

MODELOS LONGITUDINAIS MISTOS COM CORRELAÇÃO SERIAL NOS ERROS

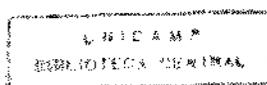
Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Raul Yukihiro Matsushita e aprovada pela comissão julgadora.

Campinas, 20 de Outubro de 1994



Prof. Dr. Luiz Koodi Hotta
(orientador)

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do título de Mestre em Estatística.



UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Ciência da Computação - IMECC

**MODELOS LONGITUDINAIS MISTOS
COM CORRELAÇÃO SERIAL NOS ERROS**

Raul Yukihiro Matsushita

Prof. Dr. Luiz Koodi Hotta
orientador

Campinas - São Paulo
1994

BANCA EXAMINADORA:

PROF. DR. LUIZ KOODI HOTTA (ORIENTADOR, IMECC - UNICAMP)

PROF. DR. PEDRO LUIZ VALLS PEREIRA (IME - USP)

PROF. DR. AIPORÉ RODRIGUES DE MORAES (ICEX - UFMG)

PROF^ª. DR^ª. CLARICE A. DE L. FREIRE (SUPLENTE, IMECC - UNICAMP)

*aos meus pais,
Masaru e Yasuko,
dedico*

AGRADECIMENTOS

AO PROF. DR. LUIZ KOODI HORTA pela sua dedicação, paciência, amizade e trabalho para a minha orientação,

AOS PROFESSORES DO DEPARTAMENTO DE ESTATÍSTICA pelos ensinamentos e amizade,

AO CNPQ, FAEP E UNICAMP pelo apoio financeiro,

AOS FUNCIONÁRIOS DA UNICAMP pelo atendimento e atenção,

AOS COLEGAS E AMIGOS,

AOS MEUS PAIS E MEU IRMÃO RUBENS,

A EMÍLIA,

A DEUS.

RESUMO

O presente trabalho apresenta alguns modelos lineares para dados longitudinais observados no tempo, enfocando-se o modelo de efeitos mistos. A matriz de covariância associada a esse modelo pode ser representada de forma estruturada. O interesse do trabalho é estudar formas da matriz de covariância que requerem poucos parâmetros a serem estimados (parcimoniosidade). Como os dados são observados ao longo do tempo, uma forma de parametrização parcimoniosa pode ser obtida através de estruturas de correlação temporal nos erros ou nos efeitos aleatórios. Apesar das variedades de estruturas temporais existentes, o que nos interessa estudar são estruturas simples como o AR(1) e algumas variações, pois, em séries curtas, essas estruturas simples podem ser aproximações razoáveis de estruturas mais complexas.

Discute-se também algumas questões sobre o problema da escolha do modelo adequado e alguns critérios de seleção do modelo. Essencialmente, a abordagem de estimação dos parâmetros do modelo de efeitos mistos é via máxima verossimilhança (MV) e MV restringida. Para o cálculo numérico das estimativas, alguns algoritmos numéricos são apresentados brevemente como o método de Newton-Raphson e o filtro de Kalman (este último é um algoritmo útil para o cálculo da verossimilhança e das predições dos efeitos aleatórios). Algumas técnicas de diagnóstico são apresentadas como, por exemplo, o gráfico de probabilidade normal ponderado para checar a hipótese de normalidade dos efeitos aleatórios e o uso do semivariograma empírico para checar existência ou não de estrutura de correlação nos resíduos.

O trabalho fornece apenas uma visão geral sobre modelos lineares em dados longitudinais. Dada a abrangência do tema é impossível tratar de todos os aspectos de forma detalhada. Desta forma, certas abordagens não serão consideradas, como a abordagem Bayesiana, ou foram dadas de forma superficial, como o algoritmo EM, por exemplo.

ABSTRACT

This work presents some linear models for longitudinal data observed at time points, specially the mixed effects model. The covariance matrix associated with this model can be represented in a structured form. This work's interest is to study some covariance forms that require few parameters to be estimate. A way of parsimonious parametrization can be obtained through time correlation structures on the errors or time-varying effects. In despite of the variety of time structures, we are concerned to study simple structures as the AR(1) and some variations, because, in short series, these simple structures can approximate more complex structures.

Some questions about the adequate model choice and some model selection criterions also are discussed. The parameters estimation approaches of the random effects model considered are essentially the maximum likelihood (ML) and restricted ML (ReML). To compute estimates, some numeric algorithms are briefly presented as Newton-Raphson method and Kalman filter (this is a useful algorithm to compute exact likelihood and make predictions). Some diagnostics are presented, e.g., weighted normal probability plot to check normality of the random effects and the use of the empirical semivariogram to check residuals correlation structure.

The work gives only an overview about linear models for longitudinal data. Because very broadly, it is impossible to take detailed the all features of this theme. So, certain approaches will not be considered, as the Bayesian approach, or will be superficially, as the EM algorithm, for example.

SUMÁRIO

1. Estudos Longitudinais	
1.1. Introdução.....	1
1.2. Exemplo.....	6
1.3. Esboço do Trabalho.....	13
2. Modelos Lineares para Dados Longitudinais	
2.1. Introdução.....	15
2.2. Modelo de Curva de Crescimento Polinomial.....	17
2.3. Modelo de Regressão.....	21
2.4. Modelo Misto Univariado.....	22
2.5. Modelo de Efeitos Aleatórios.....	25
2.6. Modelo Linear Generalizado.....	32
3. Modelagem Dinâmica: Correlação Temporal	
3.1. Introdução.....	34
3.2. Erros Correlacionados Serialmente.....	35
3.2.1. Estrutura AR(1).....	37
3.2.2. Estrutura ARMA(p,q).....	42
3.2.3. Estrutura AR(1) Exponencialmente Amortecida.....	43
3.3. Coeficientes Correlacionados Serialmente.....	49
4. Escolha do Modelo Adequado	
4.1. Introdução.....	54
4.2. A Estrutura de Covariância.....	57
4.3. Efeitos Fixos ou Aleatórios?.....	59
4.4. Critério de Informação de Akaike.....	63
5. Estimação	
5.1. Introdução.....	68
5.1.1. Terminologia.....	72
5.2. Métodos de Estimação.....	72
5.2.1. Estimador de Máxima Verossimilhança.....	73
5.2.2. Máxima Verossimilhança Restringida.....	84
5.2.3. Máxima Verossimilhança ou MV Restringida?.....	89
5.2.4. BLUP.....	90
5.2.5. Equações Gerais de Estimação.....	92
5.3. Algoritmos de Estimação.....	98
5.3.1. O Algoritmo EM.....	101
5.3.2. O Método de Newton-Raphson.....	104
5.3.3. O Método do <i>Scoring</i> de Fisher.....	107
5.3.4. O Filtro de Kalman.....	108

6. Testes de Hipóteses e Diagnóstico	
6.1. Introdução.....	115
6.2. Testes de Hipóteses.....	116
6.2.1. Teste da Razão de Verossimilhança.....	116
6.2.2. Teste de Wald.....	119
6.2.3. Teste do Escore de Fisher.....	121
6.2.4. Teste F da ANOVA.....	122
6.3. Diagnóstico.....	127
6.3.1. Resíduos.....	128
6.3.2. O Gráfico de Probabilidade Normal.....	132
6.3.3. O Semivariograma.....	135
7. Aplicação	
7.1. Introdução.....	139
7.2. Análise I.....	141
7.2.1. Investigação Inicial.....	143
7.2.2. Transformação dos Dados.....	144
7.2.3. Retirada dos Ratos com Valores Aberrantes.....	149
7.2.4. Ajuste e Seleção dos Modelos.....	152
7.2.5. Interpretação dos Efeitos Aleatórios.....	155
7.2.6. Perfis Individuais Estimados.....	156
7.3. Análise II.....	162
7.4. Algumas Considerações.....	166
Apêndices	
A. Algoritmo EM.....	168
B. Filtro de Kalman.....	171
C. Modelos ARMA.....	178
D. Tempo Contínuo.....	180
Referências Bibliográficas.....	183

1

Estudos Longitudinais

1.1. Introdução

O *Estudo longitudinal* (E.L.) é uma importante estratégia de pesquisa que pode ser aplicado, por exemplo, nas áreas médicas, biológicas, sociais e econômicas. Geralmente, os E.L. são caracterizados como delineamentos onde a resposta de cada unidade de investigação (ou indivíduo) é observada ao longo de várias ocasiões (condições de avaliação). Estas podem, por exemplo, ser tempo, dosagem (grau de concentração química) e distância. A natureza da resposta é discreta (categorizada) ou contínua, multidimensional (mais de uma variável resposta) ou unidimensional. Nesse trabalho nos restringiremos ao estudo de dados de natureza contínua e unidimensional (apenas uma variável resposta) observados ao longo do tempo, como no exemplo da seção 1.2, onde a resposta é o nível de ácido láctico.

De modo geral, os indivíduos podem ser classificados em subpopulações de acordo com fatores ou tratamentos. Os fatores podem ser constituídos de níveis; assim, por exemplo, se o "sexo" for um fator de interesse, então os níveis associados a esse fator serão: "masculino" e "feminino". O interesse imediato quando se leva em consideração um determinado fator é avaliar como os diferentes níveis desse fator afetam ou influenciam na variável de interesse; a isso chamamos de efeito de um nível do fator na variável de interesse. Os efeitos podem ser divididos em duas classes: fixos e aleatórios. A natureza e a escolha pela classe do efeito adequado serão discutidas no capítulo 4. Informações adicionais à resposta podem ser incorporadas através de covariáveis (variáveis secundárias) variantes ou invariantes ao longo das condições de avaliação.

Em um estudo longitudinal, destaca-se dentre seus objetivos: avaliações do comportamento das variáveis de resposta e dos efeitos das covariáveis nos padrões da resposta ao longo das ocasiões, comparações entre as subpopulações e ajuste de modelos. No estudo longitudinal é possível não somente descobrir certas características de uma população, como também acompanhar a evolução dessas características ao longo das condições de avaliação. Ware (1985) destaca que os estudos longitudinais em geral têm um

potencial superior aos estudos *cross-section* ou a uma seqüência de estudos *cross-section* de uma mesma população, como será discutido rapidamente na seção 4.1.

O conjunto de dados pode ser caracterizado através do número de observações disponíveis: quando houver o mesmo número de ocasiões para todos os indivíduos, os dados serão balanceados quanto à ocasião (por exemplo, balanceado no tempo). Se além disso todos forem observados nas mesmas condições de avaliação então os dados serão completamente balanceados (ou, simplesmente, balanceados). No exemplo da seção 1.2, a Tabela 1.2.1 ilustra um conjunto de dados balanceados.

É comum, entretanto, a ocorrência de dados não balanceados nos estudos longitudinais. No caso de dados experimentais, isso pode ocorrer, por exemplo, acidentalmente, resultando em valores perdidos (*missing values*); planejadamente; e pela falta ou até impossibilidade de se ter um planejamento adequado. Outro caso de estruturação não balanceada pode ocorrer de forma natural em estudos observacionais, pela própria impossibilidade de se executar um planejamento específico na coleta dos dados. Podem existir muitos fatores ou mecanismos que geram dados não balanceados. Nesse trabalho, entretanto, assumiremos que o conjunto incompleto de dados é resultado de informações perdidas de forma completamente aleatória (*missing completely at random, MCAR*, segundo Little e Rubin, 1987, pp. 14). No exemplo da seção 1.2, a Tabela 1.2.2 ilustra um conjunto não balanceado, com muitas observações censuradas (a censura para um determinado rato nesse caso significa que ele simplesmente deixou de ser observado antes do término do estudo para evitar seu afogamento). Rigorosamente, nesse exemplo o mecanismo gerador dos dados incompletos não deveria ser ignorado, pois o rato em iminência de afogamento teve seu período observacional abreviado. Por isso esse conjunto será utilizado no capítulo 7 apenas como ilustração de um conjunto não balanceado, aplicando-se algumas das técnicas de análise que serão apresentadas no trabalho.

Uma distinção deve ser feita entre delineamentos longitudinais e medidas repetidas (Ware, 1985). Pode-se dizer que o planejamento longitudinal é um caso particular do planejamento em medidas repetidas, pois esta é caracterizada pela observação de duas ou mais medidas em cada indivíduo. No entanto, o que distingue a estrutura Longitudinal dentro das Medidas Repetidas é a forma de associação entre as ocasiões e os indivíduos. Por exemplo, em Medidas Repetidas do tipo *Split-plot* as ocasiões são alocadas de forma aleatória às unidades de investigação; em Planejamentos *Cross-over* as unidades de investigação são submetidas em uma dentre as diversas seqüências de tratamentos que definem as condições de avaliação. Em dados longitudinais, por exemplo, os indivíduos são acompanhados durante um determinado período de tempo, podendo ser observados em conjuntos não homogêneos de tempo (i. e., os indivíduos não necessariamente devem ser

observados nos mesmos instantes), os intervalos de tempo entre as observações podem ser diferentes (a esse tipo de conjunto chamaremos de *dados desigualmente espaçados*), e podem envolver covariáveis que variam no decorrer do estudo.

Por outro lado, as medidas repetidas, quando se considera o tempo como condição de avaliação, são caracterizadas por observações colhidas em um espaço de tempo relativamente curto, em um conjunto homogêneo de tempo (i.e., todos os indivíduos devem ser observados nos mesmos instantes que, em geral, são *dados igualmente espaçados* no tempo), e com covariáveis que não variam no decorrer do estudo.

No **capítulo 2** (a abordagem discutida nesse trabalho restringe-se a dados contínuos observados ao longo do tempo, assumido-se normalidade nos dados) são apresentados brevemente alguns modelos lineares para dados longitudinais. Entre os modelos multivariados serão apresentados os modelos de regressão e de curva de crescimento polinomial (e.g., Rao, 1965; Potthoff e Roy, 1964; Grizzle e Allen, 1969). Nesses modelos a matriz de covariância assume uma forma geral (também chamada de forma não estruturada), ou seja, a única restrição à matriz de covariância é que ela seja positiva definida e simétrica.

Muitos modelos estatísticos para analisar medidas repetidas dependem dos padrões assumidos na matriz de covariância. Duas abordagens comuns são a ANOVA e a MANOVA (análise de variância multivariada). A primeira abordagem, que trata de modelos univariados, possuem um grau de estruturação na matriz de covariância. Por exemplo, Geisser e Greenhouse (1958) e Huynh e Feldt (1970) consideraram um modelo ANOVA de efeitos mistos caracterizado por uma estrutura de covariância do tipo *H* (que representa uma forma de simetria composta, onde são necessários dois elementos de covariância: um para as variâncias ao longo da diagonal principal da matriz e o outro para as posições fora da diagonal). A abordagem via MANOVA com efeitos fixos foi descrita por Cole e Grizzle (1966) e Timm (1980). Nesse caso, não se impõe nenhuma restrição ou nenhum padrão à matriz de covariância, sendo portanto, necessário estimar $T(T+1)/2$ elementos da covariância, onde *T* representa o número de observações individuais num conjunto de dados balanceado. Grizzle e Allen (1969) estenderam esses procedimentos para o caso onde um padrão particular de crescimento é postulado sobre o tempo (padrões polinomiais, como por exemplo, linear e quadrático).

Formas mais estruturadas podem ser encontradas no modelo de efeitos mistos (modelos com efeitos fixos e aleatórios) e no modelo linear generalizado. Laird e Ware (1982) apresentaram uma família de modelos chamados hierárquicos (ou modelos de duplo estágio segundo Laird e Ware) para analisar dados longitudinais ou medidas repetidas, baseando-se no trabalho de Harville (1977). Isso resultou numa estrutura característica da

matriz de covariância individual, incluindo como casos especiais, por exemplo, estruturas de covariância resultantes da ANOVA e o padrão de uniformidade ou simetria composta.

Muitos são os fatores que influem na escolha do modelo adequado. Por exemplo, para um número pequeno de observações individuais (poucas ocasiões) ou dados balanceados, pode-se aplicar modelos gerais como ANOVA ou modelos de regressão. Por outro lado, com número grande de ocasiões ou dados não balanceados, a modelagem via métodos gerais requer muitos parâmetros a estimar (parâmetros que descrevem a estrutura de covariância). Como Laird e Ware (1982) observaram, quando isso ocorre, muitos desses parâmetros poderão ser pobremente estimados. Uma alternativa é construir uma estrutura de correlação para obter modelos parcimoniosos, i.e., modelos com poucos parâmetros a serem estimados. Pelo estudo da dependência temporal dos dados é possível construir esse tipo de estrutura.

Em delineamentos longitudinais é razoável esperar que a correlação entre duas observações de um indivíduo seja maior na medida que o intervalo de tempo entre essas observações diminua. Assim, um modo de introduzir parametrização parcimoniosa é incorporar uma estrutura temporal ao modelo. Como nesse trabalho não serão utilizados modelos de séries temporais para as observações, a correlação serial (i.e., correlação temporal associado a um determinado indivíduo) é incorporada ao modelo através dos erros ou através de coeficientes que variam no tempo seguindo um determinado processo.

No capítulo 3 discute-se como a modelagem dinâmica pode ser introduzida em modelos lineares, enfatizando-se a estrutura de correlação dos erros associados aos modelos apresentados. Apesar das variedades de estruturas temporais que podem ser adotadas, o que nos interessa estudar são estruturas simples. Considerando-se as suposições de modelagem, estruturas mais complicadas poderiam ser estudadas e utilizadas. Entretanto, um motivo para adotar modelos simples é evitar parametrização excessiva. Além disso, a quantidade de observações pode não ser suficiente para suportar estruturas complicadas. Outro motivo foi observado por Louis (1988):

"...I had a student who did some work on autocorrelation, autoregressive moving average models, developed the likelihood estimators, and likelihood ratio tests. Everything worked beautifully with artificial data. Then we tried it on some real data...We looked around and we couldn't find any data in a biological setting where an autoregressive moving average covariant structure fit nicely".

Evidentemente o comentário acima deve ser discutido mais profundamente, pois em geral o processo que geram os erros são complicados. Mas modelos simples como o AR(1)

podem aproximar bem a realidade, produzindo bons ajustes. Entretanto, deve-se estudar modelos mais complexos como o ARMA, porque eles fornecem uma variedade de estruturas de covariância que são economicamente parametrizadas. Em geral, nos modelos mais usuais, o número desses parâmetros não deve exceder a 4 (Rochon e Helms, 1989).

No **capítulo 4** discute-se o problema da escolha do modelo adequado. Dentre uma variedade de opções é necessário decidir adequadamente por algum modelo e aplicá-lo em um contexto particular. Nessa discussão, considera-se o tipo de efeitos (fixos ou aleatórios?), a quantidade de parâmetros da matriz de covariância a serem estimados e alguns critérios de seleção do modelo como o critério de informação de Akaike (AIC). Demais temas relacionados com esse capítulo como os diagnósticos do modelo escolhido e os testes de hipóteses serão discutidos no capítulo 6.

São considerados a estimação dos parâmetros e os algoritmos de estimação no **capítulo 5**, e os testes de hipóteses e diagnóstico no **capítulo 6**. Os métodos de estimação são técnicas estatísticas úteis na modelagem dos dados. Para um conjunto de dados, a análise depende, entre outras coisas, das suposições do modelo, por exemplo, o tipo de efeitos (fixos ou aleatórios), a estrutura de covariância e a distribuição de probabilidade dos dados. A abordagem adotada é a inferência clássica, embora a abordagem Bayesiana produza interpretações interessantes. Os métodos de estimação considerados no capítulo 5 são o de máxima verossimilhança, máxima verossimilhança modificada, a abordagem via equações generalizadas de estimação e o BLUP. Duas abordagens importantes de estimação, a ANOVA e o MINQUE (*minimum-norm quadratic unbiased equation*), não serão descritas nesse trabalho devido à grande extensão dessas técnicas, embora a aplicação das estatísticas F seja discutida no trabalho (seções 2.4 e 6.2.5); apenas um breve comentário sobre o MINQUE será feito na seção 5.1. Searle, Casella e McCulloch (1992, cap. 2 a 5) fornecem uma revisão histórica da ANOVA e descrevem essa técnica para dados balanceados e não balanceados. Searle *et alii*. (1992, cap. 11) também descrevem o MINQUE que é um método que não requer suposição de normalidade dos dados.

Numericamente as estimativas de MV e MVRe devem ser calculadas através de algoritmos iterativos. Entre os algoritmos considerados estão o algoritmo EM, Newton-Raphson, o *scoring* e o filtro de Kalman. Os algoritmos menos conhecidos, o EM e o filtro de Kalman, estão descritos nos apêndices A e B, respectivamente. O filtro de Kalman é um algoritmo útil para o cálculo da verossimilhança exata (Jones e Boateng, 1991) e pode ser utilizado juntamente com os métodos de otimização não-linear (como o Newton-Raphson e o *scoring*). Laird (1982) e Laird, Lange e Stram (1987), por exemplo, discutiram a estimação via máxima verossimilhança (EMV) e estimação via máxima verossimilhança restringida (EMVRe) utilizando o algoritmo EM. Jennrich e Schluchter (1986) apresentaram

uma variedade de estruturas de covariância e estudaram alguns dos métodos de estimação para esses casos.

Essencialmente as técnicas de inferência e diagnóstico descritas no capítulo 6 são as mesmas daquelas encontradas na análise de variância e na análise de regressão. Descreveremos algumas formas de diagnóstico como por exemplo a utilização do gráfico de probabilidade normal ponderado para checar a hipótese de normalidade dos efeitos aleatórios e o uso do semivariograma empírico para checar existência ou não de estrutura de correlação nos resíduos.

Alguns conceitos menos conhecidos serão expostos nos apêndices. Neles são expostas algumas noções sobre o algoritmo EM (**apêndice A**), filtro de Kalman (**apêndice B**), modelagem ARMA para os erros, considerando-se tempo discreto e contínuo (**apêndices C e D**, respectivamente).

1.2. Exemplo

Nessa seção descreveremos um conjunto de dados longitudinais estudado pelo Prof. José Roberto M. de Azevedo, professor assistente da UNESP (Rio Claro). Esse conjunto experimental é formado por dois grupos de ratos: no grupo 1 encontram-se ratos treinados, i.e., ratos que estão sujeitos a atividades físicas periódicas; no grupo 2 estão os ratos sedentários, i.e., ratos sem preparo físico. Os ratos são colocados em uma piscina com água, sendo forçados a nadarem. A atividade física dos ratos durante a natação produz o ácido láctico, que é o resultado do estresse muscular. Durante 1 hora, o nível de ácido láctico de cada rato é observado a cada 10 minutos. O nível de ácido é medido a partir de amostra de sangue retirada da cauda do animal; a coleta é feita com a mínima interferência possível na atividade do rato.

Os dados dos ratos do grupo 1 e 2 estão registrados, respectivamente, nas Tabelas 1.2.1 e 1.2.2 (a unidade de medida não foi informada pelo pesquisador). Cada i -ésima linha dessas tabelas representa as observações do i -ésimo rato tomadas no tempo ($j = 1, \dots, 7$) de forma igualmente espaçada (a cada 10 minutos). Assim, esse conjunto de dados é um caso de dados longitudinais.

Os ratos na iminência de afogamento foram resgatados e seus dados faltantes são considerados simplesmente como dados perdidos: assume-se como ilustração que o mecanismo gerador desses valores perdidos seja completamente aleatório. A rigor, nesse exemplo, o mecanismo gerador dos dados incompletos não deveria ser ignorado, pois o rato em iminência de afogamento teve seu período observacional abreviado. Por isso esse

conjunto será utilizado apenas como ilustração de um caso não balanceado, aplicando-se algumas das técnicas de análise apresentadas no trabalho. Entre os objetivos na análise podemos considerar:

- a) modelagem e comparação das curvas de resposta (nível de ácido láctico) dos 2 grupos. O modelo de efeitos mistos permite estudar tanto o padrão de resposta dos grupos e compará-los, como permite estudar o padrão individual de resposta e compará-lo com os outros indivíduos. Um interesse, por exemplo, é comparar as curvas de resposta (nível de ácido láctico) dos 2 grupos e verificar se são iguais ou diferem no tempo.
- b) estudo e comparação das várias estruturas de correlação possíveis para a matriz de covariância individual.

Esse exemplo será utilizado ao longo do trabalho para ilustrar técnicas de modelagem (exemplo 2.5.6), heterogeneidade temporal das observações (exemplo 3.3.3) e critérios de seleção (exemplo 4.4.1). **No capítulo 7** analisaremos com mais detalhes esse conjunto dividindo-o em duas partes: análise I e análise II. Na primeira parte serão ajustados os casos particulares do modelo misto descrito no exemplo 2.5.6. Nessa etapa, a estratégia de seleção do modelo consiste basicamente de 3 partes descrita por Diggle (1988):

- i) Estimção de um modelo preliminar (em geral um modelo saturado) com erros independentes ou uma matriz de covariância geral para os erros. O objetivo nesse item é a investigação inicial do modelo. Por isso, esse item pode envolver uma transformação apropriada dos dados para se obter um padrão homocedástico nos resíduos. Pelos gráficos dos perfis individuais (Fig. 1.2.1 e 1.2.2), nota-se que as respostas de ácido são mais estáveis no início do experimento, tornando-se cada vez mais variáveis ao longo do experimento. Uma forma de se contornar isso é transformar os dados, por exemplo, tomando-se o logaritmo dos níveis de ácido láctico. A transformação deve produzir um padrão mais satisfatório nos resíduos, atenuando-se o efeito da heterocedasticidade temporal. Uma alternativa à transformação é incorporar a heterocedasticidade ao modelo, introduzindo-o como uma covariável dependente do tempo como será estudado na seção 3.3 e na análise II.
- ii) Seleção inicial das estruturas de covariância para o modelo.

iii) Esse estágio utiliza os critérios de seleção de modelo para a escolha do modelo e a análise de diagnóstico e resíduos que serão apresentados nos capítulos 4 e 6.

Na segunda parte (Análise II) ajustaremos o modelo misto que será descrito no exemplo 3.3.3, considerando-se os efeitos aleatórios dependentes do tempo com uma matriz de covariância estruturada. Nessa parte, os dados não serão transformados porque a heterocedasticia temporal pode ser incorporada ao modelo como será discutido no capítulo 3.

Os gráficos dos perfis individuais para cada grupo estão nas figuras 1.2.1 e 1.2.2. Como os dados são coletados nos mesmos instantes, é possível plotar a partir da Tabela 1.2.3 os gráficos dos perfis médios do grupos (nível médio de ácido láctico \times tempo) com as variações observadas representadas pelas barras verticais (± 1 erro padrão). Desses gráficos (Fig. 1.2.3 e 1.2.4) não se pode retirar inferências, mas eles auxiliam na seleção do modelo e na indicação de uma possível heterogeneidade nas variâncias. Possivelmente, os erros padrões obtidos para o segundo grupo serão mais imprecisos após 30 minutos por causa da retirada de muitos ratos do experimento.

Tabela 1.2.1 - Nível de ácido láctico¹ em ratos do Grupo 1
(ratos treinados)

rato	tempo (minutos)						
	0	10	20	30	40	50	60
1	1.5	8.7	11.4	10.7	7.3	4.8	3.6
2	1.2	5.8	7.5	10.2	11.6	10.5	8.1
3	0.9	4.5	3.7	2.6	2.4	2.1	1.8
4	1.5	6.3	6.0	6.4	5.1	3.9	3.9
5	1.2	9.9	12.6	15.1	15.6	15.9	14.4
6	1.2	8.7	9.9	7.2	5.2	3.6	2.7
7	1.2	8.7	10.5	12.3	14.0	14.1	15.6
8	1.2	7.8	9.9	8.4	5.9	4.2	3.9
9	1.5	5.1	3.6	3.5	3.8	3.6	2.7
10	2.7	6.3	10.8	11.9	11.2	10.5	9.9
11	0.9	5.3	3.5	2.6	2.0	2.5	2.1
12	1.3	8.2	4.3	4.3	4.9	6.7	7.7
13	1.3	11.6	14.6	10.5	6.3	2.4	3.2
14	1.5	11.9	11.2	4.5	2.4	3.1	1.5
15	1.1	7.6	9.2	7.7	7.0	5.5	5.0
16	1.3	8.9	7.9	4.2	3.3	2.8	3.1
17	1.5	7.9	5.0	3.0	4.8	5.9	4.9
18	1.1	9.8	10.5	9.2	6.8	5.7	5.3
19	0.9	9.4	12.5	11.8	10.7	9.6	8.2
20	0.9	9.4	11.8	10.7	10.5	9.4	9.4

¹ A unidade de medida não foi informada pelo pesquisador.

Tabela 1.2.2 - Nível de ácido láctico em ratos do Grupo 2

(ratos sedentários)

o símbolo "-" indica que o rato foi retirado da piscina por afogamento iminente

rato	tempo (minutos)						
	0	10	20	30	40	50	60
1	2.3	9.3	-	-	-	-	-
2	1.7	9.3	13.5	-	-	-	-
3	1.5	12.5	13.8	15.4	15.4	-	-
4	2.0	5.7	6.3	6.1	3.9	3.0	4.2
5	1.5	11.1	15.6	9.6	-	-	-
6	1.2	5.1	-	-	-	-	-
7	1.8	7.8	9.9	8.4	9.1	6.3	5.4
8	1.2	6.1	9.9	-	-	-	-
9	2.3	9.6	13.9	13.0	10.2	8.5	9.0
10	1.9	10.0	10.6	19.0	-	-	-
11	1.5	9.6	15.9	-	-	-	-
12	2.3	11.3	16.5	-	-	-	-
13	1.3	11.6	12.9	14.2	16.0	17.0	18.6
14	1.5	10.0	13.8	-	-	-	-
15	1.1	15.2	3.7	4.0	5.3	4.8	6.5
16	1.7	12.0	11.6	9.8	5.5	4.1	3.3

Tabela 1.2.3 - Nível médio de ácido láctico em ratos ± 1 erro padrão

Grupo	tempo (minutos)						
	0	10	20	30	40	50	60
1	1,30	8,09	8,82	7,84	7,04	6,34	5,85
	$\pm 0,39$	$\pm 2,05$	$\pm 3,42$	$\pm 3,79$	$\pm 3,95$	$\pm 4,02$	$\pm 4,04$
2	1,68	9,76	11,99	11,06	9,34	7,28	7,83
	$\pm 0,40$	$\pm 2,65$	$\pm 3,66$	$\pm 4,75$	$\pm 4,87$	$\pm 5,13$	$\pm 5,63$

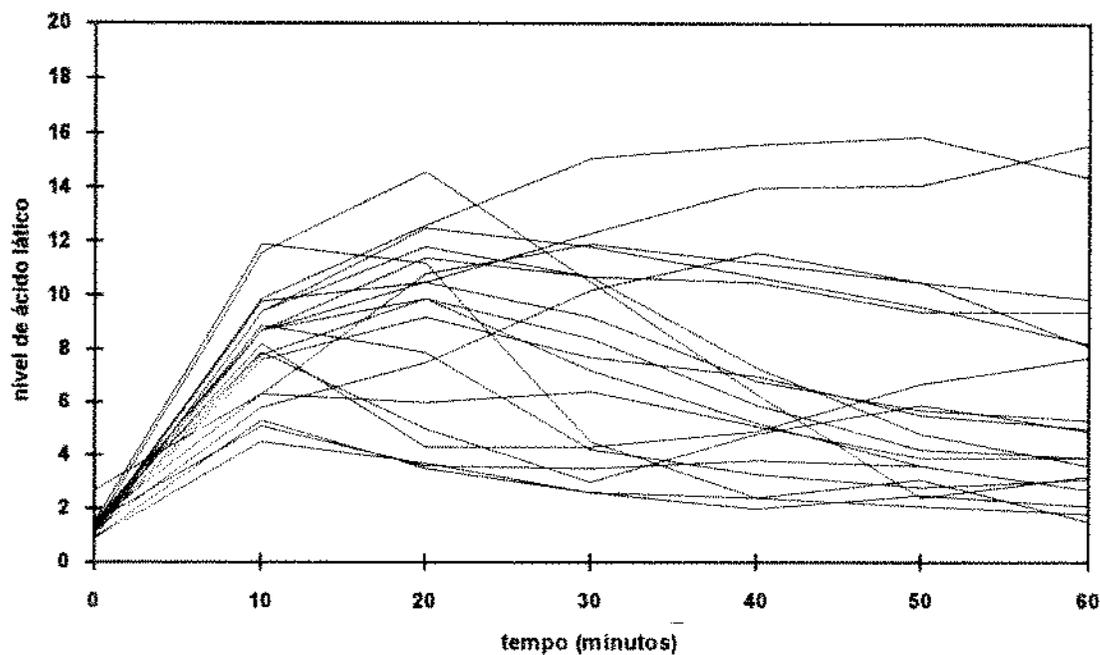


Fig. 1.2.1 - gráfico de *perfis individuais para ratos do grupo 1*

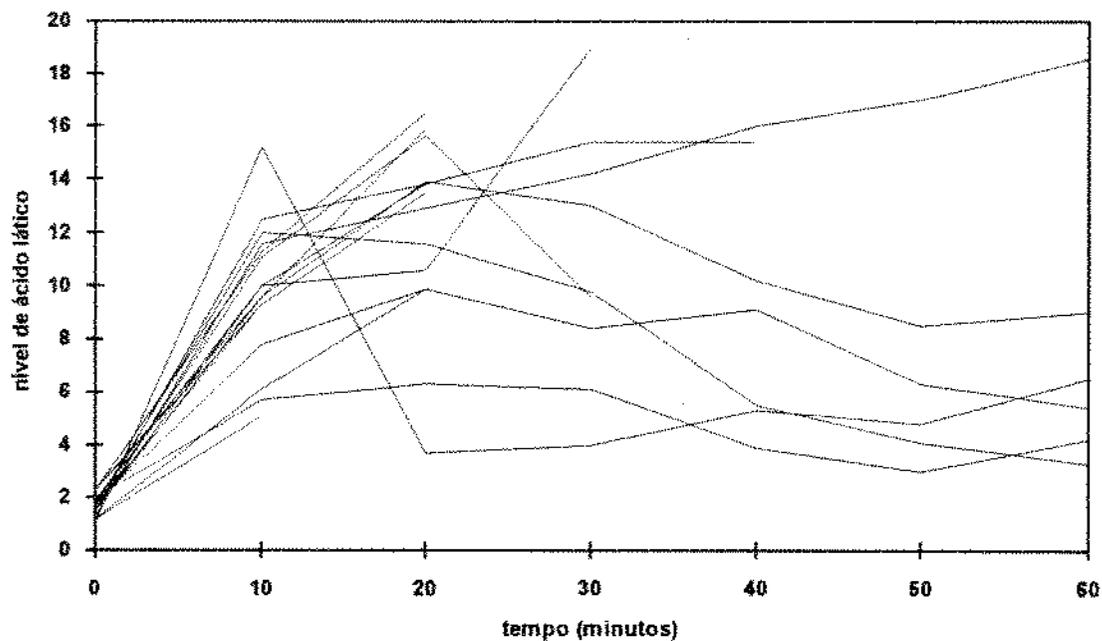


Fig. 1.2.2 - gráfico de *perfis individuais para ratos do grupo 2*

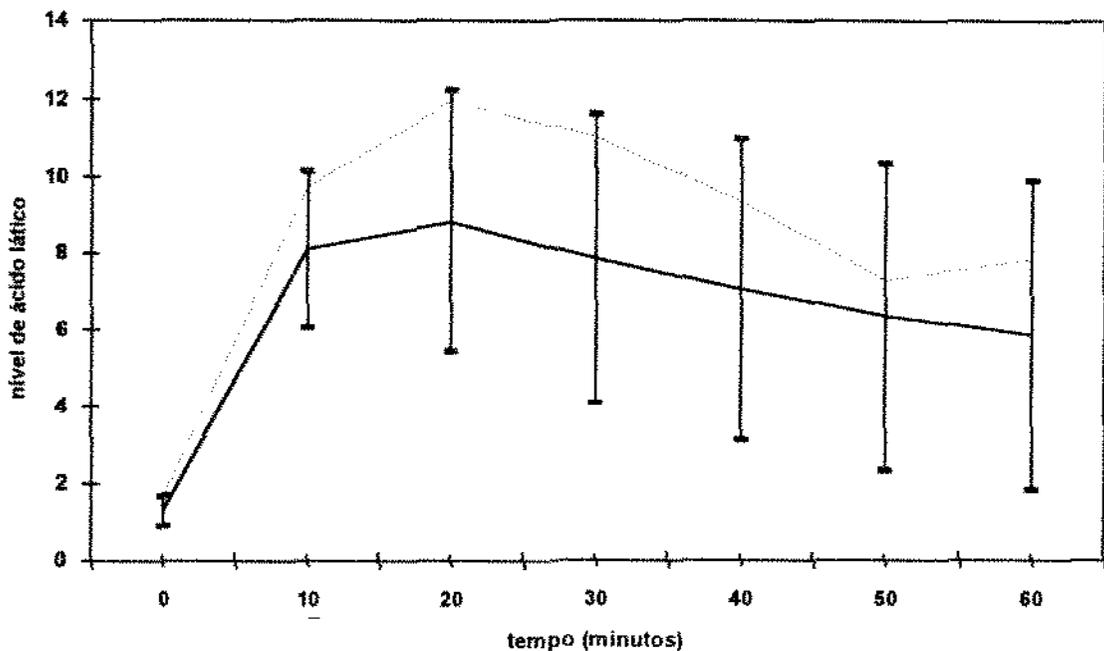


Fig. 1.2.3 - gráfico de perfis médios observados
 (— perfil médio observado do grupo 1 com ± 1 erro padrão, - - perfil médio observado do grupo 2)

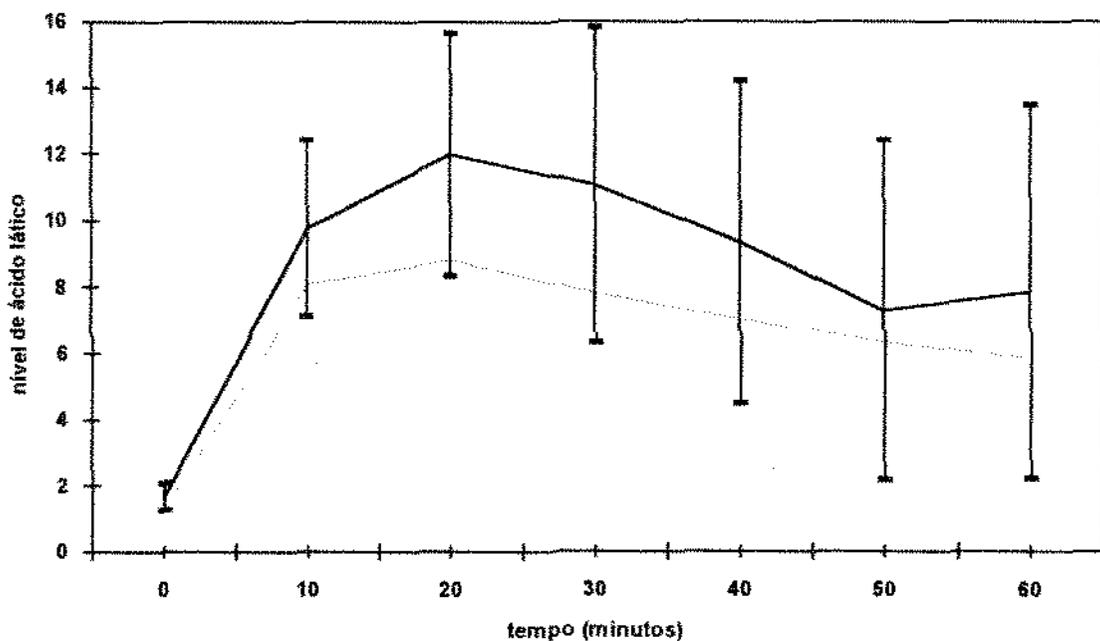


Fig. 1.2.4 - gráfico de perfis médios observados
 (— perfil médio observado do grupo 2 com ± 1 erro padrão, - - perfil médio observado do grupo 1)

1.3. Esboço do Trabalho

No **capítulo 2** serão apresentados brevemente alguns modelos lineares para dados longitudinais. Os modelos apresentados são: curva de crescimento polinomial, o modelo de regressão (modelos multivariados), o modelo univariado de efeitos mistos, o modelo de efeitos fixos e aleatórios e o modelo linear generalizado (GLM). Os dois últimos são modelos bastante genéricos, mas o enfoque nesse trabalho é o modelo de efeitos fixos e aleatórios. Esse modelo tem duas características importantes: permite representar vários modelos como o modelo componente de variâncias, modelos univariados e também multivariados; e permite estruturar a matriz de covariância arbitrariamente. Apesar de arbitrária, o interesse do trabalho é estudar formas estruturadas da matriz de covariância que requerem poucos parâmetros para serem estimados (parcimoniosidade). Como consideramos no estudo que os dados são longitudinais ao longo do tempo, uma forma de estruturar a matriz é através de modelos de correlação temporal.

