$

em que q_R é o número de parâmetros desconhecidos em \mathbf{R} , é assintoticamente não enviesado para (3.12) e é a generalização de (3.17). Assim como em (3.18), na prática, o AIC condicional utilizado é

$$cAIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}|\hat{\mathbf{b}}) + 2(\hat{\rho}+q_R). \quad (3.20)$$

[Müller et al. \(2013\)](#) reúnem um conjunto de critérios propostos por diversos autores, com premissas e estruturas mais complexas que (3.18) e (3.20). Por exemplo, contrariamente ao critério (3.18), [Burnham & White \(2002\)](#) e [Burnham & Anderson \(2002\)](#) recomendam a inclusão de todos os parâmetros de \mathbf{R} e \mathbf{G} no termo de penalização.

Capítulo 4

Critério de Informação de Schwarz

O Critério de Informação de Schwarz, mais conhecido como *BIC* (*Bayesian Information Criterion*), é uma ferramenta para seleção de modelos frequentemente utilizada em conjunto com o *AIC*. A expressão do *BIC* é muito semelhante a (3.5), porém suas premissas e obtenção são distintas. A seguir apresentamos os fundamentos para a definição do *BIC*. Para mais detalhes, o leitor pode consultar [Bhat & Kumar \(2010\)](#) ou [Delattre et al. \(2014\)](#).

Sejam $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_l$ modelos candidatos com probabilidades a priori $p(\mathcal{M}_1), p(\mathcal{M}_2), \dots, p(\mathcal{M}_l)$, respectivamente. O *BIC* utilizado como critério de informação se baseia no fator de Bayes [[Kass & Raftery \(1995\)](#)]

$$\mathcal{B}_{jk}(\mathbf{y}) = \frac{p(\mathbf{y}|\mathcal{M}_j)}{p(\mathbf{y}|\mathcal{M}_k)} = \frac{p(\mathcal{M}_j|\mathbf{y})p(\mathcal{M}_k)}{p(\mathcal{M}_k|\mathbf{y})p(\mathcal{M}_j)}, \quad (4.1)$$

em que \mathbf{y} é uma amostra aleatória de tamanho n . Se $\mathcal{B}_{jk}(\mathbf{y}) > 1$, \mathcal{M}_j é mais fortemente apoiado pelos dados do que \mathcal{M}_k , isto é, o critério favorece a escolha de \mathcal{M}_j . Supondo, a priori, que os modelos são equiprováveis, tem-se

$$\mathcal{B}_{jk}(\mathbf{y}) = \frac{p(\mathcal{M}_j|\mathbf{y})}{p(\mathcal{M}_k|\mathbf{y})}. \quad (4.2)$$

Logo, entre os modelos candidatos, o critério seleciona aquele cuja probabilidade a posteriori é máxima. Com base nisso, o objetivo é encontrar uma expressão para o máximo de $p(\mathcal{M}_j|\mathbf{y})$. Pelo Teorema de Bayes,

$$p(\mathcal{M}_j|\mathbf{y}) \propto p(\mathbf{y}|\mathcal{M}_j)p(\mathcal{M}_j). \quad (4.3)$$

Maximizar (4.3) é o mesmo que maximizar

$$p(\mathbf{y}|\mathcal{M}_j) = \int_{\Theta_j} f(\mathbf{y}|\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j = \int_{\Theta_j} \exp \{ \log [f(\mathbf{y}|\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j)] \} d\boldsymbol{\theta}_j \quad (4.4)$$

em que $f(\mathbf{y}|\boldsymbol{\theta}_j)$ é a função de verossimilhança e $\pi_j(\boldsymbol{\theta}_j)$ é a distribuição a priori dos parâmetros $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_j \in \Theta_j$, sob \mathcal{M}_j . Denotando por $\tilde{\boldsymbol{\theta}}$ a moda a posteriori de $Q = \log [f(\mathbf{y}|\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j)]$ e considerando uma expansão de Taylor de segunda ordem de Q ao redor de $\tilde{\boldsymbol{\theta}}$, tem-se

$$Q \approx \log [f(\mathbf{y}|\tilde{\boldsymbol{\theta}}_j)\pi_j(\tilde{\boldsymbol{\theta}}_j)] + (\boldsymbol{\theta}_j - \tilde{\boldsymbol{\theta}}_j) \nabla Q \Big|_{\boldsymbol{\theta}_j=\tilde{\boldsymbol{\theta}}_j} + \frac{1}{2} (\boldsymbol{\theta}_j - \tilde{\boldsymbol{\theta}}_j)^\top \mathbf{H}_{\boldsymbol{\theta}_j} (\boldsymbol{\theta}_j - \tilde{\boldsymbol{\theta}}_j), \quad (4.5)$$

em que $\nabla Q = \partial Q / \partial \boldsymbol{\theta}_j$ e $\mathbf{H}_{\boldsymbol{\theta}_j}$ é a matriz quadrada com dimensão igual ao número de elementos de $\boldsymbol{\theta}_j$, tal que $H_{mn} = \partial^2 Q / \partial \theta_m \partial \theta_n \Big|_{\boldsymbol{\theta}_j=\tilde{\boldsymbol{\theta}}_j}$. Denotando $\tilde{\mathbf{H}}_{\boldsymbol{\theta}_j} = -\mathbf{H}_{\boldsymbol{\theta}_j}$, é possível mostrar, por meio do método de Laplace [Barndorff-Nielsen & Cox (1989)] para aproximação da integral [4.4] e de manipulações algébricas, que

$$2 \log p(\mathbf{y}|\mathcal{M}_j) \approx 2 \log f(\mathbf{y}|\tilde{\boldsymbol{\theta}}_j) + 2 \log \pi_j(\tilde{\boldsymbol{\theta}}_j) + \dim(\boldsymbol{\theta}_j) \log 2\pi + \log \det (\tilde{\mathbf{H}}_{\boldsymbol{\theta}_j}^{-1}). \quad (4.6)$$

Adotando $\pi_j(\boldsymbol{\theta}_j) = 1$, isto é, distribuição a priori não informativa para $\boldsymbol{\theta}_j$, $Q = \log f(\mathbf{y}|\boldsymbol{\theta}_j)$ e $\tilde{\boldsymbol{\theta}}_j = \hat{\boldsymbol{\theta}}_j$, em que $\hat{\boldsymbol{\theta}}_j$ é o MLE de $\boldsymbol{\theta}_j$, temos

$$\tilde{H}_{mn} = - \frac{\partial^2 \log f(\mathbf{y}|\boldsymbol{\theta}_j)}{\partial \theta_m \partial \theta_n} \Big|_{\boldsymbol{\theta}_j=\hat{\boldsymbol{\theta}}_j}, \quad (4.7)$$

ou seja, $\tilde{\mathbf{H}}_{\boldsymbol{\theta}_j}$ é a matriz de Informação de Fisher observada. Supondo verdadeiras algumas condições de regularidade como, por exemplo, que os dados independentes e identicamente distribuídos e o tamanho amostral é suficientemente grande [Delattre et al. (2014), Bhat & Kumar (2010)],

$$\det (\tilde{\mathbf{H}}_{\boldsymbol{\theta}_j}) = n^{\dim(\boldsymbol{\theta}_j)} \det (\mathcal{I}_{\boldsymbol{\theta}_j}) \quad (4.8)$$

em que $\mathcal{I}_{\boldsymbol{\theta}_j}$ é a matriz de Informação de Fisher para uma única observação y_i . Substituindo (4.8) em (4.6) e ignorando os termos que não envolvem n , tem-se, para n suficientemente grande,

$$2 \log p(\mathbf{y}|\mathcal{M}_j) \approx 2 \log f(\mathbf{y}|\hat{\boldsymbol{\theta}}_j) - \dim(\boldsymbol{\theta}_j) \log n. \quad (4.9)$$

Em (4.9), fazendo $k = \dim(\boldsymbol{\theta}_j)$ e multiplicando ambos os membros da aproximação por -1 , obtém-se o critério

$$BIC = -2 \log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + k \log n. \quad (4.10)$$

O modelo a ser selecionado é aquele cujo valor do *BIC* é mínimo. Para selecionar modelos como (2.2),

$$BIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}) + (p + q) \log N. \quad (4.11)$$

Como os critérios *AIC* e *BIC* possuem regras similares em relação à seleção de modelos, no sentido de es-

colherem aquele com valor mínimo, independentemente do valor propriamente dito, [Burnham & Anderson \(2002\)](#) definiram o indicador de *diferenças de BIC*, que é análogo às diferenças de *AIC* (3.6) tanto na forma como é construído, calculando as diferenças de *BIC* dos modelos candidatos em relação ao *BIC* mínimo, quanto na regra de escolha do melhor modelo, apresentada na Tabela 3.1.

A formulação (4.11) é a versão do *BIC* mais utilizada na prática e geralmente está implementada em pacotes computacionais, como o `nlme()` do repositório R [[Pinheiro et al. \(2019\)](#)]. No Exemplo 3.1, o *AIC* marginal e o *BIC* são ferramentas úteis para seleção do melhor modelo e, no Exemplo 3.2, podemos utilizar o *AIC* condicional ou, também, o *BIC*. Portanto, o *AIC* utilizado na seleção de modelos mistos pode ser marginal ou condicional, dependendo do objetivo do estudo (estimação dos efeitos fixos ou previsão dos efeitos aleatórios, respectivamente). O *BIC*, por outro lado, não faz essa distinção.

Capítulo 5

Simulação

Neste capítulo apresentamos um estudo de simulação com o objetivo de mostrar em quais situações os critérios *mAIC* (3.8), *BIC* (4.11) e *cAIC* (3.20) são mais adequados. A ideia é comparar modelos em um contexto em que os erros aleatórios são independentes (*primeiro contexto*) e em outro contexto, em que os erros aleatórios são autorregressivos (*segundo contexto*). Os valores dos parâmetros utilizados como base foram obtidos a partir do Exemplo 6.3 de [Singer et al. \(2017\)](#).

Em cada contexto, os dados foram gerados sob o modelo misto

$$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \beta_2 t_{ij}^2 + e_{ij} \quad (5.1)$$

em que

- $i = 1, \dots, n$,
- $j = 1, \dots, m$,
- $\mathbf{t}_i = (t_{i1}, \dots, t_{im})$ centralizados segundo $t_{ij} = (t_{ij}^* - \bar{t})/4,29$, com $\mathbf{t}_i^* = (26, 27, \dots, 26 + m - 1)$,
- $\boldsymbol{\beta} = (8, 0 \ 1,39 \ 0,55)^\top$,
- $\mathbf{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}_0)$, com $\mathbf{G}_0 = \begin{pmatrix} \tau_0^2 & \rho_{01} \tau_0 \tau_1 \\ \rho_{01} \tau_0 \tau_1 & \tau_1^2 \end{pmatrix}$,
- $\mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_0)$, com $\mathbf{R}_0 = \mathbf{R}_0(\sigma^2) = \sigma^2 \mathbf{I}_m$ no primeiro contexto e $\mathbf{R}_0 = \mathbf{R}_0(\sigma^2, \phi) = (r_{jk})_{1 \leq j, k \leq m}$, $r_{jk} = \sigma^2 \phi^{|t_{ij} - t_{ik}|}$, no segundo contexto, em que σ^2 é a variância condicional e ϕ é o coeficiente de correlação autorregressivo.

No primeiro contexto, consideramos os cenários obtidos por meio da variação de n (10, 20, 30 e 40), m (8, 10, 13, 15 e 20), σ^2 (0,8 e 1,6) e dos parâmetros de \mathbf{G}_0 , a saber, τ_0^2 (0,4 e 0,8), τ_1^2 (0,4 e 0,8) e ρ_{01} (0,4 e 0,8). Para cada cenário, geramos $S = 1.000$ conjuntos de dados e, a cada um destes, ajustamos 7 modelos mistos com a mesma estrutura de independência condicional e diferentes efeitos fixos e aleatórios, apresentados na Tabela 5.1. Em seguida, calculamos os valores dos critérios de informação *mAIC*, *BIC* e *cAIC* para os

respectivos modelos. Do mesmo modo, no segundo contexto, consideramos os cenários obtidos por meio da variação de n (10, 30 e 50), m (13, 16, 19 e 22), σ^2 (0,8 e 1,6), τ_0^2 (0,4 e 0,8), τ_1^2 (0,4 e 0,8) e ϕ (0,3, 0,5 e 0,7). Para cada cenário, foram gerados $S = 1.000$ conjuntos de dados e, cada um destes, foram ajustados 4 modelos mistos com diferentes estruturas de covariância condicional e efeitos aleatórios, mas com os mesmos efeitos fixos do processo gerador dos dados, conforme apresentamos na Tabela 5.5. Em seguida, calculamos os mesmos critérios de informação do primeiro contexto.

O desempenho dos critérios de informação foi avaliado individualmente por meio da frequência de seleções corretas para cada 1.000 simulações, em porcentagem, levando em conta a regra empírica proposta por Burnham & Anderson (2002), conforme discutimos no Capítulo 3. Em cada simulação, ajustamos todos os modelos candidatos e obtivemos os respectivos valores de $mAIC$, BIC e $cAIC$. Consideramos como o modelo selecionado, por exemplo, aquele cujo valor de $mAIC$ é o menor e que possui uma diferença absoluta maior que 2 em relação ao modelo com segundo menor $mAIC$, garantindo um nível de discriminação razoável. Se a diferença entre os dois menores $mAIC$ for menor ou igual a 2, selecionamos o modelo mais parcimonioso (com menor quantidade de parâmetros). Aplicamos esta regra para o $mAIC$, o BIC e o $cAIC$ individualmente.

A seguir apresentamos, para cada critério de informação abordado, os detalhes e os resultados das simulações obtidos em cada contexto.

5.1 Modelo misto com erros aleatórios independentes

A Tabela 5.1 contém os 7 modelos candidatos ajustados.

Tabela 5.1: Modelos candidatos utilizados na simulação (\mathcal{M}_1 é o modelo verdadeiro)

\mathcal{M}_1 :	$y_{ij} =$	$(\beta_0 + b_{0i})$	$+$	$(\beta_1 + b_{1i})t_{ij}$	$+$	$\beta_2 t_{ij}^2$	$+$	e_{ij} ,	$\mathbf{R}_0 = \mathbf{R}_0(\sigma^2)$
\mathcal{M}_2 :	$y_{ij} =$	$(\beta_0 + b_{0i})$	$+$	$(\beta_1 + b_{1i})t_{ij}$	$+$		$+$	e_{ij} ,	$\mathbf{R}_0 = \mathbf{R}_0(\sigma^2)$
\mathcal{M}_3 :	$y_{ij} =$	$(\beta_0 + b_{0i})$	$+$	$b_{1i}t_{ij}$	$+$	$\beta_2 t_{ij}^2$	$+$	e_{ij} ,	$\mathbf{R}_0 = \mathbf{R}_0(\sigma^2)$
\mathcal{M}_4 :	$y_{ij} =$	b_{0i}	$+$	$(\beta_1 + b_{1i})t_{ij}$	$+$	$\beta_2 t_{ij}^2$	$+$	e_{ij} ,	$\mathbf{R}_0 = \mathbf{R}_0(\sigma^2)$
\mathcal{M}_5 :	$y_{ij} =$	β_0	$+$	$(\beta_1 + b_{1i})t_{ij}$	$+$	$\beta_2 t_{ij}^2$	$+$	e_{ij} ,	$\mathbf{R}_0 = \mathbf{R}_0(\sigma^2)$
\mathcal{M}_6 :	$y_{ij} =$	$(\beta_0 + b_{0i})$	$+$	$\beta_1 t_{ij}$	$+$	$\beta_2 t_{ij}^2$	$+$	e_{ij} ,	$\mathbf{R}_0 = \mathbf{R}_0(\sigma^2)$
\mathcal{M}_7 :	$y_{ij} =$	$(\beta_0 + b_{0i})$	$+$	$\beta_1 t_{ij}$	$+$		$+$	e_{ij} ,	$\mathbf{R}_0 = \mathbf{R}_0(\sigma^2)$

O modelo \mathcal{M}_1 é o verdadeiro, ou seja, a frequência de seleções corretas obtida pelo critério ($mAIC$, BIC ou $cAIC$) é a quantidade de vezes (%) que \mathcal{M}_1 foi selecionado para cada conjunto de 1.000 simulações, levando em conta que o valor do respectivo critério para este modelo deve ser o menor e apresentar uma

diferença absoluta de pelo menos 2 em relação aos valores obtidos para os demais modelos.

Apresentamos nas Tabelas 5.2, 5.3 e 5.4 a frequência (%) de seleções do modelo correto para os critérios $mAIC$, BIC e $cAIC$, respectivamente. Observamos que o desempenho de cada critério na escolha do modelo correto está associado ao tamanho amostral (n), ao número de observações por indivíduo (m) e à razão entre a variância total dos efeitos aleatórios ($\tau_0^2 + \tau_1^2$) e a variância do erro aleatório (σ^2), denotada por λ .

Supondo que a escolha dos modelos se baseia na capacidade dos mesmos em modelar a média populacional, concluímos por meio da Tabela 5.2 que a porcentagem de escolhas corretas ao utilizar o $mAIC$ está associada principalmente ao tamanho amostral (n), ao número de observações por indivíduo (m) e à razão entre variâncias λ . Para tamanho amostral reduzido ($n = 10$), o $mAIC$ seleciona o modelo correto com frequência elevada ($\geq 79\%$) se o número de observações por indivíduo for relativamente alto ($m \geq 15$). Para tamanho amostral mais elevado ($n \geq 20$), a porcentagem de seleções corretas é praticamente máxima se o número de observações por indivíduo também for elevado ($m \geq 15$). Observamos também que, quanto maior a variância total dos efeitos aleatórios (entre unidades amostrais) em relação à variância dos erros aleatórios (dentro da mesma unidade amostral), maior a porcentagem de escolhas do modelo correto. Por exemplo, para $n = 20$ e $m = 15$, a porcentagem de acertos aumenta de 52% para 93% quando λ aumenta de 0,5 para 2. Por outro lado, em alguns casos, o desempenho do AIC marginal não é razoável. Por exemplo, para $n \leq 20$ e $m \leq 8$, a porcentagem de acertos não ultrapassa os 50%.

Tabela 5.2: Frequência de seleções do modelo correto via AIC marginal (%)

n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$					n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$				
		0,5	0,75	1,0	1,5	2,0			0,5	0,75	1,0	1,5	2,0
10	8	2	5	9	17	21	30	8	24	29	48	61	62
	10	18	20	37	54	59		10	75	81	89	98	99
	13	63	73	87	97	98		13	100	100	100	100	100
	15	79	89	96	99	99		15	100	100	100	100	100
	20	96	98	99	100	99		20	100	100	100	100	100
20	8	9	17	28	43	46	40	8	40	41	59	75	76
	10	52	56	76	92	93		10	90	92	97	100	100
	13	97	98	99	100	100		13	100	100	100	100	100
	15	99	99	100	100	100		15	100	100	100	100	100
	20	100	100	100	100	100		20	100	100	100	100	100

Em relação ao *BIC*, observamos um comportamento similar ao do *AIC* marginal, conforme mostra a Tabela 5.3. A principal diferença é que, quando comparamos a frequência de escolhas corretas dos dois critérios para valores iguais de n , m e λ (ou seja, o mesmo cenário), a porcentagem de escolhas do modelo correto ao utilizar o *BIC* é, em geral, menor que a do *mAIC*, sobretudo quando n é pequeno ($n = 10$). Para tamanho de amostra e número de observações por indivíduo mais elevados ($n \geq 20$ e $m \geq 13$), o *BIC* apresenta porcentagem de escolhas do modelo correto tão alto quanto o *AIC* marginal e ambos os critérios são igualmente indicados para a seleção de modelos mistos quando o interesse está centrado no valor esperado (populacional) da variável resposta.

Tabela 5.3: Frequência de seleções do modelo correto via *BIC* (%)

n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$					n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$				
		0,5	0,75	1,0	1,5	2,0			0,5	0,75	1,0	1,5	2,0
10	8	0	0	2	5	5	30	8	5	7	19	30	33
	10	4	4	13	25	30		10	39	49	71	91	92
	13	31	41	64	82	88		13	95	98	100	100	100
	15	59	68	83	93	93		15	99	100	100	100	100
	20	84	88	93	94	94		20	100	100	100	100	100
20	8	1	2	7	17	20	40	8	7	14	29	43	44
	10	15	24	48	68	74		10	57	66	85	98	98
	13	82	88	98	100	100		13	100	100	100	100	100
	15	94	97	100	100	100		15	100	100	100	100	100
	20	99	99	100	100	100		20	100	100	100	100	100

Por outro lado, supondo que a escolha dos modelos se baseia na capacidade dos mesmos em modelar a média individual (ou condicional), concluímos por meio da Tabela 5.4 que a porcentagem de escolhas do modelo correto ao utilizar o *cAIC* está associada ao tamanho amostral (n), ao número de observações por indivíduo (m) e à razão entre variâncias (λ). Para tamanho amostral reduzido ($n = 10$), o *cAIC* seleciona o modelo correto com porcentagem praticamente máxima se o número de observações por indivíduo for relativamente alto ($m \geq 13$). Para tamanho amostral mais elevado ($n \geq 20$), a porcentagem de seleções do modelo correto é elevada mesmo se o número de observações por indivíduo não é tão alto ($m = 8$), ou seja, neste tipo de abordagem (condicional), o *AIC* condicional é uma ferramenta com desempenho considerável, desde que o tamanho amostral seja suficientemente grande ($n \geq 20$). Observamos também que, assim como no caso do *AIC* marginal, quanto maior a variância total dos efeitos aleatórios (entre unidades amostrais)

em relação à variância dos erros aleatórios (dentro da mesma unidade amostral), maior a porcentagem de escolhas do modelo correto. Por exemplo, para $n = 10$ e $m = 10$, a porcentagem de acertos aumenta de 57% para 86% quando λ aumenta de 0,5 para 2. Em poucos casos, o desempenho do *AIC* condicional não é razoável. Por exemplo, para $n \leq 10$ e $m \leq 8$, a porcentagem de acertos não ultrapassa os 52%.

Tabela 5.4: Frequência de seleções do modelo correto via *AIC* condicional (%)

n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$					n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$				
		0,5	0,75	1,0	1,5	2,0			0,5	0,75	1,0	1,5	2,0
10	8	37	34	44	50	52	30	8	64	64	72	84	85
	10	57	61	72	85	86		10	95	94	99	100	100
	13	97	97	99	100	100		13	100	100	100	100	100
	15	100	100	100	100	100		15	100	100	100	100	100
	20	100	100	100	100	100		20	100	100	100	100	100
20	8	52	53	62	74	75	40	8	68	72	85	94	95
	10	86	87	94	99	99		10	97	100	100	100	100
	13	100	100	100	100	100		13	100	100	100	100	100
	15	100	100	100	100	100		15	100	100	100	100	100
	20	100	100	100	100	100		20	100	100	100	100	100

Em resumo, observamos que em grande parte dos cenários os três critérios de informação estudados (*AIC* marginal, *BIC* e *AIC* condicional) acertam sempre, sobretudo quando $n \geq 20$ e $m \geq 13$.

5.2 Modelo misto com erros aleatórios autorregressivos

A Tabela 5.5 contém os 4 modelos ajustados no segundo contexto.

Tabela 5.5: Modelos candidatos utilizados na simulação (\mathcal{M}_1 é o modelo verdadeiro)

\mathcal{M}_1 :	$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \beta_2 t_{ij}^2 + e_{ij}, \mathbf{R}_0 = \mathbf{R}_0(\sigma^2, \phi)$
\mathcal{M}_2 :	$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \beta_2 t_{ij}^2 + e_{ij}, \mathbf{R}_0 = \mathbf{R}_0(\sigma^2)$
\mathcal{M}_3 :	$y_{ij} = (\beta_0 + b_{0i}) + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + e_{ij}, \mathbf{R}_0 = \mathbf{R}_0(\sigma^2, \phi)$
\mathcal{M}_4 :	$y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + \beta_2 t_{ij}^2 + e_{ij}, \mathbf{R}_0 = \mathbf{R}_0(\sigma^2, \phi)$

O modelo \mathcal{M}_1 é o verdadeiro, ou seja, a frequência de seleções corretas obtida pelo critério (*mAIC*, *BIC* ou *cAIC*) é a quantidade de vezes (%) que \mathcal{M}_1 foi selecionado para cada conjunto de 1.000 simulações, assim como nas simulações do primeiro contexto.

Apresentamos nas Tabelas 5.6, 5.7 e 5.8 a frequência (%) de seleções do modelo correto para os critérios *mAIC*, *BIC* e *cAIC*, respectivamente. Observamos que o desempenho de cada critério na escolha do modelo correto está associado ao tamanho amostral (n), ao número de observações por indivíduo (m), à razão entre a variância total dos efeitos aleatórios ($\tau_0^2 + \tau_1^2$) e a variância condicional (σ^2), denotada por λ , e ao coeficiente de correlação autorregressivo (ϕ).

Supondo que a escolha dos modelos se baseia na capacidade dos mesmos em modelar a média populacional, concluímos por meio da Tabela 5.6 que a porcentagem de escolhas do modelo correto ao utilizar o *mAIC* está associada principalmente ao tamanho amostral (n), ao número de observações por indivíduo (m), à razão entre variâncias (λ) e, em menor intensidade, ao coeficiente autorregressivo (ϕ). Para tamanho amostral reduzido ($n = 10$), o *mAIC* seleciona o modelo correto com frequência moderada se $\lambda > 1$ (entre 33% e 69%) mesmo quando o número de observações por indivíduo é elevado ($m = 22$), e essa porcentagem diminui se o valor de ϕ aumenta. Por exemplo, para $n = 10$, $m = 22$ e $\lambda = 2$, a porcentagem de acertos diminui de 69% para 44% quando o coeficiente de correlação ϕ aumenta de 0,3 para 0,7.

Tabela 5.6: Frequência de seleções do modelo correto via AIC marginal (%)

ϕ	n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$					ϕ	n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$					ϕ	n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$				
			0,5	0,75	1	1,5	2				0,5	0,75	1	1,5	2				0,5	0,75	1	1,5	2
0,3	10	13	13	18	29	44	56	10	10	13	5	20	28	35	46	10	10	13	17	17	24	33	42
		16	26	34	43	53	65			16	12	26	34	42	59			16	20	26	28	36	43
		19	29	39	52	64	68			19	21	30	40	43	49			19	25	28	29	37	43
		22	30	42	53	65	69			22	22	32	45	46	54			22	27	28	29	39	44
	30	13	46	60	81	89	94	0,5	30	13	43	48	67	79	94	0,7	30	13	35	49	69	84	91
		16	55	75	87	87	98			16	54	67	88	92	99			16	37	52	73	79	89
		19	70	75	89	89	100			19	60	79	90	95	99			19	40	64	78	82	92
		22	69	79	94	95	100			22	73	85	91	98	98			22	52	67	79	90	98
	50	13	69	82	96	96	100	50	50	13	57	58	86	99	100	50	50	13	51	74	86	94	98
		16	78	88	95	98	100			16	73	87	97	99	100			16	56	76	88	95	100
		19	90	95	99	100	100			19	88	94	99	100	100			19	84	94	97	99	100
		22	94	99	100	100	100			22	92	98	99	100	100			22	88	94	98	100	100

Para tamanho amostral mais elevado ($n = 30$), a porcentagem de seleções do modelo correto é alta

($\geq 70\%$) se o número de observações por indivíduo também é elevado ($m \geq 19$). Para $n = 50$, a porcentagem de seleção do modelo correto é consideravelmente alta mesmo para $m = 13$. Conseqüentemente, no caso em que os erros são condicionalmente autorregressivos, o *AIC* marginal é indicado para a seleção de modelos mistos se tanto o tamanho amostral quanto o número de observações por indivíduo são suficientemente grandes ($n \geq 30$ e $m \geq 16$, respectivamente).

Em relação ao *BIC*, observamos que a porcentagem de escolhas do modelo correto ao utilizar este critério é, em geral, menor que a do *mAIC*, sobretudo quando $n \leq 30$ e $\phi \geq 0,7$. Além disso, para que o *BIC* selecione o modelo correto com uma porcentagem considerável, é necessário que, além do tamanho amostral, o número de observações por indivíduo seja mais elevado que no caso de independência condicional ($m \geq 22$), conforme mostramos anteriormente na Tabela 5.3. Portanto, estes resultados mostram que, no caso em que os erros aleatórios são autorregressivos, o *BIC* é indicado para a seleção de modelos mistos apenas se $n \geq 30$ e $m \geq 22$, uma condição mais rígida quando comparada ao caso de independência condicional.

Tabela 5.7: Frequência de seleções do modelo correto via *BIC* (%)

ϕ	n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$					ϕ	n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$					ϕ	n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$				
			0,5	0,75	1	1,5	2				0,5	0,75	1	1,5	2				0,5	0,75	1	1,5	2
10	10	13	3	4	6	10	24	10	10	13	0	5	6	8	9	10	10	13	2	3	6	7	15
		16	5	6	12	19	21			16	1	6	7	13	18			16	4	3	7	15	14
		19	6	12	13	20	29			19	1	6	5	12	11			19	1	4	4	9	10
		22	7	12	17	19	28			22	2	9	10	11	12			22	3	5	6	8	10
0,3	30	13	13	19	33	43	61	0,5	30	13	2	6	21	32	61	0,7	30	13	3	11	18	28	49
		16	15	27	44	49	81			16	13	18	30	56	71			16	4	10	24	33	46
		19	21	30	49	67	97			19	14	29	44	55	80			19	10	14	26	34	60
		22	18	34	78	86	98			22	26	34	47	65	78			22	11	14	30	39	61
50	50	13	12	30	56	65	86	50	50	13	4	15	52	74	80	50	50	13	7	19	35	55	78
		16	24	39	53	84	99			16	16	38	62	80	97			16	13	20	34	56	86
		19	43	55	65	88	100			19	41	53	60	82	97			19	38	50	58	79	89
		22	48	61	91	98	100			22	45	57	85	92	98			22	44	55	83	90	95

Por outro lado, supondo que a escolha dos modelos se baseia na capacidade dos mesmos em modelar a média individual (ou condicional), concluímos por meio da Tabela 5.8 que a porcentagem de escolhas do modelo correto ao utilizar o *cAIC* está associada ao tamanho amostral (n) e ao número de observações por indivíduo (m). Para tamanho amostral reduzido ($n = 10$), o *cAIC* seleciona o modelo correto com

porcentagem praticamente máxima se o número de observações por indivíduo for relativamente alto ($m \geq 16$).

Para tamanho amostral mais elevado ($n \geq 30$), a porcentagem de seleções do modelo correto é elevada mesmo se o número de observações por indivíduo não é tão alto ($m = 13$), ou seja, neste tipo de abordagem (condicional), o AIC condicional é uma ferramenta com desempenho considerável, mesmo quando o tamanho amostral não é tão elevado ($n \geq 10$). Além disso, em geral, quanto maior o valor de ϕ , menor a porcentagem de seleções do modelo correto, sobretudo se n ou m são reduzidos.