No **capítulo 3** discute-se como a modelagem dinâmica pode ser introduzida em modelos lineares, enfatizando-se a estrutura de correlação dos erros associados aos modelos apresentados. Apesar das variedades de estruturas temporais que podem ser adotadas, o que nos interessa estudar são estruturas simples como o AR(1), AR(2), ARMA(1,1).

No **capítulo 4** discute-se o problema da escolha do modelo adequado. Dentre uma variedade de opções é necessário decidir adequadamente por algum modelo e aplicá-lo em um contexto particular. Nessa discussão, considera-se o tipo de efeitos (fixos ou aleatórios?), a quantidade de parâmetros da matriz de covariância a serem estimados e alguns critérios de seleção do modelo como o critério de informação de Akaike (AIC).

No **capítulo 5** será estudada a estimação dos parâmetros e alguns algoritmos de estimação encontrados na literatura. Os métodos de estimação são técnicas estatísticas úteis na modelagem dos dados. A abordagem adotada é a inferência clássica, embora a abordagem Bayesiana produza interpretações interessantes. Os métodos de estimação considerados no capítulo 5 são o de máxima verossimilhança, máxima verossimilhança modificada, a abordagem via equações generalizadas de estimação e o BLUP (com maior enfoque aos métodos baseados na verossimilhança). Numericamente as estimativas de MV e MVRe devem ser calculadas através de algoritmos iterativos. Entre os algoritmos considerados estão o algoritmo EM, o método de Newton-Raphson, o método do *scoring* e o filtro de Kalman. Os algoritmos menos conhecidos, o EM e o filtro de Kalman, estão descritos nos apêndices A e B, respectivamente. O filtro de Kalman é um algoritmo auxiliar que permite o cálculo da verossimilhança exata e pode ser utilizado juntamente com os

métodos de otimização não-linear (Newton-Raphson e o *scoring*); também é útil para fazer predições dos efeitos aleatórios e calcular os resíduos do ajuste.

No capítulo 6 serão descritas algumas formas de diagnóstico como por exemplo a utilização do gráfico de probabilidade normal ponderado para checar a hipótese de normalidade dos efeitos aleatórios e o uso do semivariograma empírico para checar existência ou não de estrutura de correlação nos resíduos. Outras técnicas mais conhecidas não serão apresentadas. Nesse capítulo também são descritos brevemente alguns testes de hipóteses.

No capítulo 7 analisaremos com mais detalhes o conjunto apresentado na seção 1.2, dividindo-o em duas partes: análise I e análise II. Na primeira parte serão ajustados os casos particulares do modelo misto (polinomial) descrito no exemplo 2.5.6. Na segunda parte (Análise II) ajustaremos um modelo misto semelhante ao que será descrito no exemplo 3.3.3, considerando-se os efeitos aleatórios dependentes do tempo com uma matriz de covariância estruturada. Nessa parte, os dados não serão transformados porque a heterocedasticia temporal pode ser incorporada ao modelo como será discutido no capítulo 3.

O presente trabalho tem como objetivo dar uma visão geral sobre modelos lineares em dados longitudinais. Dada a abrangência do tema é impossível tratar de todos os aspectos de forma detalhada. Desta forma, certas abordagens não serão consideradas, como a abordagem Bayesiana, ou foram dadas de forma superficial, como o algoritmo EM, por exemplo.

2

Modelos Lineares para Dados Longitudinais

2.1. Introdução

Nesse capítulo serão apresentados brevemente alguns modelos lineares abordados na literatura estatística no estudo de dados longitudinais. Esses modelos, que em geral também podem ser chamados de modelos de *Curva de Crescimento*, são úteis na análise de dados baseados em medidas repetidas sobre o tempo (ou dados longitudinais) de alguma característica observada em cada indivíduo. Como o próprio nome indica, a análise de *curva de crescimento* dá ênfase à explicação da variação intra-individual através de um desenvolvimento natural ou um processo de envelhecimento (Laird e Ware, 1982). Nesse tipo de estudo é comum, por exemplo, comparar as características do crescimento para diferentes grupos de indivíduos ou populações, enfatizando-se a contribuição das condições experimentais para a variabilidade entre os indivíduos. O termo *Curva de Crescimento* pode englobar várias técnicas de modelagem como a MANOVA e modelos de efeitos aleatórios (Ware, 1983).

Em contraste, nos modelos de medidas repetidas, tipicamente assume-se que os efeitos individuais se mantenham constantes durante o período de tempo em questão. O que se altera durante o experimento é a condição experimental, de modo que os efeitos experimentais contribuem para a variação intra-individual.

Dentro do contexto de dados longitudinais, apresentaremos os modelos por ordem de estruturação da matriz de covariância individual. Os dois primeiros, o Modelo de Curva de Crescimento Polinomial (seção 2.2) e o Modelo de Regressão (seção 2.3), são modelos multivariados que adotam uma estrutura geral para a matriz de covariância individual. Nos modelos multivariados, em geral, assume-se que o vetor de respostas tenha distribuição normal multivariada com uma matriz de dispersão arbitrária; o vetor de médias pode depender tanto do padrão das observações como também das covariáveis. Quando o conjunto de dados possui observações perdidas aleatoriamente, o modelo multivariado pode ser aplicado utilizando-se técnicas e métodos multivariados para observações perdidas (e.g., Dempster *et alii.*, 1977; Ware, 1985). Entretanto, quando os indivíduos são observados em tempos arbitrários, ou em um único instante, ou quando a dimensão da matriz de dispersão

se torna grande, essa abordagem torna-se pouco atrativa. Isso ocorre porque o método multivariado com uma matriz de dispersão irrestrita requer a proliferação dos parâmetros de variância, ou seja, requer muitos parâmetros a estimar. Isso pode resultar em estimação ineficiente, i.e., as estimativas dos parâmetros poderão estar comprometidas (Ware, 1985). Por isso, é importante estudar outras estruturas de covariância para obter modelos parcimoniosos (modelos com poucos parâmetros). Além disso, o modelo multivariado não permite a definição e estimação das características (aleatórias) individuais.

O Modelo Misto Univariado (seção 2.4) possui uma estrutura de covariância conhecida como padrão de uniformidade ou simetria composta (2.4.3). O termo misto significa que o modelo contém tanto efeitos fixos com efeitos aleatórios. Nesse capítulo, os efeitos fixos e aleatórios serão especificados e descritos em termos gerais. As discussões sobre essas duas classes de efeitos serão detalhadas no capítulo 4. O Modelo de Efeitos Mistos (seção 2.5), também conhecido como modelo de duplo estágio (segundo Laird e Ware, 1982) ou modelo hierárquico (segundo Seale, Casella e McCulloch, 1992), é mais flexível quanto à estrutura de covariância, que pode assumir várias formas. O modelo de efeitos mistos identifica explicitamente as características individuais e da população. Sua forma estende-se naturalmente para o caso não balanceado (devido à incorporação dos efeitos aleatórios ao modelo). Outro aspecto desse modelo é que, escolhendo-se formas apropriadas para as matrizes de delineamento e de covariâncias dos efeitos aleatórios, é possível obter outros modelos, inclusive aqueles já citados: curvas de crescimento polinomiais, modelos de regressão e o modelo misto univariado. Muitos dos modelos de duplo estágio encontrados na literatura podem ser descritos como modelos de crescimento ou de medidas repetidas.

O modelo linear generalizado (GLM) apresentado na seção 2.6 também é abrangente, pois engloba modelos como ANOVA, regressão e componente de variância. Sua utilidade em dados longitudinais será discutida no capítulo 5, embora não seja totalmente explorado; o enfoque desse trabalho está no modelo misto e no estudo das parametrizações da matriz de covariância associada ao modelo misto utilizando a modelagem dinâmica (capítulo 3).

A incorporação da estrutura temporal através dos erros ou dos coeficientes é uma forma de se introduzir uma parametrização parcimoniosa para a estrutura de correlação. Algumas estruturas temporais serão apresentadas no capítulo 3, considerando-se os casos de observações igualmente ou desigualmente espaçadas.

Nesse capítulo, nas seções 2.4 e 2.5, os efeitos aleatórios foram especificados e descritos em termos gerais. No capítulo 4 será detalhada a discussão sobre as duas classes de efeitos (fixos e aleatórios), as estruturas de correlação e critérios de escolha do modelo.

Para ilustrar a questão sobre a escolha do efeito como fixo ou aleatório alguns exemplos encontrados na literatura serão apresentados. A estimação do modelo de efeitos mistos será considerada no capítulo 5.

2.2. Modelo de Curva de Crescimento Polinomial

No cálculo matemático as funções podem ser aproximadas por polinômios. Uma curva polinomial, por exemplo, tem a forma

$$y_t = \xi_0 + \xi_1 t + \xi_2 t^2 + \dots$$

Para polinômios simples, é possível fazer interpretações imediatas dos coeficientes. Por exemplo, na equação linear

$$y_t = \xi_0 + \xi_1 t$$

o coeficiente ξ_1 mede a razão média de crescimento. Na equação quadrática

$$y_t = \xi_0 + \xi_1 t + \xi_2 t^2$$

o coeficiente ξ_2 representa metade da aceleração média de crescimento.

A equação polinomial de crescimento pode ser usada para descrever aproximadamente outras curvas. A precisão será maior ou menor, adicionando-se ou retirando-se termos de maior ordem.

Rao (1965, 1975), Potthoff e Roy (1964) e Grizzle e Allen (1969), entre outros, consideraram o problema da análise de curva de crescimento polinomial em medidas seriais de um conjunto de dados balanceados. Considere o seguinte:

Exemplo 2.2.1 conjunto de dados balanceados

Suponha que temos n indivíduos distribuídos em g grupos, cada um com n_k indivíduos ($k = 1, \dots, g$). Em cada grupo é administrado um tratamento diferente. Os indivíduos são observados no decorrer do tempo ($t = t_1, \dots, t_T$) de modo que os dados sejam balanceados no tempo, isto é, todos os n indivíduos são observados nos T instantes de tempo (t_1, \dots, t_T) e o intervalo de tempo entre cada observação é constante (dados igualmente espaçados). Assuma que os valores para indivíduos diferentes sejam estatisticamente independentes.

Dado o exposto no início, pode-se ajustar uma curva de grau $c-1$ associada ao i -ésimo indivíduo ($i = 1, \dots, n_k$) do k -ésimo grupo, denotado por $i(k)$. A medida (resposta) do indivíduo $i(k)$ no tempo t , denotado por $y_{i(k)t}$, é:

$$y_{i(k)t} = \xi_{k,0} + \xi_{k,1}t + \xi_{k,2}t^2 + \dots + \xi_{k,c-1}t^{c-1} + \varepsilon_{i(k)t} \quad (2.2.1)$$

onde $\varepsilon_{i(k)t}$ representa os erros aleatórios com média zero e a correlação temporal dada por

$$E(\varepsilon_{i(k)t}) = 0 \text{ e } E(\varepsilon_{i(k)t} \varepsilon_{i'(k')t'}) = \begin{cases} \sigma_t^2, & \text{se } i' = i, k' = k \text{ e } t' = t; \\ \sigma_u^2, & \text{se } i' = i, k' = k \text{ e } t' \neq t; \\ 0, & \text{se caso contrário.} \end{cases}$$

Em palavras, assume-se que a variância do erro no instante t seja igual para todos os indivíduos (σ_t^2), que a correlação dentro do indivíduo é dada por σ_u^2 e que não há correlação dos erros entre os indivíduos (correlação contemporânea).

Para representarmos todos os indivíduos considere:

$$\begin{aligned} Y_i &= (y_{i(1)t}, \dots, y_{n_1(1)t}, y_{1(2)t}, \dots, y_{n_2(2)t}, \dots, y_{n_g(g)t}), \\ \varepsilon_i &= (\varepsilon_{i(1)t}, \dots, \varepsilon_{n_1(1)t}, \varepsilon_{1(2)t}, \dots, \varepsilon_{n_2(2)t}, \dots, \varepsilon_{n_g(g)t}), \\ \xi &= (\xi_{kk'}, k = 1, \dots, g; k' = 0, \dots, c-1); \end{aligned}$$

então (2.2.1) pode ser escrita como

$$y_i = A \xi (1 \quad t \quad \dots \quad t^{c-1})' + \varepsilon_i,$$

onde $A = \text{diag}(1_{n_1} \quad \dots \quad 1_{n_g})$,

com $n = \sum_{k=1}^g n_k$, $1_{n_k} = (1 \quad \dots \quad 1)_{n_k \times 1}$; $E(\varepsilon_i \varepsilon_i') = \sigma_u^2 I_n$ e $E(\varepsilon_i) = 0$.

O modelo acima pode ser escrito na forma matricial:

$$Y = A \xi \begin{bmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_T \\ t_1^2 & t_2^2 & \dots & t_T^2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1^{c-1} & t_2^{c-1} & \dots & t_T^{c-1} \end{bmatrix} + \varepsilon \quad (2.2.2)$$

onde $A = \text{diag}(1_{n_1} \quad \dots \quad 1_{n_g})$, $\varepsilon = (\varepsilon_1 \quad \dots \quad \varepsilon_T)$ e $Y = (y_1 \quad \dots \quad y_T)$, ou seja,

$$Y = A \xi B + \varepsilon, \quad (2.2.3)$$

$\begin{matrix} n \times T & n \times m & m \times c & c \times T & n \times T \end{matrix}$

onde as colunas de B são potências do tempo ou os polinômios (ou coeficientes) ortogonais definidos pelos tempos de observação. Se o grau do polinômio da curva de crescimento for c-1, então o posto da matriz B será c. As linhas de ε são independentes com distribuição T-variada com média zero e matriz de covariância Σ. A estrutura dessa matriz tem a forma geral, isto é,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1T} & \sigma_{2T} & \cdots & \sigma_T^2 \end{bmatrix}, \quad (2.2.4)$$

□

O modelo (2.2.3) pode ser reduzido à forma usual do modelo MANOVA que é

$$E(Y) = A \beta, \quad (2.2.5)$$

$\begin{matrix} n \times T & n \times m & m \times T \end{matrix}$

onde m é o número de variáveis explanatórias.

Com a redução pode-se utilizar resultados conhecidos da MANOVA em inferências sobre os parâmetros do modelo. A redução não é única. Uma forma de redução pode ser feita através de transformação de variáveis (Potthoff e Roy, 1964):

$$Y_0 = YG^{-1}B'(BG^{-1}B')^{-1}, \quad (2.2.6)$$

onde G é matriz T×T arbitrária positiva definida tal que BG⁻¹B' seja positiva definida.

Rao (1965), para contornar o problema da arbitrariedade da matriz G, propôs outro procedimento de redução. Seja $Y = \begin{bmatrix} X & Z \end{bmatrix}$, onde X é a matriz de observações e Z a matriz das covariáveis. Considere que $E(X) = A \xi$ e $E(Z) = 0$, então

$\begin{matrix} n \times c & n \times (T-c) \end{matrix}$

$$E[X|Z] = A \xi + Z \eta, \quad (2.2.7)$$

$\begin{matrix} n \times m & m \times c & n \times (T-c) & (T-c) \times c \end{matrix}$

onde η é matriz dos parâmetros das covariáveis.

Para aplicar o procedimento de redução de Rao, seguem-se os passos:

- (a) constrói-se uma matriz não singular $H = \begin{bmatrix} H_1 & H_2 \end{bmatrix}$ tal que $BH_2 = 0$ e as colunas de H_1 formem uma base de espaço gerada pelas linhas de B, por exemplo, façamos $BH_1 = I_c$;

(b) multiplica-se 2.2.3 à direita, obtendo-se $E(YH_1) = A\xi$ e $E(YH_2) = 0$ e

(c) considere $X = YH_1$ como as variáveis de interesse e $Z = YH_2$ como as "covariáveis".

Quando assume-se que $G = I$ em (2.2.6), o estimador de ξ por Rao pode ser mais eficiente do que Potthoff e Roy, pois nessa redução a informação da matriz Z que representa as covariáveis foi ignorada no processo de estimação de ξ . Entretanto, a inclusão de covariáveis deve ser criteriosa, porque o uso de todas as covariáveis, por exemplo, nem sempre resulta em estimador mais eficiente (Rao, 1965, 1966). Mas se $G = S$ (matriz de covariância amostral), o estimador obtido pela redução de Rao é equivalente ao estimador de máxima verossimilhança de ξ através de Potthoff e Roy (Khatri, 1966).

Os próximos exemplos extraídos de Potthoff e Roy (1964) ilustram a de aplicação do modelo de curva de crescimento polinomial:

Exemplo 2.2.2 dois conjuntos de tratamentos

Considere que indivíduos sejam submetidos a dois conjuntos de tratamentos simultaneamente: o primeiro conjunto reúne três tipos de dietas e o segundo dois níveis de temperatura. Assim, cada indivíduo exposto à temperatura β é submetido à dieta α ($\alpha = 1, 2, 3$ e $\beta = 1, 2$). Assumindo-se que não haja interação entre dieta e temperatura, a curva de crescimento polinomial para um indivíduo da α -ésima dieta na β -ésima temperatura é

$$\xi_{\alpha,0} + \xi_{\alpha,1}t + \xi_{\alpha,2}t^2 + \dots + \xi_{\alpha,c-1}t^{c-1} + \psi_{\beta,0} + \psi_{\beta,1}t + \psi_{\beta,2}t^2 + \dots + \psi_{\beta,c-1}t^{c-1}.$$

Cada um dos seis grupos possui $n_{\alpha\beta}$ indivíduos ($\alpha = 1, 2, 3$ e $\beta = 1, 2$). Suponha que todos os indivíduos sejam observados nos mesmos T instantes de tempo e tenham a mesma matriz de variância (desconhecida) $\Sigma_{T \times T}$. Assim, o (a,b) -ésimo elemento de $\xi_{S,c}$ será $\xi_{a,b-1}$ se $a = 1, 2, 3$ e será $\psi_{a-3,b-1}$ se $a = 4$ ou 5 . A matriz $A_{n \times 5}$ conterà $n_{\alpha\beta}$ linhas com 1 (um) na α -ésima e $(\beta + 3)$ -ésima colunas e 0 (zero) nas demais.

□

Exemplo 2.2.3 multi-resposta

Suponha que temos g grupos de indivíduos, onde todos são observados nas mesmas T ocasiões. Considere que observamos mais de uma característica associada ao crescimento ao invés de somente uma; por exemplo, o peso e altura dos indivíduos. Para o k -ésimo grupo, a curva associada à altura é

$$\xi_{k,0} + \xi_{k,1}t + \xi_{k,2}t^2 + \dots + \xi_{k,c-1}t^{c-1},$$

e a curva associada ao peso

$$\psi_{k,0} + \psi_{k,1}t + \psi_{k,2}t^2 + \dots + \psi_{k,c_2-1}t^{c_2-1}.$$

No total, há $2T$ observações para cada indivíduo; logo, a matriz Σ será de dimensão $2T \times 2T$ e espera-se que haja correlação entre peso e altura. Fazendo $c = c_1 + c_2$ e $m = g$, o (a,b) -ésimo elemento de $\xi_{g \times c}$ será $\xi_{a,b-1}$ se $b \leq c_1$ e será $\psi_{a,b-c_1-1}$ se $b > c_1$. A matriz $Y_{n \times 2T}$ é arranjada de modo que as primeiras T colunas contenham medidas de altura e as últimas T colunas as de peso. A matriz $B_{c \times 2T}$ terá a forma

$$\left[\begin{array}{c|c} B_1_{c_1 \times T} & 0 \\ \hline 0 & B_2_{c_2 \times T} \end{array} \right],$$

onde B_1 e B_2 são matrizes semelhantes àquela apresentada em (2.2.2), isto é, as colunas são potências do tempo com graus c_1 e c_2 , respectivamente.

□

2.3. Modelo de Regressão

Na seção anterior apresentamos o modelo de curva de crescimento polinomial sob um conjunto de dados balanceado. No modelo de regressão as observações individuais y_i ($i = 1, \dots, n_k$) têm a mesma distribuição $N_T(B' \xi', \Sigma)$, onde as colunas da matriz B são potências do tempo ou polinômios ortogonais definidos pelos tempos de observações. O caso do modelo de regressão considerado, por exemplo, por Crowder e Hand (1990, pp. 73) consiste de uma generalização do modelo de regressão de Rao (1959). O modelo permite que a matriz B possa diferir entre os indivíduos, i. e., cada vetor de observações individuais y_i tem tamanho T_i . A matriz de delineamento B_i pode também ser construída a partir das variáveis explanatórias ou dos indicadores do delineamento (variáveis *dummies*). Por isso, o modelo (2.2.3) é um caso particular onde se considera dados balanceados $T_i = T$ com uma forma especial de B_i a partir do delineamento dentro e entre grupos.

O modelo de regressão pode ser representado por:

$$y'_{T_i,1} = B'_{T_i \times c} \xi'_{c,1} + \varepsilon'_{T_i,1} \quad (2.3.1)$$

onde os ε'_i ($i = 1, \dots, n$) são independentes e têm distribuição $N_{T_i}(0, \Sigma_i)$.

A estrutura Σ_j é uma submatriz de Σ (2.2.4) correspondente aos valores observados de y_j . Portanto, ela é obtida de Σ , removendo-se as linhas e colunas correspondentes aos $(T-T_j)$ valores perdidos de y_j .

Kleinbaum (1973) investigou métodos multivariados (seções 2.2 e 2.3) para estimação do vetor de médias e matriz de covariância para dados não balanceados. Entretanto, quando o número de ocasiões (T) se torna relativamente maior que o número de indivíduos (n), a qualidade dos estimadores poderá ser ruim, porque haverá perda de eficiência (Ware, 1985).

Outra abordagem de regressão é considerar as observações dos n indivíduos no instante j como a variável dependente:

$$Y_j = X_j \beta_j + \varepsilon_j \quad (2.3.2)$$

$\begin{matrix} n \times 1 & n \times m_j & m_j \times 1 & n \times 1 \end{matrix}$

onde m_j é o número de variáveis explanatórias no instante j com $E(\varepsilon_j) = 0$ e $E(\varepsilon_j \varepsilon_j') = \sigma_{jj} I_n$ para $j = 1, \dots, T$. Esse modelo é chamado de Regressões Aparentemente Não-relacionadas (Zellner, 1962) ou SUR (*Seemingly Unrelated Regression*), porque aparentemente as equações de regressão para os T vetores de variáveis dependentes não possuem nenhum termo em comum. Entretanto, dada a correlação através dos erros, pode-se aumentar a eficiência da estimação quando se considera as correlações. Stanek e Koch (1985) mostraram a equivalência entre uma forma especial do modelo SUR e o modelo de Curva de Crescimento (2.2.3). Park e Woolson (1992) propuseram uma extensão do modelo SUR para dados longitudinais. Esse modelo é aplicável em conjuntos de dados não-balanceados ou em dados incompletos. O modelo tem aplicação em uma variedade de problemas, envolvendo, por exemplo, covariáveis variando no tempo e curvas de crescimento. Outra possível aplicação é o ajuste de curvas polinomiais com graus diferenciados para cada grupo ou indivíduo.

2.4. Modelo Misto Univariado

Esse modelo foi considerado por Elston e Grizzle (1962) como alternativa univariada aos modelos multivariados (seções 2.2 e 2.3). Sob o ponto de vista univariado, um modelo para a curva média pode ser obtido diretamente definindo-se a esperança de cada observação como função do tempo da respectiva observação e de covariáveis. O modelo misto univariado adota uma estrutura mais restritiva para a matriz de covariância das observações, como pode ser visto em (2.4.3). Ware (1985) destaca algumas vantagens como

a não obrigatoriedade dos indivíduos serem observados nos mesmos tempos, ou com o mesmo número de ocasiões e permite o uso de covariáveis variando no tempo, cujas contribuições à resposta esperada podem ser escritas linearmente. Andrade e Singer (1986, pp. 24) comentam que o método univariado produz testes mais sensíveis - com maior poder - do que aqueles produzidos por métodos multivariados.

Considere a situação descrita no exemplo 2.2.1. Um modelo associado a esse exemplo pode ser escrito como:

$$y_{i(k)j} = \mu + \alpha_k + \pi_{i(k)} + \varepsilon_{i(k)j}, \quad (2.4.1)$$

onde k denota o grupo, $i(k)$ o indivíduo dentro do grupo k e j denota a observação associada ao indivíduo $i(k)$; μ representa a média geral, α_k é o efeito fixo do tratamento no grupo k , $\pi_{i(k)}$ é o efeito aleatório hierárquico do indivíduo i pertencente ao grupo k , $\varepsilon_{i(k)j}$ é o efeito aleatório. O efeito fixo é restrito a

$$\sum_k \alpha_k = 0,$$

e para os efeitos aleatórios supõe-se que sejam independentes e normalmente distribuídos com média 0, $\text{var}(\pi_{i(k)}) = \sigma_\pi^2$ (componente de variação entre indivíduos) e $\text{var}(\varepsilon_{i(k)j}) = \sigma_\varepsilon^2$ (componente de variação intra-individual).

A variância de cada observação é:

$$\text{var}(y_{i(k)j}) = \text{var}(\pi_{i(k)} + \varepsilon_{i(k)j}) = \sigma_\pi^2 + \sigma_\varepsilon^2 = \sigma^2, \text{ e} \quad (2.4.2.a)$$

$$\text{cov}(y_{i(k)j}, y_{i(k)j'}) = \text{cov}\left\{(\pi_{i(k)} + \varepsilon_{i(k)j})(\pi_{i(k)} + \varepsilon_{i(k)j'})\right\} = \sigma_\pi^2, \quad (2.4.2.b)$$

logo, a matriz de covariância para as observações individuais é dada por

$$\Sigma = \text{var}(y_{i(k)}) = \begin{bmatrix} \sigma^2 & \sigma_\pi^2 & \dots & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma^2 & \dots & \sigma_\pi^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\pi^2 & \sigma_\pi^2 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}, \quad (2.4.3)$$

onde $\sigma^2 = \sigma_\pi^2 + \sigma_\varepsilon^2$ e $\rho = \sigma_\pi^2 / \sigma^2$.

O modelo (2.4.1) é um caso do *modelo componente de variâncias*. Ele recebe esse nome porque a variância da variável observada é composta por uma soma (2.4.2.a), sendo seus elementos σ_π^2 e σ_ε^2 chamados de componentes de variância. No capítulo 4 discutiremos

sobre as duas classes de efeitos, fixos e aleatórios, e ilustraremos a natureza de cada classe com alguns exemplos encontrados na literatura.

Outro exemplo de modelo misto univariado é:

$$y_{i(k)j} = \mu + \alpha_k + \pi_{i(k)} + \beta_j + \alpha\beta_{kj} + \pi\beta_{i(k)j} + \varepsilon_{i(k)j}, \quad (2.4.4)$$

onde β_j é o efeito da j -ésima ocasião, $\alpha\beta_{kj}$ é a interação entre o k -ésimo grupo e a j -ésima ocasião e $\pi\beta_{i(k)j}$ é a interação entre a j -ésima ocasião e o indivíduo $i(k)$. Os demais efeitos foram descritos em 2.4.1. Os efeitos fixos são restritos a

$$\sum_k \alpha_k = \sum_j \beta_j = \sum_j \alpha\beta_{kj} = \sum_k \alpha\beta_{kj} = 0,$$

e os efeitos aleatórios são independentes com distribuições:

$$\pi_{i(k)} \sim N(0, \sigma_\pi^2), \quad \pi\beta_{i(k)j} \sim N(0, \sigma_{\pi\beta}^2) \quad \text{e} \quad \varepsilon_{i(k)j} \sim N(0, \sigma_\varepsilon^2).$$

Assim, a matriz de covariância Σ tem a forma de 2.4.3 com $\sigma^2 = \sigma_\pi^2 + \sigma_{\pi\beta}^2 + \sigma_\varepsilon^2$ e $\rho = \sigma_\pi^2 / \sigma^2$.

Matricialmente pode-se escrever 2.4.3 como

$$\Sigma = \sigma_\pi^2 \mathbf{1}_p \mathbf{1}'_p + \sigma_\varepsilon^2 \mathbf{I}_p = \left[\rho \mathbf{1}_p \mathbf{1}'_p + (1-\rho) \mathbf{I}_p \right] \sigma^2,$$

onde $\mathbf{1}$ é vetor $p \times 1$ de uns e \mathbf{I} é matriz identidade $p \times p$. Essa estrutura de covariância é conhecida como padrão de uniformidade ou simetria composta e ρ é denominado coeficiente de correlação intraclassa. Nota-se através de (2.4.2) que a covariância individual é constante e não há dependência em função da defasagem temporal. Ela é igual ao componente de variação entre os indivíduos, produzindo a matriz de variância (2.4.3).

O padrão de uniformidade é adequado quando os indivíduos são alocados ao acaso nas diferentes ocasiões (delineamentos de medidas repetidas). Em estudos longitudinais é razoável esperar que a correlação entre duas observações tomadas em tempos próximos seja maior do que quando tomadas em tempos distantes. Assim, a aplicação direta não é, em geral, recomendada para dados gerados por esses tipos de estudos. Por outro lado, a simplicidade na obtenção e interpretação das estatísticas F da ANOVA que serão discutidos no capítulo 6 (Tabelas 6.2.5.1 a 6.2.5.3) fazem com que elas sejam bastante utilizadas em estudos desse tipo. Desse modo, autores como Box (1954) e Geisser e Greenhouse (1958)

propuseram correções no número de graus de liberdade das estatísticas F para acomodar os desvios do padrão de uniformidade de Σ .

O modelo misto univariado pode ser considerado como um caso particular do modelo de efeitos aleatórios apresentado a seguir:

2.5. Modelo de Efeitos Aleatórios

O Modelo de Efeitos Aleatórios, também conhecido como modelo de Duplo Estágio ou modelo de efeitos mistos, foi estudado por Rao (1975), Harville (1977) e Laird e Ware (1982). Este modelo possui uma estrutura de covariância mais flexível do que os modelos anteriores. Ela é útil em casos onde os dados não são balanceados ou possuem tamanhos amostrais pequenos.