Tabela 5.8: Frequência de seleções do modelo correto via AIC condicional (%)

ϕ	n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$					ϕ	n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$					ϕ	n	m	$\lambda = (\tau_0^2 + \tau_1^2) / \sigma^2$				
			0,5	0,75	1	1,5	2				0,5	0,75	1	1,5	2				0,5	0,75	1	1,5	2
10	10	13	54	56	55	62	64	10	10	13	51	60	60	59	61	10	10	13	50	49	50	49	52
		16	80	87	92	95	96			16	72	83	88	93	93			16	81	84	84	88	88
		19	85	88	95	95	97			19	78	84	92	92	95			19	80	83	89	90	95
		22	84	92	97	95	100			22	79	87	92	93	95			22	82	84	86	89	96
0,3	30	13	84	85	93	94	94	0,5	30	13	83	72	77	76	77	0,7	30	13	52	53	50	51	50
		16	93	99	100	100	100			16	90	94	100	100	100			16	93	97	98	98	100
		19	97	99	100	100	100			19	96	100	100	100	100			19	95	97	100	100	100
		22	96	98	100	100	100			22	98	100	100	100	100			22	97	97	100	100	100
50	50	13	95	94	95	93	95	50	50	13	85	82	76	78	76	50	50	13	53	52	53	48	49
		16	98	99	100	100	100			16	97	99	100	100	100			16	97	99	100	100	100
		19	99	100	100	100	100			19	98	100	100	100	100			19	96	99	100	100	100
		22	99	100	100	100	100			22	99	100	100	100	100			22	99	99	100	100	100

Em resumo, observamos que, em ambos os cenários, o desempenho dos critérios de informação está associado a alguns fatores, como o tamanho amostral, o número de observações por indivíduo, a relação entre as variâncias inter e intraunidades amostrais e o coeficiente de correlação. Em ambos os casos, verificamos que, quanto maior n , m e λ , maior a porcentagem de acertos. Por outro lado, quanto maior o coeficiente de correlação ϕ , menor a porcentagem de seleções corretas para os três critérios de informação, sobretudo o BIC, que apresentou porcentagem de acerto, em geral, mais baixa que no caso de erros independentes.

Capítulo 6

Exemplos

Analisamos dois exemplos de conjuntos de dados com o intuito de ilustrar a utilização dos critérios de informação para seleção de modelos mistos, combinando-os a técnicas descritivas e de diagnóstico.

6.1 Dados de crescimento de bezerros

Os dados deste exemplo foram obtidos a partir de um estudo sobre o crescimento de bezerros conduzido pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) e analisado por [Singer et al. \(2017\)](#). Os dados são fornecidos na Tabela A.1 (em anexo) e correspondem a coletas quinzenais do peso, em *kg*, de 28 bezerros submetidos a uma dieta alimentar por um período de 26 semanas após o nascimento. O objetivo é estudar o comportamento do valor esperado (populacional) do peso como uma função do tempo.

A Figura 6.1 apresenta o gráfico de perfis correspondente. A linha pontilhada descreve o peso médio nos instantes de observação e sugere que um polinômio de 2º grau pode ser utilizado como modelo.

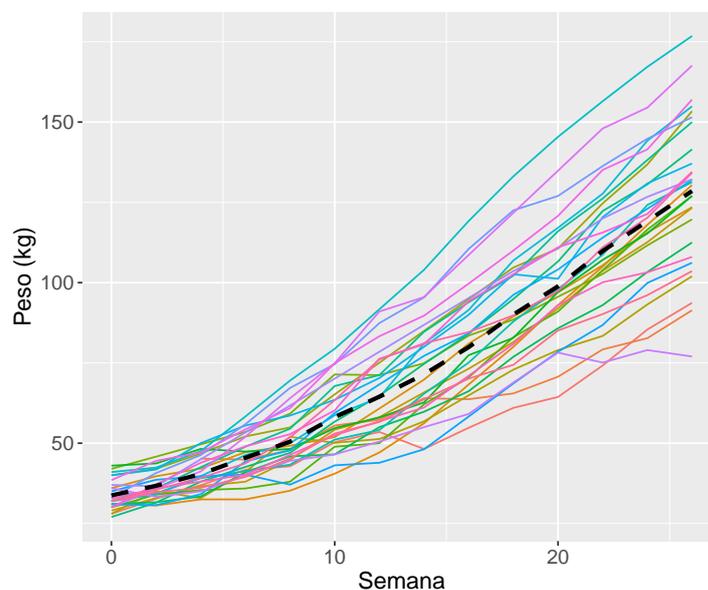


Figura 6.1: Perfis de peso dos bezerros ao longo das semanas (linha pontilhada: perfil médio)

É possível observar, também, um “paralelismo” nas curvas de crescimento individuais, isto é, bezerros que nasceram com peso menor tendem a ter o peso menor ao longo do tempo em relação a bezerros mais pesados. Este tipo de comportamento sugere a inclusão de efeitos aleatórios no modelo, dada a aparente especificidade de peso de cada bezerro. Outra característica interessante é que a variabilidade do peso aumenta ao longo do tempo, isto é, a variância da variável resposta é uma função do tempo, conforme mostra a Figura 6.2.

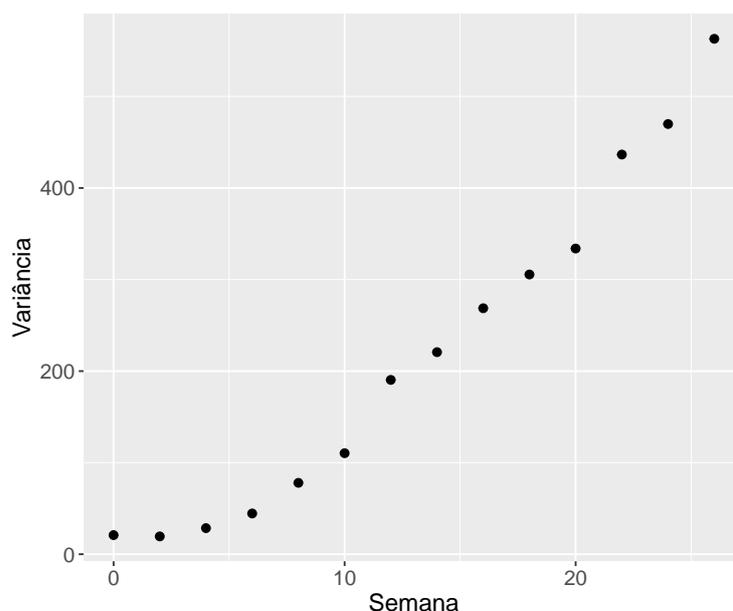


Figura 6.2: Gráfico de variância do peso em função do tempo

A Figura 6.3 apresenta o gráfico do desenhista (*Draftman's plot*) dos dados. Os elementos da diagonal principal representam os instantes de tempo (*semana*) e os elementos abaixo da diagonal principal denotam os coeficientes de correlação quando comparamos o peso dos bezerros entre dois instantes de tempo. A Figura 6.4 contém o gráfico de correlação de Grady-Helms, que mostra as correlações da Figura 6.3 em função da defasagem (*lag*) temporal de uma maneira mais clara e direta. Os gráficos sugerem a presença de correlação serial nos dados, dado que as curvas com as correlações decaem lentamente — aparentemente, de maneira exponencial.

Com base na análise descritiva, consideramos os seguintes modelos candidatos:

$$\mathcal{M}_1 : y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + e_{ij}$$

$$\mathcal{M}_2 : y_{ij} = (\beta_0 + b_{0i}) + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + e_{ij}$$

$$\mathcal{M}_3 : y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) t_{ij} + \beta_2 t_{ij}^2 + e_{ij}$$

$$\mathcal{M}_4 : y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) t_{ij} + (\beta_2 + b_{2i}) t_{ij}^2 + e_{ij}$$

$$\mathcal{M}_5 : y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) t_{ij} + (\beta_2 + b_{2i}) t_{ij}^2 + e_{ij}^*$$

em que y_{ij} é o peso do i -ésimo bezerro no j -ésimo instante, $i = 1, 2, \dots, 28$, $j = 1, \dots, m_i$ (m_i é o número de



Figura 6.3: Gráfico do desenhistas do peso ao longo do tempo (semana na diagonal principal, correlações na diagonal inferior)

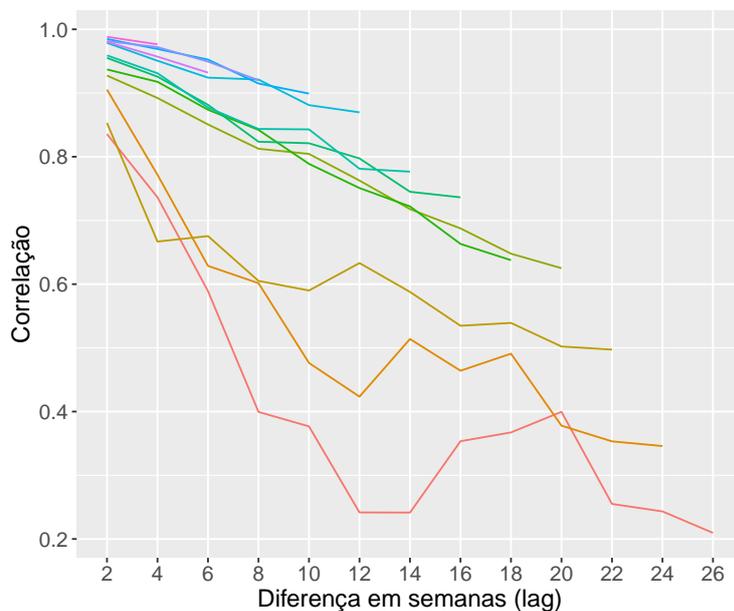


Figura 6.4: Gráfico de correlação de Grady-Helms

observações do bezerro i), $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$, $e_{ij} \sim \mathcal{N}(0, \sigma^2)$ independentes, $e_{ij}^* \sim \mathcal{N}(0, \sigma^2)$ correlacionados segundo $\text{Corr}(e_{ij}^*, e_{ik}^*) = \phi^{|t_{ij} - t_{ik}|}$, $0 \leq \phi \leq 1$, $b_{0i} \sim \mathcal{N}(0, \tau_0^2)$, $b_{1i} \sim \mathcal{N}(0, \tau_1^2)$, $b_{2i} \sim \mathcal{N}(0, \tau_2^2)$, $\text{Corr}(b_{0i}, b_{1i}) = \rho_{01}$, $\text{Corr}(b_{0i}, b_{2i}) = \rho_{02}$ e $\text{Corr}(b_{1i}, b_{2i}) = \rho_{12}$, com $(b_{0i}, b_{1i}, b_{2i})^\top$ independente de e_{ij} e de e_{ij}^* .

Na Tabela 6.1 apresentamos os resultados do ajuste dos 5 modelos candidatos obtidos por Máxima Verossimilhança. Utilizamos o pacote computacional `nlme()` do repositório R [Pinheiro et al. (2019)] e a codificação da análise é apresentada no Apêndice D.

Tabela 6.1: Estimativas (e erros padrões) dos modelos candidatos ajustados

Parâmetro	Modelo				
	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4	\mathcal{M}_5
β_0	32,8 (2,0)	32,1 (2,5)	32,9 (1,2)	32,7 (0,9)	33,5 (1,0)
β_1	7,3 (1,5)	8,1 (0,9)	7,7 (0,9)	8,0 (1,2)	7,8 (1,0)
β_2	1,4 (0,2)	1,3 (0,1)	1,4 (0,1)	1,3 (0,2)	1,3 (0,1)
τ_0^2	-	137,9	30,2	18,0	0,7
τ_1^2	-	-	16,6	35,4	18,4
τ_2^2	-	-	-	0,60	0,26
ρ_{01}	-	-	-0,31	-0,32	0,98
ρ_{02}	-	-	-	0,16	-0,59
ρ_{12}	-	-	-	-0,73	-0,54
σ^2	217,9	77,1	14,7	8,0	24,7
λ	-	1,8	3,2	6,8	0,8
ϕ	-	-	-	-	0,61
<i>N° de parâmetros</i>	4	5	7	10	11
<i>N° efetivo de parâmetros</i>	4	29,9	54,6	77,8	46,8
<i>mAIC</i>	3.055,4	2.764,5	2.269,3	2.136,4	2.039,3
<i>BIC</i>	3.071,1	2.784,1	2.296,7	2.175,6	2.082,4
<i>cAIC</i>	-	2.697,9	2.107,5	1.905,4	1.952,4
$\Delta mAIC_i$	1.016,1	725,2	230,0	97,1	0
ΔBIC_i	988,7	701,7	214,3	93,2	0
$\Delta cAIC_i$	-	792,5	202,1	0	47,0

λ : variância total dos efeitos aleatórios / variância dos erros aleatórios

Como o objetivo da modelagem é estimar o peso médio populacional (*marginal*), comparamos os

modelos por meio do $mAIC$ (3.8) e do BIC (4.11) ou, alternativamente, por meio das diferenças $\Delta mAIC_i$ e ΔBIC_i (3.6). Por outro lado, se o objetivo fosse selecionar o melhor modelo para prever o peso esperado de algum bezerro específico (*condicional*), a comparação seria feita por meio do $cAIC$ (3.20) ou das diferenças $\Delta cAIC_i$.

Nesse caso, \mathcal{M}_5 é o modelo com menor valor de AIC marginal e BIC entre os modelos candidatos e apresentou considerável diferença ΔAIC e ΔBIC em relação aos valores dos demais modelos, sobretudo \mathcal{M}_1 ($\Delta AIC_{\mathcal{M}_1} = 1.016,1$ e $\Delta BIC_{\mathcal{M}_1} = 988,7$). Além disso, observamos que $\hat{\phi} = 0,61$, indicando que um modelo com erros aleatórios autorregressivos deve ser considerado. Consequentemente, esse é o modelo selecionado por meio de ambos os critérios.

Na Figura 6.5 apresentamos os gráficos de resíduos para o modelo selecionado (\mathcal{M}_5). Observamos que os resíduos não apresentam tendências nem variabilidade expressivamente diferente em relação aos valores ajustados, e apresentam poucos valores atípicos (apenas 3% dos resíduos marginais e 7% dos resíduos condicionais). Além disso, a suposição de normalidade dos erros aleatórios parece ser satisfeita, como mostram o $QQplot$ e o histograma dos resíduos condicionais. Em relação aos efeitos aleatórios, observamos que a hipótese de normalidade parece satisfeita, conforme mostra o $QQPlot$ da distância de Mahalanobis dos efeitos aleatórios estimados, e não foram observados indivíduos atípicos. Consequentemente, levando em conta os resultados da análise descritiva, da análise de diagnóstico e dos critérios de informação, concluímos que \mathcal{M}_5 é o modelo selecionado. Adicionalmente, nas Figuras B.1, B.2, B.3 e B.4 (em anexo) apresentamos os gráficos de resíduos dos demais modelos candidatos. Para maiores detalhes sobre as técnicas de diagnóstico utilizadas, o leitor pode consultar Singer et al. (2017).

É importante observar que, se o objetivo fosse selecionar o melhor modelo para prever o peso esperado de algum bezerro específico, o AIC condicional selecionaria \mathcal{M}_4 , que é o modelo com menor valor de $cAIC$. Neste modelo, assumimos que os dados são condicionalmente independentes, o que contraria os resultados da análise descritiva (Figuras 6.3 e 6.4).

Uma possível explicação para esta divergência pode ser encontrada na conclusão das simulações do Capítulo 5, em que os modelos comparados são similares aos modelos candidatos deste exemplo. Verificamos no Capítulo 5 que, em linhas gerais, quando maior o coeficiente de correlação autorregressivo, menor a porcentagem de acertos, sobretudo para o BIC (Tabelas 5.6, 5.7 e 5.8). Nesse cenário, a utilização dos critérios de informação para seleção dos modelos é indicada apenas se tanto o tamanho amostral quanto o número de observações por indivíduo é elevado ($n = 50$ e $m \geq 19$). Entretanto, o conjunto de dados deste exemplo apresenta tamanho amostral ($n = 28$) e quantidade de observações por indivíduo ($m = 14$) relativamente moderados (com base nos padrões das simulações), ou seja, os critérios de informação têm uma porcentagem de acerto mais baixa. Por exemplo, observamos na Tabela 5.8 que, quando $n = 30$, $m = 13$ e $\phi = 0,7$, a porcentagem de seleções corretas do modelo misto com erros aleatórios autorregressivos utilizando o $cAIC$ é da ordem de 50%, um valor relativamente baixo. Isso mostra o porquê da divergência observada entre a análise descritiva e a seleção via AIC condicional.

6.2 Dados de palatos de sapos

Os dados deste exemplo foram obtidos a partir de um estudo analisado por [Singer & Andrade \(2000\)](#) e são apresentados na Tabela A.2. Eles correspondem à velocidade relativa de transporte mucociliar do palato de 10 sapos, observada 5, 10, 15, 20, 25, 30 e 35 minutos após a sua imersão em uma solução de peróxido de hidrogênio com concentração de $16 \mu\text{M}$. Para efeitos didáticos, admitamos que o objetivo seja fazer a previsão da velocidade relativa para sapos específicos.

Na Figura 6.6 apresentamos o gráfico de perfis correspondente que sugere que um polinômio de 2º grau poderia representar a evolução da velocidade relativa individual (condicional) ao longo do tempo. A linha pontilhada em destaque descreve a velocidade relativa média em cada instante de tempo.

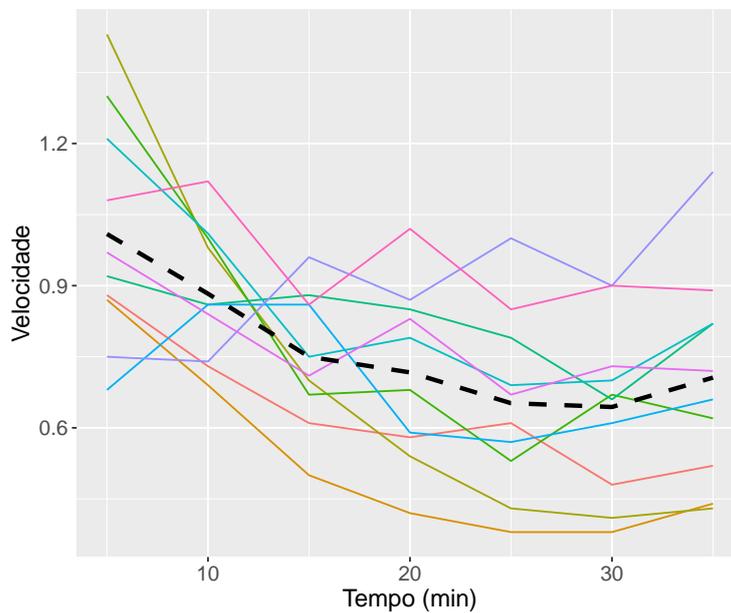


Figura 6.6: Perfis de velocidade relativa de transporte mucociliar ao longo do tempo (linha pontilhada: perfil médio)

Assim como no exemplo anterior, é possível observar um “paralelismo” nas curvas individuais, mesmo com um número pequeno de observações por indivíduo ($m = 7$). Esse comportamento sugere a inclusão de efeitos aleatórios no modelo.

Analogamente ao exemplo anterior, na Figura 6.7 apresentamos o gráfico do desenhista (*Draftman’s plot*) dos dados e na Figura 6.8 mostramos o gráfico de correlação de Grady-Helms, que não sugere a presença de correlação serial nos dados (uma vez que temos um número reduzido de observações por indivíduo).

Com base na análise descritiva e no objetivo do estudo, consideramos os seguintes modelos candidatos:

$$\mathcal{M}_1 : y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + \beta_2 t_{ij}^2 + e_{ij}$$

$$\mathcal{M}_2 : y_{ij} = (\beta_0 + b_{0i}) + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + e_{ij}$$

$$\mathcal{M}_3 : y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \beta_2 t_{ij}^2 + e_{ij}$$

$$\mathcal{M}_4 : y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + (\beta_2 + b_{2i})t_{ij}^2 + e_{ij}$$

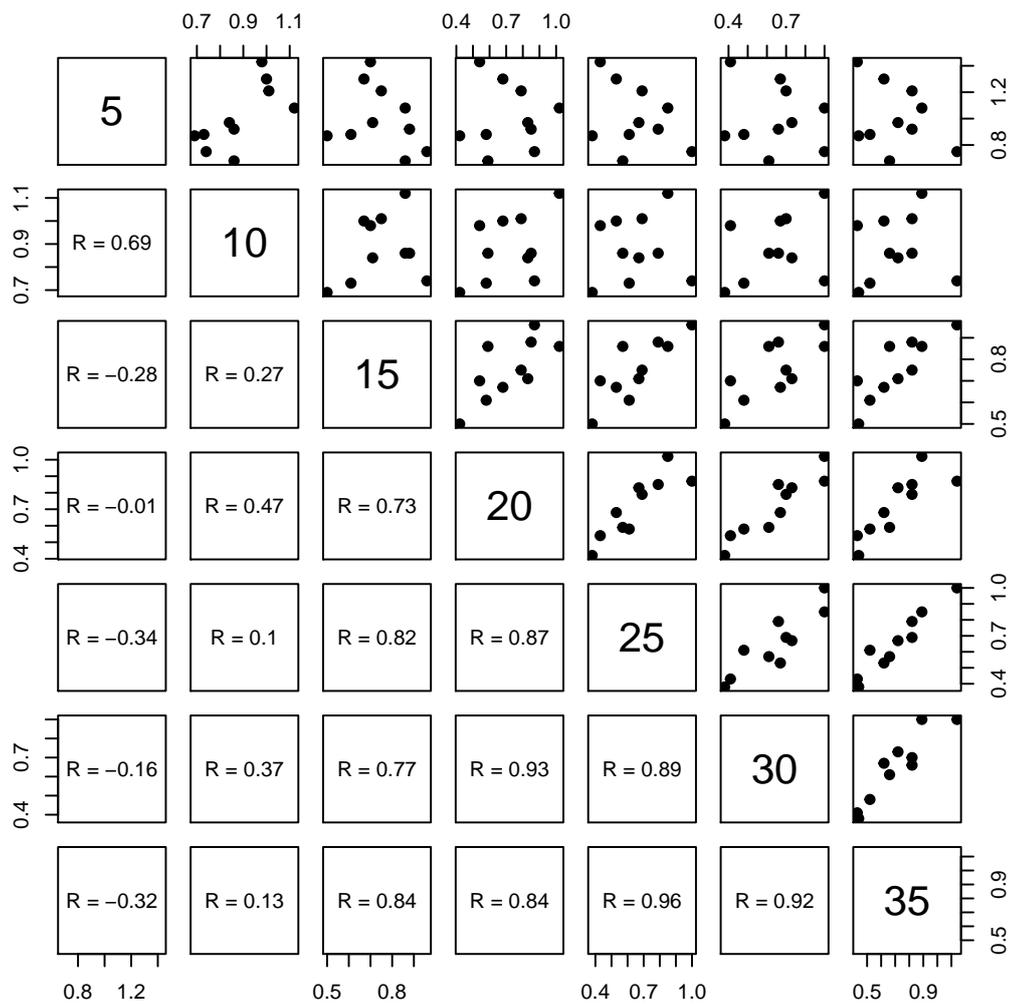


Figura 6.7: Gráfico do desenhista da velocidade ao longo do tempo (minutos na diagonal principal, correlações na diagonal inferior)

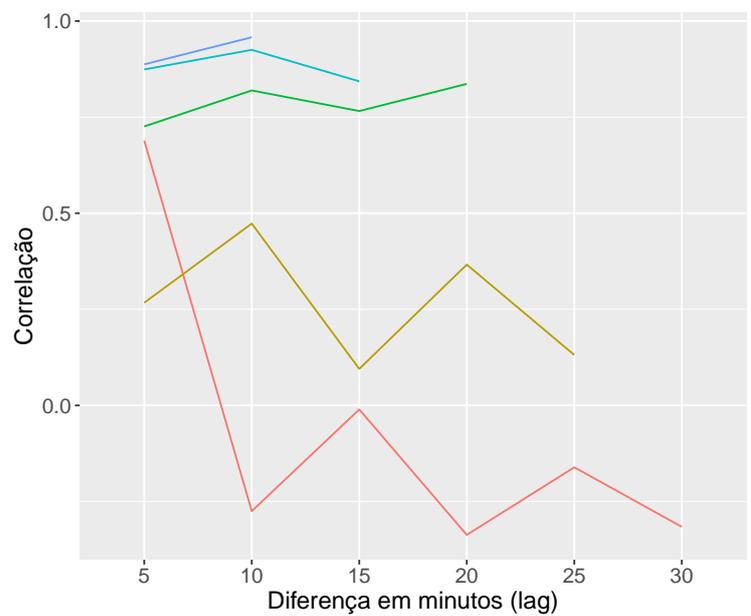


Figura 6.8: Gráfico de correlação de Grady-Helms

em que y_{ij} é a velocidade relativa de transporte do i -ésimo sapo no j -ésimo instante, $i = 1, 2, \dots, 10$, $j = 1, \dots, 7$, β_0 , β_1 e β_2 são coeficientes de regressão populacionais fixos, $e_{ij} \sim \mathcal{N}(0, \sigma^2)$ independentes, $b_{0i} \sim \mathcal{N}(0, \tau_0^2)$, $b_{1i} \sim \mathcal{N}(0, \tau_1^2)$, $b_{2i} \sim \mathcal{N}(0, \tau_2^2)$, $\text{Corr}(b_{0i}, b_{1i}) = \rho_{01}$, $\text{Corr}(b_{0i}, b_{2i}) = \rho_{02}$ e $\text{Corr}(b_{1i}, b_{2i}) = \rho_{12}$, com $(b_{0i}, b_{1i}, b_{2i})^\top$ independente de e_{ij} .

A Tabela 6.2 apresenta os resultados do ajuste dos 4 modelos candidatos obtidos por Máxima Verossimilhança. Utilizamos o pacote computacional `nlme()` do repositório R [Pinheiro et al. (2019)] e a codificação da análise é apresentada no Apêndice D.

Tabela 6.2: Estimativas (e erros padrões) dos modelos candidatos ajustados

Parâmetro	Modelo			
	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4
β_0	1,19 (0,06)	1,19 (0,08)	1,19 (0,07)	1,19 (0,11)
β_1	-2,35 (0,44)	-2,35 (0,47)	-2,35 (0,36)	-2,35 (0,66)
β_2	2,57 (0,62)	2,57 (0,70)	2,57 (0,46)	2,57 (0,72)
τ_0^2	-	0,013	0,030	0,113
τ_1^2	0,122	-	0,347	3,707
τ_2^2	-	-	-	3,957
ρ_{01}	-	-	-0,797	-0,955
ρ_{02}	-	-	-	0,967
ρ_{12}	-	-	-	-0,999
σ^2	0,015	0,020	0,008	0,005
λ	7,9	0,6	45,0	1.515,2
Nº de parâmetros	5	5	7	10
Nº (efetivo) de parâmetros	11,9	11,4	19,9	20,9
$mAIC$	-61,7	-49,3	-76,5	-92,8
BIC	-50,5	-38,1	-60,7	-70,4
$cAIC$	-78,3	-61,9	-113,8	-147,4
$\Delta mAIC_i$	31,1	43,5	16,4	0
ΔBIC_i	19,9	32,3	9,6	0
$\Delta cAIC_i$	69,1	85,5	33,5	0

λ : variância total dos efeitos aleatórios / variância dos erros aleatórios

Como o objetivo é estimar a velocidade relativa individual (*condicional*), comparamos os modelos por

meio do $cAIC$ (3.20) ou das diferenças $\Delta cAIC_i$. Por outro lado, se o objetivo fosse selecionar o melhor modelo para estimar a velocidade relativa esperada populacional (*marginal*), a comparação seria feita por meio do $mAIC$ (3.8) e do BIC (4.11) ou, alternativamente, por meio das diferenças $\Delta mAIC_i$ e ΔBIC_i (3.6).

Nesse caso, \mathcal{M}_4 é o modelo com menor valor de AIC condicional entre os modelos candidatos e apresentou considerável diferença $\Delta cAIC$ em relação aos valores dos demais modelos, sobretudo \mathcal{M}_2 ($\Delta cAIC_{\mathcal{M}_2} = 85,5$).

Na Figura 6.9 apresentamos os gráficos de resíduos para o modelo selecionado (\mathcal{M}_4). Observamos que os resíduos não apresentam tendências nem variabilidade expressivamente diferente em relação aos valores ajustados, e apresentam poucos valores atípicos (apenas 6% dos resíduos marginais e 4% dos resíduos condicionais). Além disso, a suposição de normalidade dos erros aleatórios parece ser satisfeita, como mostram o $QQplot$ e o histograma dos resíduos condicionais. Em relação aos efeitos aleatórios, observamos que a hipótese de normalidade parece satisfeita, conforme mostra o $QQPlot$ da distância de Mahalanobis dos efeitos aleatórios estimados, e não foram observados indivíduos atípicos. Consequentemente, levando em conta os resultados da análise descritiva, da análise de diagnóstico e dos critérios de informação, concluímos que \mathcal{M}_4 é o modelo selecionado. Adicionalmente, nas Figuras B.5, B.6 e B.7 (em anexo) apresentamos os gráficos de resíduos dos demais modelos candidatos.

Neste exemplo, apesar da limitação no tamanho amostral ($n = 10$) e no número de observações por indivíduo ($m = 7$), observamos que, em \mathcal{M}_4 , a razão de variâncias estimada ($\hat{\lambda} = 1.515,2$) é expressivamente alta. Mostramos no Capítulo 5 que, em geral, quanto maior o valor de λ , maior a proporção de seleções corretas, sobretudo para o $cAIC$ (Tabela 5.4).