Na formulação de Rao, o primeiro estágio consiste de um modelo linear condicionado ao parâmetro individual da curva de crescimento t_i , isto é,

$$(1^\circ \text{ estágio}) \quad y_i | t_i = A_i t_i + \varepsilon_i, \quad (2.5.1)$$

onde $\varepsilon_i \sim N(0, \Omega_i)$, A_i é a matriz de delineamento $T_i \times m$ do i -ésimo indivíduo. No segundo estágio assume-se que

$$(2^\circ \text{ estágio}) \quad t_i \sim N(\tau, \Gamma). \quad (2.5.2)$$

Laird e Ware representaram o modelo numa forma menos restritiva:

$$(1^\circ \text{ estágio}) \quad y_i | b_i = X_i \beta + Z_i b_i + \varepsilon_i, \quad (2.5.3)$$

onde $\varepsilon_i \sim N(0, \Omega_i)$, X_i é a matriz $T_i \times a$ de delineamento do indivíduo i , β é o vetor dos parâmetros fixos, Z_i é a matriz $T_i \times m$ de delineamento individual para o vetor de efeitos (não observáveis) b_i que nesse estágio é considerado fixo, mas no segundo estágio,

$$(2^\circ \text{ estágio}) \quad b_i \sim N(0, \Gamma). \quad (2.5.4)$$

A matriz de covariância é dada por:

$$\Sigma_i = \text{var}(y_i) = Z_i \Gamma Z_i' + \Omega_i. \quad (2.5.5)$$

As matrizes Γ e Ω_i são arbitrárias e em geral Ω_i contém poucos parâmetros. Por exemplo, $\Omega_i = \sigma^2 \mathbf{I}$, ou Ω_i pode ter a estrutura de processo AR(1) (3.2.1.2). Quando $\Omega_i = \sigma^2 \mathbf{I}$, o modelo é chamado de "modelo de independência condicional", pois as T_i respostas do i -ésimo indivíduo condicionadas em b_i e β são independentes (Laird e Ware, 1982).

Do ponto de vista Bayesiano, os modelos de efeitos mistos são chamados de modelos hierárquicos (Searle *et alii.*, 1992, cap. 9). Um modelo hierárquico é especificado em estágios, onde cada estágio é construído sobre outro estágio. A vantagem de se construir um modelo em estágios é que cada estágio pode ser relativamente simples e fácil de ser entendido, enquanto que o modelo inteiro (com todos os estágios reunidos) possa ser complicado. Os métodos Bayesianos estão fortemente ligados aos modelos hierárquicos. A estimação de Bayes baseia-se no cálculo de uma distribuição a posteriori que é resultado da combinação da distribuição a priori com a distribuição amostral. A especificação da distribuição amostral e da priori é um exemplo de um modelo hierárquico. Por exemplo, se observarmos uma variável aleatória X com distribuição $f(x|\theta)$, e supondo que θ tenha uma distribuição a priori $\pi(\theta)$, então,

$$X|\theta \sim f(x|\theta) \text{ e } \theta \sim \pi(\theta)$$

é uma especificação hierárquica. Nesse caso há dois níveis de hierarquia: o primeiro nível lida com a distribuição da variável X (por exemplo, o conjunto de dados) condicional a θ (por exemplo, o conjunto de parâmetros). Então, o modelo de efeitos mistos é um modelo hierárquico com dois estágios hierárquicos representados por (2.5.3) e (2.5.4).

Exemplo 2.5.1 modelo misto univariado

O modelo misto univariado (§ 2.4) pode ser obtido a partir de (2.5.3) e (2.5.4). Para cada indivíduo $i(k)$,

$$y_{i(k)} = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}_{T_i \times 2} \begin{bmatrix} \mu \\ \alpha_k \end{bmatrix} + \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{T_i \times 1} \pi_{i(k)} + \begin{bmatrix} \varepsilon_{i(k)1} \\ \vdots \\ \varepsilon_{i(k)T_i} \end{bmatrix}, \quad (2.5.6)$$

onde $\pi_{i(k)} \sim N(0, \sigma_\pi^2)$ e $\varepsilon_{i(k)} \sim N(0, \sigma_\varepsilon^2 \mathbf{I})$.

Assim,

$$\text{var}(y_{i(k)}) = Z_i \Gamma Z_i' + \Omega_i = 1 \sigma_\pi^2 1' + \sigma_\varepsilon^2 \mathbf{I} = \sigma_\pi^2 1 1' + \sigma_\varepsilon^2 \mathbf{I}$$

é a mesma estrutura de covariância em (2.4.3). Essencialmente, o modelo de efeitos aleatórios é uma generalização do modelo misto univariado (2.4.1). Os efeitos fixos daquele estendem-se para $X_i\alpha$ e os efeitos aleatórios para $Z_i b_i$. A estrutura de covariância resultante (2.5.5) é, por um lado, mais geral do que a simetria composta (2.4.2), e por outro, mais estruturada do que aquelas no modelo de curva de crescimento polinomial (seção 2.2) e de regressão (seção 2.3).

□

Exemplo 2.5.2 modelo de regressão univariada

Tome uma reta de regressão (Crowder e Hand, 1990, pp. 87)

$$y_{ij} = a_i + b_i x_j + \varepsilon_{ij}$$

onde y_{ij} é a medida do i -ésimo indivíduo no tempo t . Suponha que o coeficiente de inclinação da reta b_i varie entre os indivíduos, ou seja,

$$b_i = b + v_i$$

onde b é a média da população dos coeficientes de inclinação e v_i é o desvio individual da média. Assuma que $E(v_i) = 0$ e que os erros ε_{ij} não sejam correlacionados entre si, nem com v_i . Sejam $\sigma_\varepsilon^2 = \text{var}(\varepsilon_{ij})$ e $\sigma_b^2 = \text{var}(v_i)$ e, por enquanto, tome $a_i = a$. Então, $E(y_{ij}) = a + bx_j$, $\text{var}(y_{ij}) = \sigma_b^2 x_j^2 + \sigma_\varepsilon^2$, e $\text{cov}(y_{ij}, y_{ij'}) = \sigma_b^2 x_j x_{j'}$, para $j \neq j'$. Nota-se que a variação no coeficiente b_i incorpora tanto a heterogeneidade da variância ($\text{var}(y_{ij})$ cresce com x_j) como a correlação das observações y_{ij} com $y_{ij'}$.

Suponha agora que o intercepto também varie entre os indivíduos, isto é,

$$a_i = a + \eta_i$$

onde $E(\eta_i) = 0$ e $\text{var}(\eta_i) = \sigma_a^2$. Seja $\sigma_{ab} = \text{cov}(\eta_i, v_i)$ e assumamos que $\text{cov}(\eta_i, \varepsilon_{ij}) = 0$.

Então,

$$E(y_{ij}) = a + bx_j$$

$$\text{var}(y_{ij}) = (\sigma_a^2 + 2x_j\sigma_{ab} + x_j^2\sigma_b^2) + \sigma_\varepsilon^2$$

$$\text{cov}(y_{ij}, y_{ij'}) = \sigma_a^2 + (x_j + x_{j'})\sigma_{ab} + x_j x_{j'}\sigma_b^2, \text{ para } j \neq j'$$

Para representarmos as equações acima na forma matricial, façamos

$y_i = (y_{i1}, \dots, y_{iT})'$; $\alpha = (a, b)'$; $b_i = (\eta_i, v_i)'$; $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'$ e X matriz $T \times 2$ com a j -ésima linha $(1, x_j)$.

Logo, $y_i = X(\alpha + b_i) + \varepsilon_i = X\alpha + Xb_i + \varepsilon_i$, com $E(y_i) = X\alpha$ e $\text{var}(y_i) = X\Gamma X' + \Omega_i$,

onde $\Gamma = \text{var}(b_i) = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}$ e $\Omega_i = \text{var}(\varepsilon_i) = \sigma^2 I_{T_i}$. □

Outros modelos também podem ser obtidos:

Exemplo 2.5.3 modelo componente de variância

Com $\Omega_i = 0$ e $\Gamma_{m \times m} = \text{diag}(\sigma_1^2 I_{c_1}, \dots, \sigma_w^2 I_{c_w})$, onde $m = \sum_{u=1}^w c_u \geq T_i$, o modelo resultante é chamado "variance components" (Dielman, 1989):

$$y_i = X_i \alpha + Z_i b_i, \text{ onde } b_i \sim N(0, \Gamma_{m \times m}); \tag{2.5.7}$$

por exemplo, escrevendo-se (2.5.6) na forma acima temos:

$$y_{i(k)} = \begin{bmatrix} 1_{T_i} & 1_{T_i} \end{bmatrix}_{T_i \times 2} \begin{bmatrix} \mu \\ \alpha_k \end{bmatrix} + \begin{bmatrix} 1_{T_i} & 1_{T_i} \end{bmatrix}_{T_i \times (T_i+1)} \begin{bmatrix} \pi_{i(k)} \\ \varepsilon_{i(k)1} \\ \vdots \\ \varepsilon_{i(k)T_i} \end{bmatrix},$$

com $\Gamma_{(T_i+1) \times (T_i+1)} = \text{diag}(\sigma_\pi^2, \sigma_\varepsilon^2 I_{T_i})$; os parâmetros σ_π^2 e σ_ε^2 são as componentes da variância (na análise de componente de variância, a idéia é que o erro experimental pode originar-se de diferentes fontes de variação, e a importância está na identificação dessas fontes).

Outra forma de representar o modelo componente de variâncias é escrever o modelo linear misto (2.5.3) com $b_i \sim N(0, \sigma_b^2 D)$ e $\varepsilon_i \sim N(0, \sigma_\varepsilon^2 I)$, onde D é uma matriz simétrica positiva definida. Por exemplo, $b' = (b'_1, \dots, b'_g)$, onde b_k contém os efeitos aleatórios do k -ésimo tratamento. Tipicamente, assume-se que b_1, \dots, b_g não são correlacionados e que $\text{var}(b) = \text{diag}(\sigma_k^2 I)$. Os parâmetros $\sigma_1^2, \dots, \sigma_g^2$ e σ_ε^2 são as componentes da variância. É possível, entretanto, considerar que haja correlação entre os coeficientes aleatórios b_1, \dots, b_g e assumir, por exemplo, uma forma arbitrária para $\text{var}(b)$. □

Exemplo 2.5.4 modelo de regressão

No exemplo anterior, fazendo-se $Z_i = I_{T_i}$ em (2.5.7), obtemos

$$y_i = X_i \alpha + b_i, \text{ onde } b_i \sim N(0, \Gamma_{T_i \times T_i}), \tag{2.5.8}$$

que é um modelo de regressão (seção 2.3) com o número de linhas de X_i igual ao número de coeficientes do polinômio ($a = c$). O valor de a depende da parametrização adotada. Considere que no exemplo 2.2.1, temos $g = 2$ grupos, cujos indivíduos são observados em

$t = t_1, t_2$ e t_3 . Com $a = 6$ constrói-se por exemplo o modelo saturado, onde cada grupo é representado por uma curva arbitrária (2.5.9), ou o modelo (2.5.10):

$$\begin{cases} y_{i(1)} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ 3 \times 3 & 3 \times 3 \end{bmatrix} \alpha_{6 \times 1} \\ y_{i(2)} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ 3 \times 3 & 3 \times 3 \end{bmatrix} \alpha_{6 \times 1} \end{cases}, \quad (2.5.9)$$

$$\begin{cases} y_{i(1)} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ 3 \times 3 & 3 \times 3 \end{bmatrix} \alpha_{6 \times 1} \\ y_{i(2)} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ 3 \times 3 & 3 \times 3 \end{bmatrix} \alpha_{6 \times 1} \end{cases}, \quad (2.5.10)$$

onde $\alpha = (\alpha_1 \dots \alpha_6)'$; os três primeiros correspondem às observações t_1, t_2 e t_3 do grupo 1 e os restantes do grupo 2. Com $a = 4$, pode-se construir um modelo polinomial de grau 1,

$$\begin{aligned} y_{i(1)} &= \begin{bmatrix} 1 & t_1 & 0 & 0 \\ 1 & t_2 & 0 & 0 \\ 1 & t_3 & 0 & 0 \end{bmatrix} \alpha_{4 \times 1} \\ y_{i(2)} &= \begin{bmatrix} 1 & t_1 & 1 & t_1 \\ 1 & t_2 & 1 & t_2 \\ 1 & t_3 & 1 & t_3 \end{bmatrix} \alpha_{4 \times 1} \end{aligned},$$

que pode ser representado da forma 2.2.1:

$$\begin{aligned} y_{i(1)} &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \end{bmatrix} \\ y_{i(2)} &= \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \end{bmatrix} \end{aligned},$$

onde $\alpha = (\alpha_1 \dots \alpha_4)'$; α_1 e α_2 correspondem, respectivamente, ao intercepto e inclinação da reta associada ao grupo 1 e α_3 e α_4 representam as diferenças dos coeficientes entre os grupos.

□

Exemplo 2.5.5 demanda de gasolina nos EUA

Como exemplo de aplicação em Econometria do modelo de efeitos aleatórios, Mehta, Narasimham e Swamy (1978) utilizaram esse modelo para examinar a demanda de

gasolina nos Estados Unidos. O modelo adotado $y_i = X_i\beta + \varepsilon_i$, com $\beta_i = \bar{\beta} + \tau_i$, pode ser reescrito na forma de 2.5.1

$$y_i = X_i(\bar{\beta} + \tau_i) + \varepsilon_i \Rightarrow y_i = X_i\bar{\beta} + X_i\tau_i + \varepsilon_i.$$

Os vetores dos coeficientes β_i representam os efeitos de várias variáveis explanatórias em demanda para os diversos Estados. Eles notaram que os coeficientes poderiam estar próximos a um valor médio $\bar{\beta}$ com alguma variação. Kraft e Rodekohr (1978) justificaram o uso de coeficientes aleatórios na função de demanda de gasolina. Segundo eles, a variação se deve à infraestrutura e possibilidades de transporte de massa que afetam as alternativas das pessoas quando deparados com mudanças de preço da gasolina e renda.

□

Exemplo 2.5.6 nível de ácido láctico

Considere o conjunto de dados experimentais apresentado na seção 1.2. Esse conjunto de dados é formado por dois grupos: ratos treinados (grupo 1) e ratos sedentários (grupo 2). Os ratos são submetidos à atividade física durante 1 hora. Nesse período, observou-se a cada 10 minutos (incluindo o instante inicial) o nível individual de ácido láctico.

O nível observado de ácido do i -ésimo indivíduo do grupo k ($k = 1, 2$) no tempo t será denotado por $y_{i(k)t}$, onde $i(1) = 1, \dots, 20$ (ratos do grupo 1); $i(2) = 1, \dots, 16$ (ratos do grupo 2) e $t = 0, 10, 20, 30, 40, 50, 60$ minutos.

Para se ajustar por exemplo um modelo polinomial, as matrizes de delineamento dos efeitos fixos para cada grupo podem ser dadas por:

para o grupo 1:
$$X_{i(1)} = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{c-1} & 0 & \dots & 0 \\ 1 & t_2 & t_2^2 & \dots & t_2^{c-1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & t_7 & t_7^2 & \dots & t_7^{c-1} & 0 & \dots & 0 \end{bmatrix} \mathbf{e}$$

para o grupo 2:
$$X_{i(2)} = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{c-1} & 1 & t_1 & t_1^2 & \dots & t_1^{d-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{c-1} & 1 & t_2 & t_2^2 & \dots & t_2^{d-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & t_7 & t_7^2 & \dots & t_7^{c-1} & 1 & t_7 & t_7^2 & \dots & t_7^{d-1} \end{bmatrix},$$

onde t_1 a t_7 denotam, respectivamente, os instantes $t = 0$ a 60 minutos; $c-1 \leq 6$ representa o grau do polinômio da curva ajustada para o grupo 1 (grupo de referência) e $d-1 \leq c-1$

representa os desvios dos respectivos parâmetros polinomiais do grupo 2 em relação ao grupo 1.

Por exemplo, para fazer uma investigação inicial aos dados, pode-se ajustar o modelo saturado. Nesse caso, a matriz de delineamento dos efeitos fixos terá dimensão 7×14 , i. e., as linhas dessa matriz formam um polinômio de grau $(c-1) = 6$ e cada coeficiente polinomial associada ao grupo 2 será representada pelo desvio em relação ao grupo 1 ($d-1 = 6$). Assim, no vetor de coeficientes fixos β do modelo saturado, 14×1 , os parâmetros β_1 a β_7 são os coeficientes polinomiais associados ao grupo 1, e β_8 a β_{14} ao grupo 2.

Para ajustarmos um modelo sob suposição de "paralelismo" entre as curvas, a matriz de delineamento dos efeitos fixos poderá ter dimensão, por exemplo, 7×6 , onde as linhas formam um polinômio de grau 4 ($c-1 = 4$) e o desvio do grupo 2 em relação ao grupo 1, por causa do paralelismo, será avaliada pelo intercepto ($d-1 = 0$). Assim, nesse modelo, os coeficientes β_1 a β_5 do vetor β , 6×1 , são os coeficientes polinomiais associados ao grupo 1 e β_6 é a diferença vertical entre as curvas (intercepto).

As linhas da matriz de delineamento dos efeitos aleatórios também podem ser polinômios até de grau 6, ou seja, para $m-1 \leq 6$:

$$Z_i = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{m-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{m-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_7 & t_7^2 & \dots & t_7^{m-1} \end{bmatrix}$$

Por exemplo, o modelo saturado com um efeito aleatório ($m-1=1$) pode ser descrito individualmente como

$$y_{i(1)t} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \beta_5 t^5 + \beta_6 t^6 + b_i + \varepsilon_{it}$$

$$y_{i(2)t} = (\beta_0 + \beta_7) + (\beta_1 + \beta_8)t + (\beta_2 + \beta_9)t^2 + (\beta_3 + \beta_{10})t^3 + (\beta_4 + \beta_{11})t^4 + (\beta_5 + \beta_{12})t^5 + (\beta_6 + \beta_{13})t^6 + b_i + \varepsilon_{it}$$

onde, dos 14 parâmetros lineares, β_0 a β_{13} , os 7 primeiros descrevem a curva o grupo de referência (ratos treinados) e os 7 restantes descrevem a diferença dos respectivos coeficientes entre os grupos; b_i é a componente aleatória que descreve a variação entre indivíduos no intercepto: $\beta_0 + b_i$ (para o grupo 1) e $\beta_0 + \beta_7 + b_i$ (para o grupo 2).

A matriz de covariância individual Σ_i pode ser escrita como $\sigma^2(Z_i B Z_i' + W_i)$, onde σ^2 é a variância residual, B é a matriz $m \times m$ das covariâncias dos efeitos aleatórios e W_i é a matriz de covariância 7×7 dos erros. A matriz W_i em geral assume formas estruturadas. Por exemplo, se os erros forem independentes, então ela será dada por $W_i = I$. Caso seja

considerada a existência de uma correlação temporal nos erros, como por exemplo a estrutura AR(1) nos erros a ser discutido no capítulo 3, i.e.,

$$\varepsilon_{it} = \phi \varepsilon_{it-1} + a_{it},$$

então, W_i assumirá a forma autoregressiva:

$$W_i = \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^6 \\ \phi & 1 & \phi & \dots & \phi^5 \\ \phi^2 & \phi & 1 & \dots & \phi^4 \\ \vdots & \vdots & \vdots & & \vdots \\ \phi^6 & \phi^5 & \phi^4 & \dots & 1 \end{bmatrix},$$

onde ϕ é o parâmetro de autoregressão e a_{it} é chamado de ruído branco. No capítulo 3 estudaremos mais detalhadamente algumas estruturas para modelar a correlação serial nos erros.

No capítulo 7 aplica-se a abordagem do modelo de efeitos mistos descrita nesse exemplo para analisar os dados dos ratos descrito no capítulo 1.

□

2.6. Modelo Linear Generalizado

O modelo linear generalizado (GLM) introduzido por Nelder e Wedderburn (1972) tem sido utilizado em dados longitudinais por vários pesquisadores (e.g., McCullagh e Nelder (1983), Liang e Zeger (1986) e Zeger e Liang (1986)). Esse modelo difere dos anteriores porque ele engloba vários modelos como ANOVA, regressão, log linear, logito, probito e componente de variância. Todos esses modelos são especificados por uma variável aleatória, um conjunto de variáveis explanatórias e a especificação de uma distribuição de probabilidade para a variável aleatória em termos das variáveis explanatórias. Nos modelos citados, a especificação da distribuição tem uma forma particular, permitindo a simplificação e unificação teórica. Os modelos que têm essa forma particular são conhecidos como GLM.

O GLM é definido pelas seguintes características:

- a) A distribuição das observações y_i dada as variáveis explanatórias x_{i1}, \dots, x_{in} pertence à família exponencial

$$f(y_i) = \exp\left[\frac{y_i \lambda_i - a(\lambda_i)}{\psi} + b(y_i, \psi)\right], \quad (2.6.1)$$

onde λ_i é uma função de x_{i1}, \dots, x_{in} que envolve parâmetros desconhecidos e y_i são conjuntamente independentes. O parâmetro λ_i é chamado de parâmetro natural e ψ é o parâmetro de escala. A média e variância de y_i são, respectivamente, $a'(\lambda_i)$ e $\psi a''(\lambda_i)$.

b) O preditor linear é dado por

$$\eta_i = \sum_{q=1}^a x_{iq} \beta_q, \quad (2.6.2)$$

onde β são os parâmetros lineares.

c) O valor esperado de y_i pode ser expressado como função do seu preditor linear, i.e.,

$$E(y_i) = g(\eta_i), \quad (2.6.3)$$

onde $g(\cdot)$ é chamada de função de ligação.

É importante destacar a diferença entre o modelo linear generalizado e o modelo linear geral. No modelo linear geral, assume-se que $y = X\beta + e$, onde $E(e) = 0$ e $\text{var}(e) = V$ (matriz de covariância positiva definida). No GLM assume-se as condições (2.6.1), (2.6.2) e (2.6.3).

Como o GLM permite modelar tanto dados contínuos como dados discretos, ele pode ser utilizado em análise de resposta binária e Poisson. Apesar dessa generalidade, consideraremos apenas o caso de normalidade e de especificação somente da média e variância da observação. Neste último, a estimação está relacionada com a quase-verossimilhança que será discutida no capítulo 5. Na seção 5.2.4 discutiremos o emprego do GLM em dados longitudinais introduzido por Liang e Zeger (1986).

3

Modelagem Dinâmica

3.1. Introdução

Em Dados Longitudinais observados no tempo, cada conjunto individual representa uma série temporal o que, em geral, leva ao aparecimento de correlação serial. Assim, é razoável esperar que a correlação entre duas observações individuais tomadas em tempos próximos seja maior do que quando tomadas em tempos distantes. Este capítulo apresenta algumas estruturas de correlação, mostrando como elas podem aparecer e como podem ser incorporadas ao modelo de efeitos mistos. O estudo dessas estruturas é interessante porque permite, por um lado, a redução do número de parâmetros da matriz de covariância em comparação à forma não estruturada, e por outro lado, é uma opção possivelmente mais adequada do que simplesmente assumir independência dos erros ou o padrão de uniformidade para as observações. Nesse trabalho, a correlação serial será incorporada a um determinado modelo através dos erros ou dos coeficientes, i.e., os erros ou os coeficientes ou ambos poderão ser modelados como séries temporais.

É importante conhecer a estrutura de covariância dos vários modelos propostos na literatura para que se possa escolher o mais adequado para modelar o conjunto de dados sob consideração. Como é comum haver um número pequeno de observações seriais, torna-se necessário, sempre que possível, a aplicação do princípio da parcimoniosidade no ajuste de modelos, i.e., determinar e modelar a estrutura de correlação, levando-se em consideração esse princípio. Assim, os modelos para a correlação serial dos erros não devem exigir um número grande de parâmetros para serem estimados. Em séries curtas, muitas vezes, é difícil determinar exatamente a estrutura de correlação real. Entretanto, modelos simples como os que serão apresentados nesse capítulo podem aproximar-se bem da realidade, produzindo bons ajustes.

Dentre os modelos que podem ser utilizados para ajustar a estrutura serial, serão expostos nesse capítulo o processo autoregressivo discreto e contínuo de ordem 1 - AR(1), (seção 3.2.1), o processo misto autoregressivo e média-móvel discreto de ordens, respectivamente, p e q - ARMA(p,q), (seção 3.2.2) e uma parametrização que permite

aceleração ou atenuação do decaimento da função de autocorrelação (fac) do AR(1) (seção 3.2.3), o AR(1) exponencialmente amortecido - AR(1)EA.

A estrutura de correlação do AR(1) é simples e útil, porém a sua fac em geral tem um decaimento rápido nas primeiras correlações comparado a processos como o AR(2) e o ARMA(1,1). O modelo AR(1)EA permite atenuar ou acelerar o decaimento da fac. No caso da atenuação do decaimento em séries curtas, o AR(2), por exemplo, produz autocorrelações muito próximas do AR(1)EA, com discrepância às vezes menor do que 0,05.

A modelagem dinâmica também pode ser aplicada aos coeficientes do modelo. Nem sempre é correto ou suficiente assumir que os coeficientes são fixos (constantes no tempo) ou aleatórios da forma vista anteriormente (seção 2.5). Por exemplo, em economia, a restrição de que as relações econômicas sejam sempre estáveis no tempo pode ser muito forte, porque elas estão sujeitas a mudanças. Em ensaios clínicos, pode ser útil incorporar ao modelo informações de covariáveis que dependem do tempo. Através da modelagem dos coeficientes pode-se também incorporar a heterocedasticia temporal das observações; por exemplo, quando as respostas de um determinado grupo de indivíduos sujeitos a um certo tratamento são relativamente estáveis no início do ensaio clínico, tornando-se cada vez mais variáveis ao longo do experimento. Por isso é importante estudar modelos que permitem variações nos parâmetros através do tempo (seção 3.3).

Essas mudanças podem ocorrer de forma estática ou dinâmica. No caso estático, por exemplo, elas podem ocorrer em determinados pontos do tempo, conhecidos ou não. Outro exemplo são coeficientes cujos valores em cada período são retirados de uma distribuição com média e variância comum a todos os períodos, assumindo-se que as retiradas sejam independentes (modelo de efeitos aleatórios na seção 2.5). No caso dinâmico (seção 3.3), relaxa-se a suposição de independência, modelando os coeficientes como um processo estocástico; por exemplo, Rosenberg (1973), modela a variação dos parâmetros através de um processo AR(1).

3.2. Erros Correlacionados Serialmente

Considere que as observações para o indivíduo i sejam tomadas nos instantes t_{ij} , para $j = 1, \dots, T_i$. Seja $\delta_{i(j)} = t_{i,j} - t_{i,j-1}$, para $j = 2, \dots, T_i$. $\delta_{i(j)}$ representa a defasagem (intervalo de tempo) entre duas observações consecutivas, i. e., o espaçamento entre os dados.

Os dados são igualmente espaçados quando existe um intervalo amostral básico. Por exemplo, observações diárias, como número total de chamadas telefônicas, podem ser

representadas por um processo temporal (discreto) onde o dia é o intervalo amostral básico. Por outro lado, dados desigualmente espaçados (Jones, 1985; Chi e Reinsel, 1989; Jones e Boateng, 1991) podem ocorrer de duas maneiras distintas: (i) existência de observações perdidas no conjunto de dados igualmente espaçados; e (ii) não há um intervalo amostral básico. O primeiro caso pode ser tratado como dados igualmente espaçados com valores perdidos (falta de informação), ou como (ii) caso exista uma proporção muito grande de dados perdidos. O caso (ii) também pode ser tratado da mesma forma de (i), tomando-se um intervalo tão pequeno que seja o mínimo múltiplo comum dos intervalos, i. e., obtendo-se um intervalo amostral básico. No entanto, como isso pode produzir muita falta de informação, a estrutura dos erros deverá, nesse caso, ser baseada em um modelo temporal contínuo. Em outras palavras, em ambos os casos, em situações onde o intervalo amostral básico é tal que muitas das observações sejam consideradas perdidas, é razoável considerar um processo contínuo.

Cabe notar que apenas em alguns casos um processo ARMA contínuo amostrado em intervalos de tempo igualmente espaçados resulta num processo ARMA discreto da mesma ordem (Pandit e Wu, 1983, pp. 220 e Jones, 1993, pp. 53). Um desses casos é o processo AR(1): um AR(1) contínuo amostrado em intervalos de tempo igualmente espaçados produz um processo discreto AR(1).

A escolha de um modelo adequado depende da estrutura dos erros. Basicamente, o modelo deve ser capaz de gerar a estrutura esperada dos erros. Como a escolha normalmente é realizada para ajustar a matriz de covariância, é fundamental conhecer a estrutura de covariância dos vários modelos propostos na literatura. É importante que o modelo seja parcimonioso, principalmente quando o número de observações for pequeno. Dentre as estruturas seriais, serão expostos nessa seção o processo AR(1) (seção 3.2.1), o ARMA(p,q) (seção 3.2.2) e o AR(1) Exponencialmente Amortecido - AR(1)EA, que é uma parametrização onde se pode acelerar ou atenuar o decaimento da função de autocorrelação (fac) do AR(1) (seção 3.2.3).

3.2.1. Estrutura AR(1)

Em primeiro lugar, considere que as observações sejam igualmente espaçadas com intervalos unitários ($\delta_{i(j)}=1, \forall j$). O processo AR(1) discreto para descrever os erros é:

$$\varepsilon_{i,t} = \phi \varepsilon_{i,t-1} + a_{i,t}, \quad (3.2.1.1)$$

onde $a_{i,t}$ são independentes e identicamente distribuídos com média zero e variância σ_a^2 , e ϕ é o parâmetro de autoregressão. O efeito aleatório $a_{i,t}$ é chamado de inovação ou choque aleatório (Box e Jenkins, 1976, pp. 53). O processo é estacionário quando $|\phi| < 1$.

A função de covariância é $cov(j) = E(\varepsilon_{i,t} \varepsilon_{i,t-j})$; se $j = 0$, então $cov(0) = E(\varepsilon_{i,t}^2)$ é a variância do processo. Multiplicando-se (3.2.1.1) por $\varepsilon_{i,t-j}$, $j > 0$, e tomando-se as esperanças, obtemos a equação de Yule-Walker para o AR(1):

$$E(\varepsilon_{i,t} \varepsilon_{i,t-j}) = \phi E(\varepsilon_{i,t-1} \varepsilon_{i,t-j}),$$

pois $E(a_{i,t} \varepsilon_{i,t-j}) = 0 \quad \forall j > 0$. A função de covariância decai exponencialmente em função da defasagem, i.e., $cov(j) = \phi^j cov(0)$. Logo, a função de autocorrelação (fac) é dada por

$$corr(j) = \frac{cov(j)}{cov(0)} = \phi^j, \quad (3.2.1.2.a)$$

e a variância do processo é dada por $\sigma_\varepsilon^2 = cov(0) = \frac{\sigma_a^2}{1-\phi^2}$.

Observa-se que, em dados longitudinais, não se espera que haja correlação negativa nas observações individuais. Portanto, em geral, considera-se que $0 < \phi < 1$. A seguir, ilustramos graficamente (Fig. 3.2.1.1) algumas funções de autocorrelação do AR(1) para diferentes ϕ :

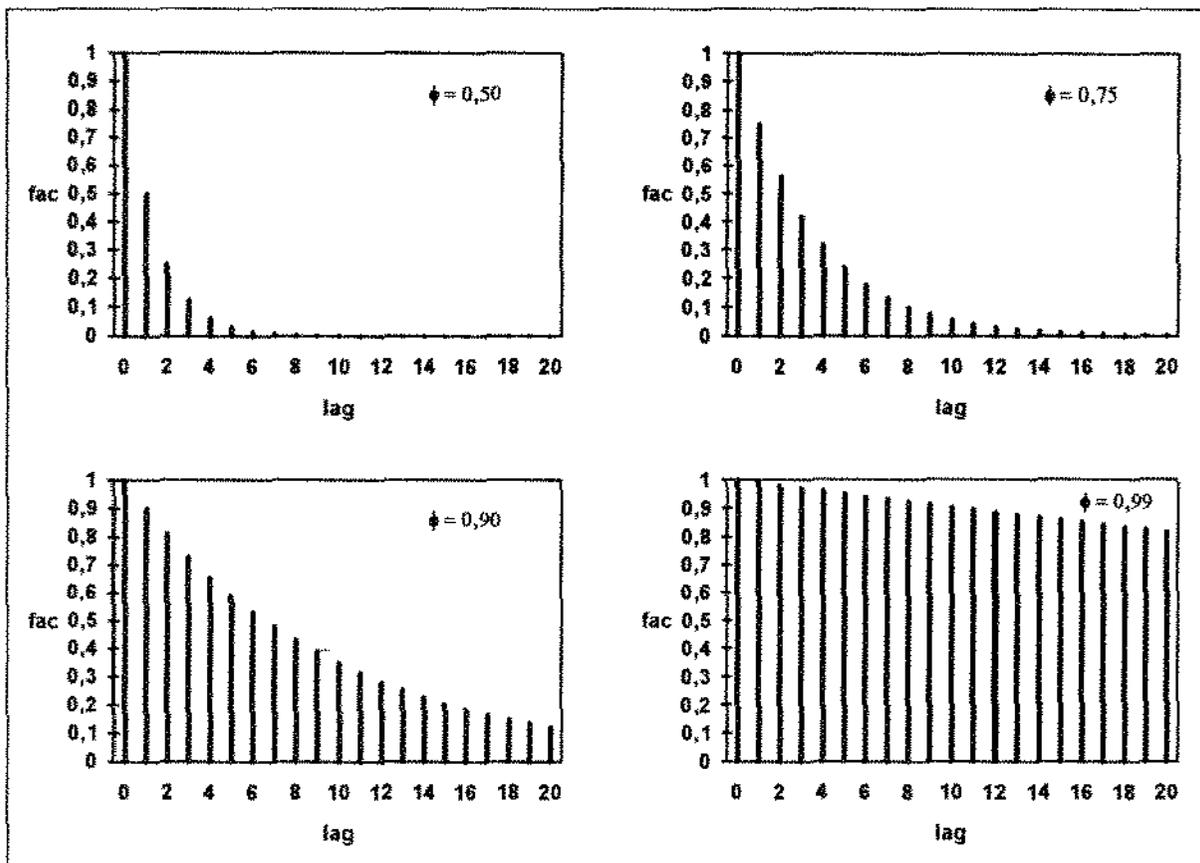


Fig. 3.2.1.1 - exemplos de funções de autocorrelação do AR(1)

No modelo de efeitos mistos (2.5), quando assume-se a estrutura AR(1) nos erros, a matriz Ω , terá a forma autoregressiva

$$\sigma_{\varepsilon}^2 \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T_1-1} \\ \phi & 1 & \phi & \dots & \phi^{T_1-2} \\ \phi^2 & \phi & 1 & \dots & \phi^{T_1-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{T_1-1} & \phi^{T_1-2} & \phi^{T_1-3} & \dots & 1 \end{bmatrix} \quad (3.2.1.2)$$

onde ϕ é determinado pela correlação no lag 1, i.e., $\phi = \gamma = \text{corr}(1)$.

Considerando-se agora um processo AR(1) contínuo (Apêndice D):

$$\varepsilon_i(t_j) = \phi \varepsilon_i(t_{j-1}) + a_i(t_j) \quad , \quad (3.2.1.3)$$

a matriz Ω_i dependerá dos intervalos entre os lags

$$\sigma_e^2 \begin{bmatrix} 1 & \exp(-\alpha(t_{1,2} - t_{1,1})) & \cdots & \exp(-\alpha(t_{1,T_i} - t_{1,1})) \\ \exp(-\alpha(t_{1,2} - t_{1,1})) & 1 & \cdots & \exp(-\alpha(t_{1,T_i-1} - t_{1,2})) \\ \vdots & \vdots & \ddots & \vdots \\ \exp(-\alpha(t_{1,T_i} - t_{1,1})) & \exp(-\alpha(t_{1,T_i-1} - t_{1,2})) & \cdots & 1 \end{bmatrix}, \quad (3.2.1.4)$$

onde α é um parâmetro não linear (descrito no apêndice D). O parâmetro de autoregressão no caso contínuo dependerá de α e do intervalo do tempo:

$$\phi(\delta t) = \exp(-\alpha \delta t),$$

onde δt é o intervalo de tempo.