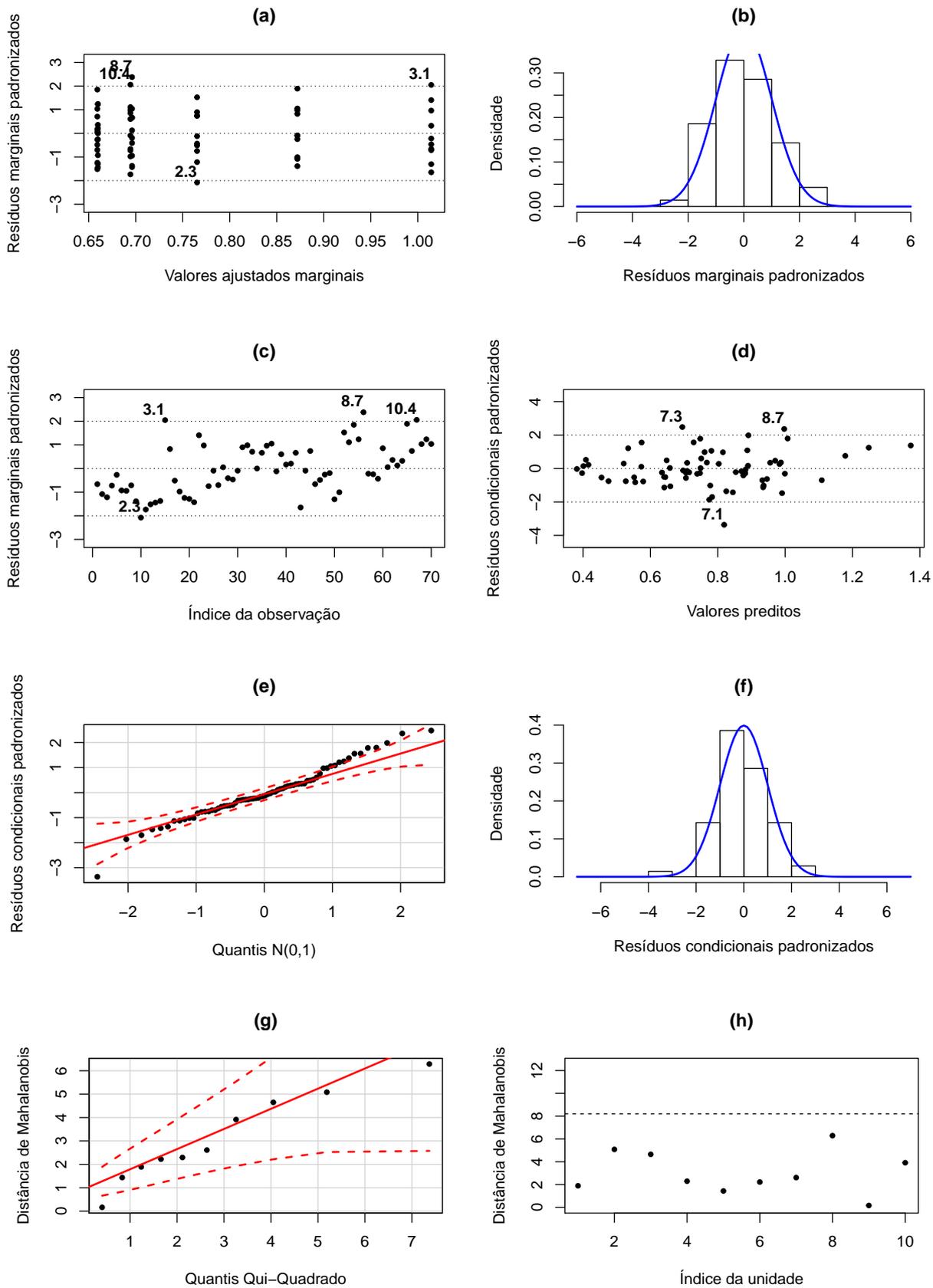


Figura 6.9: Análise de resíduos para o modelo M_4 : (a) Valores ajustados marginais versus Resíduos marginais padronizados, (b) Histograma dos resíduos marginais padronizados, (c) Resíduo marginal padronizado por observação, (d) Valores preditos versus Resíduos condicionais padronizados, (e) QQplot dos resíduos condicionais padronizados, (f) Histograma dos resíduos condicionais padronizados, (g) QQplot da distância de Mahalanobis, (h) Distância de Mahalanobis por unidade

Capítulo 7

Conclusão

CrITÉRIOS de informaÇo so ferramentas de seleÇo de modelos especialmente teis e indicadas quando a quantidade de modelos candidatos  elevada, uma vez que permitem a comparaÇo da adequabilidade dos modelos aos dados de maneira simples e direta. Entretanto, apresentam limitaÇes ocasionadas pela forma como foram definidos ou pela complexidade dos modelos que buscam selecionar. Especialmente no caso da seleÇo de modelos mistos, vimos que alguns fatores influenciam o poder de seleÇo do *AIC* (marginal e condicional) e do *BIC*, como tamanho amostral reduzido, estruturas de correlaÇo mais complexas ou o tipo de abordagem do modelo misto selecionado (marginal ou condicional).

Os estudos de simulaÇo que realizamos mostraram que, no caso de independncia condicional, o *AIC* marginal, o *AIC* condicional e o *BIC* apresentaram frequncia de seleÇo do modelo correto elevada mesmo para tamanho amostral reduzido ($n = 10$), desde que a quantidade de observaÇes por indivduo seja elevada ($m \geq 13$). Alm disso, verificamos que, quanto maior o quociente entre a varincia total dos efeitos aleatrios e a varincia dos erros aleatrios, maior a porcentagem de seleÇes corretas. Em comparaÇo, no contexto de erros condicionalmente autorregressivos, as simulaÇes mostraram que  necessrio tanto tamanho amostral quanto nmero de observaÇes por indivduo maiores ($n \geq 30$ e $m \geq 16$, respectivamente) para detecÇo do modelo correto com porcentagem considervel, mas que esta porcentagem decresce com o aumento do coeficiente de correlaÇo autorregressivo. Portanto, quanto mais correlacionados os dados, menor  a frequncia de escolhas do modelo correto pelos crITÉrios de informaÇo estudados.

A anlise dos dois exemplos mostrou que, na prtica,  aconselhvel combinar diferentes tcnicas para seleÇo de modelos. No exemplo da subseÇo 6.1, vimos que os crITÉrios de informaÇo (*AIC* marginal e *BIC*) sugerem a mesma deciso que a anlise de resduos, ou seja, ambas as tcnicas validaram o mesmo modelo como o melhor entre os candidatos. A deciso foi sustentada, tambm, por estimativas significantes de alguns parmetros — como o coeficiente de autorregresso —, pelo tamanho amostral ($n = 28$) alm do nmero de observaÇes por indivduo ($m = 14$). Da mesma forma, no exemplo da subseÇo 6.2, tanto a anlise dos resduos como o *AIC* condicional indicaram um nico modelo como o melhor, mesmo com

tamanho amostral ($n = 10$) como o número de observações por indivíduo ($m = 7$) relativamente pequenos. Nesse caso, a magnitude da estimativa da razão entre a variância total dos efeitos aleatórios e a variância dos erros aleatórios foi um fator importante para a consolidação da escolha do melhor modelo.

Em resumo, o cenário em que o tamanho amostral e o número de observações por indivíduo são ambos pequenos é o menos indicado para a utilização dos critérios de informação. Por isso, é sempre conveniente utilizá-los juntamente com técnicas descritivas e de diagnóstico.

Apêndice A

Tabelas

Tabela A.1: *Peso de bezerros (em kg)*

Bezerro	Semana													
	0	2	4	6	8	10	12	14	16	18	20	22	24	26
1	28,0	32,7	36,9	41,7	42,8	50,3	53,5	48,2	54,8	61,0	64,4	74,2	85,4	93,7
2			45,2	44,9	47,8	55,5	57,3	64,0	63,7	65,5	70,7	79,2	82,7	91,4
3	31,0	30,6	32,5	32,5	35,2	40,5	47,1	56,2	68,3	80,0	93,0	105,0	118,0	130,4
4	36,0	39,7	42,1	47,2	49,2	51,9	60,8	69,8	81,0	89,5	97,0	105,4	115,6	123,5
5	29,0	32,9	36,6	38,0	44,8	54,2	58,1	65,8	73,2	81,5	92,0	103,7	112,6	123,3
6	28,0	35,1	32,9	41,1	52,0	50,1	51,3	56,8	64,9	72,9	79,0	83,4	93,2	102,0
7	40,0	42,0	46,8	52,1	54,9	65,2				104,7	110,6	124,7	136,8	153,4
8	42,0	45,8		53,5	59,2	71,4	71,2	74,9	83,5	88,5		102,7	111,6	119,7
9	30,0	34,0	35,4	35,9	38,0	48,9	50,0	62,1	70,5	83,0	91,0	103,9	116,5	127,0
10	43,0	43,7	48,2	47,4	48,2	54,7	58,0	62,7	77,4	82,8	97,1	107,1	115,3	127,0
11	31,0	31,6	33,4	40,9	45,9	46,5	55,0	59,8	66,2	76,9	85,8	93,0	103,3	112,5
12	27,0	31,5	38,0	42,6	47,0	56,7	64,8	74,7	84,1	94,9	106,6	122,3	130,6	141,5
13	32,0	35,9	42,5	48,9	54,5	67,8	71,2	84,7	93,9	102,2	116,0	126,2	138,2	150,0
14	34,0	34,4	39,1	41,3	43,3	51,1	54,1	65,4	75,1	87,9	98,0	108,7	124,2	131,3
15	41,0	42,4	47,5	58,3	69,5	79,4	91,7	104,0	119,2	133,1	145,4	156,5	167,2	176,8
16	31,9	34,1	39,7	46,3	50,0	58,7	64,2	81,0	91,5	106,9	117,1	127,7	144,2	154,9
17	40,0	41,8	50,0	55,5	58,6	63,5	70,3	80,0	90,0	102,6	101,2	120,4	130,9	137,1
18	31,0	30,7	34,1	44,3	47,8	59,4	68,3	77,2	84,2	96,2	104,1	114,0	123,0	132,0
19	35,0	38,6	39,7	40,2	37,1	43,1	43,9	48,1	58,3	68,6	78,5	86,8	99,9	106,2
20	33,0	40,7	46,3	55,9	67,2	74,6	87,4	95,4	110,5	122,5	127,0	136,3	144,8	151,5
21	37,0	36,6								103,6	110,7	120,0	126,7	132,2
22	36,0	33,3	35,0	39,9	44,5	46,5	50,5	55,0	59,1	68,9	78,2	75,1	79,0	77,0
23	30,0	35,8	41,2	52,0	64,0	74,9	91,0	95,5				148,0	154,5	167,6
24	38,5	44,5	47,1	54,3	60,9	74,9	83,3	89,7	99,7	110,0	120,8	135,1	141,5	157,0
25	30,0	37,2	43,9	49,0	52,8	60,2	76,3	80,8	94,2	102,6	111,0	115,6	121,4	134,5
26	32,7	35,9		40,2		52,5		61,1		80,9	93,0	100,1	103,2	108,0
27	32,0	35,5	39,2	43,7	51,8	63,4	76,1	81,1	84,6	89,8	97,4	111,0	120,2	134,2
28	32,0	34,7	36,1	39,5	45,2	53,0	56,6	63,7	70,1	74,4	85,1	90,2	96,1	103,6

Tabela A.2: *Velocidade relativa de transporte mucociliar*

Sapo	Tempo (min)						
	5	10	15	20	25	30	35
1	0,88	0,73	0,61	0,58	0,61	0,48	0,52
2	0,87	0,69	0,50	0,42	0,38	0,38	0,44
3	1,43	0,98	0,70	0,54	0,43	0,41	0,43
4	1,30	1,00	0,67	0,68	0,53	0,67	0,62
5	0,92	0,86	0,88	0,85	0,79	0,66	0,82
6	1,21	1,01	0,75	0,79	0,69	0,70	0,82
7	0,68	0,86	0,86	0,59	0,57	0,61	0,66
8	0,75	0,74	0,96	0,87	1,00	0,90	1,14
9	0,97	0,84	0,71	0,83	0,67	0,73	0,72
10	1,08	1,12	0,86	1,02	0,85	0,90	0,89

Apêndice B

Figuras

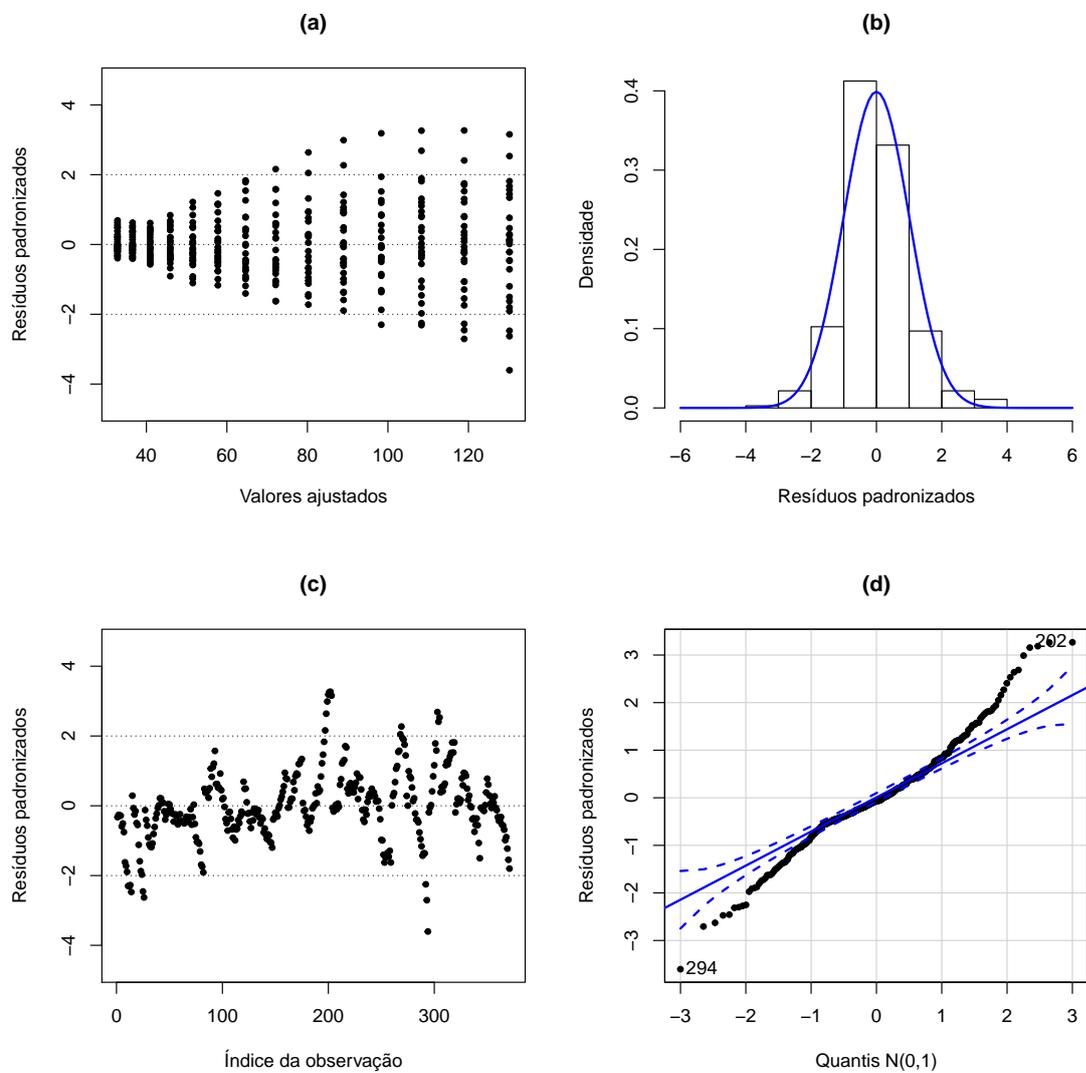


Figura B.1: Análise de resíduos para o modelo M_1 (Exemplo 1): (a) Valores ajustados versus Resíduos padronizados, (b) Histograma dos resíduos padronizados, (c) Resíduo padronizado por observação, (d) QQplot dos resíduos padronizados