Nos algoritmos de estimação que serão descritos no capítulo 5, é importante expressar a matriz Ω_i em uma forma fatorada, por exemplo, através da decomposição de Cholesky. Chi e Reinsel (1989) considerando (3.2.1.3) expressaram Ω_i na forma fatorada

$$\sigma_e^2 (R_i')^{-1} V_i (R_i)^{-1} \quad (3.2.1.5)$$

onde $V_i = \text{diag}(v_{ij})$, $j = 1, \dots, T_i$; com $v_{ij} = \frac{1}{1-\phi^2}$, se $j = 1$ e $v_{ij} = \frac{1-\phi^{2\delta_{ij}}}{1-\phi^2}$, se $j > 1$;

e R_i' é matriz triangular inferior (chamada por Chi e Reinsel, 1989, de matriz inovadora de Ω_i), tendo a seguinte forma:

$$R_i' = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\phi^{\delta_{u(2)}} & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\phi^{\delta_{u(3)}} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\phi^{\delta_{k(T_i)}} & 1 \end{bmatrix}$$

A estrutura AR(1) pode ser empregada nos modelos do capítulo 2. Por exemplo, no modelo 2.1.1 Potthoff e Roy (1964) usaram a estrutura AR(1) discreta em um conjunto de dados balanceado. No modelo misto univariado 2.4.1, Pantula e Pollock (1985) consideraram

$$\sum_{T_1, T_2} \Sigma_j = \sigma_\pi^2 1_{T_1} 1_{T_2}' + \sigma_e^2 \Omega_i, \quad (3.2.1.6)$$

onde Ω_i tem a forma 3.2.2. Nesse caso, a função de autocorrelação (fac), tem a forma:

$$\text{corr}(j) = \rho + (1-\rho) \gamma^j \quad (3.2.1.7)$$

onde $\rho = \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma_{\epsilon}^2}$ é o coeficiente de correlação intraclasse visto em seção 2.4, as variâncias σ_{π}^2 e σ_{ϵ}^2 são, respectivamente, a componente de variação entre indivíduos e a componente de variação dentro dos indivíduos. A estrutura (3.2.1.6) será chamada de AR(1)RE (AR(1) with Random Effects: estrutura do modelo de efeitos aleatórios com erros AR(1) e com $Z_i = 1_{T_i}$).

Na função (3.2.1.7), os parâmetros ρ e γ podem ser determinados pelas duas primeiras autocorrelações:

$$\rho = \frac{\text{corr}(2) - [\text{corr}(1)]^2}{1 - 2\text{corr}(1) + \text{corr}(2)} \text{ e}$$

$$\gamma = \frac{\text{corr}(1) - \text{corr}(2)}{1 - \text{corr}(1)}$$

Graficamente, a fac (3.2.1.7) pode ser ilustrada com os seguintes exemplos:

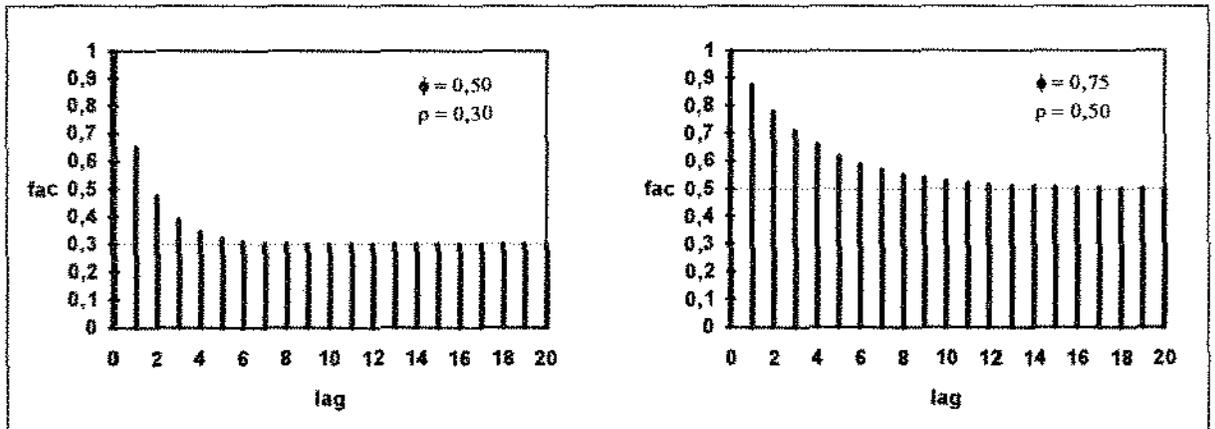


Fig. 3.2.1.2 - exemplos de funções de autocorrelação do AR(1)RE

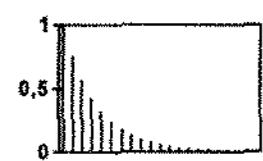
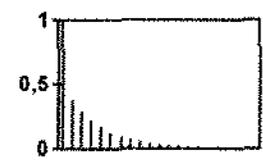
Diggle (1988) adicionou um efeito aleatório z_{it} à equação (2.4.1). Este efeito pode representar, por exemplo, uma variação dentro das observações individuais. Com $z_{it} \sim N(0, \sigma_z^2)$,

$$\Sigma_{T_i \times T_i} = \sigma_{\pi}^2 1_{T_i} 1_{T_i} + \sigma_z^2 1_{T_i} + \sigma_{\epsilon}^2 \Omega_i$$

onde Ω , tem a forma (3.2.4) ou (3.2.5). Chi e Reinsel (1989) e Jones e Boateng (1991) utilizaram o AR(1) contínuo para modelar os erros no modelo (2.5.1).

A tabela abaixo ilustra alguns exemplos que utilizam o AR(1) como parte da estrutura de correlação:

Tabela 3.2.1 - Exemplos de estruturas de covariância

Nome	Exemplo	Forma da fac
AR(1)	$\sigma_e^2 \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix}$	
AR(1) com erros observacionais	$\sigma_e^2 \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix} + \begin{bmatrix} \sigma_0^2 & 0 & 0 & 0 \\ 0 & \sigma_0^2 & 0 & 0 \\ 0 & 0 & \sigma_0^2 & 0 \\ 0 & 0 & 0 & \sigma_0^2 \end{bmatrix}$	
modelo de efeitos aleatórios com $Z_i = 1_{T_i}$ e erros AR(1) - (AR(1)RE)	$\sigma_e^2 \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix} + \begin{bmatrix} \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 \end{bmatrix}$	

3.2.2. Estrutura ARMA(p,q)

O modelo ARMA(p,q) discreto para os erros individuais, com p e q fixados ($p + q < T_t$), pode ser descrito como:

$$\varepsilon_{i,t} - \phi_1 \varepsilon_{i,t-1} - \phi_2 \varepsilon_{i,t-2} - \dots - \phi_p \varepsilon_{i,t-p} = a_{i,t} - \theta_1 a_{i,t-1} - \theta_2 a_{i,t-2} - \dots - \theta_q a_{i,t-p}, \quad (3.2.2.1)$$

onde os coeficientes ϕ_1 a ϕ_p e θ_1 a θ_q representam, respectivamente, as componentes autoregressivas (AR(p)) e média-móveis (MA(q)). O termo $a_{i,t}$ é o ruído branco com média 0, variância 1 e $E(a_{i,t} a_{i,t'}) = 0$ para $t \neq t'$.

No apêndice C apresentaremos o modelo ARMA discreto e noções sobre estacionariedade e invertibilidade do processo. Essas duas suposições (estacionário e invertível) gera um sistema de restrições ao espaço dos parâmetros (ϕ e θ). Por exemplo, no caso do ARMA(1,1), a escolha dos coeficientes (ϕ e θ) deve seguir às restrições: $|\phi| < 1$ e $|\theta| < 1$; além dessas restrições, em estudos longitudinais, espera-se que a correlação seja positiva como no caso do AR(1). Assim, as restrições serão dadas por $0 < \phi < 1$ e $0 < \theta < 1$. No caso do AR(2) as restrições serão: $\phi_1 + \phi_2 < 1$; $\phi_2 - \phi_1 < 1$ e $|\phi_2| < 1$; e para se considerar apenas as correlações positivas deve-se fazer $0 < \phi_1 < 1$.

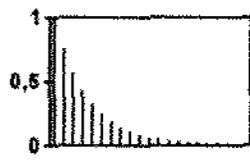
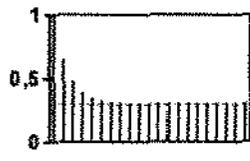
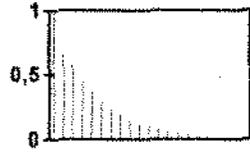
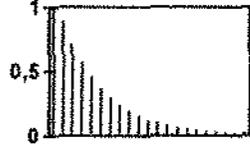
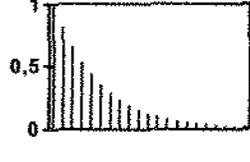
A matriz de covariância $\Omega_{i,t}$, conhecida com estrutura geral autoregressiva ou *banded*, é dada por

$$\sigma_e^2 \begin{bmatrix} 1 & \gamma_1 & \gamma_2 & \dots & \gamma_{T_t-1} \\ \gamma_1 & 1 & \gamma_1 & \dots & \gamma_{T_t-2} \\ \gamma_2 & \gamma_1 & 1 & \dots & \gamma_{T_t-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{T_t-1} & \gamma_{T_t-2} & \gamma_{T_t-3} & \dots & 1 \end{bmatrix} \quad (3.2.2.2)$$

onde $\gamma_1, \gamma_2, \dots, \gamma_{T_t-1}$ são funções de $\{\phi_1, \dots, \phi_p\}$ e $\{\theta_1, \dots, \theta_q\}$. O procedimento para calcular essas funções está descrito no Apêndice B.

Em geral, é conveniente utilizar modelos simples como o AR(1), o AR(2) ou ARMA(1,1), pois estes requerem poucos parâmetros a estimar. A simplicidade se torna obrigatória quando os dados individuais são séries curtas (lembrando-se que $p + q < T_t$). A Tabela 3.3.2 apresenta algumas estruturas de correlação:

Tabela 3.2.2 - Algumas estruturas de correlação

Modelo:	Estrutura de Correlação	Forma da fac
AR(1)	$\text{corr}(0) = 1$ $\text{corr}(h) = \phi^h, h \geq 1$	
AR(1)RE	$\text{corr}(h) = \rho + (1-\rho)\phi^h, h \geq 0$	
AR(2)	$\text{corr}(0) = 1$ $\text{corr}(1) = \frac{\phi_1}{1-\phi_2}$, e para $h > 1$: $\text{corr}(h) = \phi_1 \text{corr}(h-1) + \phi_2 \text{corr}(h-2)$	
ARMA(1,1)	$\text{corr}(0) = 1$ $\text{corr}(1) = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta}$ $\text{corr}(h) = \phi^{h-1} \text{corr}(1), h > 1$	
ARMA(2,1)	$\text{corr}(0) = 1$ $\text{corr}(1) = \frac{(1-\phi_1\theta)(\phi_1-\theta)+\theta\phi_2^2}{(1-\phi_2)(1+\theta^2)-2\phi_1\theta}$ $\text{corr}(h) = \phi_1 \text{corr}(h-1) + \phi_2 \text{corr}(h-2), h \geq 2$	

3.2.3. Estrutura AR(1) Exponencialmente Amortecida

Apesar do AR(1) possuir uma estrutura de correlação simples e útil, sua função de autocorrelação tem um decaimento relativamente rápido nas primeiras correlações (Fig. 3.2.1.1, exceto para o caso de ϕ muito próximo de 1) comparado a certos processos como o AR(2) (Tabela 3.2.2) e ARMA(p,q), e relativamente lento comparado ao MA(q), por exemplo. É necessário, às vezes, utilizar estruturas intermediárias que têm decaimento mais lento (entre a simetria composta e o AR(1)) ou mais rápido (entre o AR(1) e o MA(1)).

Diggle (1988) e Muñoz *et alii.* (1992) utilizaram uma estrutura de correlação flexível e parcimoniosa para produzir uma fac com decaimento intermediário: o AR(1) Exponencialmente Amortecido, AR(1)EA, cuja forma da fac está ilustrada na Fig. 3.2.3.1.

A correlação entre duas observações separadas por h unidades de tempo é modelada como

$$\text{corr}(h) = \gamma^{h^\omega} \quad (3.2.3.1)$$

onde γ ($-1 \leq \gamma \leq 1$) é a correlação entre observações separadas por uma unidade de tempo, i.e., $\text{corr}(1) = \gamma$; e ω é o parâmetro de amortecimento (ou de "escala"). Esse parâmetro permite atenuação ($0 \leq \omega \leq 1$) ou aceleração ($\omega \geq 1$) do decaimento exponencial de uma fac que define um AR(1) como pode ser ilustrado abaixo:

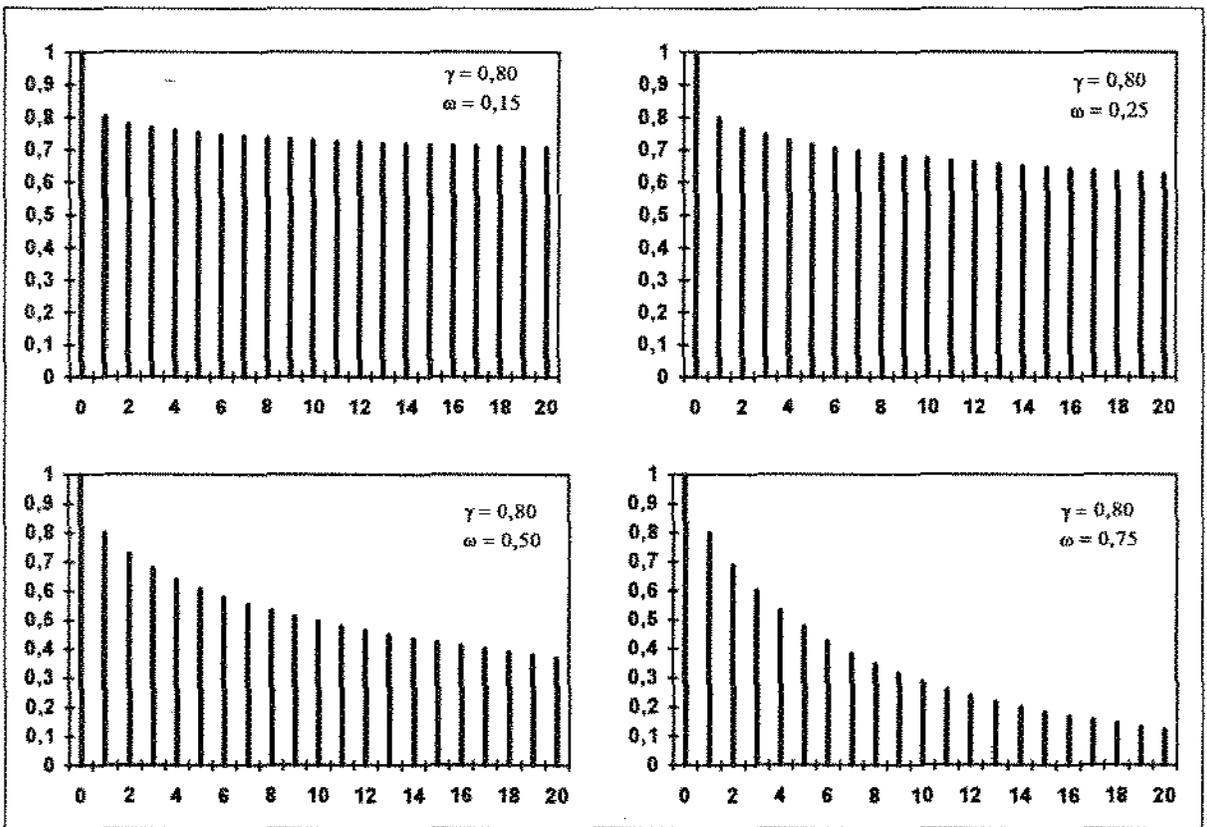


Fig. 3.2.3.1 - exemplos de AR(1)EA

Na maioria das aplicações, e.g. em bioestatística e epidemiologia (Muñoz *et alii.*, 1992), observa-se que a correlação γ é não-negativa ($0 \leq \gamma \leq 1$). Similarmente, valores negativos para o parâmetro de amortecimento ω também podem produzir funções de autocorrelação sem entretanto aplicação prática. Por isso, a situação usual é modelar os dados restringindo os parâmetros γ e ω como

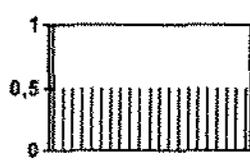
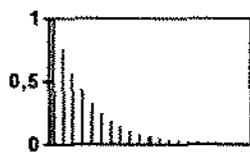
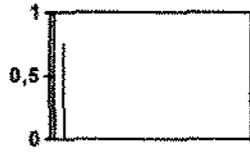
$$0 \leq \gamma \leq 1 \text{ e } \omega \geq 0.$$

(3.2.3.2)

Detalhes das propriedades teóricas do modelo com fac exponencialmente amortecida podem ser verificados considerando-se uma estrutura um tanto mais complexa desse espaço paramétrico (Muñoz *et alii.*, 1992). Nota-se, por exemplo, que o teste para verificar a hipótese $\gamma = 0$ pode ser realizado dentro da família paramétrica (3.2.3.2) se ω for restrito a valores inteiros e γ restrito apenas aos valores que produzem fac válidas. O teste $\omega = 0$ versus $\omega > 0$ pode ser feito comparando-se a verossimilhança otimizada do modelo (3.2.3.2) com o padrão de uniformidade. Detalhes sobre verossimilhança e testes de hipóteses serão apresentados nos capítulos 5 e 6.

Para γ não negativo, a estrutura de correlação pode produzir uma variedade de estruturas conhecidas, fixando-se o parâmetro de amortecimento ω . Por exemplo, para $\omega = 0$, $\omega = 1$ e $\omega \rightarrow +\infty$, as estruturas de correlação resultantes são, respectivamente, o padrão de uniformidade, o AR(1) e o MA(1). A forma dessas estruturas estão na Tabela 3.2.2.

Tabela 3.2.2 - Algumas estruturas resultantes do AR(1) amortecido

ω	fac	Modelo Resultante	forma da fac
$\omega = 0$	$\text{corr}(0) = 1$ $\text{corr}(h) = \gamma, h > 0$	Padrão de Uniformidade	
$\omega = 1$	$\text{corr}(h) = \gamma^h, h \geq 0$	AR(1)	
$\omega \rightarrow +\infty$	$\text{corr}(0) = 1$ $\text{corr}(1) = \gamma$ $\text{corr}(h) = 0, h > 1$	MA(1)	

Para a situação $0 \leq \omega \leq 1$, obtém-se uma família de estruturas de correlação com decaimento entre a simetria composta (padrão de uniformidade) e o AR(1). Essa situação que significa atenuação do decaimento exponencial da fac é mais comum, com exemplos de aplicação citados por Muñoz *et alii.* (1992). Para $\omega > 1$ obtém-se uma estrutura de correlação com decaimento mais rápido que o AR(1) (aceleração do decaimento exponencial da fac).

Existem outras estruturas que também possuem decaimento intermediário entre o AR(1) e o modelo de simetria composta (ou padrão de uniformidade). Por exemplo, o AR(2) e o modelo de efeitos aleatórios com erros AR(1) (AR(1)RE). A fac do AR(2) é dada pela Tabela 3.2.2 e do AR(1)RE, com $Z_i = 1_{T_i}$ (3.2.1.6), é dado pela equação (3.2.1.7).

Embora a estrutura de dependência do AR(2) e AR(1)RE não sejam casos particulares do AR(1) amortecido, essas estruturas podem ser aproximadas quando se tem séries curtas (Muñoz et. al. (1992)).

Uma forma de se obter um membro da família AR(1)RE que aproxima uma dada função de autocorrelação $\text{corr}(h)$ é selecionar γ e ω para determinar a distância minimax, isto é, minimizar $\max|\text{corr}(h) - \gamma^{h^\omega}|$ para $0 \leq h \leq T$, onde T é o tamanho da série (T pequeno):

Exemplo 3.2.3.1 aproximação do AR(1)EA para o AR(2) e AR(1)RE

Nesse exemplo ilustraremos como o AR(1)EA pode ser aproximado do AR(2) e do AR(1)RE para uma série curta. Tanto o AR(2) como o AR(1)RE são determinados pelas duas primeiras correlações ($\text{corr}(1)$ e $\text{corr}(2)$), i.e.,

para o AR(2),

$$\phi_1 = \frac{\text{corr}(1)[\text{corr}(2) - 1]}{[\text{corr}(1)]^2 - 1} \text{ e}$$

$$\phi_2 = \frac{[\text{corr}(1)]^2 - \text{corr}(2)}{[\text{corr}(1)]^2 - 1}$$

e para o AR(1)RE,

$$\rho = \frac{\text{corr}(2) - [\text{corr}(1)]^2}{1 - 2\text{corr}(1) + \text{corr}(2)} \text{ e}$$

$$\gamma = \frac{\text{corr}(1) - \text{corr}(2)}{1 - \text{corr}(1)}$$

Então, considere a família $\text{corr}(h)$ determinado pelas duas primeiras autocorrelações ($\text{corr}(1)$ e $\text{corr}(2)$), onde $\text{corr}(1) = \{0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8 \text{ e } 0,9\}$ e $\text{corr}(2) = [\text{corr}(1)]^k$, para $k = \{1,0; 1,25; 1,5; 1,75; 2,0; 2,25 \text{ e } 2,5\}$ e $T = 10$.

Muñoz *et alii.* (1992) determinaram a distância minimax quando a fac pré-especificada é a AR(2) e verificaram que em 91% dos casos para os quais o AR(2) está bem definido, a discrepância máxima foi menor do que 0,05. A maior discrepância encontrada foi de 0,069 e resultou de uma estrutura AR(2) não usual, com $\text{corr}(1) = \text{corr}(2) = 0,5$. Similarmente, os autores determinaram a distância minimax quando a estrutura de correlação pré-especificada é o AR(1)RE. Nesse caso, 96% dos casos para os quais o

AR(1)RE está bem definido tiveram discrepância máxima menor do que 0,05, a maior discrepância encontrada foi de 0,056 para o caso com $\text{corr}(1) = 0,6$ e $k = 1,75$.

□

Exemplo 3.2.3.2 aproximação do AR(2) e AR(1)RE para o AR(1)EA

Nesse exemplo procedemos de forma inversa do exemplo anterior, i.e., especifica-se o AR(1)EA (fixando-se os parâmetros γ e ω) e determina-se a melhor aproximação do AR(2) e AR(1)RE pelo critério do minimax para $T = 8$. Os resultados estão ilustrados na Figura 3.2.3.2.

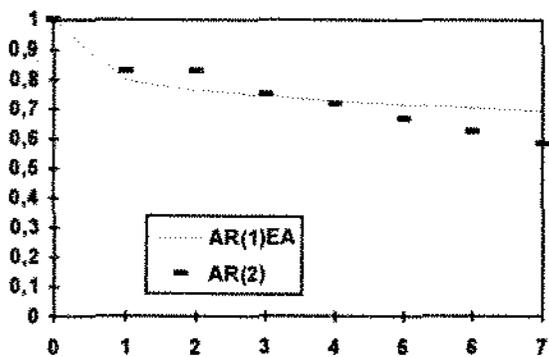
No primeiro caso, onde $\gamma = 0,80$ e $\omega = 0,25$, a máxima discrepância encontrada foi 0,11 para o AR(2) e 0,02 para o AR(1)RE. No segundo, com $\gamma = 0,80$ e $\omega = 0,5$, a discrepância encontrada foi 0,065 para o AR(2) e 0,025 para o AR(1)RE. No último, fixando-se o AR(1)EA com $\gamma = 0,80$ e $\omega = 0,75$, a máxima discrepância foi de 0,023 para o AR(2) e 0,022 para o AR(1)RE. Obviamente a aproximação se tornará melhor à medida que T diminui. Por exemplo, no caso extremo com $T = 3$ todas as estruturas de correlação determinadas por $\text{corr}(1)$ e $\text{corr}(2)$ produzirão a mesma fac. Os resultados encontrados estão ilustrados graficamente na página seguinte.

Esse dois exemplos de aproximações (3.2.3.1 e 3.2.3.2) de funções de autocorrelação servem para ilustrar a questão do uso ou não de modelos mais complicados. Nos dois exemplos as aproximações foram para uma fac teórica. Na prática, ao se ajustar uma fac observada, o AR(1) também pode ser uma boa opção de aproximação produzindo bons ajustes.

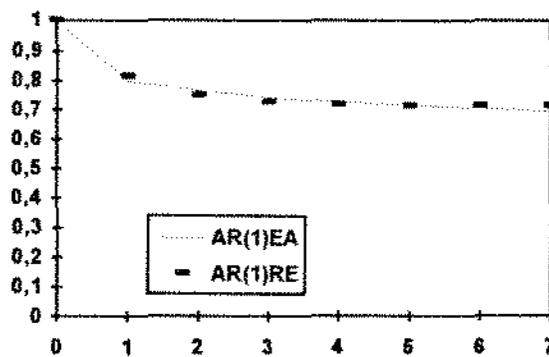
Como já existem programas computacionais que tratam de modelos mistos como o SAS (proc MIXED), permitindo a modelagem dos erros como um processo AR(1), o AR(1)RE é uma boa alternativa de ajuste da fac amortecida em séries curtas.

□

aproximação do AR(2) e do AR(1)RE para o AR(1)EA com $\gamma = 0,80$ e $\omega = 0,25$:

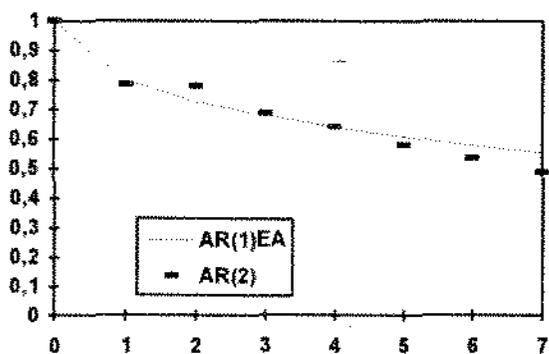


AR(2) com $\phi_1 = \phi_2 = 0,45355$

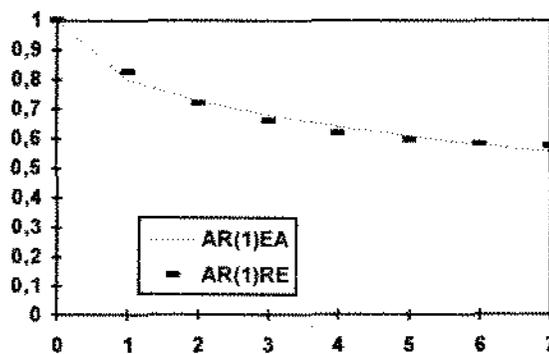


$\rho = 0,71479$ e $\gamma = 0,35135$

aproximação do AR(2) e do AR(1)RE para o AR(1)EA com $\gamma = 0,80$ e $\omega = 0,50$:

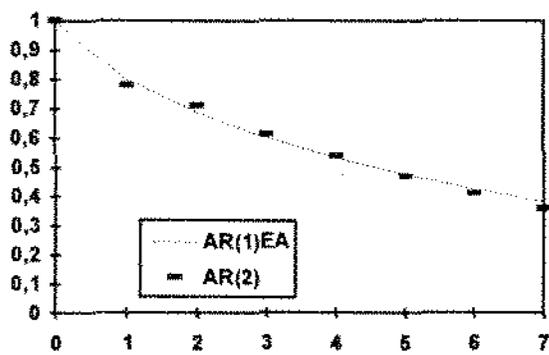


$\phi_1 = 0,45$ e $\phi_2 = 0,42675$

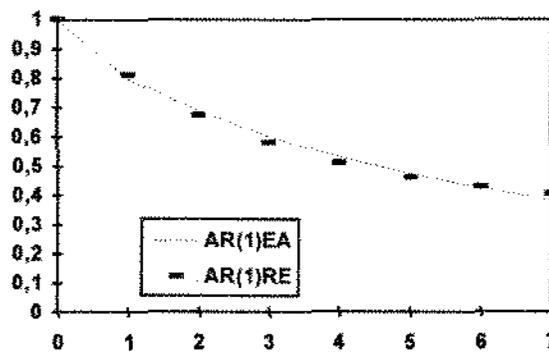


$\rho = 0,5625$ e $\gamma = 0,60$

aproximação do AR(2) e do AR(1)RE para o AR(1)EA com $\gamma = 0,80$ e $\omega = 0,75$:



$\phi_1 = 0,57763$ e $\phi_2 = 0,25945$



$\rho = 0,34364$ e $\gamma = 0,71053$

Fig. 3.2.3.2 - aproximações do AR(2) e AR(1)RE para o AR(1)EA

3.3. Coeficientes Correlacionados Serialmente

No modelo de efeitos mistos descrito na seção 2.5, a matriz de covariância individual é dada por duas matrizes arbitrárias Γ e Ω_i (2.5.5) que representam, respectivamente, a matriz de variância dos efeitos aleatórios individuais e a matriz de covariância dentro do indivíduo. Na seção anterior discutiu-se sobre algumas parametrizações possíveis para Ω_i através da modelagem dinâmica. Já a parametrização de Γ depende do modelo adotado. Por exemplo, no modelo componente de variâncias (como no exemplo 2.5.3), ou a matriz Γ assume a forma diagonal e cada efeito aleatório representa uma fonte independente de variação, ou a matriz assume a forma não estruturada e os efeitos aleatórios serão correlacionados.

Nos casos onde se considera a modelagem dinâmica de Γ os coeficientes serão denotados por b_{it} . A modelagem dinâmica dos coeficientes tem aplicação, entre outras áreas, em economia (Harvey, 1978) e bioestatística (Rochon, 1992; Hastie e Tibshirani, 1993). Como ilustrações, considere os seguintes exemplos:

Exemplo 3.3.1 estrutura AR(1) para Γ

Em economia, as relações econômicas estão sujeitas a mudanças no decorrer do tempo. Um modelo de parâmetro variando no tempo pode ser escrito como (Harvey, 1978)

$$y_{it} = Z_{it}b_{it} + \varepsilon_{it}, \quad (3.3.1)$$

com b_{it} vetor $k \times 1$ de coeficientes no tempo t modelados por um AR(1)

$$b_{it} = A_0 + A_1 b_{i,t-1} + a_{it}, \quad (3.3.2)$$

onde $i = 1, \dots, n$; $t = 1, \dots, T$; ε_{it} e a_{it} ruídos independentes com média 0; A_0 e A_1 respectivamente vetor $k \times 1$ e matriz $k \times k$. Quando $A_1 = 0$, o modelo resultante é o de coeficientes aleatórios visto na seção 2.5

Em (3.3.2), para que o processo seja estacionário, é necessário que as raízes de A_1 caiam dentro do círculo unitário. A média desse processo é dada por $(I - A_1)^{-1}A_0$. Harvey (1978) utiliza o termo "retorno à normalidade" para (3.3.2). Ele é utilizado quando os coeficientes do modelo têm valores próximos ao valor médio fixado, isto é, gerados por um processo estacionário. Em (3.3.2), fazendo-se $A_1 = \Phi$ e $A_0 = (I - \Phi)\bar{b}$ temos:

$$b_t = (I - \Phi)\bar{b} + \Phi b_{t-1} + \varepsilon_t \Rightarrow b_t - \bar{b} = \Phi(b_{t-1} - \bar{b}) + \varepsilon_t \Rightarrow$$

$$\delta_t = \Phi\delta_{t-1} + \varepsilon_t,$$

onde $\delta_t = b_t - \bar{b}$, $\varepsilon_t \sim N(0, \sigma^2 Q)$ e $E(\varepsilon_t \varepsilon_s') = 0$, $\forall t \neq s$; e Φ é matriz $k \times k$ de coeficientes. Em geral, Φ é diagonal cujos elementos têm valores absolutos menores do que 1 para assegurar estacionariedade. É mais simples do que assumir $k \times k$ parâmetros desconhecidos.

A estimação pode ser feita pelo método de duplo estágio ou por máxima verossimilhança, ambos utilizando o filtro de Kalman.

A equação 3.3.1 pode ser reescrita na forma

$$Y_t = x_t' \bar{b} + x_t'(b_t - \bar{b}) + u_t \Rightarrow Y_t = x_t' \bar{b} + x_t' \delta_t + u_t \Rightarrow$$

$$Y_t = x_t' \bar{b} + \zeta_t, \quad (3.3.4)$$

onde $\zeta_t = x_t' \delta_t + u_t$, $E(\zeta_t \zeta_s') = \sigma^2 V$, $E(u_t u_s') = 0$ e $E(u_t \delta_s') = 0 \forall t \neq s$ e $E(u_t u_t') = \sigma^2 H_t$, $t = 1, \dots, T$. Essa matriz não é necessariamente diagonal, considerando portanto a correlação contemporânea (entre os indivíduos).

A equação (3.3.4) está na forma de um modelo linear com coeficientes fixos e erros ζ_t heterocedásticos e serialmente correlacionados. Como será detalhado no capítulo 5, para se evitar a inversão da matriz V no processo de estimação, o filtro de Kalman que está descrito no apêndice B é uma forma atrativa de se calcular a verossimilhança. O filtro de Kalman envolve um número menor de multiplicações e possibilita facilmente previsões de valores futuros de Y_t .

Tomando-se $A_0 = 0$ e $A_1 = I$ em 3.3.2 teremos

$$b_t = b_{t-1} + \varepsilon_t$$

onde $\varepsilon_t \sim N(0, \sigma^2 Q)$. Este modelo difere do anterior por não ser estacionário (passeio aleatório). Essa estrutura permite incorporar ao modelo mudanças às relações estruturais. A matriz Q é $k \times k$ e determina a extensão na qual os parâmetros podem variar. Se $Q = 0$ o modelo é um caso de regressão, pois os parâmetros serão iguais em todos os períodos.

□

Exemplo 3.3.2 estrutura ARMA(p,q) para Γ

Em ensaios clínicos, pode ser útil incorporar ao modelo informações de covariáveis dependentes do tempo, ou incorporar a heterocedasticia temporal das observações. Rochon (1992) considerou o modelo

$$y_{it} = x_{it}'\beta + z_{it}b_{it}, \quad (3.3.3)$$

onde x_{it} é um vetor $a \times 1$ de variáveis explanatórias não-estocásticas para o i -ésimo indivíduo no tempo t ; β é o vetor de parâmetros fixos; Z_{it} é um vetor de covariáveis que dependem do tempo e b_{it} segue um processo ARMA(p,q) como descrito em seção 3.2.2, i.e.,

$$b_{i,t} - \phi_1 b_{i,t-1} - \phi_2 b_{i,t-2} - \dots - \phi_p b_{i,t-p} = a_{i,t} - \theta_1 a_{i,t-1} - \theta_2 a_{i,t-2} - \dots - \theta_q a_{i,t-p}, \quad (3.3.4)$$

com $p + q < T_i$. Os termos $a_{i,t}$ são os ruídos aleatórios e são assumidos independentes e identicamente distribuídos com média 0 e variância 1. Assim, $E(b_{it}) = 0$ e $\text{var}(b_{it}) = \sigma_b^2 P$, onde $\sigma_b^2 = \text{var}(b_{i,t})$ e P é a matriz autoregressiva geral dada por (3.2.2.2).

A matriz de covariância individual pode ser escrita como

$$\Sigma_i = \begin{bmatrix} \sigma_1^2 & \gamma_1 \sigma_1 \sigma_2 & \dots & \gamma_{T_i-1} \sigma_1 \sigma_{T_i} \\ \gamma_1 \sigma_1 \sigma_2 & \sigma_2^2 & \dots & \gamma_{T_i-2} \sigma_2 \sigma_{T_i} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{T_i-1} \sigma_1 \sigma_{T_i} & \gamma_{T_i-2} \sigma_2 \sigma_{T_i} & \dots & \sigma_{T_i}^2 \end{bmatrix}, \quad (3.3.5)$$

onde $\sigma_t = (\sigma_b^2)^{1/2} z_t$ é o desvio padrão da observação no tempo t , e γ_t é a autocorrelação no tempo t . Então, z_t é um valor proporcional ao desvio padrão de uma observação no tempo t .