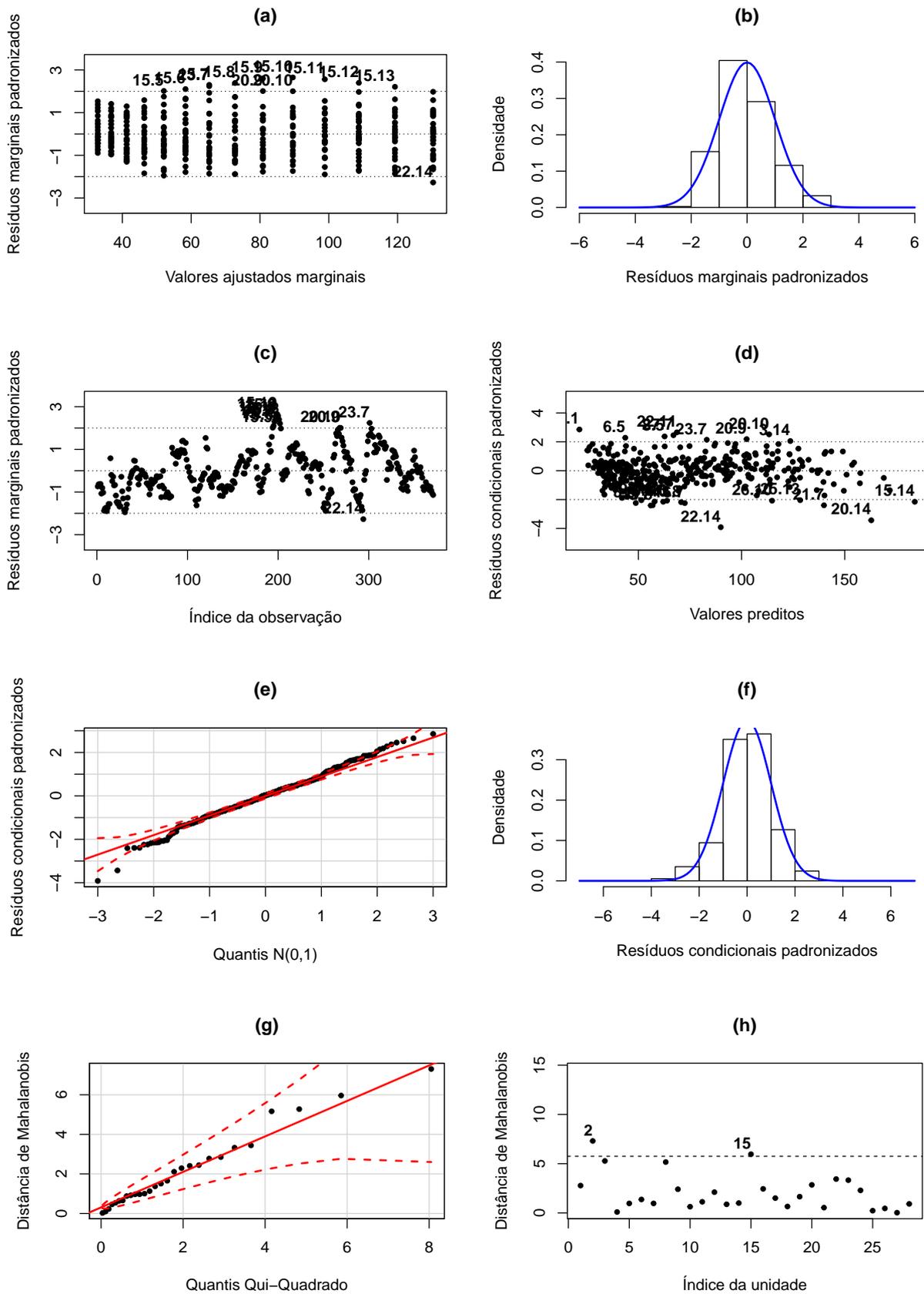


Figura B.3: Análise de resíduos para o modelo M_3 (Exemplo 1): (a) Valores ajustados marginais versus Resíduos marginais padronizados, (b) Histograma dos resíduos marginais padronizados, (c) Resíduo marginal padronizado por observação, (d) Valores preditos versus Resíduos condicionais padronizados, (e) QQplot dos resíduos condicionais padronizados, (f) Histograma dos resíduos condicionais padronizados, (g) QQplot da distância de Mahalanobis, (h) Distância de Mahalanobis por unidade

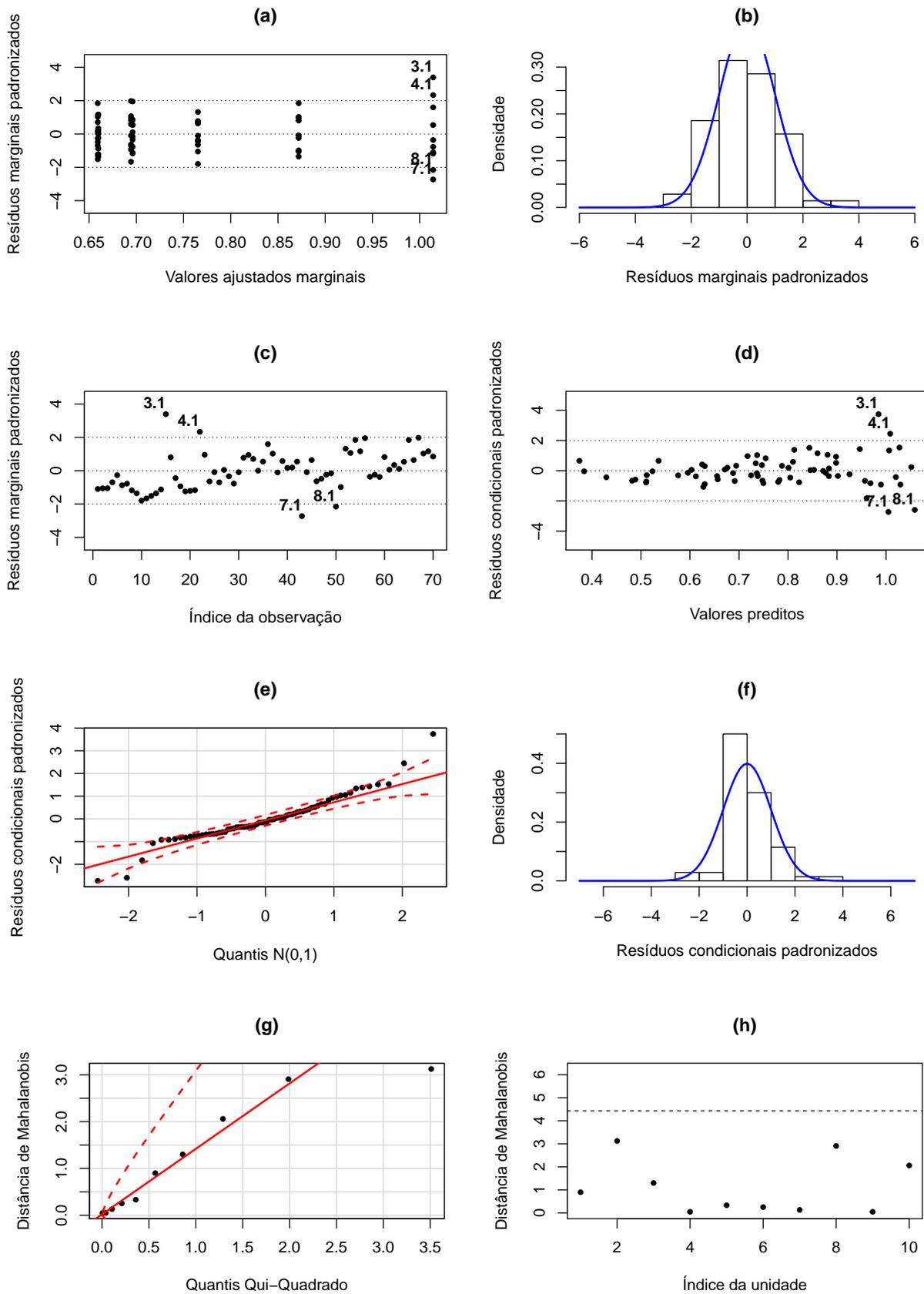


Figura B.5: Análise de resíduos para o modelo M_1 (Exemplo 2): (a) Valores ajustados marginais versus Resíduos marginais padronizados, (b) Histograma dos resíduos marginais padronizados, (c) Resíduo marginal padronizado por observação, (d) Valores preditos versus Resíduos condicionais padronizados, (e) QQplot dos resíduos condicionais padronizados, (f) Histograma dos resíduos condicionais padronizados, (g) QQplot da distância de Mahalanobis, (h) Distância de Mahalanobis por unidade

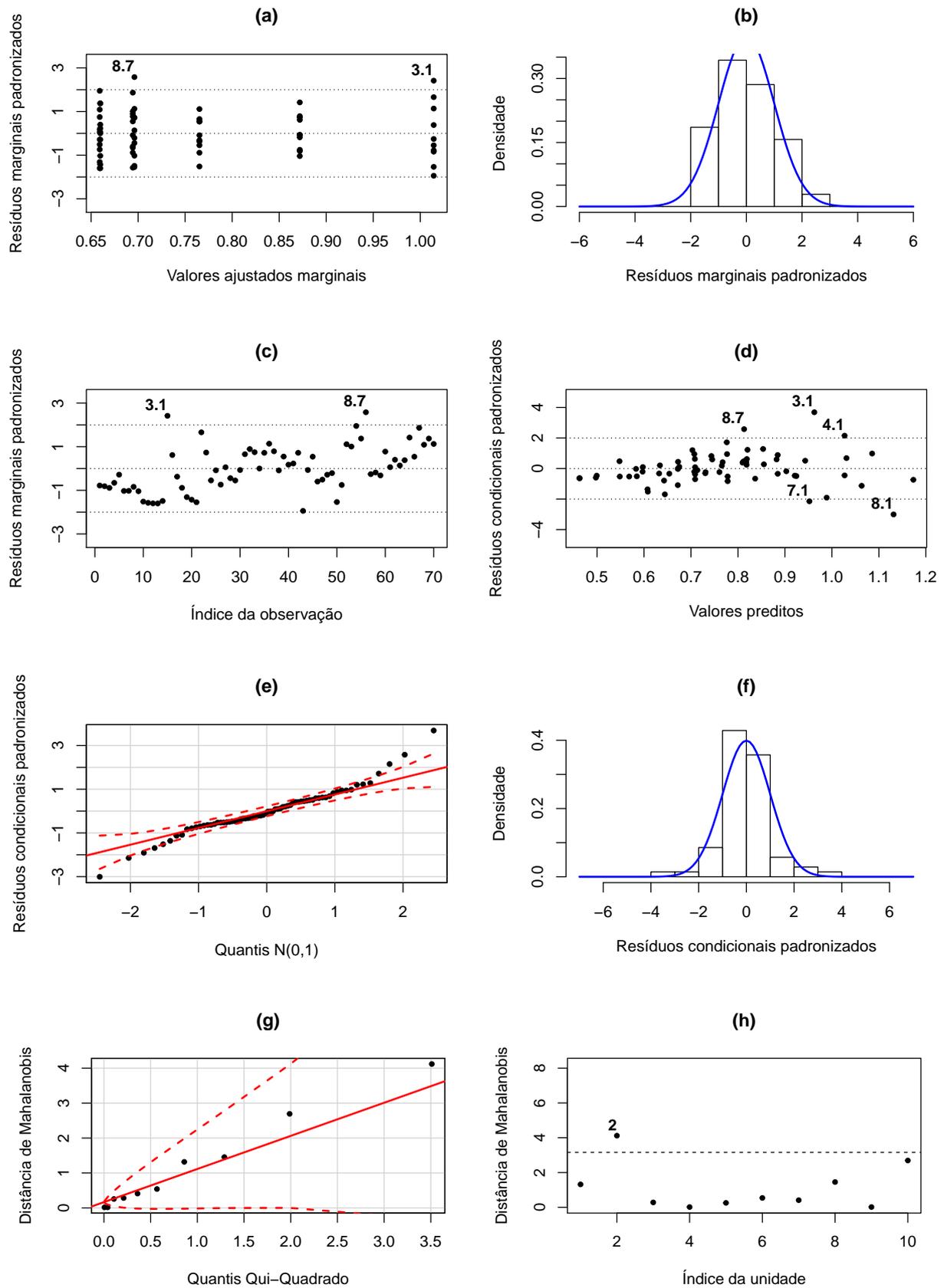


Figura B.6: Análise de resíduos para o modelo M_2 (Exemplo 2): (a) Valores ajustados marginais versus Resíduos marginais padronizados, (b) Histograma dos resíduos marginais padronizados, (c) Resíduo marginal padronizado por observação, (d) Valores preditos versus Resíduos condicionais padronizados, (e) QQplot dos resíduos condicionais padronizados, (f) Histograma dos resíduos condicionais padronizados, (g) QQplot da distância de Mahalanobis, (h) Distância de Mahalanobis por unidade

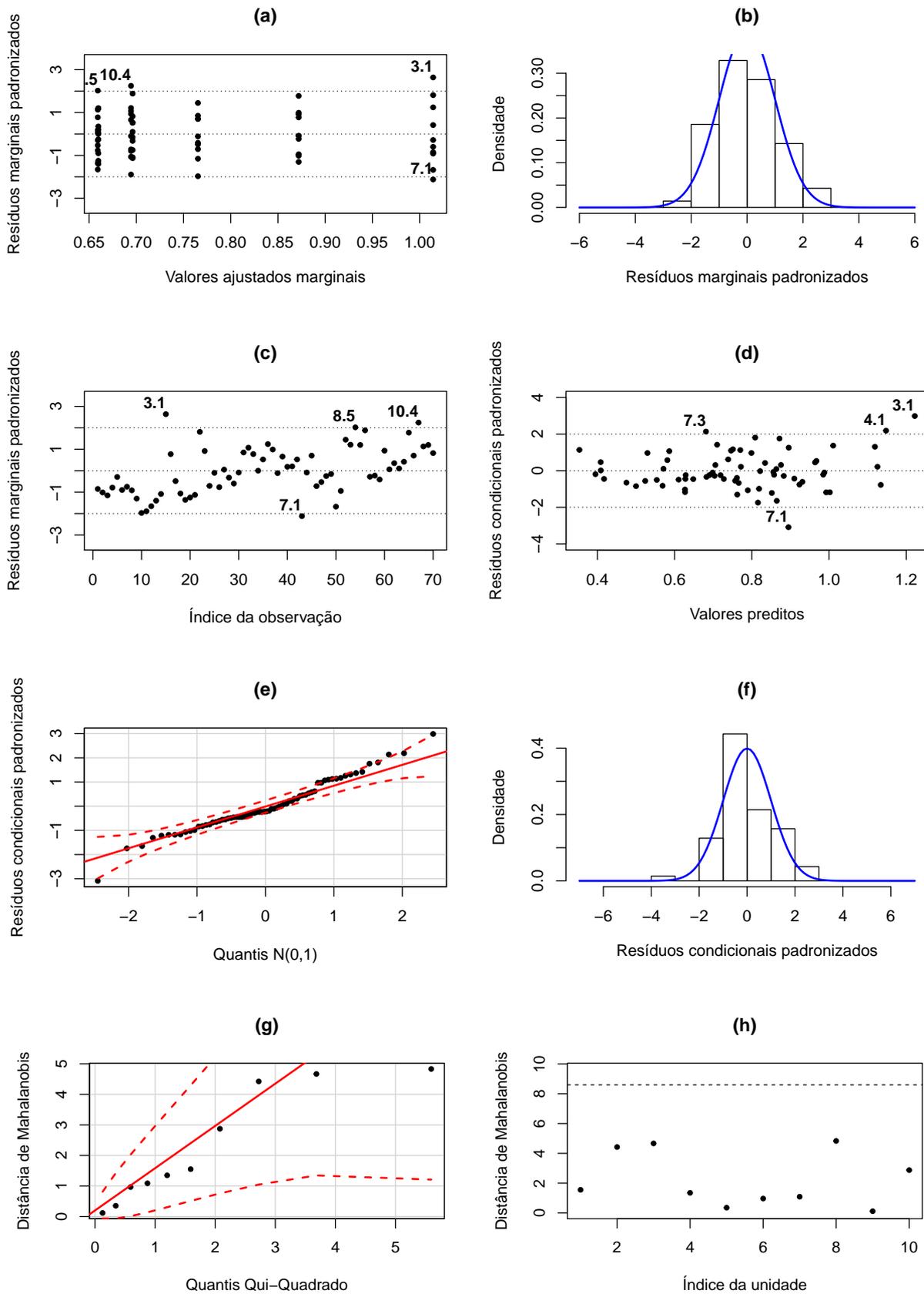


Figura B.7: Análise de resíduos para o modelo M_3 (Exemplo 2): (a) Valores ajustados marginais versus Resíduos marginais padronizados, (b) Histograma dos resíduos marginais padronizados, (c) Resíduo marginal padronizado por observação, (d) Valores preditos versus Resíduos condicionais padronizados, (e) QQplot dos resíduos condicionais padronizados, (f) Histograma dos resíduos condicionais padronizados, (g) QQplot da distância de Mahalanobis, (h) Distância de Mahalanobis por unidade

Apêndice C

Graus de liberdade efetivos

No modelo condicional, $\hat{y} = E(y|\hat{\beta}, \hat{\mathbf{b}}) = \mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\mathbf{b}}$. Com base nisso, temos:

$$\begin{aligned}\hat{y} &= \mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\mathbf{b}} \\ &= \begin{pmatrix} \mathbf{X} & \mathbf{Z} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\mathbf{b}} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{X} & \mathbf{Z} \end{pmatrix} \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \end{pmatrix} \mathbf{y} \\ &= \mathbf{H} \mathbf{y}\end{aligned}$$

em que $\mathbf{H} = \begin{pmatrix} \mathbf{X} & \mathbf{Z} \end{pmatrix} \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \end{pmatrix}$.