□

Exemplo 3.3.3 nível de ácido láctico: heterogeneidade temporal

Considere os dados da seção 1.2. Pelos gráficos dos perfis individuais dos grupos (Fig. 1.2.1 e Fig. 1.2.2), nota-se que as respostas de ácido são mais estáveis no início do experimento, tornando-se cada vez mais variáveis ao longo do experimento. Uma forma de se contornar isso é transformar os dados, por exemplo, tomando-se o logaritmo dos níveis de ácido láctico. Como será analisado na seção 7.2, a transformação deve produzir um

padrão mais satisfatório nos resíduos, atenuando o efeito da heterocedasticia temporal. Outra alternativa é incorporar a heterocedasticia ao modelo, introduzindo-o como uma covariável dependente do tempo.

Como uma aplicação do exemplo anterior, um modelo para os dados de ácido do grupo 1 pode ser escrito como

$$y_{i(t)} = \beta_{0i} + \beta_{1i}t + \dots + \beta_{6i}t^6 + z_i b_{it},$$

onde b_{it} segue um processo AR(1) na forma

$$b_{i,t} = \phi b_{i,t-1} + \alpha_{i,t}.$$

A matriz de covariância individual será dada por

$$\Sigma_i = \begin{bmatrix} \sigma_1^2 & \phi \sigma_1 \sigma_2 & \dots & \phi^{T_i-1} \sigma_1 \sigma_{T_i} \\ \phi \sigma_1 \sigma_2 & \sigma_2^2 & \dots & \phi^{T_i-2} \sigma_2 \sigma_{T_i} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{T_i-1} \sigma_1 \sigma_{T_i} & \phi^{T_i-2} \sigma_2 \sigma_{T_i} & \dots & \sigma_{T_i}^2 \end{bmatrix}, \quad (3.3.6)$$

onde $\sigma_t = (\sigma_b^2)^{\frac{1}{2}} z_t$ é o desvio padrão da observação no tempo t , e ϕ é o parâmetro de autoregressão dos efeitos aleatórios.

□

Em termos mais gerais, a correlação serial dos coeficientes e dos erros segundo um processo ARMA(p,q) pode ser incorporada ao modelo de efeitos mistos (2.5) escrito na forma

$$y_{it} = x'_{it} \beta + z_t b_{it} + \varepsilon_{it}, \quad (3.3.7)$$

tomando-se matrizes Γ e Ω_i com a forma da estrutura autoregressiva geral dada por (3.2.2.2).

Assim, como pode ser visto na Tabela 3.3.1, a representação da matriz de covariância individual do modelo de efeitos mistos, $\Sigma_i = Z_i \Gamma Z_i' + \Omega_i$, permite uma vasta opção de modelagem da estrutura de covariância combinando-se várias formas para Γ e Ω_i . Algumas dessas combinações serão estudadas na análise II do capítulo 7 utilizando-se os dados da seção 1.2 (sem transformar os dados).

Tabela 3.3.1 - Exemplos de estruturas para $\Sigma_t = Z_t \Gamma Z_t' + \Omega_t$

Γ	Ω_t	# parâmetros para Σ_t
0	geral	$T(T-1)/2$
geral	geral	$m(m-1)/2 + T(T-1)/2$
$\sigma^2 I$	$\sigma^2 I$	1
$\sigma^2 I$	AR(1)	3
geral	AR(1)	$m(m-1)/2 + 2$
AR(1) _b	AR(1) _v	4
AR(1)	$\sigma^2 I$	3
AR(1)	geral	$2 + T(T-1)/2$
geral	0	$m(m-1)/2$

4

Escolha do Modelo Adequado

4.1. Introdução

Nos capítulos anteriores foram apresentados alguns modelos para a análise de medidas seriais. Dentre uma certa variedade de opções, é preciso decidir por algum modelo e aplicá-lo em um contexto particular.

Como foi introduzido no capítulo 1, algumas características dos dados longitudinais são:

- i) a resposta para cada indivíduo (unidade experimental) é uma sequência no tempo de medidas em uma escala contínua;
- ii) as respostas entre os indivíduos são estatisticamente independentes;
- iii) a resposta média depende do tratamento e do instante de tempo da observação;
- iv) um dos objetivos é fazer inferências sobre os efeitos dos tratamentos experimentais no perfil médio de resposta;

No contexto acima, excluem-se desse trabalho modelos como curvas de sobrevivência, regressões aparentemente não-relacionadas (SUR) considerada por Zellner (1962) que permite modelar a correlação contemporânea (entre indivíduos) e MANOVA (análise de variância multivariada). A comparação entre a análise *cross-section* dos dados e a longitudinal também não será abordada; como o vetor médio do modelo de efeitos mistos (X, β) independe funcionalmente de Σ_i (2.5.5) e esta não depende dos valores de β , seria possível levantar a questão discutida por Louis (1988b): o que é essencialmente *cross-section* (corte transversal) e o que é longitudinal. Basicamente, as questões referentes à análise *cross-section* são aquelas associadas às características da população, fornecendo informações sobre a estrutura marginal das médias; e a longitudinal com as mudanças individuais, fornecendo informações sobre as correlações entre respostas medidas em tempos distintos. Fora os casos discutidos por Louis (1988b) onde há discrepância entre as duas abordagens, os parâmetros que determinam a média podem ser estimados via análise *cross-section* apropriada. Entretanto, os parâmetros de covariância somente podem ser estimados longitudinalmente. Se os parâmetros de covariância não forem de interesse, deve-

se mostrar o porquê da necessidade de um estudo longitudinal; entre as justificativas estão, por exemplo, o aumento de precisão das covariáveis e das respostas. Nesse trabalho os parâmetros de covariância e a modelagem das respostas individuais são fatores de interesse, justificando-se a análise longitudinal dos dados.

Embora não seja tão simples, às vezes, decidir pelo modelo mais adequado, deve-se considerar fatores como a parcimoniosidade e a bondade de ajuste do modelo escolhido:

- a) a especificação do perfil médio de resposta deve ser suficientemente flexível para considerar o padrão temporal dentro de cada grupo e as diferenças desses padrões entre os grupos;
- b) a especificação da estrutura de covariância dentro de cada seqüência temporal deverá ser flexível e econômica (poucos parâmetros). Parametrização excessiva poderá resultar em estimação ineficiente (Diggle, 1988) e potencialmente as estimativas dos erros padrões dos perfis médios de resposta estimados serão ruins. Por outro lado, uma especificação muito restritiva poderá invalidar inferências sobre a resposta média quando a suposição sobre a matriz de covariância não for correta. A Tabela 4.1.1 ilustra a diversidade de estruturas de covariâncias com os respectivos número de parâmetros desconhecidos.

Tabela 4.1.1 - Algumas estruturas de covariâncias

Nome da Estrutura	# parâmetros	descrição
observações independentes	1	$\Sigma_i = \sigma^2 I$
padrão de uniformidade	2	$\Sigma_i = \sigma_\pi^2 11' + \sigma^2 I$
independência condicional	$m(m+1)/2 + 1$	$\Sigma_i = Z_i \Gamma Z_i' + \sigma^2 I$
AR(1)	2	$\Sigma_i = \sigma^2 W_i$, onde $w_{jk} = \gamma^{ j-k }$
Banded	T_i	$\Sigma_i = \sigma^2 W_i$, onde $w_{jk} = \gamma_q$, $q = j - k + 1$
efeitos aleatórios com erros AR(1)	$m(m+1)/2 + 2$	$\Sigma_i = Z_i \Gamma Z_i' + \sigma^2 W_i$, onde $w_{jk} = \gamma^{ j-k }$
não estruturada	$T_i(T_i+1)/2$	Σ_i , onde $\sigma_{jk} = \sigma_{kj}$ e $\sigma_{jj} = \sigma_j^2$

Como discutiu-se no capítulo 2, o modelo de efeitos mistos abrange as estruturas apresentadas na Tabela 4.1.1 como casos especiais. Embora geral, o modelo de efeitos mistos não incorpora, por exemplo, efeitos aleatórios para os parâmetros de covariância (e.g., autocorrelação aleatória). Muitas formas híbridas são possíveis, mas elas podem não ser práticas por causa de dificuldades de interpretação, computacional e da compatibilidade dessas estruturas híbridas com o próprio conjunto de dados.

- c) o método de análise deverá acomodar padrões mais ou menos arbitrários nas sequências de tempo irregularmente espaçados dentro das unidades experimentais. Isso pode ocorrer acidentalmente através de valores perdidos, ou por planejamento; por exemplo, quando se espera que a partir de um determinado período a resposta média seja mais estável, as observações podem ser tomadas menos frequentemente. É comum também ocorrer o término prematuro da sequência de tempo para algum indivíduos. Mesmo que a razão para o término não seja relacionada com a quantidade sendo medida, essas sequências abreviadas contêm informações e deverão ser analisadas.
- d) o método de análise deverá contar com alguma forma de diagnóstico que permita, no mínimo, fornecer alguma idéia de bondade de ajuste para a estrutura de covariância que foi adotada.

A decisão em se considerar os efeitos como fixos ou aleatórios (seção 4.3) é um passo importante na modelagem, pois, em alguns casos, é possível haver muita diferença entre as estimativas dos parâmetros desses modelos. Basicamente, os efeitos serão considerados fixos quando for possível atribuir a eles um conjunto finito de níveis de um fator em interesse. Por exemplo, o fator sexo possui dois níveis: "masculino" e "feminino"; assim, os efeitos para o fator sexo são considerados efeitos fixos. Já para os efeitos aleatórios, atribui-se um conjunto infinito (ou um conjunto finito, porém muito grande) de níveis de um fator de onde se toma apenas uma amostra aleatória. Por exemplo, quatro cestas de pães são tomadas dentre seis fornadas em 3 temperaturas diferentes. O efeito devido à temperatura pode ser considerado como fixo, pois o interesse está em um conjunto particular de temperaturas; já o efeito devido à fornada pode ser considerado aleatório porque tomou-se uma amostra aleatória de fornadas dentre uma população hipoteticamente infinita de fornadas. Na seção 4.3 apresentaremos alguns exemplos para ilustrar essa questão.

Na seção 4.2 discute-se que a correlação serial nos erros pode ser parcialmente confundida com os efeitos aleatórios individuais. Em dados reais, muitas vezes, é difícil determinar se há correlação serial ou efeitos aleatórios ou ambos. Jones (1990) sugere o uso de alguns critérios de seleção de modelo descritos na seção 4.4, comentando-se sobre o critério de Informação de Akaike (AIC). Os testes de hipóteses e o diagnóstico do modelo, que são assuntos relacionados com a escolha do modelo, serão detalhados no capítulo 6.

4.2. A estrutura de covariância

Como já se discutiu anteriormente, dados incompletos e não balanceados são comuns em estudos longitudinais por vários motivos: retirada de um indivíduo antes do término previsto do experimento, observações perdidas devido a falhas técnicas ou humanas, e dados desigualmente espaçados no tempo. Outra característica dos dados longitudinais que pode complicar a análise surge quando as variáveis independentes de interesse para um determinado indivíduo incluem covariáveis que mudam ao longo do tempo. Uma análise apropriada desses dados deve levar em consideração as correlações entre as medidas tomadas em um mesmo indivíduo ao longo do tempo. Técnicas analíticas clássicas desenvolvidas para dados balanceados e completos centralizam-se ou no modelo univariado, onde assume-se que as observações de um mesmo indivíduo têm uma variância constante e o valor da correlação em comum, ou no modelo multivariado, que permite a adoção de uma estrutura de variância arbitrária.

No desenvolvimento da análise longitudinal de dados não balanceados, tomaram-se modelos lineares gerais para as respostas esperadas, com a capacidade de modelar a estrutura da matriz de covariância como o modelo de efeitos mistos. A escolha por um modelo para a estrutura de covariância depende da compreensão do processo gerador dos dados e da forma com que os dados foram coletados no tempo. É importante também considerar se os indivíduos são medidos em um mesmo conjunto de tempo (e.g., dados mensais, semanais, diários) ou em tempos que variam irregularmente de um indivíduo para outro.

A Tabela 4.1.1 apresenta algumas das estruturas de covariância que podem resultar naturalmente. Considere Θ o conjunto de parâmetros da matriz de covariância Σ . A forma não estruturada, ou forma completamente parametrizada para Σ , com $T(T+1)/2$ parâmetros para Θ , permite uma generalização de muitas formas de análise multivariada convencionais para dados incompletos. O exemplo mais simples, talvez, é o caso onde os dados consistem de uma amostra simples incompleta retirada de uma distribuição normal T -variada. Com Σ não estruturada, os delineamentos entre e dentro dos indivíduos são especificados através de uma escolha apropriada para a matriz B_i em (2.3.1).

Quando T for grande e os dados forem altamente não balanceados ou incompletos, o uso da forma não estruturada para Σ poderá levar à perda de eficiência na estimação dos parâmetros de regressão β devido ao grande número de parâmetros Θ a serem estimados. Portanto, deve-se considerar modelos como séries temporais e de efeitos aleatórios que são classes de modelos parcimoniosos.

O modelo de séries temporais surge quando cada seqüência individual dos erros for visto como uma série temporal (capítulo 3). Os modelos autoregressivos ou os médias móveis são intuitivos quando for possível associar as correlações entre as observações individuais com as distâncias de tempo (defasagens) que separam essas observações. Todos os modelos de séries temporais estacionárias são casos especiais da estrutura autoregressiva geral, chamada de *banded*, que tem um parâmetro separado para cada uma das defasagens.

O modelo de efeitos mistos possui a forma estruturada (2.5.5) para Σ . Essa forma é bastante flexível, englobando vários casos como curvas de crescimento polinomiais e modelos de regressão. A forma mais simples é a estrutura simétrica (padrão de uniformidade), onde Z é um vetor $(1, \dots, 1)'$. Nesse modelo assume-se que as medidas têm variância constante e uma correlação em comum. Essa forma corresponde a uma forma simples do modelo componente de variância com uma componente de variação entre indivíduos e outra dentro dos indivíduos. A simetria composta aparece em modelos de curva de crescimento polinomial quando as curvas individuais são todas paralelas e diferem aleatoriamente em elevação sobre a curva média populacional esperada.

Uma questão que surge quando se trabalha com estruturas de covariância é relacionar o ganho no ajuste com o número de parâmetros Θ introduzido no modelo. No exemplo de Jennrich e Schluchter (1986), 5 dentre 8 possíveis estruturas de covariância (com 1 a 13 parâmetros) ajustaram bem os dados e levaram a inferências essencialmente indistinguíveis. No exemplo de Rochon e Helms (1989), um AR(1) (com 1 parâmetro) reproduziu bem uma estrutura de covariância geral com 10 parâmetros.

Jones (1990) observou que mesmo havendo correlação serial das observações individuais, ela pode ser parcialmente confundida com os efeitos aleatórios individuais. Em dados reais, muitas vezes, não está claro se há correlação serial, efeitos aleatórios ou ambos.

Usando inferência baseada na função de verossimilhança (capítulo 5), não é sempre possível identificar o modelo correto, especialmente em pequenas amostras. Entretanto, é importante que algum esforço seja feito para tentar ajustar um bom modelo ao invés de fazer suposições. Isso significa, em geral, tentar modelos com efeitos aleatórios, com correlação serial e com ambos. Podem ser usados critérios como testes de razão de verossimilhança (capítulo 6) e o critério de Akaike (AIC) (Akaike, 1973) que será apresentado na seção 4.3. No exemplo investigado por Jones (1990), os efeitos aleatórios explicaram tanto a correlação serial como a heterogeneidade das variâncias.

Finalmente, a capacidade para modelar a estrutura de covariância é um fator importante na análise em pelo menos dois momentos: no primeiro, quando o interesse principal for a resposta média esperada (i.e., os parâmetros β), o ajuste de estruturas de covariância parcimoniosas podem produzir estimativas mais eficientes para β , especialmente

com dados altamente não balanceados ou incompletos. No segundo, em casos onde a estrutura de covariância é de interesse principal; por exemplo, descrever como as respostas individuais se desviam da média populacional ao longo do tempo. A análise dos resíduos longitudinais que também dependem da estrutura de covariância será discutida no capítulo 6. Os testes de razão de verossimilhança para comparar uma estrutura de covariância mais geral com uma estrutura mais simples (um caso especial da estrutura geral), como por exemplo, comparar a estrutura *banded* com o AR(1) podem ajudar na escolha do modelo e serão discutidos no capítulo 6.

4.3. Efeitos Fixos ou Aleatórios?

Teoricamente a decisão de tomar um efeito como fixo ou aleatório é um problema já resolvido. Para ilustrar essa questão, considere os seguintes exemplos considerados por Searle *et alii*. (1992):

Exemplo 4.3.1 variedade de tomates (efeitos fixos)

Considere que um pequeno jardineiro faça um pequeno experimento com 6 pés de tomate para cada uma das 4 variedades de interesse do jardineiro (no total, 24 pés de tomate). A comparação entre as 4 variedades é feita em um jardim com 12 m × 8 m. Cada planta é alocada aleatoriamente em um dos quadrados 2 m × 2 m. Se y_{ij} é a produção de tomates da planta j da variedade i , um modelo possível será

$$y_{ij} = \mu + \gamma_i + \varepsilon_{ij},$$

onde μ é a média geral de tomates produzidos; γ_i é o efeito da i -ésima variedade na produção dos tomates e ε_{ij} é o erro aleatório.

A maneira com que os dados foram obtidos sempre influencia nas inferências que podem ser extraídas dos dados. Por isso descrevemos todo o processo de amostragem pertinente a esse modelo de efeitos fixos.

□

Exemplo 4.3.2 médicos (efeitos aleatórios)

Suponha que uma nova forma de insulina injetável é testada por 15 médicos. É razoável tomar os 15 médicos como uma amostra aleatória de uma população de médicos que administram injeções. Se o médico i possuir n_i pacientes no ensaio, e a resposta do paciente j do médico i for y_{ij} , então um modelo possível será

$$y_{ij} = \mu + \gamma_i + \varepsilon_{ij},$$

onde μ é a média geral de resposta à insulina; γ_i é o efeito (aleatório) do i -ésimo médico e ε_{ij} é o erro aleatório.

Apesar do modelo descrito aqui ter a mesma forma algébrica do exemplo anterior, algumas das suposições básicas são diferentes. Nesse exemplo, γ_i é o efeito no nível de açúcar no sangue do paciente sob os cuidados do médico i . Como os médicos foram escolhidos aleatoriamente dentre uma população, o rótulo i não é de interesse particular no experimento ao contrário do exemplo anterior, onde a variedade de tomate era de interesse particular do jardineiro. Assim, uma das características dos efeitos aleatórios é que eles são utilizados como base para fazer inferências sobre a população de onde vieram. Portanto, γ_i é um efeito aleatório.

Mais precisamente, γ_i é uma realização da variável aleatória do efeito médico. Entretanto, como discutiremos no capítulo 5, por conveniência de notação, ignora-se a distinção entre a variável aleatória e o seu valor realizado.

Como γ_i é uma variável aleatória, torna-se necessário atribuir a ela algumas propriedades de probabilidade. Duas suposições comuns são: *a*) os efeitos são independentes e identicamente distribuídos (iid); *b*) eles têm média zero e a mesma variância; assim as duas suposições podem ser resumidas como $\gamma_i \sim \text{iid}(0, \sigma_\gamma^2)$, $\forall i$. A Tabela 4.3.1 ilustra comparativamente o modelo de efeito aleatório descrito nesse exemplo e o de efeito fixo descrito no exemplo anterior.

□

Exemplo 4.3.3 técnicos de uma fábrica (efeitos fixos ou aleatórios)

Em uma fábrica, dentre todos os técnicos que cuidam de máquinas, alguns técnicos são escolhidos aleatoriamente para um determinado experimento. Tome o modelo:

$$y_{ij} = \mu + \gamma_i + \varepsilon_{ij},$$

onde y_{ij} denota certa medida relativa ao i -ésimo técnico na j -ésima máquina; μ é a média geral; γ_i é o efeito do i -ésimo técnico e ε_{ij} é o erro aleatório. Na situação descrita, os efeitos dos técnicos são tomados como aleatórios, pois os técnicos são escolhidos aleatoriamente. Entretanto, se ao contrário, fosse necessário tomar todos eles e se o objetivo fosse avaliar diferenças entre técnicos específicos, então o modelo de efeitos fixos é mais apropriado. Similarmente, se um experimento envolve centenas de indivíduos que são considerados

como uma amostra aleatória de uma população, então efeitos aleatórios são apropriados. Mas se houver poucos indivíduos, digamos 5 ou 6, e o interesse estiver unicamente nesses indivíduos, então os efeitos individuais deverão ser fixos.

□

Na prática, no entanto, a decisão pela inferência condicional (efeitos fixos) ou marginal (efeitos aleatórios) não é tão imediata. Tome como exemplo o efeito do ano no estudo da plantação de trigo: será que o efeito dos anos na plantação devem ser considerados fixos ou aleatórios? os anos, por si só, dificilmente serão aleatórios, pois provavelmente considera-se o conjunto de anos consecutivos sobre a qual os dados foram colhidos. Por outro lado, os efeitos dos anos na plantação poderiam razoavelmente ser considerados aleatórios porque estariam sujeitos, talvez, à correlação entre as plantações em anos sucessivos. É claro que se fosse de interesse comparar anos específicos, então o tratamento dos anos como aleatórios não seria apropriado (Searle, 1971).

Assim, a decisão entre efeitos fixos e aleatórios também dependerá de fatores como o contexto dos dados, a maneira com que eles foram coletados, o tipo de inferência e o meio de onde eles vieram. Sobre isso, Searle (1971, pp. 382) comenta:

"In endeavoring to decide whether a set of effects is fixed or random, the context of the data, the manner in which they were gathered and the environment from which they came are the determining factors. In considering these points the important question is that of inference: are inferences going to be drawn from these data about just these levels of the factors? 'Yes' - then the effect are to be considered as fixed effects. 'No' - then (...) the effects are considered as being random."

Uma solução ao problema é, ou partir do modelo geral para o particular, ou do particular para o geral. No primeiro caso, uma forma de unificar os dois modelos (efeitos fixos e efeitos aleatórios) é assumir inicialmente que os efeitos são aleatórios. Para esse modelo, pode-se fazer inferências marginais ou incondicionais com respeito à população de efeitos. O modelo de efeitos fixos é obtido fazendo-se inferências condicionais nos efeitos que estão na amostra. Outra forma é fazer ao contrário: considerar os efeitos como fixos em um modelo condicional, aplicando-se uma estratégia semelhante ao modelo de duplo estágio (2.5.3) para os efeitos aleatórios.

Outra solução é estudar detalhadamente os fatores citados anteriormente que influem na escolha do modelo. Logo, a situação para o qual um modelo se aplica e as inferências baseadas nele são fatores que determinam os efeitos como fixos ou aleatórios. Quando os

efeitos forem uma amostra aleatória de uma população de efeitos, e desejando-se fazer inferências sobre essa população, então eles serão considerados aleatórios. Mas, quando as inferências são limitadas aos efeitos no modelo eles serão considerados como fixos.

Outra forma de responder à questão é verificar as seguintes perguntas: "os efeitos vêm de uma distribuição de probabilidade?" e "há informação suficiente sobre um fator de modo que os níveis desse fator possam ser considerados como uma amostra aleatória?" Caso as respostas sejam negativas, os efeitos devem ser considerados como fixos e serão estimados os níveis desses efeitos. Em caso afirmativo, os efeitos serão aleatórios e as componentes de variância devido aos efeitos aleatórios serão estimados; caso seja de interesse, pode-se também obter os valores realizados dos efeitos aleatórios.

A tabela seguinte é um resumo ilustrativo e comparativo das características entre o modelo de efeitos fixos e aleatórios:

Tabela 4.3.1 - Características dos efeitos fixos e dos efeitos aleatórios no modelo univariado $y_{ij} = \mu + \beta_i + e_{ij}$

característica	modelo de efeitos fixos	modelo de efeitos aleatórios
equação	$y_{ij} = \mu + \beta_i + e_{ij}$	$y_{ij} = \mu + \beta_i + e_{ij}$
média de y_{ij}	$E(y_{ij}) = \mu + \beta_i$	$E(y_{ij} \beta_i) = \mu + \beta_i$ $E(y_{ij}) = \mu$
β_i	efeito fixo e desconhecido	$\beta_i \sim \text{iid}(0, \sigma_\beta^2)$
e_{ij}	$e_{ij} = y_{ij} - E(y_{ij}) = y_{ij} - (\mu + \beta_i)$, $e_{ij} \sim \text{iid}(0, \sigma_e^2)$	$e_{ij} = y_{ij} - E(y_{ij} \beta_i) = y_{ij} - (\mu + \beta_i)$, $e_{ij} \sim \text{iid}(0, \sigma_e^2)$
$E(e_{ij}\beta_i)$	$E(e_{ij}\beta_i) = \beta_i E(e_{ij}) = 0$	$E(e_{ij}\beta_i) = 0$
$\text{var}(y_{ij})$	$\text{var}(y_{ij}) = \sigma_e^2$	$\text{var}(y_{ij}) = \sigma_\beta^2 + \sigma_e^2$
$\text{cov}(y_{ij}, y_{i'j'})$	$= \begin{cases} \sigma_e^2, & \text{para } i=i' \text{ e } j=j' \\ 0, & \text{caso contrário} \end{cases}$	$= \begin{cases} \sigma_\beta^2 + \sigma_e^2, & \text{para } i=i' \text{ e } j=j' \\ \sigma_\beta^2, & \text{para } i=i' \text{ e } j \neq j' \\ 0, & \text{caso contrário} \end{cases}$

A decisão sobre o tratamento dos efeitos como fixos ou aleatórios é importante. Em alguns casos, é possível haver muita diferença entre as estimativas dos parâmetros desses modelos, principalmente quando T for pequeno. Por exemplo, Hausman (1978), comparando o modelo de efeitos fixos com o de efeitos aleatórios, encontrou resultados significativamente diferentes na estimação de uma equação de salários. Ele utilizou uma amostra de 629 pessoas que completaram a escola (*high school*). Essas pessoas foram

seguidas por seis anos através do MIDS (*Michigan Income Dynamics Study*). Sobre essa questão Hsiao (1986, pp. 43) observa:

"(...) the question is why sometimes the fixed-effects and random-effects approaches yield vastly different estimates of the common slope coefficients that are not supposed to vary across individuals. (...) This brings us to the (...) important issue of whether or not the model is properly specified."

4.4. Critério de Informação de Akaike

Jones (1990) observou que a correlação serial das observações individuais pode ser parcialmente confundida com os efeitos aleatórios individuais. Em dados reais, muitas vezes, não está claro se há correlação serial, efeitos aleatórios ou ambos.

Usando inferência baseada na função de verossimilhança (a ser descrito no capítulo 5), nem sempre possível identificar o modelo correto, especialmente em pequenas amostras. Entretanto, é importante que algum esforço seja feito para tentar ajustar um bom modelo ao invés de fazer suposições. Isso significa, em geral, tentar modelos com efeitos aleatórios, com correlação serial e com ambos. Para isso, podem ser usados testes como a razão de verossimilhança (capítulo 6) e critérios de escolha como o de Akaike (AIC) (Akaike, 1973). Nessa seção apresentaremos o critério AIC que não deve ser confundido com os testes de hipóteses que serão apresentados no capítulo 6.

O AIC é baseado na teoria de decisão e penaliza $l = -2 \ln \text{verossimilhança}$ pelo número de parâmetros ajustados aos dados (r_a) para evitar excesso de parametrização:

$$\text{AIC} = l + 2r_a \quad (4.4.1)$$

Calcula-se o AIC para todos os modelos considerados e escolhe-se aquele que tiver o menor valor do AIC. Esse critério tem sido utilizado em várias áreas da estatística por vários autores como Bozdogan (1987), Jones (1993) e Wolfinger (1993).

Exemplo 4.4.1 nível de ácido láctico

Para ilustrar o AIC, considere os casos do modelo de efeitos mistos descritos no exemplo 2.5.6. Dentre os modelos polinomiais com $d-1 \leq c-1 \leq 6$ e $m-1 \leq 1$, ajustados

através do método de máxima verossimilhança (a ser discutido no próximo capítulo), o que produziu o menor AIC foi o modelo reduzido com $c-1 = 4$, $d-1 = 1$ e $m-1 = 1$, e erros AR(1). A Tabela 4.4.1 contém os resultados de alguns dos modelos ajustados:

Tabela 4.4.1 - Sumário dos ajustes de modelos do exemplo 2.5.6

modelo	$c-1$	$d-1$	$m-1$	r	l	AIC
saturado	6	6	0	16	1047,48	1079,48
	6	6	1	18	959,08	995,08
reduzido ("paralelismo") (com erros independentes)	6	0	0	10	1055,13	1075,13
	4	0	0	8	1055,40	1071,40
	4	0	1	10	967,38	987,38
reduzido ("paralelismo") (com erros AR(1))	6	0	0	11	1055,13	1077,13
	4	0	0	10	963,97	983,97
	4	0	1	11	932,36	954,36
reduzido ("paralelismo") (com erros AR(1), mas sem efeitos aleatórios)	6	0	0	10	963,95	983,95
	4	0	0	8	965,37	981,37
reduzido (com erros independentes)	5	1	1	12	962,02	986,02
	4	1	1	11	962,55	984,55
	4	2	1	12	961,66	985,66
reduzido (com erros AR(1))	5	1	1	13	925,54	951,54
	4	1	1	12	927,06	951,06
	4	2	1	13	926,19	952,19

$c-1$ é o grau do polinômio ajustado, r representa o número de parâmetros estimados, l é o valor de $-2 \ln$ verossimilhança e AIC é o critério de informação de Akaike: $AIC = l + 2r$.

Os resultados dos ajustes dos modelos reduzidos com erros AR(1) estão ilustrados a seguir:

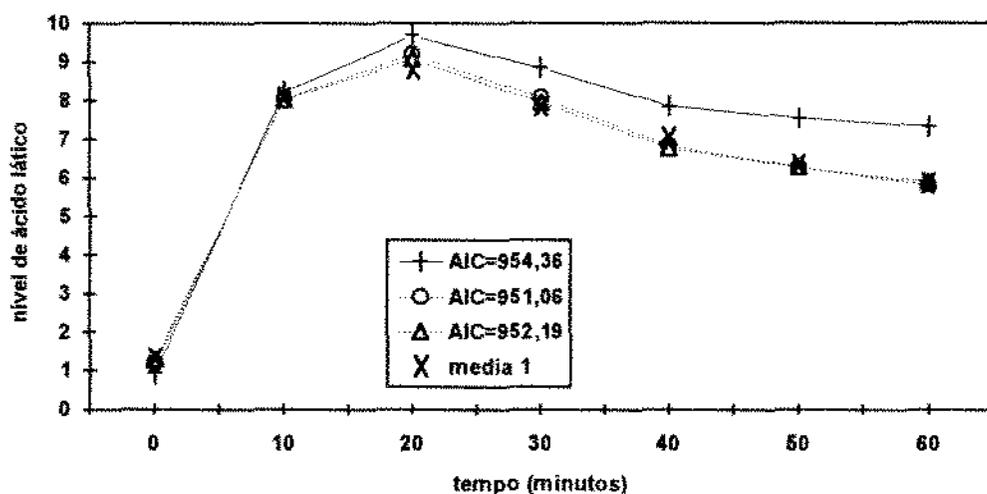


Fig. 4.1.1 - gráfico dos perfis ajustados para o grupo 1

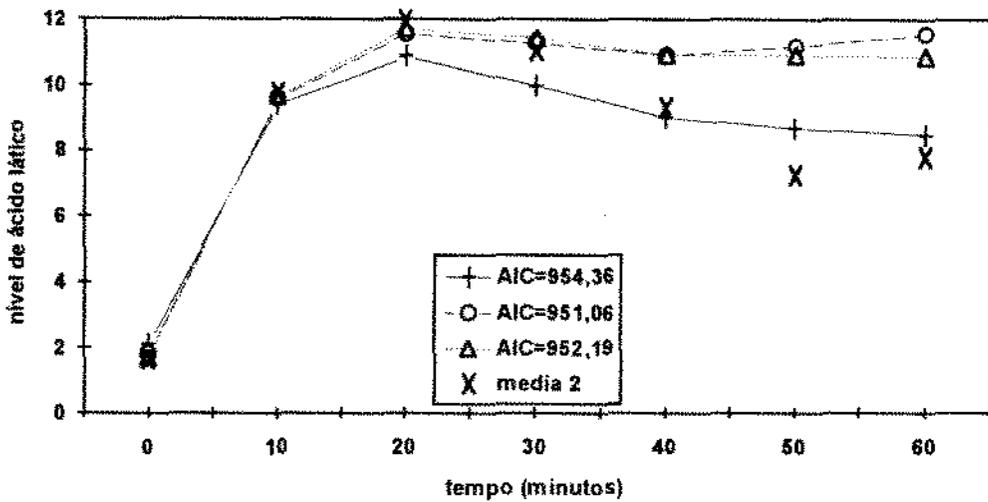


Fig. 4.1.2 - gráfico dos perfis ajustados para o grupo 2

Os detalhes de cálculos e análises desse conjunto de dados serão discutidos no capítulo 7. Um aspecto a ser considerado é que a modelagem direta dos dados possivelmente é prejudicada por causa da heterogeneidade das variâncias que pode ser percebida nas Figuras 1.2.3 e 1.2.4 (onde a variância inicial é pequena se comparada com os outros instantes).

□

O AIC tem sido criticado em análise de séries temporais por não ser um estimador consistente; por exemplo, se um modelo autoregressivo tem ordem verdadeira p_0 e o número de observações na série for para o infinito, o AIC nem sempre escolherá p_0 , mas um valor maior, sendo que algumas vezes ele escolhe uma ordem muito grande.

Por isso existem modificações como o SC (4.4.2) sugerida por Schwarz (1978) e o CAIC (4.4.3) proposto por Bozdogan (1987). As modificações em geral têm a forma

$$l + f(n, r_a), \quad (4.4.3)$$

onde f é uma função de penalidade.

Por exemplo, os critérios mais comuns são:

$$\text{AIC} = l + 2r_a$$

$$\text{HQC} = l + \ln(\ln(n))2r_a \quad (4.4.2)$$

$$\text{SC} = l + (\ln n)r_a \quad (4.4.3)$$

$$\text{CAIC} = l + (1 + \ln n)r_a \quad (4.4.4)$$

onde n é o número total de observações em todos os indivíduos. O SC tem uma penalidade crescente para r_a comparada ao AIC. O CAIC, chamado de AIC Consistente, possui uma penalidade ligeiramente maior. Assim, os critérios (4.4.1) ao (4.4.4) estão apresentados em ordem crescente de penalidade, ou seja, ordem crescente na preferência de parcimônia como está ilustrado no exemplo 4.4.2.

Outra forma de contornar as críticas é modificar ligeiramente o critério de decisão através do AIC (Jones 1990). Como o AIC não é utilizado como um critério absoluto, a escolha pelo modelo pode ser feita dentre os competitivos. Por exemplo, todos os modelos que possuírem o AIC dentro de 2 unidades do menor AIC encontrado podem ser considerados como competitivos.