Utilizando a propriedade cíclica do traço do produto de três matrizes, temos que:

$$\begin{aligned}\rho &= \text{tr}\{\mathbf{H}\} \\ &= \text{tr} \left\{ \begin{pmatrix} \mathbf{X} & \mathbf{Z} \end{pmatrix} \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \end{pmatrix} \right\} \\ &= \text{tr} \left\{ \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{X} & \mathbf{Z} \end{pmatrix} \right\} \\ &= \text{tr} \left\{ \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} \end{pmatrix} \right\}.\end{aligned}$$

■

Apêndice D

Códigos

Exemplo 1 - Ajuste do modelo e análise de diagnóstico

```
require(nlme)
require(AICcmodavg)
require(car)

# Função que calcula o AIC condicional
source("AIC_condicional_Exemplo1.R")

# O arquivo da análise de diagnóstico pode ser encontrado no sítio:
# http://www.ime.usp.br/~jmsinger/GLMMdiagnostics.zip
source("residdiag_nlme.R")

### Leitura da base de dados
data_calf = read.csv("Dados_Bezerros.csv",
                    header = T, sep = ';', dec = '.')
data_calf$bezerro = as.factor(data_calf$bezerro)
# Transformar a variável 'semana' em 'meses'
data_calf$meses = (data_calf$semana)*7/30

### Ajuste dos modelos candidatos
modex1 = list()
ctrl <- lmeControl(opt='optim', maxIter=1e8, msMaxIter = 1e8)
modex1[[1]] <- lm(peso ~ 1 + meses + I(meses^2),
                 data=data_calf)
modex1[[2]] <- lme(peso ~ 1 + meses + I(meses^2),
                  random = ~ 1|bezerro, data=data_calf,
                  method="ML", control=ctrl, na.action=na.omit)
```

```

modex1[[3]] <- lme(peso ~ 1 + meses + I(meses^2),
                 random = ~ 1 + meses|bezerro, data=data_calf,
                 method="ML", control=ctrl, na.action=na.omit)
modex1[[4]] <- lme(peso ~ 1 + meses + I(meses^2),
                 random = ~ 1 + meses + I(meses^2)|bezerro,
                 data=data_calf,
                 method="ML", control=ctrl, na.action=na.omit)
modex1[[5]] <- lme(peso ~ 1 + meses + I(meses^2),
                 random = ~ 1 + meses + I(meses^2)|bezerro,
                 correlation=corCAR1(form=~meses|bezerro),
                 data=data_calf,
                 method="ML", control=ctrl, na.action=na.omit)

### Resultados

# Efeitos fixos (beta)
summary.lm(modex1[[1]])$coefficients
round(as.vector(fixed.effects(modex1[[2]])),1)
round(as.vector(fixed.effects(modex1[[3]])),1)
round(as.vector(fixed.effects(modex1[[4]])),1)
round(as.vector(fixed.effects(modex1[[5]])),1)

# Erro padrão de beta
round(sqrt(diag(vcov(modex1[[1]]))),1)
round(sqrt(diag(vcov(modex1[[2]]))),1)
round(sqrt(diag(vcov(modex1[[3]]))),1)
round(sqrt(diag(vcov(modex1[[4]]))),1)
round(sqrt(diag(vcov(modex1[[5]]))),1)

# componentes de variância
summary(modex1[[1]])$sigma^2
VarCorr(modex1[[2]])
VarCorr(modex1[[3]])
VarCorr(modex1[[4]])
VarCorr(modex1[[5]])

# Coeficiente de correlação autorregressivo (modelo 5)
as.numeric(coef(modex1[[5]])$model$corStruct, unconstrained=FALSE)[1])

```

```

# mAIC, BIC e cAIC
mAIC = unlist(lapply(modex1, AIC))
BIC = unlist(lapply(modex1, BIC))
cAIC = c(AIC(modex1[[1]]),
        cAIC1(modex1[[2]], phi=0)$cAIC,
        cAIC1(modex1[[3]], phi=0)$cAIC,
        cAIC1(modex1[[4]], phi=0)$cAIC,
        cAIC1(modex1[[5]], phi=0.614791)$cAIC)

# Número de parâmetros do modelo (p + q)
k = c(4,
      cAIC1(modex1[[2]], phi=0)$df,
      cAIC1(modex1[[3]], phi=0)$df,
      cAIC1(modex1[[4]], phi=0)$df,
      cAIC1(modex1[[5]], phi=0.614791)$df)

# Número efetivo de parâmetros do modelo (rho + 1 ou rho + 2)
k_efetivo = c(4,
              cAIC1(modex1[[2]], phi=0)$rho + 1,
              cAIC1(modex1[[3]], phi=0)$rho + 1,
              cAIC1(modex1[[4]], phi=0)$rho + 1,
              cAIC1(modex1[[5]], phi=0.614791)$rho + 2)

m1 = modex1[[1]]
m2 = modex1[[2]]
m3 = modex1[[3]]
m4 = modex1[[4]]
m5 = modex1[[5]]

### Análise de Resíduos – Modelo 1
residM1 = data.frame(index= 1:371,
                    fit1 = fitted(m1),
                    resid1 = residuals(m1)/(summary(modex1[[1]])$sigma))

# Gráfico 1
par(mfrow = c(1,2), mar = c(5.1, 4.1, 4.1, 2.1))
limit = 2
plot(residM1$fit1, residM1$resid1,
     xlab = expression(paste("Valores ajustados")), main = "(a)",

```

```

    ylab = expression(paste("Resíduos padronizados")),
    pch = 20, cex = 1.0, cex.lab = 1.0, cex.axis = 1.0,
    ylim = c(-1.3*max(abs(range(residM1$resid1))),
              1.3*max(abs(range(residM1$resid1))))
abline(h = 0, lty = 3)
abline(h = -limit, lty = 3)
abline(h = limit, lty = 3)
# Gráfico 2
hist(residM1$resid1, breaks = c(-6:6), freq = F, main = "(b)",
     ylab = "Densidade",
     xlab = expression(paste("Resíduos padronizados")),
     cex = 1.0, cex.lab = 1.0, cex.axis = 1.0)
curve(dnorm(x, mean=0, sd=1), col="blue", lwd=2, add=TRUE, yaxt="n")
# Gráfico 3
plot(residM1$resid1, xlab = "Índice da observação",
     ylab = expression(paste("Resíduos padronizados")),
     main = "(c)",
     pch = 20, cex = 1.0, cex.lab = 1.0, cex.axis = 1.0,
     ylim = c(-1.3*max(abs(range(residM1$resid1))), 1.3*max(abs(range(residM1$resid1
))))))
abline(h = 0, lty = 3)
abline(h = -limit, lty = 3)
abline(h = limit, lty = 3)
# Gráfico 4
qqPlot(residM1$resid1, ylab = "Resíduos padronizados", main = "(d)",
       xlab = "Quantis N(0,1)", pch = 20)

### Análise de Resíduos – Modelo 2
residM2<-residdiag.nlme(m2, limit=2, plotid=2:6)

### Análise de Resíduos – Modelo 3
residM3<-residdiag.nlme(m3, limit=2, plotid=2:6)

### Análise de Resíduos – Modelo 4
residM4<-residdiag.nlme(m4, limit=2, plotid=2:6)

### Análise de Resíduos – Modelo 5
residM5<-residdiag.nlme(m5, limit=2, plotid=2:6)

```

Exemplo 1 - AIC condicional

```

require(matrixcalc)

cAIC1 = function(modelo, phi) {

  beta = as.vector(fixed.effects(modelo))
  n = as.numeric(modelo[[2]]$ngrps[1])
  N = modelo[[2]]$'N'
  tempos = data_calf[!is.na(data_calf$peso),c(1,4)]
  X_temp = Z_temp = list()

  # Matrizes de especificação por indivíduo completas
  # (supondo que o modelo inclui todos os efeitos fixos/aleatórios)
  for(k in 1:n) {
    X_temp[[k]] = cbind(rep(1, length(tempos[tempos$bezerro==k,2])),
                        tempos[tempos$bezerro==k,2],
                        tempos[tempos$bezerro==k,2]^2)
    Z_temp[[k]] = X_temp[[k]]
  }

  Xi = list()
  Zi = list()

  # Selecionar as colunas das matrizes de especificação Xi
  # (de acordo com os efeitos fixos incluídos no modelo)
  if(modelo$call$fixed == "peso ~ 1 + meses + I(meses^2)") {
    for(k in 1:n) {
      Xi[[k]] = X_temp[[k]][,c(1,2,3)]
    }
  } else if (modelo$call$fixed == "peso ~ 1 + meses") {
    for(k in 1:n) {
      Xi[[k]] = X_temp[[k]][,c(1,2)]
    }
  } else if (modelo$call$fixed == "peso ~ 1 + I(meses^2)") {
    for(k in 1:n) {
      Xi[[k]] = X_temp[[k]][,c(1,3)]
    }
  } else if (modelo$call$fixed == "peso ~ 0 + meses + I(meses^2)") {

```

```

for(k in 1:n) {
  Xi[[k]] = X_temp[[k]][,c(2,3)]
}
} else if (modelo$call$fixed == "peso ~ 0 + I(meses^2)") {
for(k in 1:n) {
  Xi[[k]] = as.matrix(X_temp[[k]][,3])
}
} else if (modelo$call$fixed == "peso ~ 0 + meses") {
for(k in 1:n) {
  Xi[[k]] = as.matrix(X_temp[[k]][,2])
}
} else if (modelo$call$fixed == "peso ~ 1") {
for(k in 1:n) {
  Xi[[k]] = as.matrix(X_temp[[k]][,1])
}
}

# Selecionar as colunas das matrizes de especificação Zi
# (de acordo com os efeitos aleatórios incluídos no modelo)
if(modelo$call$random == "~1 + meses + I(meses^2) | bezerro") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = Z_temp[[k]][,c(1,2,3)]
}
} else if (modelo$call$random == "~1 + meses | bezerro") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = Z_temp[[k]][,c(1,2)]
}
} else if (modelo$call$random == "~1 + I(meses^2) | bezerro") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = Z_temp[[k]][,c(1,3)]
}
} else if (modelo$call$random == "~0 + meses + I(meses^2) | bezerro") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = Z_temp[[k]][,c(2,3)]
}
} else if (modelo$call$random == "~0 + I(meses^2) | bezerro") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = as.matrix(Z_temp[[k]][,3])
}
}

```

```

} else if (modelo$call$random == "~0 + meses | bezerro") {
  for(k in 1:length(X_temp)) {
    Zi[[k]] = as.matrix(Z_temp[[k]][,2])
  }
} else if (modelo$call$random == "~1 | bezerro") {
  for(k in 1:length(X_temp)) {
    Zi[[k]] = as.matrix(Z_temp[[k]][,1])
  }
}

Ri_temp = getVarCov(modelo, type="conditional")[[1]]
Gi = getVarCov(modelo, type='random.effects') # Dimensão de Gi é fixa
Gi_inv = solve(Gi)
Ri = Vi = list() # Dimensão de Ri e Vi varia entre indivíduos
Ri_inv = Vi_inv = list()

# redimensionar as tabelas para levar em conta as observações faltantes
for(k in 1:n) {

  obs = Xi[[k]][,2]*30/14 + 1 # observações presentes por indivíduo
  Ri[[k]] = Ri_temp[obs,obs]
  Vi[[k]] = Zi[[k]]%*%Gi%*%t(Zi[[k]]) + Ri[[k]]
  Ri_inv[[k]] = solve(Ri[[k]])
  Vi_inv[[k]] = solve(Vi[[k]])
}

R_inv = Reduce(direct.sum, Ri_inv)
V_inv = Reduce(direct.sum, Vi_inv)

# Blocos da matriz utilizada para estimar rho
A_temp = B_temp = C_temp = D_temp = E_temp = list()

for(k in 1:n) {
  A_temp[[k]] = t(Xi[[k]])%*%Ri_inv[[k]]%*%Xi[[k]]
  B_temp[[k]] = t(Xi[[k]])%*%Ri_inv[[k]]%*%Zi[[k]]
  C_temp[[k]] = t(Zi[[k]])%*%Ri_inv[[k]]%*%Xi[[k]]
  D_temp[[k]] = t(Zi[[k]])%*%Ri_inv[[k]]%*%Zi[[k]]+Gi_inv
  E_temp[[k]] = t(Zi[[k]])%*%Ri_inv[[k]]%*%Zi[[k]]
}

```

```

A = Reduce( '+', A_temp)
B = Reduce( cbind, B_temp)
C = Reduce( rbind, C_temp)
D = Reduce( direct.sum, D_temp)
E = Reduce( direct.sum, E_temp)

rhoA = rbind( cbind(A,B), cbind(C,D))
rhoB = rbind( cbind(A,B), cbind(C,E))

# Inversa de uma matriz de blocos
D_orig_inv = solve(D)
A_inv = solve(A-B%D_orig_inv%C)
B_inv = -A_inv%B%D_orig_inv
C_inv = -D_orig_inv%C%A_inv
D_inv = D_orig_inv+D_orig_inv%C%A_inv%B%D_orig_inv
rhoA_inv = rbind( cbind(A_inv, B_inv), cbind(C_inv, D_inv))

# Simplificar o cálculo do determinante
logdetV = sum(unlist(lapply(lapply(Vi, FUN=det), FUN=log)))
logdetR = sum(unlist(lapply(lapply(Ri, FUN=det), FUN=log)))

resMarg = residuals(modelo, level=0)
resCond = residuals(modelo, level=1)

logLik_m = -0.5*(N*log(2*pi) + logdetV + t(resMarg)%V_inv%resMarg)
logLik_c = -0.5*(N*log(2*pi) + logdetR + t(resCond)%R_inv%resCond)
rho = sum(diag(rhoA_inv%rhoB))

p = length(beta)
q = ifelse(phi==0,
           nrow(Gi)+choose(nrow(Gi), 2) + 1,
           nrow(Gi)+choose(nrow(Gi), 2) + 2)
mAIC = -2*logLik_m + 2*(p+q)
mAICc = -2*logLik_m + 2*(N/(N-p-q-1))*(p+q)
BIC = -2*logLik_m + log(N)*(p+q)

cAIC = ifelse(phi==0,
              -2*logLik_c + 2*(rho+1),

```

```

        -2*logLik_c + 2*(rho+2))
cAICc = ifelse(phi==0,
              -2*logLik_c + 2*(N*(N-p-1)*(rho+1)+N*(p+1))/((N-p)*(N-p-2)),
              -2*logLik_c + 2*(N*(N-p-1)*(rho+1)+N*(p+2))/((N-p)*(N-p-2)))

return (list(logLik_m=logLik_m, logLik_c=logLik_c,
            p=round(p), q=round(q), df=round(p+q), rho=rho,
            mAIC=mAIC, mAICc=mAICc, BIC=BIC, cAIC=cAIC, cAICc=cAICc))
}