Outra alternativa de interpretação dos contrastes entre os vários critérios de seleção do modelo utiliza argumentos Bayesianos (Poskitt e Tremayne, 1987). Não serão discutidos detalhes teóricos, mas apenas serão apresentados os resultados principais. Utilizando o critério AIC, por exemplo, tome a quantidade

$$\mathfrak{R} = \exp\left\{-\frac{1}{2} [AIC(r_1) - AIC(r)]\right\},$$

onde $AIC(r)$ e $AIC(r_1)$ denotam os critérios AIC para modelos com r e r_1 parâmetros, respectivamente. Considere que o modelo com r_1 parâmetros tenha o menor AIC e que o modelo com r seja um candidato. Poskitt e Tremayne (1987) discute que os modelos com

$$1 < \mathfrak{R} < \sqrt{10}$$

são modelos competitivos com o modelo que possui o menor AIC. O conjunto de modelos competitivos formam um *portfolio* dos modelos. Como ilustração das comparações entre os critérios, considere o exemplo 4.4.2.

O conjunto de dados da seção 1.2 contém 36 ratos com 214 observações (38 consideradas perdidas). A partir da Tabela 4.4.1, tomando-se os 4 menores valores do AIC, obtém-se os seguintes resultados:

Tabela 4.4.2 - Comparação entre os critérios de escolha

l	r	AIC	HQC	SC	CAIC
932,36	11	954,36	969,32	991,38	1002,38
925,54	13	951,54	969,22	995,29	1008,29
927,06	12	951,06	967,38	991,45	1003,45
926,19	13	952,19	969,87	995,94	1008,94

Os critérios estão em ordem crescente na preferência de parcimônia (do que penaliza menos ao que tem maior penalidade): o AIC escolhe o modelo que possui entre 12 e 13 parâmetros; o HQC escolhe o modelo com 12 parâmetros; o SC o modelo entre 11 e 12 parâmetros e o CAIC escolhe o modelo com 11 parâmetros.

Para ilustrar o *portfolio*, tome da Tabela 4.4.2 o modelo com menor AIC ($r_1 = 12$) como referência, e como candidato o modelo com AIC = 951,54. Assim, calculando-se a quantidade

$$\mathfrak{R} = \exp\left\{-\frac{1}{2} [951,06 - 951,54]\right\},$$

obtemos $\mathfrak{R} = 1,27 < \sqrt{10}$, indicando que os modelos são competitivos. A Tabela 4.4.3 ilustra o resultado do *portfolio*: nesse critério, os modelos com AIC dentro de 2,3 unidades do modelo com menor AIC podem ser considerados como competitivos.

Tabela 4.4.3 - *Portfolio* dos modelos

l	r	AIC	\mathfrak{R}	$\sqrt{10}$
932,36	11	954,36	5,21	>
925,54	13	951,54	1,27	<
927,06	12	951,06	1	<
926,19	13	952,19	1,76	<

□

5

Estimação

5.1. Introdução

Este capítulo apresenta alguns métodos e algoritmos de estimação de parâmetros dos modelos de efeitos aleatórios apresentados no capítulo 2. Viu-se que esse modelo possui uma estrutura de covariância flexível, podendo assumir várias formas e também ser aplicado em delineamentos não balanceados ou com tamanhos amostrais pequenos.

Os métodos de estimação são técnicas estatísticas úteis na modelagem dos dados. Eles podem ser baseados, por exemplo, em verossimilhança ou em critérios como os mínimos quadrados. Para um conjunto de dados, a análise depende das suposições do modelo, entre outros, o tipo de efeitos (fixos ou aleatórios), a estrutura de covariância e suposições sobre a distribuição dos dados. Basicamente, a abordagem adotada nesse trabalho é a inferência clássica, embora a abordagem Bayesiana permita fazer interpretações e justificativas interessantes (e.g., Harville, 1976; Laird e Ware, 1982 e Robinson, 1991). As relações entre ambas abordagens serão, eventualmente, citadas no trabalho, mas não exploradas.

Os estimadores de ANOVA não serão tratados detalhadamente nesse trabalho, mas uma breve introdução será dada a seguir. Um breve histórico e comentários sobre esses estimadores podem ser encontrados em Searle *et alii.* (1992, cap. 2). Os métodos de estimação via ANOVA foram largamente desenvolvidos e utilizados pelo menos até a década de 70. Basicamente, o nome ANOVA é dado ao método que estima as componentes de variância equacionando-se as somas de quadrados de seus valores esperados. Rigorosamente esta só ocorre se o conjunto de dados for balanceado, quando as somas de quadrados formam um conjunto único de equações, o que não é verdade no caso não balanceado. Quando isto ocorre, encontram-se na literatura diversas variações de ANOVA, como por exemplo, os três métodos de Henderson (Searle *et alii.*, 1992, pp. 38) para dados não balanceados que são conhecidos como métodos I, II e III. Uma característica comum a qualquer estimador ANOVA é que ele sempre será não viciado. Para dados balanceados, os estimadores ANOVA, dentro de uma classe de funções quadráticas das observações que são estimadores não viciados das componentes de variância, têm mínima variância (Searle *et*

alii., 1992, pp. 129). Em outras palavras, os estimadores ANOVA são BQUE (*Best Quadratic Unbiased Estimators*). Adicionando-se a suposição de normalidade, os estimadores ANOVA serão BUE (*Best Unbiased estimators*). Em contraste com dados balanceados, no caso não balanceado não existe um estimador que seja BUE. Há algumas variações do método ANOVA para dados não balanceados mas nenhuma resolveu completamente os problemas (algumas modificações podem ser encontradas em Searle *et alii.*, 1992, pp. 130; bem como um estudo detalhado desse assunto, pp. 168).

Apesar dos vários desenvolvimentos dos estimadores ANOVA, o método continuou com suas principais deficiências: negatividade (possibilidade em estimar valores negativos para as variâncias), falta de propriedades distribucionais (principalmente no caso não balanceado, o estudo dessas propriedades se torna complicado) e falta de um meio prático para comparar as diferentes aplicações da metodologia ANOVA. Em face desses problemas, desenvolveram-se métodos alternativos como a máxima verossimilhança para a estimação no modelo componente de variâncias.

Os métodos baseados em verossimilhança, que serão discutidos nas seções 5.2.1 e 5.2.2, requerem suposições sobre a distribuição dos dados. Sob a suposição de normalidade nos dados observados, serão apresentadas as técnicas de estimação de Máxima Verossimilhança (MV) e de Máxima Verossimilhança Restringida (MVRe), também conhecida como máxima verossimilhança modificada, marginal ou residual. Harville (1977) considerou a estimação via máxima verossimilhança. Uma desvantagem dessa abordagem é que, com dados balanceados, suas estimativas não levam em consideração os graus de liberdade perdidos na estimação do efeito fixo, resultando em estimação viciada e diferindo dos estimadores de ANOVA (que são BQUE e, caso haja normalidade, são BUE). Uma alternativa é a modificação da MV sugerida por Patterson e Thompson (1971), a EMVRe. Nessa modificação, a função de verossimilhança é fatorada em duas partes; uma parte é totalmente livre dos parâmetros de locação (os efeitos fixos) e quando otimizada produz os estimadores MVRE. Efetivamente, esse método maximiza a verossimilhança em certas combinações lineares dos dados (restrições). Nessa metodologia, as soluções das equações de verossimilhança para dados balanceados são idênticas aos estimadores ANOVA. Além disso, ao contrário da MV, ela considera os graus de liberdade associados aos efeitos fixos.

No caso mais geral, onde cada observação é considerada pertencente à família exponencial, poderá ser utilizada a análise via Equações Generalizadas de Estimação (GEE) proposta por Liang e Zeger (1986). Esse método é uma abordagem de estimação através do modelo linear generalizado (GLM), onde não é necessário especificar completamente a distribuição conjunta das observações individuais. Especifica-se somente a verossimilhança para as distribuições marginais e uma matriz de correlação individual, chamada de *working*

correlation matrix. Essa abordagem permite modelar tanto dados contínuos como discretos dentro da formulação de família exponencial, podendo ser utilizada em dados de resposta binária e Poisson. Apesar dessa generalidade, trataremos apenas dos casos de normalidade e de especificação somente da média e variância da observação. Neste último, a estimação resulta em quase-verossimilhança. Com normalidade, as GEE se reduzem à equação de escore de máxima verossimilhança apenas no caso de haver dados completos com matriz de correlação não estruturada.

Particularmente, a aplicação dos métodos de estimação do modelo de efeitos aleatórios viabiliza, entre outros objetivos de inferência, a predição e a análise das componentes de variância. O problema de estimar ou predizer um dado futuro sob esse modelo pode ser formulado como um problema de estimar uma combinação linear dos efeitos fixos e aleatórios (Harville, 1977). O estimador que será considerado é chamado de BLUP (*Best Linear Unbiased Prediction*), porque, dentre todos os estimadores lineares não-viciados, possui o menor erro quadrático médio. Na seção seguinte, 1.1, discutiremos brevemente sobre a terminologia "predição" utilizada para designar a "estimação" do efeito aleatório. A expressão "estimação dos parâmetros" empregada nesse trabalho deve ser entendida de forma genérica, considerando-se a discussão a ser apresentada.

Henderson (1950, citado por Robinson, 1991) obteve estimadores para os efeitos fixos e aleatórios sob suposição de normalidade dos dados utilizando a função de "verossimilhança" conjunta (5.2.1.20). Como será discutido, essa função, que é otimizada com respeito aos parâmetros do modelo, não é de fato uma função de verossimilhança. Entretanto, a combinação linear desses estimadores é o BLUP (Henderson, 1963, citado por Robinson, 1991). Uma extensão do teorema de Gauss-Markov foi apresentada por Harville (1976) para incluir a estimação dos efeitos aleatórios. Como será visto na seção 5.2.4, esse resultado é um pouco mais geral do que o resultado de Henderson.

Os métodos baseados em mínimos quadrados foram considerados por autores como Harvey (1978), Dielman (1989, pp. 49) e Hsiao (1986, pp. 34). Essa abordagem não será tratada nesse trabalho. Outra classe de estimadores, conhecidos como estimadores MINQUE (*minimum-norm quadratic unbiased equation*) obtidos por Rao (1973, pp. 303) também não serão considerados. O MINQUE resultou da preocupação em se desenvolver estimadores para as componentes de variância que fossem não viciadas de mínima norma quadrática, uma idéia análoga aos estimadores BLUE da média em modelos lineares. Em termos gerais, os estimadores resultantes têm uma variância generalizada minimizada e são estimadores não viciados. A estimação MINQUE não requer propriedades de distribuição dos efeitos aleatórios ou dos dados, nem requer cálculos iterativos para obter as estimativas (os estimadores são soluções de equações lineares). Entretanto, os estimadores obtidos pelo

MINQUE são funções de valores a priori, utilizadas em lugar das componentes de variância no próprio procedimento de estimação. Assim, a propriedade minimal se aplica apenas a esses valores a priori, o que constitui de certo modo uma deficiência do método. Uma consequência disso é: a partir de um mesmo conjunto de dados e ajustando-se o mesmo modelo, se N diferentes pessoas escolherem cada um seu próprio conjunto a priori de valores, cada um produzirá N diferentes conjuntos de estimadores, embora os estimadores sejam não viciados, independentemente dos valores a priori.

Em geral, as equações de estimação de máxima verossimilhança não produzem expressões explícitas para os estimadores. A maximização da verossimilhança, em vários métodos, é realizada iterativamente, consistindo essencialmente de dois passos. No primeiro, estima-se os efeitos do modelo com um valor fixado para o conjunto de parâmetros da matriz de covariância. Na inicialização do algoritmo iterativo, esse valor pode ser arbitrário, por exemplo, a matriz identidade. No segundo passo, estima-se os parâmetros da matriz de covariância com base nas estimativas do primeiro passo. Assim, os novos valores estimados são utilizados no primeiro passo para dar continuidade ao algoritmo. Para a construção desses passos é importante estudar os algoritmos de estimação que são procedimentos destinados ao cálculo numérico das estimativas dos parâmetros. Entre os algoritmos considerados estão o algoritmo EM (apêndice A), o método de Newton-Raphson, o *scoring* e o filtro de Kalman (apêndice B). A aplicação dos algoritmos de estimação para o modelo de efeitos aleatórios será apresentada na seção 5.3.

O algoritmo EM, descrito no Apêndice A, é útil quando há dados perdidos. Entretanto, ele recebe algumas críticas, por exemplo, quando se utiliza uma estrutura de correlação temporal nos erros. O método de Newton e o *Scoring* são alternativas para contornar essas críticas. O filtro de Kalman é uma opção atrativa pois suporta dados com estrutura não balanceada, permite o cálculo da verossimilhança exata, generaliza situações onde a verossimilhança não pode ser expressa em uma forma fechada e é numericamente rápido, mesmo que haja um número grande de observações em alguns dos indivíduos (Harvey, 1978; Jones e Boateng, 1991).

5.1.1. Terminologia

Os "estimadores" dos efeitos aleatórios, em geral, são chamados de preditores, enquanto que os dos efeitos fixos são estimadores. Autores como Henderson (1984, p. 37) citado por Robinson (1991) expressaram dúvidas a respeito do uso indiscriminado dessa terminologia:

"Which is the more logical concept, prediction of a random variable or estimation of the realized value of a random variable? If we have an animal already born, it seems reasonable to describe the evaluation of its breeding value as an estimation problem. On the other hand, if we are interested in evaluating the potential breeding value of a mating between two potential parents, this would be a problem in prediction. If we are interested in future records, the problem is clearly one of prediction".

O uso do termo "preditor" pode ser inadequado, por exemplo, em casos onde o objeto a ser estimado já tenha ocorrido. Também é inadequado em casos onde o valor presente não é relevante para realizações futuras, não sendo, portanto, de interesse.

De forma geral, costuma-se "estimar" efeitos quando são considerados fixos, e "predizer" quando aleatórios. Alguns autores preferem usar termos como "estimativa de um valor realizado" ou simplesmente "estimativa". Por simplicidade, a expressão "estimação dos parâmetros" empregada nesse trabalho deve ser entendida de forma genérica, considerando-se a discussão apresentada nessa subseção.

5.2. Métodos de Estimação

Nessa seção serão considerados os métodos baseados na verossimilhança (MV e MVRe), o BLUP e o GEE. Em contraste aos métodos ANOVA, MINQUE e BLUP que não requerem suposições sobre a distribuição dos dados, os métodos baseados em verossimilhança, que serão discutidos nas seções 5.2.1 e 5.2.2, requerem essas suposições; nesse trabalho, assumiremos a normalidade nos dados. No método via GEE não é necessário especificar completamente a distribuição conjunta das observações individuais. Especifica-se somente a verossimilhança para as distribuições marginais e uma matriz de correlação individual (*working correlation matrix*). Apesar da generalidade desse método, trataremos apenas dos casos de normalidade e de especificação somente da média e variância da observação.

5.2.1. Estimador de Máxima Verossimilhança

5.2.1.a. Verossimilhança Marginal

Reconsidere o modelo de efeitos aleatórios apresentado na seção 2.5, i. e.,

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i, \quad (5.2.1.1)$$

onde $\varepsilon_i \sim N(0, \Omega_i)$, $b_i \sim N(0, \Gamma)$, com b_i e ε_i não correlacionados, X_i é a matriz $T_i \times a$ de delineamento do indivíduo i ; β é o vetor dos parâmetros fixos; Z_i é a matriz $T_i \times m$ de delineamento individual para o vetor de efeitos aleatórios (não observáveis) b_i .

Portanto os vetores individuais y_i , $i = 1, \dots, n$, têm distribuições normais multivariadas independentes, ou seja,

$$y_i \sim \text{NID}(X_i\beta, \Sigma_i), \quad (5.2.1.2)$$

onde

$$\Sigma_i = \text{var}(y_i) = Z_i\Gamma Z_i' + \Omega_i. \quad (5.2.1.3)$$

Denote por Θ_i o vetor $f \times 1$ de parâmetros das matrizes Γ e Ω_i , i.e., $\Theta_i = \{\gamma \in \Gamma \text{ e } \omega \in \Omega_i\}$ e considere o vetor de parâmetros $\vartheta' = \{\beta', b_1', \dots, b_n', \Theta_i'\}$.

A função de densidade marginal de y_i (5.2.1.2) é

$$\begin{aligned} f(y_i; \Theta_i, \beta) &= |2\pi\Sigma_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}Q_i(\beta, \Sigma_i)\right\} = \\ &= (2\pi)^{-\frac{T_i}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}Q_i(\beta, \Sigma_i)\right\}, \end{aligned} \quad (5.2.1.4)$$

onde

$$Q_i(\beta, \Sigma_i) = (y_i - X_i\beta)' \Sigma_i^{-1} (y_i - X_i\beta). \quad (5.2.1.5)$$

A função de verossimilhança é dada por

$$L(\Theta_i, \beta; y_1, \dots, y_n) = \prod_{i=1}^n f(y_i; \Theta_i, \beta). \quad (5.2.1.6)$$

Assim,

$$l = -2\ln(\text{verossimilhança}) = -2\ln\{L(\Theta_i, \beta; y_1, \dots, y_n)\} = -2\ln\left\{\prod_{i=1}^n f(y_i; \Theta_i, \beta)\right\} =$$

$$= -2 \sum_{i=1}^n \ln f(y_i; \Theta_i, \beta) = \sum_{i=1}^n T_i \ln 2\pi + \sum_{i=1}^n \ln |\Sigma_i| + \sum_{i=1}^n Q_i(\beta, \Sigma_i).$$

Portanto,

$$l = T \ln 2\pi + \sum_{i=1}^n \ln |\Sigma_i| + \sum_{i=1}^n Q_i(\beta, \Sigma_i), \quad (5.2.1.7)$$

onde $T = \sum_{i=1}^n T_i$ é o número total de observações em todos os indivíduos.

Note que, na função (5.2.1.7), o efeito aleatório aparece apenas através de Σ_i . O modelo pode, então, ser escrito na forma do modelo componente de erro

$$y_i = X_i \beta + v_i, \quad (5.2.1.8)$$

onde $v_i = Z_i b_i + \epsilon_i$ é o termo aleatório.

Uma forma de minimizar (5.2.1.7) em função dos parâmetros é calcular as soluções das equações de máxima verossimilhança determinadas por:

$$(a) \text{ desenvolver } \frac{\partial l}{\partial \beta} \text{ e fazer } \left. \frac{\partial l}{\partial \beta} \right|_{\beta=\hat{\beta} \text{ e } \Theta_i=\hat{\Theta}_i} = 0; \quad (5.2.1.9.a)$$

$$(b) \text{ desenvolver } \frac{\partial l}{\partial \Theta_i} \text{ e fazer } \left. \frac{\partial l}{\partial \Theta_i} \right|_{\beta=\hat{\beta} \text{ e } \Theta_i=\hat{\Theta}_i} = 0, \quad (5.2.1.9.b)$$

onde $\hat{\beta}$ e $\hat{\Theta}_i$ são os estimadores de MV (EMV).

Assim, desenvolvendo-se (5.2.1.9.a) obtemos

$$\frac{\partial l}{\partial \beta} = -2X' \Sigma^{-1} (y - X\beta), \quad (5.2.1.9.c)$$

$$\text{onde } X_{T \times n} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, y_{T \times 1} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \Sigma_{T \times T} = \text{diag}(\Sigma_1 \cdots \Sigma_n), \text{ e } T = \sum_{i=1}^n T_i.$$

Fazendo-se $\left. \frac{\partial l}{\partial \beta} \right|_{\beta=\hat{\beta} \text{ e } \Theta_i=\hat{\Theta}_i} = 0$, obtemos o EMV de β :

$$\hat{\beta} = (X' \hat{\Sigma}^{-1} X)^{-1} X' \hat{\Sigma}^{-1} y. \quad (5.2.1.10)$$

Deve-se observar que as soluções $\hat{\beta}$ e $\hat{\Theta}_i$ das equações (5.2.1.9.a) e (5.2.1.9.b) não são necessariamente os estimadores de máxima verossimilhança de β e Θ_i , mas são possíveis candidatos. Para verificar se $\hat{\beta}$ e $\hat{\Theta}_i$ são os estimadores de MV (EMV) de β e Θ_i , respectivamente, é preciso checar as derivadas segundas da função de verossimilhança (que, como será visto na seção 5.3.2, formam uma matriz chamada de *Hessiana*), e também checar a função nas fronteiras do espaço paramétrico Θ (e.g., no modelo componente de variâncias, os EMV de σ_e^2 e σ_π^2 devem satisfazer as condições $\sigma_e^2 > 0$ e $\sigma_\pi^2 \geq 0$). Como em geral as equações de máxima verossimilhança não produzem soluções fechadas, tanto as soluções $\hat{\beta}$ e $\hat{\Theta}_i$ como a verificação da otimalidade são feitas numericamente através de algoritmos de cálculo a serem discutidos e apresentados na seção 5.3. Considerando-se essas observações, assumiremos a seguir que as soluções $\hat{\beta}$ e $\hat{\Theta}_i$ sejam os EMV de β e Θ_i , respectivamente.

Dado $\hat{\Theta}_i$, a estimativa de β que minimiza \bar{l} é a solução das *equações normais* definida por (5.2.1.10). O termo *equações normais* relaciona-se à teoria dos mínimos quadrados que utiliza projeções ortogonais na construção das equações. O estimador (5.2.1.10) é o estimador de mínimos quadrados ponderados quando a matriz de covariância dos erros na regressão linear é $\hat{\Sigma}_i$ (Jones, 1993, pp. 33).

Para se obter $\hat{\Theta}_i$ através de (5.2.1.9.b), considere a seguinte ilustração:

Exemplo 5.2.1 estrutura de variância conhecida

Jones (1993, pp. 32) considera o caso onde há uma componente de variação σ^2 comum aos efeitos aleatórios, i.e., $\Gamma = \sigma^2 B$ e $\Omega_i = \sigma^2 W_i$, com B e W_i conhecidos. Assim, $\Sigma_i = \sigma^2 (Z_i B Z_i' + W_i) = \sigma^2 G_i$ e $\Theta = \sigma^2$.

A função $l = -2 \ln(\text{Verossimilhança})$ é dada por (5.2.1.7):

$$\begin{aligned} l &= \sum_{i=1}^n \left[T_i \ln 2\pi + \ln |\Sigma_i| + (y_i - X_i \beta)' \Sigma_i^{-1} (y_i - X_i \beta) \right] = \\ &= \sum_{i=1}^n \left[T_i \ln 2\pi + \ln |\sigma^2 G_i| + \sigma^{-2} (y_i - X_i \beta)' G_i^{-1} (y_i - X_i \beta) \right] = \\ &= T \ln 2\pi + \sum_{i=1}^n \left[\ln |\sigma^2 G_i| + \sigma^{-2} (y_i - X_i \beta)' G_i^{-1} (y_i - X_i \beta) \right], \end{aligned}$$

onde $T = \sum_{i=1}^n T_i$ é o número total de observações em todos os indivíduos. Como σ^2 multiplica todos os elementos da matriz G_i , então $|\sigma^2 G_i| = \sigma^{2T_i} |G_i|$. Logo, $\ln |\sigma^2 G_i| = \ln \sigma^{2T_i} |G_i| = T_i \ln \sigma^2 + \ln |G_i|$. Então

$$l = T \ln 2\pi + \ln \sigma^2 + \sum_{i=1}^n \left[\ln |G_i| + \sigma^{-2} (y_i - X_i \beta)' G_i^{-1} (y_i - X_i \beta) \right] \quad (5.2.1.11)$$

Diferenciando-se l com respeito a σ^2 e fazendo $\left. \frac{\partial l}{\partial \sigma^2} \right|_{\beta = \hat{\beta} \text{ e } \sigma^2 = \hat{\sigma}^2} = 0$ obtemos

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - X_i \hat{\beta})' G_i^{-1} (y_i - X_i \hat{\beta})}{T} \quad (5.2.1.12)$$

A estimativa de β nesse exemplo tem a forma

$$\hat{\beta} = \left(\sum_{i=1}^n X_i' G_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^n X_i' G_i^{-1} y_i \right) \quad (5.2.1.13)$$

□

Genericamente, desenvolvendo-se $\frac{\partial l}{\partial \Theta_i}$ para (5.2.1.7), temos a seguinte expressão para cada elemento $\sigma_{\alpha'} \in \Theta_i$:

$$\frac{\partial l}{\partial \sigma_{\alpha'}} = \sum_{i=1}^n \text{tr} \left(\Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \sigma_{\alpha'}} \right) + (y_i - X_i \beta)' \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \sigma_{\alpha'}} \Sigma_i^{-1} (y_i - X_i \beta), \quad (5.2.1.14)$$

5.2.1.b. "Verossimilhança" Condicional

Do exemplo 5.2.1, observa-se que o EMV para σ^2 não leva em consideração os graus de liberdade perdidos na estimação de β , incorporando um certo vício ao estimador de β . Uma alternativa de estimação para contornar esse problema será abordada na próxima seção.

A função (5.1.2.7) não permite achar diretamente uma expressão para \hat{b}_i , que é útil para fazer previsões para o i -ésimo indivíduo, i.e.,

$$\hat{y}_i = X_i \hat{\beta} + Z_i \hat{b}_i. \quad (5.2.1.15)$$

É possível reescrever a função de verossimilhança de modo que o efeito aleatório esteja explícito (Robinson, 1991). Por simplicidade, assuma que Γ seja matriz não-singular.

A densidade conjunta de y_i e b_i pode ser expressa como

$$f(y_i, b_i; \Theta_i, \beta) = f(b_i; \Gamma) f(y_i | b_i; \Omega_i, \beta), \quad (5.2.1.16)$$

onde a distribuição condicional de y_i dado b_i é $N_{T_i}(X_i \beta + Z_i b_i, \Omega_i)$, com densidade

$$f(y_i | b_i; \Omega_i, \beta) = |2\pi\Omega_i|^{-1} \exp\left\{-\frac{1}{2} Q_i(\beta, b_i, \Omega_i)\right\}, \quad (5.2.1.17)$$

sendo $Q_i(\beta, b_i, \Omega_i) = (y_i - X_i \beta - Z_i b_i)' \Omega_i^{-1} (y_i - X_i \beta - Z_i b_i)$,

e a distribuição marginal de b_i é $N_m(0, \Gamma)$, com densidade

$$f(b_i; \Gamma) = |2\pi\Gamma|^{-1} \exp\left\{-\frac{1}{2} b_i' \Gamma^{-1} b_i\right\}. \quad (5.2.1.18)$$

Logo, substituindo-se (5.2.1.17) e (5.2.1.18) em (5.2.1.16), a densidade conjunta de y_i e b_i pode ser escrita como

$$f(y_i, b_i; \Theta_i, \beta) = |2\pi\Omega_i|^{-1} |2\pi\Gamma|^{-1} \exp\left\{-\frac{1}{2} [Q_i(\beta, b_i, \Omega_i) + b_i' \Gamma^{-1} b_i]\right\}, \quad (5.2.1.19)$$

e a função de "verossimilhança" conjunta é expressa como

$$L(\Theta_i, \beta, b_i; y_1, \dots, y_n) = \prod_{i=1}^n f(y_i, b_i; \Theta_i, \beta). \quad (5.2.1.20)$$

Assim,

$$\begin{aligned}
 l &= -2 \ln \{L(\Theta_1, \beta, b_1; y_1, \dots, y_n)\} = -2 \ln \left\{ \prod_{i=1}^n f(y_i; \Theta_1, \beta, b_1) \right\} = \\
 &= \sum_{i=1}^n \{ \ln |2\pi\Gamma| + \ln |2\pi\Omega_i| + Q_i(\beta, b_1, \Omega_i) + b_i' \Gamma^{-1} b_i \} = \\
 &= \sum_{i=1}^n \{ m \ln 2\pi + \ln |\Gamma| + T_i \ln 2\pi + \ln |\Omega_i| + Q_i(\beta, b_1, \Omega_i) + b_i' \Gamma^{-1} b_i \} = \\
 &= (n \cdot m + T) \ln 2\pi + \sum_{i=1}^n \{ \ln |\Gamma| + \ln |\Omega_i| + Q_i(\beta, b_1, \Omega_i) + b_i' \Gamma^{-1} b_i \}. \quad (5.2.1.21)
 \end{aligned}$$

Deve-se observar que a função de "verossimilhança" conjunta de y e b (5.2.1.20) não é de fato uma verossimilhança no sentido usual, pois os efeitos aleatórios b são variáveis aleatórias não observáveis (Henderson, 1973, pp. 16, citado por Robinson, 1991). É possível justificar seu uso em termos de distância entre duas distribuições (Maritz e Lwin, 1989, p. 212). Em particular, o critério *Kullback-Leibler* pode ser empregado. Assim, a distância entre duas distribuições empíricas de y_i e b_i é minimizado se $\sum_{i=1}^n \ln f(b_i; \Gamma)$ for minimizado. Podemos considerar b como o "ponto" intermediário entre a distribuição empírica dos y_i e a distribuição teórica dos b_i . Então podemos considerar que a distância total entre a distribuição empírica dos y_i e a distribuição teórica dos b_i é minimizada quando minimiza-se a expressão:

$$\sum_{i=1}^n \ln f(y_i | b_i, \Omega_i, \beta) + \sum_{i=1}^n \ln f(b_i; \Gamma), \quad (5.2.1.22)$$

que é o logaritmo da expressão (5.2.1.20).

Para minimizar l em função de β , b_1 e Θ_1 pode-se:

$$\text{(a) desenvolver } \frac{\partial l}{\partial \beta} \text{ e fazer } \left. \frac{\partial l}{\partial \beta} \right|_{\beta=\hat{\beta}, b_1=\hat{b}_1, \Theta_1=\hat{\Theta}_1} = 0; \quad (5.2.1.23 \text{ a})$$

$$\text{(b) desenvolver } \frac{\partial l}{\partial b_i} \text{ e fazer } \left. \frac{\partial l}{\partial b_i} \right|_{\beta=\hat{\beta}, b_1=\hat{b}_1, \Theta_1=\hat{\Theta}_1} = 0; \quad (5.2.1.23 \text{ b})$$

$$\text{(c) desenvolver } \frac{\partial l}{\partial \Theta_1} \text{ e fazer } \left. \frac{\partial l}{\partial \Theta_1} \right|_{\beta=\hat{\beta}, b_1=\hat{b}_1, \Theta_1=\hat{\Theta}_1} = 0; \quad (5.2.1.23 \text{ c})$$

Assim,

$$\frac{\partial l}{\partial \beta} = -2 \sum_{i=1}^n X_i' \Omega_i^{-1} (y_i - X_i \beta - Z_i b_i).$$

Fazendo-se $\frac{\partial l}{\partial \beta} \Big|_{\beta=\hat{\beta}, b_i=\hat{b}_i, \epsilon \Theta_i=\hat{\Theta}_i} = 0$, temos

$$\sum_{i=1}^n X_i' \hat{\Omega}_i^{-1} (y_i - X_i \hat{\beta} - Z_i \hat{b}_i) = 0 \Rightarrow$$

$$\sum_{i=1}^n (X_i' \hat{\Omega}_i^{-1} X_i \hat{\beta} + X_i' \hat{\Omega}_i^{-1} Z_i \hat{b}_i) = \sum_{i=1}^n X_i' \hat{\Omega}_i^{-1} y_i \quad (5.2.1.24)$$

$$\frac{\partial l}{\partial b_i} = \frac{\partial Q_i(\beta, b_i, \Omega_i)}{\partial b_i} + \frac{\partial b_i' \Gamma^{-1} b_i}{\partial b_i} = -2Z_i' \hat{\Omega}_i^{-1} (y_i - X_i \hat{\beta} - Z_i \hat{b}_i) + 2\Gamma^{-1} b_i;$$

Fazendo-se $\frac{\partial l}{\partial b_i} \Big|_{\beta=\hat{\beta}, b_i=\hat{b}_i, \epsilon \Theta_i=\hat{\Theta}_i} = 0$, temos

$$\hat{\Gamma}^{-1} \hat{b}_i = Z_i' \hat{\Omega}_i^{-1} (y_i - X_i \hat{\beta} - Z_i \hat{b}_i) \Rightarrow$$

$$(\hat{\Gamma}^{-1} + Z_i' \hat{\Omega}_i^{-1} Z_i) \hat{b}_i = Z_i' \hat{\Omega}_i^{-1} (y_i - X_i \hat{\beta}) \Rightarrow$$

$$Z_i' \hat{\Omega}_i^{-1} X_i \hat{\beta} + (\hat{\Gamma}^{-1} + Z_i' \hat{\Omega}_i^{-1} Z_i) \hat{b}_i = Z_i' \hat{\Omega}_i^{-1} y_i \Rightarrow \quad (5.2.1.25.a)$$

$$\hat{b}_i = (\hat{\Gamma}^{-1} + Z_i' \hat{\Omega}_i^{-1} Z_i)^{-1} Z_i' \hat{\Omega}_i^{-1} (y_i - X_i \hat{\beta}) \quad (5.2.1.25.b)$$

As equações (5.2.1.24) e (5.2.1.25.a) são chamadas de equações simultâneas e podem ser representadas matricialmente¹: os estimadores $\hat{\beta}$ e \hat{b} são as soluções do sistema

$$\begin{bmatrix} X' \Omega^{-1} X & X' \Omega^{-1} Z \\ Z' \Omega^{-1} X & \Gamma^{-1} + Z' \Omega^{-1} Z \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X' \Omega^{-1} y \\ Z' \Omega^{-1} y \end{bmatrix}, \quad (5.2.1.26)$$

onde $X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ e $\Omega = \text{diag}(\Omega_1, \dots, \Omega_n)$.

A partir dessas equações, pode-se obter expressões fechadas para $\hat{\beta}$ e \hat{b} , em função de $\hat{\Theta}_i$, como será mostrado a seguir.

¹Embora a notação totalmente matricial facilite o manuseio das fórmulas, optou-se pela representação via matrizes individuais para destacar a natureza individual dos efeitos aleatórios e a independência entre os indivíduos.