```

Exemplo 2 - Ajuste do modelo e análise de diagnóstico

```

require(nlme)
require(AICcmodavg)
require(car)

# Função que calcula o AIC condicional
source("AIC_condicional_Exemplo2.R")

# O arquivo da análise de diagnóstico pode ser encontrado no sítio:
# http://www.ime.usp.br/~jmsinger/GLMMdiagnostics.zip
source("residdiag_nlme.R")

### Leitura da base de dados
data_frog = read.csv("Dados_Sapos.csv",
                    header = T, sep = ';', dec = '.')
data_frog$sapo = as.factor(data_frog$sapo)

### Ajuste dos modelos candidatos
modex2 = list()
ctrl <- lmeControl(opt='optim', maxIter=1e8, msMaxIter = 1e8)
modex2[[1]] <- lme(speed ~ 1 + time_h + I(time_h^2),
                  random = ~ 0+time_h|sapo, data=data_frog,
                  method="ML", control=ctrl, na.action=na.omit)
modex2[[2]] <- lme(speed ~ 1 + time_h + I(time_h^2),
                  random = ~ 1|sapo, data=data_frog,
                  method="ML", control=ctrl, na.action=na.omit)
modex2[[3]] <- lme(speed ~ 1 + time_h + I(time_h^2),
                  random = ~ 1 + time_h|sapo, data=data_frog,

```

```

        method="ML", control=ctrl, na.action=na.omit)
modex2[[4]] <- lme(speed ~ 1 + time_h + I(time_h^2),
                 random = ~ 1 + time_h + I(time_h^2)|sapo, data=data_frog,
                 method="ML", control=ctrl, na.action=na.omit)

### Resultados

# Efeitos fixos (beta)
round(as.vector(fixed.effects(modex2[[1]])),2)
round(as.vector(fixed.effects(modex2[[2]])),2)
round(as.vector(fixed.effects(modex2[[3]])),2)
round(as.vector(fixed.effects(modex2[[4]])),2)

# Erro padrão de beta
round(sqrt(diag(vcov(modex2[[1]]))),2)
round(sqrt(diag(vcov(modex2[[2]]))),2)
round(sqrt(diag(vcov(modex2[[3]]))),2)
round(sqrt(diag(vcov(modex2[[4]]))),2)

# componentes de variância
VarCorr(modex2[[1]])
VarCorr(modex2[[2]])
VarCorr(modex2[[3]])
VarCorr(modex2[[4]])

# mAIC, BIC e cAIC
mAIC = unlist(lapply(modex2, AIC))
BIC = unlist(lapply(modex2, BIC))
cAIC = c(cAIC2(modex2[[1]])$cAIC,
        cAIC2(modex2[[2]])$cAIC,
        cAIC2(modex2[[3]])$cAIC,
        cAIC2(modex2[[4]])$cAIC)

# Número de parâmetros do modelo (p + q)
k = c(cAIC2(modex2[[1]])$df,
      cAIC2(modex2[[2]])$df,
      cAIC2(modex2[[3]])$df,
      cAIC2(modex2[[4]])$df)

```

```

# Número efetivo de parâmetros do modelo (rho + 1 ou rho + 2)
k_efetivo = c(cAIC2(modex2[[1]])$rho + 1,
              cAIC2(modex2[[2]])$rho + 1,
              cAIC2(modex2[[3]])$rho + 1,
              cAIC2(modex2[[4]])$rho + 1)

### Análise de Resíduos – Modelo 1
residM1<-residdiag.nlme(modex2[[1]], limit=2, plotid=2:6)

### Análise de Resíduos – Modelo 2
residM2<-residdiag.nlme(modex2[[2]], limit=2, plotid=2:6)

### Análise de Resíduos – Modelo 3
residM3<-residdiag.nlme(modex2[[3]], limit=2, plotid=2:6)

### Análise de Resíduos – Modelo 4
residM4<-residdiag.nlme(modex2[[4]], limit=2, plotid=2:6)

```

Exemplo 2 - AIC condicional

```

require(matrixcalc)

cAIC2 = function(modelo) {

  beta = as.vector(fixed.effects(modelo))
  n = as.numeric(modelo[[2]]$ngrps[1])
  N = modelo[[2]]$'N'
  tempos = data_frog[!is.na(data_frog$speed), c(1,4)]
  X_temp = Z_temp = list()

  # Matrizes de especificação por indivíduo completas
  # (supondo que o modelo inclui todos os efeitos fixos/aleatórios)
  for(k in 1:n) {
    X_temp[[k]] = cbind(rep(1, length(tempos[tempos$sapo==k,2])),
                       tempos[tempos$sapo==k,2],
                       tempos[tempos$sapo==k,2]^2)
    Z_temp[[k]] = X_temp[[k]]
  }
}

```

```

Xi = list()
Zi = list()

# Selecionar as colunas das matrizes de especificação Xi
# (de acordo com os efeitos fixos incluídos no modelo)
if(modelo$call$fixed == "speed ~ 1 + time_h + I(time_h^2)") {
  for(k in 1:n) {
    Xi[[k]] = X_temp[[k]][,c(1,2,3)]
  }
} else if (modelo$call$fixed == "speed ~ 1 + time_h") {
  for(k in 1:n) {
    Xi[[k]] = X_temp[[k]][,c(1,2)]
  }
} else if (modelo$call$fixed == "speed ~ 1 + I(time_h^2)") {
  for(k in 1:n) {
    Xi[[k]] = X_temp[[k]][,c(1,3)]
  }
} else if (modelo$call$fixed == "speed ~ 0 + time_h + I(time_h^2)") {
  for(k in 1:n) {
    Xi[[k]] = X_temp[[k]][,c(2,3)]
  }
} else if (modelo$call$fixed == "speed ~ 0 + I(time_h^2)") {
  for(k in 1:n) {
    Xi[[k]] = as.matrix(X_temp[[k]][,3])
  }
} else if (modelo$call$fixed == "speed ~ 0 + time_h") {
  for(k in 1:n) {
    Xi[[k]] = as.matrix(X_temp[[k]][,2])
  }
} else if (modelo$call$fixed == "speed ~ 1") {
  for(k in 1:n) {
    Xi[[k]] = as.matrix(X_temp[[k]][,1])
  }
}

# Selecionar as colunas das matrizes de especificação Zi
# (de acordo com os efeitos aleatórios incluídos no modelo)
if(modelo$call$random == "~1 + time_h + I(time_h^2) | sapo") {

```

```

for(k in 1:length(X_temp)) {
  Zi[[k]] = Z_temp[[k]][,c(1,2,3)]
}
} else if (modelo$call$random == "~1 + time_h | sapo") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = Z_temp[[k]][,c(1,2)]
}
} else if (modelo$call$random == "~1 + I(time_h^2) | sapo") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = Z_temp[[k]][,c(1,3)]
}
} else if (modelo$call$random == "~0 + time_h + I(time_h^2) | sapo") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = Z_temp[[k]][,c(2,3)]
}
} else if (modelo$call$random == "~0 + I(time_h^2) | sapo") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = as.matrix(Z_temp[[k]][,3])
}
} else if (modelo$call$random == "~0 + time_h | sapo") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = as.matrix(Z_temp[[k]][,2])
}
} else if (modelo$call$random == "~1 | sapo") {
for(k in 1:length(X_temp)) {
  Zi[[k]] = as.matrix(Z_temp[[k]][,1])
}
}
}

Ri_temp = getVarCov(modelo, type="conditional")[[1]]
Gi = getVarCov(modelo, type='random.effects') # Dimensão de Gi é fixa
Gi_inv = solve(Gi)
Ri = Vi = list()
Ri_inv = Vi_inv = list()

for(k in 1:n) {
  Ri[[k]] = Ri_temp[1:nrow(Xi[[k]]),1:nrow(Xi[[k]])]
  Vi[[k]] = Zi[[k]]%*%Gi%*%t(Zi[[k]]) + Ri[[k]]
  Ri_inv[[k]] = solve(Ri[[k]])
}

```

```

Vi_inv[[k]] = solve(Vi[[k]])
}

R_inv = Reduce(direct.sum, Ri_inv)
V_inv = Reduce(direct.sum, Vi_inv)

# Blocos da matriz utilizada para estimar rho
A_temp = B_temp = C_temp = D_temp = E_temp = list()

for(k in 1:n) {
  A_temp[[k]] = t(Xi[[k]])%*%Ri_inv[[k]]%*%Xi[[k]]
  B_temp[[k]] = t(Xi[[k]])%*%Ri_inv[[k]]%*%Zi[[k]]
  C_temp[[k]] = t(Zi[[k]])%*%Ri_inv[[k]]%*%Xi[[k]]
  D_temp[[k]] = t(Zi[[k]])%*%Ri_inv[[k]]%*%Zi[[k]]+Gi_inv
  E_temp[[k]] = t(Zi[[k]])%*%Ri_inv[[k]]%*%Zi[[k]]
}

A = Reduce('+',A_temp)
B = Reduce(cbind,B_temp)
C = Reduce(rbind,C_temp)
D = Reduce(direct.sum,D_temp)
E = Reduce(direct.sum,E_temp)

rhoA = rbind(cbind(A,B),cbind(C,D))
rhoB = rbind(cbind(A,B),cbind(C,E))

# Inversa de uma matriz de blocos
D_orig_inv = solve(D)
A_inv = solve(A-B%*%D_orig_inv%*%C)
B_inv = -A_inv%*%B%*%D_orig_inv
C_inv = -D_orig_inv%*%C%*%A_inv
D_inv = D_orig_inv+D_orig_inv%*%C%*%A_inv%*%B%*%D_orig_inv
rhoA_inv = rbind(cbind(A_inv,B_inv),cbind(C_inv,D_inv))

# Simplificar o cálculo do determinante
logdetV = sum(unlist(lapply(lapply(Vi,FUN=det),FUN=log)))
logdetR = sum(unlist(lapply(lapply(Ri,FUN=det),FUN=log)))

resMarg = residuals(modelo, level=0)

```

```

resCond = residuals(modelo, level=1)

logLik_m = -0.5*(N*log(2*pi) + logdetV + t(resMarg)%*%V_inv%*%resMarg)
logLik_c = -0.5*(N*log(2*pi) + logdetR + t(resCond)%*%R_inv%*%resCond)
rho = sum(diag(rhoA_inv%*%rhoB))

p = length(beta)
q = nrow(Gi)+choose(nrow(Gi),2) + 1
mAIC = -2*logLik_m + 2*(p+q)
mAICc = -2*logLik_m + 2*(N/(N-p-q-1))*(p+q)
BIC = -2*logLik_m + log(N)*(p+q)
cAIC = -2*logLik_c + 2*(rho+1)
cAICc = -2*logLik_c + 2*(N*(N-p-1)*(rho+1)+N*(p+1))/((N-p)*(N-p-2))

return (list(logLik_m=logLik_m, logLik_c=logLik_c,
             p=round(p), q=round(q), df=round(p+q), rho=rho,
             mAIC=mAIC, mAICc=mAICc, BIC=BIC, cAIC=cAIC, cAICc=cAICc))
}

```


Referências

- AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. *In B. N. Petrov and F. Csaki, Proceedings of the 2nd International Symposium on Information Theory. Budapest: Akademiai Kiado*, 267–81. [2](#), [7](#), [8](#)
- BARNDORFF-NIELSEN, O. E. & COX, D. R. (1989). *Asymptotic techniques for use in statistics*. New York: Chapman and Hall. [16](#)
- BHAT, H. & KUMAR, N. (2010). On the derivation of the bayesian information criterion. . . [15](#), [16](#)
- BURNHAM, K. P. & ANDERSON, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed.* New York: Springer. [7](#), [8](#), [9](#), [13](#), [17](#), [20](#)
- BURNHAM, K. P. & WHITE, G. C. (2002). Evaluation of some random effects methodology applicable to bird ringing data. *Journal of Applied Statistics* **29**, 245–66. [13](#)
- DELATTRE, M., LAVIELLE, M. & POURSAT, M. (2014). A note on BIC in mixed-effects models. *Electronic Journal of Statistics* **8 (1)**, 456–75. [15](#), [16](#)
- DEMIDENKO, E. (2013). *Mixed Models: Theory and Applications with R, 2th ed.* Wiley, New York. [1](#), [4](#), [5](#)
- DIGGLE, P., HEAGERTY, P., LIANG, K. & ZEGER, S. L. (2002). *Analysis of longitudinal data. Second Edition.* Oxford University Press, New York. [4](#), [5](#)
- FUNG, W.-K., ZHU, Z.-Y., WEI, B.-C. & HE, X. (2002). Influence diagnostics and outlier tests for semiparametric mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 565–79. [6](#)
- GREVEN, S. & KNEIB, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika* **97 (4)**, 773–89. [8](#), [11](#)
- HARVILLE, D. (2011). *Matrix Algebra From a Statistician's Perspective.* Springer New York. [3](#)
- HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320–40. [4](#)
- HENDERSON, C. R. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics* **21**, 309–10. [5](#)

- HILDEN-MINTON, J. A. (1995). Multilevel diagnostics for mixed and hierarchical linear models. phd thesis. *University of California, Los Angeles* . 6
- HODGES, J. & SARGENT, D. J. (2001). Counting degrees of freedom in hierarchical and other richly parameterized models. *Biometrika* **88**, 367–79. 2, 12
- HURVICH, C. M. & TSAI, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307. 9
- KASS, R. E. & RAFTERY, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795. 15
- KULLBACK, S. & LEIBLER, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics* **22**, 79–86. 7
- MÜLLER, S., SCEALY, J. L. & WELSH, A. H. (2013). Model selection in linear mixed models. *Statistical Science* **28** (2), 135–67. 11, 13
- NOBRE, J. & SINGER, J. (2007). Residual analysis for linear mixed models. *Biometrika* **49**, 863–75. 1, 6
- NOBRE, J. & SINGER, J. (2011). Leverage analysis for linear mixed models. *J. Appl. Stat.* **38**, 1063–72. 1, 6
- NOBRE, J. S. (2004). Methods of diagnostics for linear mixed models. dissertation of master's degree (in portuguese). *Departamento de Estatística, Universidade de São Paulo, Brazil* . 1, 6
- OVERHOLSER, R. & XU, R. (2008). Effective degrees of freedom and its application to conditional aic for linear mixed-effects models with correlated error structures. *Computational Statistics & Data Analysis* **52** (C), 2004–21. 13
- PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–54. 4
- PINHEIRO, J., BATES, D., DEBROY, S., SARKAR, D. & R CORE TEAM (2019). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-141. 17, 30, 35
- ROBINSON, G. (1991). That blup is a good thing: The estimation of random effects. *Statistical Science* **6**. 5
- ROCHA, F. M. M. & SINGER, J. M. (2018). Selection of terms in random coefficient regression models. *Journal of Applied Statistics* **45**, 225–42. 1, 2
- SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of statistics* **6**, 461–64. 2
- SHANG, J. & CANAUAUGH, J. E. (2008). Bootstrap variants of the akaike information criterion for mixed model selection. *Computational Statistics & Data Analysis* **132**, 160–70. 11

- SINGER, J. & ANDRADE, D. (2000). Analysis of longitudinal data. *Handbook of Statistics, Volume 18: Bio-Environmental and Public Health Statistics*, P.K. Sen and C.R. Rao, eds., North Holland, Amsterdam , 115–60. 33
- SINGER, J. M., ROCHA, F. M. M. & NOBRE, J. S. (2017). Graphical tools for detecting departures from linear mixed model assumptions and some remedial measures. *International Statistical Review* **85** (2), 290–324. 1, 2, 6, 19, 27, 31
- SUGIURA, N. (1978). Further analysis of the data by akaike's information criteria and the finite corrections. *Communications in Statistics - Theory and Methods* **A7**, 13–26. 9
- TAN, F., OUWENS, M. & BERGER, M. (2001). Detection of influential observations in longitudinal mixed effects regression models. *Journal of the Royal Statistical Society: Series D (The Statistician)* **50**, 271–84. 6
- VAIDA, F. & BLANCHARD, S. (2005). Conditional akaike information for mixed-effects models. *Biometrika* **92**, 351–70. 2, 10, 12, 13
- VERBEKE, G. & MOLENBERGHS, G. (2005). *Linear Mixed Models For Longitudinal Data*. Springer, New York. 4