Desenvolvendo-se a equação (5.2.1.24) utilizando (5.2.1.25 b), temos:

$$\begin{aligned} & \sum_{i=1}^n (\mathbf{X}'_i \hat{\Omega}_i^{-1} \mathbf{X}_i \hat{\beta}) + \sum_{i=1}^n \mathbf{X}'_i \hat{\Omega}_i^{-1} \mathbf{Z}_i \hat{b}_i = \sum_{i=1}^n \mathbf{X}'_i \hat{\Omega}_i^{-1} y_i \Rightarrow \\ \Rightarrow & \sum_{i=1}^n (\mathbf{X}'_i \hat{\Omega}_i^{-1} \mathbf{X}_i \hat{\beta}) + \sum_{i=1}^n \mathbf{X}'_i \hat{\Omega}_i^{-1} \mathbf{Z}_i (\hat{\Gamma}^{-1} + \mathbf{Z}'_i \hat{\Omega}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \hat{\Omega}_i^{-1} (y_i - \mathbf{X}_i \hat{\beta}) = \sum_{i=1}^n \mathbf{X}'_i \hat{\Omega}_i^{-1} y_i; \\ \Rightarrow & \sum_{i=1}^n \mathbf{X}'_i \left[\hat{\Omega}_i^{-1} - \hat{\Omega}_i^{-1} \mathbf{Z}_i (\hat{\Gamma}^{-1} + \mathbf{Z}'_i \hat{\Omega}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \hat{\Omega}_i^{-1} \right] \mathbf{X}_i \hat{\beta} = \\ & = \sum_{i=1}^n \mathbf{X}'_i \left[\hat{\Omega}_i^{-1} - \hat{\Omega}_i^{-1} \mathbf{Z}_i (\hat{\Gamma}^{-1} + \mathbf{Z}'_i \hat{\Omega}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \hat{\Omega}_i^{-1} \right] y_i. \end{aligned}$$

A partir de $\hat{\Omega}_i$ e $\hat{\Gamma}$ é natural fazer $\hat{\Sigma}_i = \mathbf{Z}_i \hat{\Gamma} \mathbf{Z}'_i + \hat{\Omega}_i$. Pela álgebra de matrizes (Mardia *et alii.*, 1979, p.459), se $\hat{\Sigma}_i = \mathbf{Z}_i \hat{\Gamma} \mathbf{Z}'_i + \hat{\Omega}_i$, então

$$\hat{\Sigma}_i^{-1} = \hat{\Omega}_i^{-1} - \hat{\Omega}_i^{-1} \mathbf{Z}_i (\mathbf{Z}'_i \hat{\Omega}_i^{-1} \mathbf{Z}_i + \hat{\Gamma}^{-1})^{-1} \mathbf{Z}'_i \hat{\Omega}_i^{-1} \quad (5.2.1.27)$$

Logo,

$$\sum_{i=1}^n \mathbf{X}'_i \hat{\Sigma}_i^{-1} \mathbf{X}_i \hat{\beta} = \sum_{i=1}^n \mathbf{X}'_i \hat{\Sigma}_i^{-1} y_i \Rightarrow \hat{\beta} = \left(\sum_{i=1}^n \mathbf{X}'_i \hat{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}'_i \hat{\Sigma}_i^{-1} y_i \right)$$

Pode-se simplificar a expressão (5.2.1.25.b), utilizando o fato de que o termo $(\hat{\Gamma}^{-1} + \mathbf{Z}'_i \hat{\Omega}_i^{-1} \mathbf{Z}_i)^{-1}$ pode ser escrito como

$$\hat{\Gamma} - \hat{\Gamma} \mathbf{Z}'_i (\mathbf{Z}_i \hat{\Gamma} \mathbf{Z}'_i + \hat{\Omega}_i)^{-1} \mathbf{Z}_i \hat{\Gamma} = \hat{\Gamma} - \hat{\Gamma} \mathbf{Z}'_i \hat{\Sigma}_i^{-1} \mathbf{Z}_i \hat{\Gamma}.$$

Assim,

$$\hat{b}_i = (\hat{\Gamma} - \hat{\Gamma} \mathbf{Z}'_i \hat{\Sigma}_i^{-1} \mathbf{Z}_i \hat{\Gamma}) \mathbf{Z}'_i \hat{\Omega}_i^{-1} (y_i - \mathbf{X}_i \hat{\beta}) = (\hat{\Gamma} \mathbf{Z}'_i \hat{\Omega}_i^{-1} - \hat{\Gamma} \mathbf{Z}'_i \hat{\Sigma}_i^{-1} \mathbf{Z}_i \hat{\Gamma} \mathbf{Z}'_i \hat{\Omega}_i^{-1}) (y_i - \mathbf{X}_i \hat{\beta}).$$

Como $\hat{\Sigma}_i = \mathbf{Z}_i \hat{\Gamma} \mathbf{Z}'_i + \hat{\Omega}_i$, multiplicando-se os termos da igualdade por $\hat{\Omega}_i^{-1}$ temos:

$$\hat{\Sigma}_i \hat{\Omega}_i^{-1} = \mathbf{Z}_i \hat{\Gamma} \mathbf{Z}'_i \hat{\Omega}_i^{-1} + \hat{\Omega}_i \hat{\Omega}_i^{-1} \Rightarrow \mathbf{Z}_i \hat{\Gamma} \mathbf{Z}'_i \hat{\Omega}_i^{-1} = \hat{\Sigma}_i \hat{\Omega}_i^{-1} - \mathbf{I}$$

Logo,

$$\hat{b}_i = \left[\hat{\Gamma} \mathbf{Z}'_i \hat{\Omega}_i^{-1} - \hat{\Gamma} \mathbf{Z}'_i \hat{\Sigma}_i^{-1} (\hat{\Sigma}_i \hat{\Omega}_i^{-1} - \mathbf{I}) \right] (y_i - \mathbf{X}_i \hat{\beta}) = \hat{\Gamma} \mathbf{Z}'_i \hat{\Sigma}_i^{-1} (y_i - \mathbf{X}_i \hat{\beta}),$$

Portanto, as expressões para $\hat{\beta}$ e \hat{b}_i em função de $\hat{\Theta}_i$ podem ser expressas por:

$$\hat{\beta} = \left(\sum_{i=1}^n X_i' \hat{\Sigma}_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^n X_i' \hat{\Sigma}_i^{-1} y_i \right) \quad (5.2.1.28)$$

e

$$\hat{b}_i = \hat{\Gamma} Z_i' \hat{\Sigma}_i^{-1} (y_i - X_i \hat{\beta}). \quad (5.2.1.29)$$

Alguns autores (Laird e Ware, 1982; Rao, 1975; Rosenberg, 1973) chamam o estimador \hat{b}_i de Bayes Empírico de b_i quando se considera $\hat{\Theta}_i$ como valor verdadeiro. Sem aprofundarmos na abordagem Bayesiana, o estimador Bayes Empírico conta com a existência de uma distribuição a priori de b_i , em particular $N(0, \Gamma)$, e tem a forma $E(b_i | y_i, \hat{\beta}, \hat{\Theta}_i)$.

Resta desenvolver $\frac{\partial l}{\partial \Theta_i}$ e fazer $\left. \frac{\partial l}{\partial \Theta_i} \right|_{\beta = \hat{\beta}, b_i = \hat{b}_i, \Theta_i = \hat{\Theta}_i} = 0$.

Como $\Theta_i = \{\gamma \in \Gamma \text{ e } \omega \in \Omega_i\}$ temos

$$\frac{\partial l}{\partial \gamma} = \sum_{i=1}^n \text{tr} \left(\Gamma^{-1} \frac{\partial \Gamma}{\partial \gamma} \right) + b_i' \Gamma^{-1} \frac{\partial \Gamma}{\partial \gamma} \Gamma^{-1} b_i, \quad (5.2.1.30)$$

e

$$\frac{\partial l}{\partial \omega} = \sum_{i=1}^n \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \omega} \right) + u_i' \Omega_i^{-1} \frac{\partial \Omega_i}{\partial \omega} \Omega_i^{-1} u_i, \quad (5.2.1.31)$$

onde $u_i = (y_i - X_i \hat{\beta} - Z_i \hat{b}_i)$.

Vimos que, condicionado a $\hat{\Theta}_i$, temos soluções analíticas fechadas para estimar β . Este fato sugere a utilização de um método iterativo para encontrar os EMV dos parâmetros do modelo. Numericamente, a otimização de l em função dos parâmetros é iterativa. Cada iteração consiste, essencialmente, de dois passos. No primeiro, estima-se os efeitos do modelo em função de um conjunto Θ_i fixo (calculada em um estágio anterior da iteração); este passo é dado por (5.2.1.23.a). O segundo passo, dado por (5.2.1.23.c) ou por (5.2.1.30) e (5.2.1.31), é equivalente à maximização da verossimilhança condicionada às últimas estimativas de β e b . Assim, no segundo passo, os valores de Θ_i são estimados em função das estimativas obtidas no primeiro passo. Tais valores são aplicados no primeiro passo e assim repete-se o algoritmo. Os detalhes sobre algoritmos baseados em verossimilhança serão discutidos na seção 5.3. Entre os algoritmos que serão estudados estão o algoritmo EM, o *scoring*, o método de Newton-Raphson e o filtro de Kalman.

5.2.1.c. Propriedades dos Estimadores de MV

Quando Θ_i é conhecido, pode-se obter com facilidade as respectivas variâncias de $\hat{\beta}$ e \hat{b}_i , que são funções lineares de y_i :

$$\begin{aligned} \text{var}(\hat{\beta}) &= \text{var}\left[\left(\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} y_i\right)\right] = \\ &= \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i\right)^{-1} \sum_{i=1}^n \left[X_i' \Sigma_i^{-1} \text{var}(y_i) \Sigma_i^{-1} X_i\right] \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i\right)^{-1} = \\ &= \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i\right)^{-1} \sum_{i=1}^n \left[X_i' \Sigma_i^{-1} X_i\right] \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i\right)^{-1}. \end{aligned}$$

Portanto,

$$\text{var}(\hat{\beta}) = \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i\right)^{-1} \quad (5.2.1.32)$$

$$\begin{aligned} \text{var}(\hat{b}_i) &= \text{var}\left[\Gamma Z_i' \Sigma_i^{-1} (y_i - X_i \hat{\beta})\right] = \Gamma Z_i' \Sigma_i^{-1} \text{var}\left[y_i - X_i \hat{\beta}\right] \Sigma_i^{-1} Z_i \Gamma = \\ &= \Gamma Z_i' \Sigma_i^{-1} \left[\text{var}(y_i) + \text{var}(X_i \hat{\beta}) - \text{cov}(y_i, X_i \hat{\beta}) - \text{cov}(y_i, X_i \hat{\beta})'\right] \Sigma_i^{-1} Z_i \Gamma. \end{aligned}$$

Desenvolvendo a covariância para Θ_i conhecido, obtemos:

$$\begin{aligned} \text{cov}(y_i, X_i \hat{\beta}) &= \text{cov}\left[y_i, X_i \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} y_i\right)\right] = \\ &= \text{cov}\left[y_i, X_i \text{var}(\hat{\beta}) \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} y_i\right)\right] = \text{cov}\left[y_i, X_i \text{var}(\hat{\beta}) X_i' \Sigma_i^{-1} y_i\right] = \\ &= \text{cov}\left[y_i, y_i\right] \Sigma_i^{-1} X_i \text{var}(\hat{\beta}) X_i' = \Sigma_i \Sigma_i^{-1} X_i \text{var}(\hat{\beta}) X_i' = X_i \text{var}(\hat{\beta}) X_i'. \end{aligned}$$

Logo,

$$\begin{aligned} \text{var}(\hat{b}_i) &= \Gamma Z_i' \Sigma_i^{-1} \left[\Sigma_i + X_i \text{var}(\hat{\beta}) X_i' - X_i \text{var}(\hat{\beta}) X_i' - X_i \text{var}(\hat{\beta}) X_i'\right] \Sigma_i^{-1} Z_i \Gamma = \\ &= \Gamma Z_i' \Sigma_i^{-1} \Sigma_i \Sigma_i^{-1} Z_i \Gamma - \Gamma Z_i' \Sigma_i^{-1} X_i \text{var}(\hat{\beta}) X_i' \Sigma_i^{-1} Z_i \Gamma. \end{aligned}$$

Portanto,

$$\text{var}(\hat{b}_i) = \Gamma Z_i' \left[\Sigma_i^{-1} - \Sigma_i^{-1} X_i \text{var}(\hat{\beta}) X_i' \Sigma_i^{-1}\right] Z_i \Gamma. \quad (5.2.1.33)$$

Para avaliar quão bom é o estimador \hat{b}_i de b_i , deve-se utilizar a variação em $\hat{b}_i - b_i$:

$$\text{var}(\hat{b}_i - b_i) = \text{var}(\hat{b}_i) + \text{var}(b_i) - \text{cov}(\hat{b}_i, b_i) - \text{cov}(b_i, \hat{b}_i),$$

onde

$$\begin{aligned} \text{cov}(\hat{b}_i, b_i) &= \text{cov}\left[\Gamma Z_i' \Sigma_i^{-1} (y_i - X_i \hat{\beta}), b_i\right] = \Gamma Z_i' \Sigma_i^{-1} \text{cov}\left[(y_i - X_i \hat{\beta}), b_i\right] = \\ &= \Gamma Z_i' \Sigma_i^{-1} \left[\text{cov}(y_i, b_i) - \text{cov}(X_i \hat{\beta}, b_i) \right] = \\ &= \Gamma Z_i' \Sigma_i^{-1} \left[\text{cov}(X_i \beta + Z_i b_i + \varepsilon_i, b_i) - \text{cov}(X_i \text{var}(\hat{\beta}) X_i' \Sigma_i^{-1} y_i, b_i) \right] = \\ &= \Gamma Z_i' \Sigma_i^{-1} \left[Z_i \Gamma - X_i \text{var}(\hat{\beta}) X_i' \Sigma_i^{-1} \text{cov}(y_i, b_i) \right] = \\ &= \Gamma Z_i' \Sigma_i^{-1} \left[Z_i \Gamma - X_i \text{var}(\hat{\beta}) X_i' \Sigma_i^{-1} Z_i \Gamma \right] = \text{var}(\hat{b}_i). \end{aligned}$$

Portanto, a variação em $\hat{b}_i - b_i$ é dada por

$$\text{var}(\hat{b}_i - b_i) = \Gamma - \text{var}(\hat{b}_i) \quad (5.2.1.34)$$

Quando as matrizes de covariância forem desconhecidas, mas dispendo-se de $\hat{\Theta}$, as estimativas para $\text{var}(\hat{\beta})$ e $\text{var}(\hat{b}_i)$ podem ser obtidas substituindo-se Θ por $\hat{\Theta}$ nas equações (5.2.1.32) a (5.2.1.34). Essas matrizes de covariância estimadas podem ser usadas em testes de hipóteses e intervalos de confiança que serão discutidos no próximo capítulo.

Quando $n \rightarrow \infty$, a matriz $\hat{\text{var}}(\hat{\beta}) = (\sum_{i=1}^n X_i' \hat{\Sigma}_i^{-1} X_i)^{-1}$ converge em probabilidade para $\text{var}(\beta)$, i.e., $\hat{\text{var}}(\hat{\beta}) \xrightarrow{p} \text{var}(\beta)$, e a distribuição de $\hat{\beta}$ converge à distribuição Normal com parâmetros β e $\text{var}(\beta)$, i.e., $\hat{\beta} \xrightarrow{D} N(\beta, \text{var}(\beta))$. Em outras palavras, os testes e intervalos de confiança para $\hat{\beta}$, em grandes amostras, poderão ser baseados na aproximação da distribuição $N(\beta, \hat{\text{var}}(\hat{\beta}))$ para a distribuição de $\hat{\beta}$ (Crowder e Hand, 1990, pp.90).

5.2.2. Estimador de Máxima Verossimilhança Restringida

Em modelos componentes de variância com dados balanceados, as estimativas de máxima verossimilhança não levam em consideração os graus de liberdade perdidos na estimação do efeito fixo, resultando em estimação viciada (Patterson e Thompson, 1971; Harville, 1977). O fato dos EMV diferirem dos estimadores de Análise de Variância (ANOVA) no caso balanceado é desvantajosa, pois estes são BQUE (*Best Quadratic Unbiased Estimators*) - melhor estimador quadrático não viciado - e, se houver normalidade são BUE (*Best Unbiased Estimators*) - melhor estimador não viciado - (Robinson, 1987).

Os primeiros desenvolvimentos para certas situações com dados balanceados foram feitos por Anderson e Bancroft (1952) e Russel e Bradley (1958), ambos citados por Searle *et alii.* (1992, pp. 250); o caso balanceado em geral foi estendido por Thompson (1962). Para o caso de modelos mistos em geral, o problema do vício pode ser resolvido através de uma modificação do método de máxima verossimilhança sugerida por Patterson e Thompson (1971). Esse método tem recebido muitas formas de descrições na literatura, desde considerações sobre estimativas negativas até "*maximizing that part of the likelihood which is invariant to the fixed effects*" (Searle *et alii.*, 1992, pp. 250). Por causa dessa variedade de descrições há muitos termos associados ao método como máxima verossimilhança modificada, restringida, residual ou marginal. O fato de ser restringida se deve à maximização da função de verossimilhança não de todos os dados, mas de um conjunto de contrastes de erro selecionados (que têm esperança zero). Outros o chamam de residual porque os resíduos são usados no procedimento de estimação. O termo marginal vem da maximização da verossimilhança marginal como descrito por Searle *et alii.* (1992, pp. 337).

Nessa modificação do método de máxima verossimilhança, a função de verossimilhança é fatorada em duas partes, sendo que uma delas é totalmente livre dos efeitos fixos. A otimização dessa parte, com respeito aos parâmetros restantes produz os Estimadores de Máxima Verossimilhança Restringida ou Residual (EMVRe) que, no caso balanceado, são os estimadores não viciados da ANOVA (Laird e Ware, 1982; Robinson, 1987). A outra parte é otimizada com respeito aos efeitos fixos, levando-se em conta os EMVRe. Corbeil e Searle (1976) aplicaram a MVRe na estimação do modelo componente de variância. Mesmo no caso de dados não balanceados, que foi considerado por Patterson e Thompson (1971), Thompson (1977) e Lin e McAllister (1984), os EMVRe são geralmente preferíveis se um número razoavelmente grande de graus de liberdade forem necessários para os efeitos fixos. Uma discussão sobre a comparação da MV com MVRe é feita no final dessa subseção.

A fatoração da verossimilhança pode ser feita através de uma transformação linear dos dados originais.

Lembrando-se que da seção 5.2.1 temos, matricialmente, $y \sim N(X\beta, \Sigma)$,

$$\text{onde } X_{T \times a} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \text{ e } \Sigma_{T \times T} = \text{diag}(\Sigma_1, \dots, \Sigma_n), \text{ com } T = \sum_{i=1}^n T_i,$$

seja

$$y^* = P y,$$

onde P e y^* são particionadas como

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \text{ e } y^* = \begin{bmatrix} y_1^* \\ y_2^* \end{bmatrix},$$

e as matrizes P_1 e P_2 são tais que

(i) $\text{posto}(P_1) = T-a$ e $\text{posto}(P_2) = a$;

(ii) $P_1 y$ e $P_2 y$ são estatisticamente independentes, i.e., $\text{cov}(P_1 y, P_2 y) = 0$.

Isso é satisfeito se $P_1 \Sigma P_2' = 0$;

(iii) a matriz P_1 representa um conjunto de contrastes, tal que $E(P_1 y) = 0$, i.e., $P_1 X = 0$;

(iv) a matriz $P_2 X$ tem posto a , tal que todas as funções lineares dos elementos de $P_2 y$ estimem uma função linear dos elementos de β .

Então resulta que $y^* \sim N(PX\beta, P\Sigma P')$,

$$\text{onde } PX\beta = \begin{bmatrix} 0 \\ P_2 X\beta \end{bmatrix} \text{ e } P\Sigma P' = \text{diag}(P_1 \Sigma P_1', P_2 \Sigma P_2').$$

Como $P_1 y$ e $P_2 y$ são independentes, a função $l^* = -2 \ln(\text{verossimilhança})$ é a soma de dois termos,

$$l^* = l_1 + l_2, \quad (5.2.2.1)$$

onde l_1 e l_2 são funções das verossimilhanças associadas, respectivamente, a $P_1 y$ e $P_2 y$.

Patterson e Thompson (1971) propuseram a transformação com P formada por

$$P_1 = [I - X(X'X)^{-1} X'] \text{ e } P_2 = X' \Sigma^{-1}. \quad (5.2.2.2)$$

Neste caso, P_1 é matriz simétrica $T \times T$ e idempotente com posto $T-a$. Essa transformação é singular, pois $|P_1 \Sigma P_1| = 0$. Assim, a expressão para I_1 deverá ser obtida através da distribuição Normal singular, que é expressa em termos da inversa generalizada $(P_1 \Sigma P_1)^-$ e dos autovalores não nulos de $P_1 \Sigma P_1$.

Para evitar a singularidade, há alternativas para P_1 que resultam em $|P_1 \Sigma P_1| \neq 0$. Uma delas é tomar como P_1 um conjunto qualquer de $T-a$ linhas linearmente independentes da matriz P dada em 5.2.1.

Outra forma é tomar uma decomposição de $I - X(X'X)^{-1}X'$. Como essa matriz é simétrica e idempotente, pode-se escrevê-la na forma AA' , onde A é matriz $T \times (T-a)$ e $A'A = I$. Tomando-se P formada por

$$P_1 = A' \text{ e } P_2 = X' \Sigma^{-1}, \quad (5.2.2.3)$$

resulta que $y^* \sim N \left(\begin{bmatrix} 0 \\ X' \Sigma^{-1} X \beta \end{bmatrix}, \begin{bmatrix} A' \Sigma A & 0 \\ 0 & X' \Sigma^{-1} X \end{bmatrix} \right)$,

pois $A'X = A'AA'X = A'[I - X(X'X)^{-1}X']X = 0$.

Logo, a função de verossimilhança, por (5.2.2.1), é a soma de

$$I_1 = (T-a) \ln 2\pi + \ln |A' \Sigma A| + y' A (A' \Sigma A)^{-1} A' y \quad (5.2.2.4)$$

com

$$I_2 = a \ln 2\pi + \ln |X' \Sigma^{-1} X| + (y - X\beta)' \Sigma^{-1} X (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} (y - X\beta) \quad (5.2.2.5)$$

O termo I_1 independe do efeito fixo β e sua otimização com respeito a Θ produz os EMVRe, $\hat{\Theta}_{MVRe}$. A outra parte, I_2 , é otimizada com respeito a β , considerando-se $\hat{\Theta}_{MVRe}$, ou seja,

$$\begin{aligned} \frac{\partial I_2}{\partial \beta} \Big|_{\beta = \hat{\beta} \text{ e } \Theta = \hat{\Theta}_{MVRe}} &= 0 \Rightarrow \\ &\Rightarrow (X' \hat{\Sigma}^{-1} X) (X' \hat{\Sigma}^{-1} X)^{-1} X' \hat{\Sigma}^{-1} (y - X\hat{\beta}) = 0 \\ &\Rightarrow X' \hat{\Sigma}^{-1} X \hat{\beta} = X' \hat{\Sigma}^{-1} y \\ &\Rightarrow \hat{\beta} = \left(\sum_{i=1}^n X_i' \hat{\Sigma}_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^n X_i' \hat{\Sigma}_i^{-1} y_i \right). \end{aligned}$$

Nota-se que a expressão para $\hat{\beta}$ é a mesma da MV (5.2.1.28). A diferença entre MV e MVRe está na estimação de Θ .

Antes de diferenciar I_1 , é possível simplificá-lo. Como P , dada por (5.2.2.3) é matriz não-singular, pois tem posto completo, a verossimilhança de y é igual a de y^* . Os termos

exponenciais das verossimilhanças dadas por (5.2.1.4), (5.2.2.4) e (5.2.2.3), portanto, são iguais, i.e.,

$$(y - X\beta)' \Sigma^{-1} (y - X\beta) = \\ = y' A (A' \Sigma A)^{-1} A' y + (y - X\beta)' \Sigma^{-1} X (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} (y - X\beta), \quad (5.2.2.6)$$

onde o termo do lado esquerdo da igualdade vem de (5.2.1.4) e os termos do lado direito vêm das verossimilhanças "fatoradas" (5.2.2.4) e (5.2.2.3).

Logo,

$$A (A' \Sigma A)^{-1} A' = \Sigma^{-1} [I - X (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1}] \\ = \Sigma^{-1} M, \quad (5.2.2.7)$$

onde $M = I - X (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1}$. Substituindo-se (5.2.2.7) em (5.2.2.4) obtemos

$$l_1 = (T - a) \ln 2\pi + \ln |A' \Sigma A| + y' \Sigma^{-1} M y. \quad (5.2.2.8)$$

Diferenciando-se o termo $\ln |A' \Sigma A|$ com respeito a Θ , obtemos

$$\frac{\partial \ln |A' \Sigma A|}{\partial \Theta} = \text{tr} \left[(A' \Sigma A)^{-1} A' \frac{\partial \Sigma}{\partial \Theta} A \right] = \text{tr} \left[A (A' \Sigma A)^{-1} A' \frac{\partial \Sigma}{\partial \Theta} \right] = \text{tr} \left[\Sigma^{-1} M \frac{\partial \Sigma}{\partial \Theta} \right].$$

Logo,

$$\frac{\partial l_1}{\partial \Theta} = \text{tr} \left[\Sigma^{-1} M \frac{\partial \Sigma}{\partial \Theta} \right] - \frac{\partial (y' \Sigma^{-1} M y)}{\partial \Theta}. \quad (5.2.2.9)$$

Como ilustração, considere o seguinte

Exemplo 5.2.2 estrutura de covariância conhecida

Tomando-se o exemplo 5.2.1, onde $\Sigma_i = \sigma^2 (Z_i B Z_i' + W_i) = \sigma^2 G_i$, com B e W_i conhecidos e $\Theta = \sigma^2$, seja $G = \text{diag}(G_1 \cdots G_n)$.

Então, por (5.2.2.8), o termo l_1 pode ser escrito como

$$l_1 = (T - a) \ln 2\pi + (T - a) \ln \sigma^2 + \ln |A' G A| + \sigma^{-2} y' G^{-1} M y;$$

observa-se que o termo l_2 está implícito na função l_1 por meio de $G^{-1} M$.

Por (5.2.2.9),

$$\frac{\partial l_1}{\partial \sigma^2} = \frac{T-a}{\sigma^2} - \frac{y'G^{-1}My}{\sigma^4}.$$

Assim,

$$\left. \frac{\partial l_1}{\partial \sigma^2} \right|_{\sigma^2 = \hat{\sigma}_{MVR}^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{y'G^{-1}My}{T-a} = \frac{y'M'G^{-1}My}{T-a}.$$

Como $My = (I - X(X'G^{-1}X)^{-1}X'G^{-1})y$ e, se tomarmos o estimador de β dado por (5.2.1.28), então teremos $My = y - X\hat{\beta}$. Logo a EMVRe de σ^2 é dada por

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - X_i\hat{\beta})'G_i^{-1}(y_i - X_i\hat{\beta})}{T-a} \quad (5.2.2.10)$$

□

É possível reescrever l_1 como função de l (5.2.1.7), $\hat{\beta}$ e Θ , i.e., $l(\hat{\beta}, \Theta|y)$. Pelas propriedades de determinante, temos que

$$\begin{aligned} |PP'| &= \begin{vmatrix} A'A & A'\Sigma^{-1}X \\ X'\Sigma^{-1}A & X'\Sigma^{-1}\Sigma^{-1}X \end{vmatrix} = |I| |X'\Sigma^{-1}\Sigma^{-1}X - X'\Sigma^{-1}A I^{-1}A'\Sigma^{-1}X| = \\ &= |X'\Sigma^{-1}\Sigma^{-1}X - X'\Sigma^{-1}[I - X(X'X)^{-1}X']\Sigma^{-1}X| = |X'\Sigma^{-1}X|^2 |(X'X)^{-1}|. \end{aligned}$$

Como $|P\Sigma P'| = |A'\Sigma A| |X'\Sigma^{-1}X|$, então,

$$\begin{aligned} \ln|A'\Sigma A| &= \ln|P\Sigma P'| - \ln|X'\Sigma^{-1}X| = \ln|PP'| + \ln|\Sigma| - \ln|X'\Sigma^{-1}X| = \\ &= \ln|(X'X)^{-1}| + \ln|\Sigma| + \ln|X'\Sigma^{-1}X| \end{aligned} \quad (5.2.2.10)$$

A relação (5.2.2.6) pode ser reescrita como

$$\begin{aligned} (y - X\beta)' \Sigma^{-1} (y - X\beta) &= \\ &= (y - X\hat{\beta})' \Sigma^{-1} (y - X\hat{\beta}) + (\beta - \hat{\beta})' (X'\Sigma^{-1}X) (\beta - \hat{\beta}). \end{aligned} \quad (5.2.2.12)$$

Logo, l_1 pode ser expressa substituindo-se os resultados (5.2.2.11) e (5.2.2.12) em (5.2.2.8):

$$\begin{aligned}
 l_1 &= (T - a)\ln 2\pi - \ln|X'X| + \ln|\Sigma| + \ln|X'\Sigma^{-1}X| + (y - X\hat{\beta})'\Sigma^{-1}(y - X\hat{\beta}) \\
 &= l(\hat{\beta}, \Theta|y) - \ln|X'X| + \ln|X'\Sigma^{-1}X| + a\ln 2\pi
 \end{aligned}
 \tag{5.2.2.13}$$

Alternativamente, Harville (1974) desenvolve l_1 a partir de argumentos Bayesianos resultando em (5.2.2.13). O resultado (5.2.2.13) é útil pois a matriz A não precisa ser especificada explicitamente, tornado mais simples os cálculos.

Laird e Ware desenvolveram o EMVRe de Θ a partir de uma formulação Bayesiana a fim de unificar as duas abordagens.

5.2.3. Máxima Verossimilhança (MV) ou MV Restringida (MVRe)?

Em geral, a escolha entre EMV e EMVRe se deve a argumentos intuitivos (Diggle, 1988), mas possivelmente ela deve considerar a diferença básica existente entre as duas: a segunda fornece estimativas com vícios menores do que as respectivas estimativas produzidas por EMV (Jones, 1993, pp. 39; Diggle, 1988). Esse fato torna-se importante quando o número de parâmetros fixados (a) não for relativamente pequeno ao número total de observações. Ambos os métodos são baseados no mesmo princípio e têm quase as mesmas necessidades computacionais. O método MV produz diretamente os estimadores para os efeitos fixos, enquanto que no MVRe não. Entretanto, com dados balanceados, as soluções da MVRe são idênticas aos estimadores ANOVA que são estimadores de mínima variância (Searle *et alii.*, 1992, pp. 255).

Searle *et alii.* (1992, pp. 254) discutem que, para dados não balanceados, os métodos MV ou MVRe são preferíveis ao método via ANOVA. Isso porque o princípio da máxima verossimilhança possui propriedades de consistência e normalidade assintótica dos estimadores e também porque a matriz de dispersão amostral assintótica dos estimadores é conhecida. Isso facilita a construção de intervalos de confiança e de testes de hipóteses sobre os parâmetros. Em contraste, os estimadores ANOVA possuem a característica de serem não viciados; em geral, suas matrizes de dispersão amostrais são difíceis de serem derivadas.

Apesar dos EMV e EMVRe serem geralmente baseados na suposição de normalidade dos dados, essa hipótese pode ser razoavelmente aceita em muitos casos. Deve-

se observar, porém, que as propriedades assintóticas da matriz de covariância amostral serão válidas obviamente apenas para grandes (ou razoavelmente grandes) amostras.

Um estudo de simulação para comparar EMV com EMVRe em um conjunto de dados não balanceado foi feito por Lin e McAllister (1984). Nesse estudo, cada simulação compôs-se de 480 manadas, 120 reprodutores e de 5 a 100 descendentes fêmeas por reprodutor. Parâmetros típicos foram escolhidos para compor a variância do reprodutor e hereditariedade.

Após o ajuste do efeito aleatório (reprodutor) e dos erros residuais mais um efeito fixo (manada), os resultados foram comparados com os parâmetros de entrada. Para 10 simulações, os Erros Quadráticos Médios (EQM) da variância do reprodutor estimada por MVRe e MV foram semelhantes: 10,99 e 10,9, respectivamente. Entretanto, o EQM da variância residual do EMVRe foi 1,0 contra 316,7 do EMV. Mesmo simulando um padrão assimétrico no efeito aleatório (reprodutor) através de uma distribuição qui-quadrada, encontrou-se um padrão semelhante. Os EQM para a variância do reprodutor estimada por MVRe e MV foram, respectivamente, 1,1 e 1,0, mas para o termo residual, os erros foram 0,6 e 151,7, respectivamente, para MVRe e MV.

Por causa desses problemas com a EMV da variância residual, muitos pesquisadores corrigem suas estimativas para a perda dos graus de liberdade devido aos efeitos fixos. Em um modelo simples com classificação por dois fatores, manada e reprodutor, essa correção é possível e melhora o estimador. Entretanto, para muitos modelos, as correções não são particularmente simples ou possíveis de se fazer (Robinson, 1987).

Deste modo, os EMVRe das componentes de variância são geralmente preferíveis se um número razoavelmente grande de graus de liberdade forem necessários para os efeitos fixos.

5.2.4. BLUP

Considere o modelo (5.2.2.1) e que se deseja estimar uma combinação linear entre os efeitos fixos e aleatórios

$$c_1'\beta + c_2'b,$$

onde $b = (b'_1, \dots, b'_n)'$, c_1 e c_2 são vetores de constantes fixadas.

Dizemos que o estimador linear $a'y$ de $c_1'\beta + c_2'b$ é BLUP (*Best Linear Unbiased Prediction*) quando $a'y$ tiver o menor Erro Quadrático Médio, $E(a'y - c_1'\beta - c_2'b)^2$, entre os estimadores lineares $a'y$ que satisfazem $E(a'y - c_1'\beta - c_2'b) \equiv 0$.

Para o caso onde Θ é conhecido, Harville (1976) apresentou uma extensão do teorema de Gauss-Markov para incluir a estimação dos efeitos aleatórios. A demonstração desse resultado não será tratada nesse trabalho, mas pode ser encontrada em Harville (1976) e pode ser enunciada da seguinte forma:

O BLUP de uma combinação linear $c_1'\beta + c_2'b$ dos elementos de β e b , com $c_1'\beta$ estimável, é $c_1'\hat{\beta} + c_2'\hat{b}$. Os estimadores $\hat{\beta}$ e \hat{b} são as soluções do sistema

$$\begin{bmatrix} X'\Omega^{-1}X & X'\Omega^{-1}Z\Gamma \\ Z'\Omega^{-1}X & I + Z'\Omega^{-1}Z\Gamma \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\zeta} \end{bmatrix} = \begin{bmatrix} X'\Omega^{-1}y \\ Z'\Omega^{-1}y \end{bmatrix} \quad (5.2.4.1)$$

onde, $\hat{b} = \Gamma\hat{\zeta}$,

$$\hat{\zeta} = Z'\Sigma^{-1}(y - X\hat{\beta}) = Z'Py,$$

$P = \Sigma^{-1} - \Sigma^{-1}X(X'\Sigma^{-1}X)^-X'\Sigma^{-1}$; a matriz $(X'\Sigma^{-1}X)^-$ é a inversa generalizada de $X'\Sigma^{-1}X$.

O sistema (5.2.4.1) pode ser chamada de equações normais estendidas ou equações simultâneas, e representa uma versão modificada daquela obtida por Henderson (1963, citado por Harville, 1977). O resultado de Henderson se aplica para um conjunto Θ , tal que a matriz Γ seja não-singular (5.2.1.26), enquanto que o sistema (5.2.4.1) é aplicável para qualquer conjunto de parâmetros Θ e a solução de $\hat{\zeta}$ está embutida na de \hat{b} .

Quando $a = 0$ (o modelo não possui efeitos fixos), ou equivalentemente, quando β for conhecido, o enunciado essencialmente se reduz ao resultado obtido por Rao (1973, sec. 4a.11) Quando $c_2 = 0$ ele se reduz ao teorema de Gauss-Markov e quando c_1 , para um resultado obtido por Henderson (1963, citado por Harville, 1977).

Harville (1977) observa que, quando se utiliza modelos ANOVA como casos especiais de (5.2.1.1), as matrizes Ω , Γ , em geral, e possivelmente Z e X exibem estruturas relativamente simples.

Os elementos de \hat{b} pertencem a uma classe de estimadores conhecidos como *Shrinkers*. Para $a = 0$, com Θ conhecido, com dados normais multivariadas, \hat{b} é o estimador de Bayes de b . Sem normalidade, \hat{b} é o Bayes linear descrito por Hartigan (1969). Para $a > 0$ e Θ conhecido, a abordagem de Harville pode ser caracterizada como parcialmente Bayesiana. Se Θ é desconhecido, a expressão $c_1'\hat{\beta} + c_2'\hat{b}$ não pode mais ser considerada como o BLUP de $c_1'\beta + c_2'b$ (Harville, 1977).

Um procedimento factível é utilizar o sistema (5.2.4.1) utilizando algum valor de Θ . Dependendo de como esse valor é escolhido, há uma semelhança ou com *ridge regression* ou com estimadores do tipo Stein ou Bayes empírico, considerada, e.g., por Efron e Morris

(1973). Em particular, uma prática comum é utilizar uma EMV ou EMVRe como valor de Θ para calcular $c_1'\hat{\beta} + c_2'\hat{b}$. Essa expressão, obtida a partir de $\hat{\Theta}$, é chamada de BLUP Empírico. Harville e Carriquiry (1992), utilizando o modelo (5.2.1.1) com $\Gamma = \sigma_b^2 \mathbf{I}$ e $\Omega = \sigma_e^2 \mathbf{I}$, apontaram algumas deficiências dessa prática:

- a) quando a estimativa da razão $\rho = \sigma_b^2 / \sigma_e^2$ for zero, ou próximo de zero, os preditores BLUP empírico não têm uma forma atraente. Em outras palavras, quando $\hat{\rho} = 0$, a abordagem BLUP empírico para prever o valor de $a'y$ considera que $b = 0$, ou equivalentemente, que y segue o modelo de efeitos fixos. Como a razão ρ é sempre positiva, podemos considerar que, com uma estimativa de ρ próxima de zero, o verdadeiro valor de ρ "provavelmente" seja maior que o valor estimado;
- b) o Erro Quadrático Médio do estimador BLUP empírico tende a subestimar o EQM do BLUP;
- c) as propriedades da distribuição se alteram, de modo que os testes e intervalos de predição se alteram e, portanto, não podem ser construídos sob as mesmas suposições do BLUP.

Muitas modificações têm sido propostas ao BLUP empírico para aliviar essas deficiências e uma alternativa Bayesiana foi proposta por Harville e Carriquiry (1992). Essas modificações não serão discutidas nesse trabalho, salvo aquelas que se referem à construção de intervalos de predição que serão discutidos no capítulo 6.

5.2.5. Equações Gerais de Estimação

As Equações Gerais de Estimação (GEE) estão relacionadas com o método de quase-verossimilhança considerado por Wedderburn (1974) e McCullagh e Nelder (1983). Para descrever rapidamente a quase-verossimilhança, considere o vetor de observações y_i do i -ésimo indivíduo com sua respectiva média μ_i e matriz de covariância Σ_i . Cada μ_i é função do parâmetro β . Uma abordagem para a estimação de "máxima quase-verossimilhança" (EMQV) é via mínimos quadrados. O Estimador de Mínimos Quadrados (EMQ) $\hat{\beta}$ é obtido minimizando-se a forma quadrática

$$\sum_{i=1}^n (y_i - \mu_i)' \Sigma_i^{-1} (y_i - \mu_i). \quad (5.2.5.1)$$

Para Σ_i conhecido, $\hat{\beta}$ satisfaz

$$\sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right) \Sigma_i^{-1} (y_i - \mu_i) = 0, \quad (5.2.5.2)$$

que é obtido diferenciando-se $\sum_{i=1}^n (y_i - \mu_i)' \Sigma_i^{-1} (y_i - \mu_i)$ com respeito a β . No caso linear, por exemplo, $\mu_i = X_i \beta$ produz $\partial \mu_i / \partial \beta = X_i$.

Na quase-verossimilhança, considera-se que: $g(\mu_{it}) = x_{it}' \beta$, onde $g(\cdot)$ é a função de ligação e x_{it} é um vetor de variáveis explanatórias para y_{it} ; e que $\Sigma_i = \Psi^{-1} V_i$, onde Ψ é o parâmetro de escala (possivelmente desconhecido) e V_i é função de μ_i . Agora, a equação (5.2.5.2) não equivale mais à minimização de (5.2.5.1), pois a diferenciação deveria introduzir um termo extra envolvendo $\partial \Sigma_i / \partial \beta$. A equação (5.2.5.2) é utilizada para definir o EMQV $\tilde{\beta}$, que é a solução de (5.2.5.2). No caso unidimensional ($t = 1$) a equação (5.2.5.2) pode ser identificada com as equações de verossimilhança que correspondem a uma distribuição da família exponencial.

O emprego das Equações Gerais de Estimação (GEE) em dados longitudinais foi discutido por Liang e Zeger (1986), Zeger e Liang (1986) e Zeger, Liang e Albert (1988). Nesse caso, cada observação é considerada pertencente à família exponencial. Essas equações são como uma extensão da quase-verossimilhança de resposta univariada para respostas multivariadas, quando não se especifica completamente o segundo momento em termos da esperança. Nas GEE, não é necessário especificar completamente a distribuição conjunta das observações individuais; especifica-se somente a verossimilhança para as distribuições marginais e uma matriz de correlação individual, chamada de *working correlation matrix*.

Reconsidere o modelo linear generalizado apresentado no capítulo 2, i.e., seja y_i o vetor $T_i \times 1$ das variáveis resposta e X_i a matriz $T_i \times a$ de covariáveis para o i -ésimo indivíduo. A densidade marginal de y_{it} é

$$f(y_{it}) = \exp\left\{ \left[y_{it} \lambda_{it} - a(\lambda_{it}) + b(y_{it}) \right] \psi_{it} \right\}, \quad (5.2.5.3)$$

onde $\lambda_{it} = h(\eta_{it})$ e $\eta_{it} = x_{it}' \beta$. Por essa formulação, $h(\cdot) = g^{-1}(\cdot)$, o parâmetro de escala ψ_{it} pode variar no tempo e os dois primeiros momentos de y_{it} são dados por $E(y_{it}) = \mu_{it} = a'(\lambda_{it})$ e $\text{var}(y_{it}) = \frac{a''(\lambda_{it})}{\psi_{it}}$. Sob normalidade, por exemplo, $\mu_{it} = \lambda_{it}$ e $\text{var}(y_{it}) = \frac{1}{\psi_{it}}$.

Na especificação parcial do modelo proposto por Liang e Zeger (1986) utiliza-se uma função de ligação $g(\cdot)$ e uma função de variância V_i , tal que $\sigma_{ij}^2 = (\Sigma_i)_{ij} = \Psi^{-1} V_i$. O que diferencia essa abordagem com a quase-verossimilhança é que não se assume conhecimento preciso das covariâncias $(\Sigma_i)_{jk}$. Somente os momentos marginais das observações y_{ij} são explicitamente modelados (média e variância). Seja $R_i(\alpha)$ uma matriz simétrica $T_i \times T_i$, de modo que represente uma matriz de correlação de y_i e seja α um vetor $s \times 1$ que caracterize

$R_i(\alpha)$. Essa matriz é chamada de *working correlation matrix*. Embora os tempos de observações e as matrizes de correlação possam diferir entre os indivíduos, assumamos que α seja o mesmo para todos os indivíduos. Defina

$$V_i = (A_i \Psi_i^{-1})^{\frac{1}{2}} R_i(\alpha) (A_i \Psi_i^{-1})^{\frac{1}{2}}, \quad (5.2.5.4)$$

onde a matriz $(T_i \times T_i)$ $A_i = \text{diag}\{a''(\lambda_{it})\}$, $\Psi_i = \text{diag}\{\psi_1 \cdots \psi_{T_i}\}$ e $V_i = \Sigma_i = \text{cov}(y_i)$ se $R_i(\alpha)$ for a matriz de correlação verdadeira de y_i .

As equações generalizadas de estimação (Liang e Zeger, 1986, e Zeger e Liang, 1986) são

$$\sum_{i=1}^n U_i(\beta, \alpha) = 0, \quad (5.2.5.5)$$

onde as funções escores $U_i(\beta, \alpha)$ são definidas como

$$U_i(\beta, \alpha) = D_i' V_i^{-1} S_i,$$

com $D_i = \frac{\partial \mu_i}{\partial \beta} = A_i \Delta_i X_i$, para $\mu_i = (\mu_{i1} \cdots \mu_{iT_i})'$;

$$\Delta_i = \text{diag}\left\{\frac{\partial \lambda_{it}}{\partial \eta_{it}}\right\}, \text{ e}$$

$$S_i = y_i - \mu_i.$$

Para estimar β é necessário representar (5.2.5.5) em função unicamente de β . Isso é possível, primeiro, substituindo-se α por um estimador $n^{\frac{1}{2}}$ -consistente de α , $\hat{\alpha}(y, \beta, \Psi)$, e depois substituir Ψ em $\hat{\alpha}(y, \beta, \Psi)$ por um estimador $n^{\frac{1}{2}}$ -consistente $\hat{\Psi}(y, \beta)$. Conseqüentemente, para um dado $R_i(\alpha)$, a estimativa $\hat{\beta}_G$ de β é definida como a solução de

$$\sum_{i=1}^n U_i[\beta, \hat{\alpha}(\beta, \hat{\Psi}(\beta))] = 0. \quad (5.2.5.6)$$

Assim, dadas as estimativas de $R_i(\alpha)$ e Ψ , a estimativa de β é a solução de (5.2.5.6). O processo de estimação é feito iterativamente, utilizando uma modificação do método de *scoring* de Fisher para estimar β e os estimadores de α e Ψ como funções de β (Liang e Zeger, 1986).

Sob condições moderadas de regularidade, Liang e Zeger (1986, teorema 2) mostraram que quando $n \rightarrow \infty$, $\hat{\beta}_G$ é um estimador consistente de β e que $n^{1/2}(\hat{\beta}_G - \beta)$ é assintoticamente Normal com média zero e matriz de covariância dada por

$$V_G = \lim_{n \rightarrow \infty} (\sum_{i=1}^n D_i' V_i^{-1} D_i)^{-1} (\sum_{i=1}^n D_i' V_i^{-1} \Sigma_i V_i^{-1} D_i)^{-1} (\sum_{i=1}^n D_i' V_i^{-1} D_i)^{-1} \quad (5.2.5.7)$$

A variância assintótica de $\hat{\beta}_G$ não depende da escolha dos estimadores de α entre aqueles que são $n^{1/2}$ -consistentes. Ele depende, entretanto, dos parâmetros de escala Ψ , a menos que $\psi = \psi_t, \forall t$.

Na estimação via GEE, não é necessário especificar corretamente $R_i(\alpha)$, i.e., não se assume conhecimento preciso da covariância Σ , especificando-se apenas os momentos marginais de y_{it} (média e variância). Para se obter estimativas consistentes e assintoticamente normais de β e V_G , requer-se que α e Ψ sejam estimados consistentemente e que

$$\frac{1}{n} \sum_{i=1}^n D_i' V_i^{-1} D_i \quad (5.2.5.8)$$

seja convergente quando $n \rightarrow \infty$. Logo, não é necessário que as observações dos indivíduos tenham a mesma estrutura de correlação. Entretanto, na presença de observações perdidas essa propriedade somente é válida se o número de dados perdidos for pequeno ou se houver completa casualidade na perda dos dados (Liang e Zeger, 1986).

Para um dado valor de β , α e Ψ podem ser estimados a partir dos resíduos de Pearson, definidos como

$$\hat{r}_{it} = \frac{(y_{it} - \hat{\mu}_{it})}{\sqrt{a''(\hat{\lambda}_{it})}} \quad (5.2.5.9)$$

onde $\hat{\mu}_{it} = a'(\hat{\lambda}_{it})$ e $\hat{\lambda}_{it}$ depende do valor corrente de β . Então,

$$\hat{\Psi}_t^{-1} = \sum_{i=1}^n \frac{\hat{r}_{it}^2}{n_t} \quad (5.2.5.10)$$

onde n_t é o número de indivíduos no tempo t . Park (1993) observa que, a divisão de (5.2.5.10) por n_t ao invés de n , permite que se obtenha estimativas consistentes e assintoticamente normais de β e V_G sem se especificar corretamente $R_i(\alpha)$.

Quando o parâmetro de escala é considerado fixo, i.e., $\psi = \psi_t, \forall t$, a estimativa do parâmetro comum é dada por

$$\hat{\Psi}^{-1} = \sum_{i=1}^n \sum_{t=1}^{T_i} \frac{\hat{r}_i^2}{T - a}, \quad (5.2.5.11)$$

onde $T = \sum_{i=1}^n T_i$.

Para se estimar α consistentemente, deve-se levar em consideração os n indivíduos. Como será visto nos próximos exemplos, o estimador específico depende da escolha de $R_i(\alpha)$. Quando $\psi = \psi_t, \forall t$, a forma geral é estimar α como uma função de

$$\hat{R}_{uv} = \sum_{i=1}^n \frac{\hat{r}_{iu} \hat{r}_{iv}}{T - a}. \quad (5.2.5.12)$$

Existem muitas possibilidades de escolha para $R_i(\alpha)$ e cada uma leva a uma análise distinta. Nos exemplos que se seguem nota-se que os parâmetros Ψ e o estimador de α variam de caso a caso. Nos exemplos 5.2.5.1 ao 5.2.5.5, considere que, para todos os indivíduos, $T = T_i$ e $\psi = \psi_t, \forall t$:

Exemplo 5.2.5.1 matriz $R(\alpha)$ fixada

Seja $R(\alpha) = R_0$. O caso mais simples é tomar $R_0 = I$, onde I é matriz identidade $n_i \times n_i$. Nesse caso, as observações repetidas do i -ésimo indivíduo não são correlacionadas. Entretanto, para qualquer matriz de correlação R_0 fixada, $\hat{\beta}_G$ e \hat{V}_G serão consistentes. Obviamente, quanto mais próxima for a escolha de R_0 da matriz de correlação verdadeira, maior será a eficiência do estimador. □

Exemplo 5.2.5.2 estrutura m -dependente

Seja $\alpha = (\alpha_1 \dots \alpha_{T-1})'$, onde $\alpha_t = \text{corr}(y_{i,t}, y_{i,t+1})$, para $t = 1, \dots, T-1$. Dados β e ψ , é natural estimar α por

$$\hat{\alpha}_t = \psi \sum_{j=1}^n \frac{\hat{r}_{j,t} \hat{r}_{j,t+1}}{n - a}.$$

Para o caso 1-dependente seja $R(\alpha)$ matriz tridiagonal com $[R]_{t,t+1} = \alpha_t$. Não é necessário estimar ψ , pois há seu cancelamento no cálculo de V_i . Particularmente, seja $s = 1$, i.e., $\alpha_t = \alpha$, para $t = 1, \dots, T-1$. Então, o parâmetro comum pode ser estimado por

$$\hat{\alpha} = \sum_{t=1}^{T-1} \frac{\hat{\alpha}_t}{T-1}$$

□

Exemplo 5.2.5.3 estrutura do modelo de efeitos aleatórios

Seja $s = 1$ e assumamos que $\alpha_t = \text{corr}(y_{it}, y_{it'})$, para todo $t \neq t'$. Esta é a mesma estrutura de correlação do modelo de efeitos aleatórios com um nível aleatório para cada indivíduo. Dado ψ , α pode ser estimado como

$$\hat{\alpha}_t = \psi \sum_{i=1}^n \sum_{t > t'} \frac{\hat{r}_{it} \hat{r}_{it'}}{D}$$

onde $D = \sum_{i=1}^n \frac{T_i(T_i-1)}{2}$. Como nos dois exemplos anteriores, não é necessário estimar ψ para se obter $\hat{\beta}_G$ e \hat{V}_G .

Com essa suposição, é possível haver um número arbitrário de observações e tempos de observações para cada indivíduo.

□

Exemplo 5.2.5.4 estrutura AR(1) contínuo

Seja $\text{corr}(y_{it}, y_{it'}) = \alpha^{|t-t'|}$. Para y_{it} gaussiano, esta representa a estrutura de correlação de um processo AR(1) contínuo. Como $E(\hat{r}_{it} \hat{r}_{it'}) = \alpha^{|t-t'|}$, podemos estimar α através do coeficiente de inclinação da regressão de $\ln(\hat{r}_{it} \hat{r}_{it'})$ em $\ln(|t-t'|)$. Note que um número arbitrário de espaçamentos de observações podem ser acomodados, mas é necessário estimar ψ para se obter $\hat{\beta}_G$ e \hat{V}_G .

□

Exemplo 5.2.5.5 matriz R(α) não-estruturada

Seja $R(\alpha)$ totalmente não especificada, i.e., $s = \frac{T(T-1)}{2}$. Pode-se estimar R por

$$\frac{\psi}{n} \sum_{i=1}^n A_i^{-\frac{1}{2}} S_i S_i' A_i^{-\frac{1}{2}}$$

Se y_i for Normal multivariada, a expressão acima reduz-se a

$$\frac{\Psi}{n} \sum_{i=1}^n S_i S_i'$$

e, juntamente com (5.2.5.6), produz-se as equações de verossimilhança. Não é necessário estimar ψ para se obter $\hat{\beta}_G$. A variância assintótica é dada por

$$V_G = \lim_{n \rightarrow \infty} n \left(\sum_{i=1}^n X_i' R_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^n X_i' R_i^{-1} S_i S_i' R_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^n X_i' R_i^{-1} X_i \right)^{-1}$$

□

Quando as variáveis de resposta y_i são normais multivariadas, uma questão de interesse é como os estimadores GEE se comparam com os EMV obtidos, por exemplo, por Jennrich e Schluchter (1986) e Laird *et alii.* (1987). A abordagem via GEE especifica apenas a distribuição marginal de y_{it} , enquanto a EMV especifica a distribuição conjunta das variáveis resposta. Na abordagem GEE, estima-se o parâmetro de escala ψ_i , a matriz $R_i(\alpha)$ e os trata como parâmetros de perturbação (*nuisance parameters*). Por outro lado, na abordagem via MV, estima-se a matriz de covariância. Park (1993) mostra que, quando os dados são completos e y_i é conjuntamente Normal multivariada, os estimadores obtidos por GEE se reduzem aos EMV, ou seja, as GEE se reduzem às equações de escore de MV. Mas quando os dados não forem completos, com completa casualidade na perda dos dados, a solução de (5.2.5.6) é um estimador consistente de β , mas pode diferir do EMV.

5.3. Algoritmos de Estimação

As equações de estimação que foram apresentadas, em geral, não produzem expressões explícitas para os estimadores. Nessa seção, abordaremos alguns procedimentos destinados aos cálculos das estimativas de MV e MVRe. Alguns os algoritmos de otimização de verossimilhança, como o Newton-Raphson, consistem de dois passos: no primeiro, fixando-se a matriz de covariância, estima-se os efeitos do modelo, e no segundo reavalia-se os valores dos parâmetros da matriz de covariância com base nas estimativas do primeiro passo. Assim, os novos valores são utilizados no primeiro passo para iniciar uma nova iteração. Como exceção, o algoritmo EM realiza a iteração de forma diferente: o primeiro passo representa o cálculo da esperança (passo E) e o segundo passo representa a maximização (passo M).

Para implementar esses métodos, pode-se utilizar algoritmos auxiliares para, por exemplo, implementar o passo da otimização, para calcular a verossimilhança, para calcular as predições, etc. Por exemplo, o método de Newton-Raphson (NR) requer o cálculo da verossimilhança a cada passo da iteração. Para evitar manipulação de matrizes que dependem do tamanho do conjunto de dados ($T_1 \times T_1$), o filtro de Kalman pode ser formulado como um algoritmo auxiliar para o cálculo da verossimilhança. O filtro de Kalman é um método recursivo e pode ser construído de modo que as dimensões das matrizes envolvidas na recursão dependam somente do número de parâmetros.

Ao se estudar os algoritmos de estimação, deve-se considerar algumas questões de interesse como:

- i) Qual método iterativo é o melhor para as equações de verossimilhança?
- ii) Qual é a melhor representação das equações para a aplicação do algoritmo?
- iii) Qual deve ser o critério de convergência? A convergência é sempre possível?
- iv) Quando o algoritmo converge o resultado obtido corresponde ao máximo global da verossimilhança?
- v) Os valores iniciais dos parâmetros influem no resultado do algoritmo? Será que existe um conjunto de valores iniciais que sempre leva o algoritmo a convergir para o máximo global?
- vi) Quais são os custos e benefícios em termos de equipamentos e tempo de execução do algoritmo?
- vii) O que deve ser feito quando a estimativa de uma matriz de covariância não for positiva definida? E se for singular (matriz não inversível)?

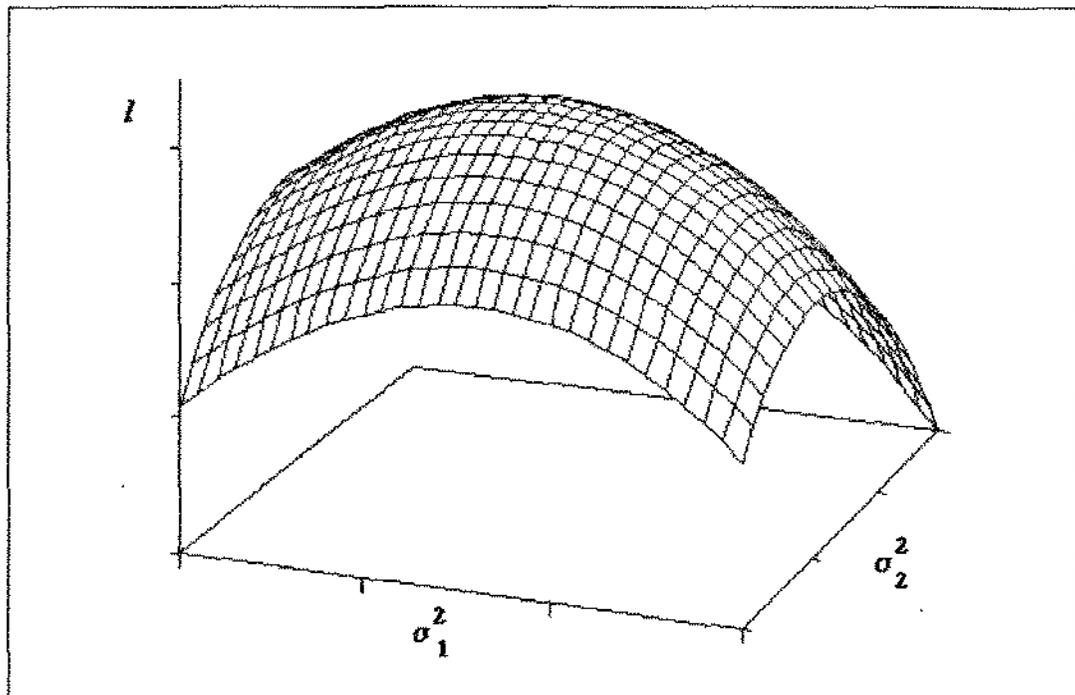


Fig. 5.3.1 - exemplo de um máximo local da função do logaritmo da verossimilhança

Esse trabalho não discute detalhadamente as questões citadas apesar de serem fundamentais para o estudo de métodos numéricos. Como as equações de MV e MVRe não produzem expressões explícitas para os estimadores, a tentativa de se maximizar o logaritmo da verossimilhança, $\ln(\text{verossimilhança})$, ou o logaritmo da verossimilhança restringida é, em geral, um problema numérico difícil. As primeiras derivadas de $\ln(\text{verossimilhança})$ é uma função não linear de parâmetros. A verossimilhança contém a inversa de uma matriz $T_i \times T_i$ para MV e $(T_i - a) \times (T_i - a)$ para MVRe, onde T_i é o número de observações do i -ésimo indivíduo e a é o número de efeitos fixos. Além disso, há problemas de maximização que pode acontecer até mesmo em casos simples (Searle *et alii.*, 1992, pp. 291), principalmente quando o máximo ocorre nas fronteiras do espaço paramétrico (e.g., estimativa de σ^2 muito próximo de zero). Outro problema que pode ocorrer são os máximos locais que muitas vezes "enganam" os algoritmos. Como ilustração, a Figura 5.3.1 apresenta um caso onde há apenas um máximo local dentro da região paramétrica especificada para as componentes de variância σ_1^2 e σ_2^2 ; pergunta-se: será que é ponto de máximo global? A Figura 5.3.2 ilustra o caso onde a função de verossimilhança cresce lentamente sem um máximo local definido na região paramétrica especificada.

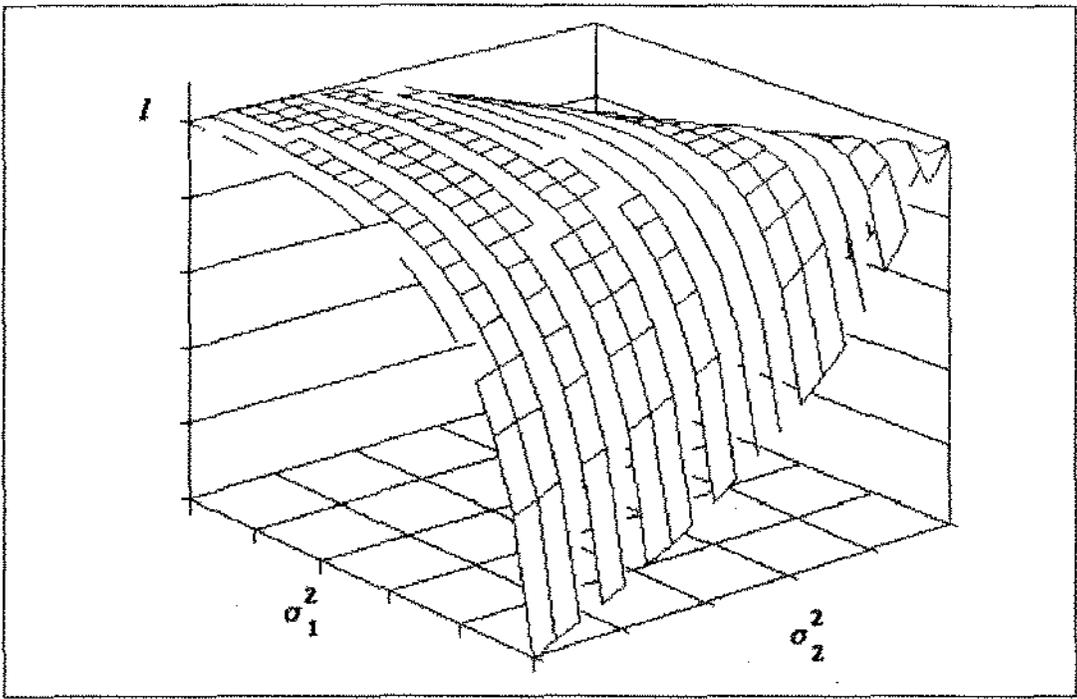


Fig. 5.3.2 - exemplo de uma superfície do logaritmo da verossimilhança sem máximo definido

O algoritmo EM, descrito no apêndice A, possui base estatística na sua formulação, envolvendo, e.g., conceitos de estatísticas suficientes, família exponencial e espaços amostrais. Ele é útil quando há dados perdidos e tem justificativa teórica quando os dados têm distribuição que pertence à família exponencial regular. Entretanto, ele recebe algumas críticas, por exemplo, quando se utiliza uma estrutura de correlação temporal nos erros (Chi e Reinsel, 1989). Existem muitas variações do algoritmo EM, e.g., Jennrich e Schluchter (1986) e Lindstrom e Bates (1988), que não serão tratadas aqui.

O método de Newton (Lindstrom e Bates, 1988) e o *Scoring* (Chi e Reinsel, 1989) são alternativas para contornar essas críticas. O filtro de Kalman é uma opção atrativa pois suporta dados com estrutura não balanceada, permite o cálculo da verossimilhança exata, generaliza situações onde a verossimilhança não pode ser expressa em uma forma fechada e é numericamente rápido, mesmo que haja um número grande de observações em alguns dos indivíduos (Jones e Boateng, 1991).

5.3.1. O Algoritmo EM

Na estimação via MV com dados incompletos, Dempster *et alii.* (1977) observaram que muitos algoritmos iterativos para o cálculo das estimativas poderiam ser considerados como casos especiais do algoritmo EM (apêndice A). Laird (1982), Laird e Ware (1982) e

Laird *et alii.* (1987), por exemplo, consideraram o algoritmo EM ("Expectation and Maximization") para estimar parâmetros no modelo (5.2.1.1) pelos métodos de MV e MVRe. Esse algoritmo é útil quando há dados perdidos; estes são tomados como se fossem efeitos aleatórios não observáveis e erros (b_i e ε_i). Nesse algoritmo, cada iteração consiste de uma etapa de Esperança (passo E) e de Maximização (passo M).

Aplica-se esse algoritmo, por exemplo, em um conjunto incompleto de dados para que ele seja completado, criando-se artificialmente um conjunto "completo" de dados. Outra aplicação é utilizar o algoritmo para estimar os parâmetros não observáveis (Laird e Ware, 1982). Cada passo do algoritmo está descrito no Apêndice A. Para ilustrarmos o algoritmo EM considere o exemplo:

Exemplo 5.3.1 estrutura de independência condicional

Assuma que, no modelo (5.2.1.1), $\Omega_i = \sigma^2 I$ e $\Gamma = \sigma^2 B$. Assim, a matriz de covariância tem a forma $\Sigma_i = \sigma^2 (Z_i B Z_i' + I) = \sigma^2 V_i$ e para um dado valor inicial de B, os coeficientes β e b_i são estimados, respectivamente, por (5.2.1.28) e (5.2.1.29). O vetor dos erros é estimado pelos resíduos

$$\begin{aligned} \hat{\varepsilon}_i &= E(\varepsilon_i | y_i, \hat{\beta}, \hat{\Theta}_i) \\ &= y_i - X_i \hat{\beta} - Z_i \hat{b}_i. \end{aligned} \quad (5.3.1.1)$$

Substituindo-se (5.2.1.29) e (5.2.1.3) na expressão anterior, pode-se representar os resíduos como

$$\hat{\varepsilon}_i = V_i^{-1} (y_i - X_i \hat{\beta}). \quad (5.3.1.2)$$

Sob normalidade, podemos utilizar as propriedades da família exponencial regular (apêndice A.6) para construir os passos do EM (apêndice A.7 e A.8). Para estimação via MV, o passo E consiste no cálculo das estatísticas suficientes $(t_1, t_2)'$. Esses valores serão utilizados na etapa de maximização (passo M) para determinar $\hat{\sigma}^2$ e \hat{B} . Assim, as seguintes equações constituem o passo E:

(passo E):
$$t_1 = E \left[\sum_{i=1}^n \varepsilon_i' \varepsilon_i \mid y_i, \hat{\beta}, \hat{\Theta}_i \right] =$$

$$\begin{aligned}
&= \sum_{i=1}^n \left\{ \hat{\varepsilon}'_i \hat{\varepsilon}_i + \text{tr} \left[\text{Var}(\varepsilon_i | y_i, \hat{\beta}, \hat{\Theta}_i) \right] \right\} = \\
&= \sum_{i=1}^n \left\{ \hat{\varepsilon}'_i \hat{\varepsilon}_i + \hat{\sigma}^2 \text{tr}(\mathbf{I} - \mathbf{V}_i^{-1}) \right\}, \quad (5.3.1.3)
\end{aligned}$$

e

$$\begin{aligned}
t_2 &= E \left[\sum_{i=1}^n \mathbf{b}_i \mathbf{b}'_i \mid y_i, \hat{\beta}, \hat{\Theta}_i \right] = \\
&= \sum_{i=1}^n \left\{ \hat{\mathbf{b}}_i \hat{\mathbf{b}}'_i + \text{Var}(\mathbf{b}_i | y_i, \hat{\beta}, \hat{\Theta}_i) \right\} = \\
&= \sum_{i=1}^n \left\{ \hat{\mathbf{b}}_i \hat{\mathbf{b}}'_i + \hat{\sigma}^2 (\hat{\mathbf{B}} - \hat{\mathbf{B}} \mathbf{Z}'_i \mathbf{V}_i^{-1} \mathbf{Z}_i \hat{\mathbf{B}}) \right\}, \quad (5.3.1.4)
\end{aligned}$$

onde $\hat{\sigma}^2$ e $\hat{\mathbf{B}}$ serão definidos no passo M. Se \mathbf{b}_i e ε_i fossem observáveis, então as estatísticas teriam, respectivamente, a forma $\sum_{i=1}^n \varepsilon'_i \varepsilon_i$ e $\sum_{i=1}^n \mathbf{b}'_i \mathbf{b}_i$. Entretanto, como esses fatores não são observáveis, utiliza-se $\hat{\varepsilon}_i$ e $\hat{\mathbf{b}}_i$ com as respectivas correções de vícios nas equações (5.3.1.3) e (5.2.1.4). O passo M consiste em determinar valores factíveis de $\hat{\sigma}^2$ e $\hat{\mathbf{B}}$ que maximizam a função de verossimilhança. Na família exponencial regular, esses valores são funções das estatísticas suficientes do passo E:

(passo M):

$$\hat{\sigma}^2 = \frac{t_1}{T} \quad (5.3.1.5)$$

e

$$\hat{\mathbf{B}} = \frac{t_2}{n\hat{\sigma}^2} \quad (5.3.1.6)$$

As modificações no algoritmo quando se considera EMVRe ocorrem no passo E. Com EMVRe, as esperanças são condicionadas apenas em y , porque β foi integrado para fora da verossimilhança:

(passo E):

$$\begin{aligned}
t_1^* &= E \left[\sum_{i=1}^n \varepsilon'_i \varepsilon_i \mid y_i, \hat{\Theta}_i \right] = \\
&= t_1 + \hat{\sigma}^2 \sum_{i=1}^n \text{tr} \left\{ \mathbf{V}_i^{-1} \mathbf{X}_i \text{var}(\hat{\alpha}) \mathbf{X}'_i \mathbf{V}_i^{-1} \right\}, \quad (5.3.1.7)
\end{aligned}$$

e

$$\begin{aligned}
t_2^* &= E \left[\sum_{i=1}^n \mathbf{b}_i \mathbf{b}'_i \mid y_i, \hat{\Theta}_i \right] = \\
&= t_2 + \hat{\sigma}^2 \mathbf{B} \sum_{i=1}^n \left\{ \mathbf{Z}'_i \mathbf{V}_i^{-1} \mathbf{X}_i \text{var}(\hat{\alpha}) \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{Z}_i \right\} \mathbf{B}, \quad (5.3.1.8)
\end{aligned}$$

O algoritmo, entretanto, recebe algumas críticas. Um problema é a lentidão da convergência em relação aos métodos quase-Newton (e.g. *Scoring*) de otimização não linear (Jones, 1993, pp. 42). Outro problema é que, quando se considera uma estrutura de correlação dos erros, a etapa de maximização (passo M) não produz uma forma explicitamente fechada do EMV; seria necessário incorporar ao algoritmo um segundo nível de iterações dentro do passo M. Quando se considera modelos com estrutura temporal nos erros ou modelos com coeficientes aleatórios, encontram-se na literatura outros procedimentos de maximização que veremos a seguir.

5.3.2. O Método de Newton-Raphson

O Newton-Raphson (NR) é um método iterativo geral para a determinação dos pontos de máximo de funções. Para a estimação de MV dos parâmetros do modelo componente de variâncias, Jennrich e Sampson (1976) destacaram algumas vantagens do NR: a velocidade de convergência de razão quadrática e a aparente robustez próxima aos valores iniciais. Jennrich e Schluchter (1986) consideraram o método NR na EMV do modelo de efeitos aleatórios. Lindstrom e Bates (1988) desenvolveram uma implementação do NR para o mesmo modelo, utilizando EMV e EMVRe, com a preocupação de aumentar a velocidade de convergência e assegurar a obtenção de uma matriz de covariância positiva definida para os efeitos aleatórios a cada iteração. A implementação de Lindstrom e Bates (1988) será discutida posteriormente.

Seja o vetor de escores definido por

$$s(\beta, \Theta) = \begin{bmatrix} s_{\beta} \\ s_{\Theta} \end{bmatrix} = \begin{bmatrix} \partial l / \partial \beta \\ \partial l / \partial \Theta \end{bmatrix}, \quad (5.3.2.1)$$

e a matriz Hessiana de l definida como

$$H(\beta, \Theta) = \begin{bmatrix} H_{\beta\beta} & H_{\beta\Theta} \\ H_{\Theta\beta} & H_{\Theta\Theta} \end{bmatrix} = \begin{bmatrix} \partial^2 l / \partial \beta \partial \beta & \partial^2 l / \partial \beta \partial \Theta \\ \partial^2 l / \partial \Theta \partial \beta & \partial^2 l / \partial \Theta \partial \Theta \end{bmatrix}.$$

Na forma geral, o passo (p) da iteração obtém os novos valores dos parâmetros ($\beta^{(p)}$ e $\Theta^{(p)}$) dados os valores anteriores ($\beta^{(p-1)}$ e $\Theta^{(p-1)}$) é dado por:

$$\begin{bmatrix} \beta^{(p)} \\ \Theta^{(p)} \end{bmatrix} = \begin{bmatrix} \beta^{(p-1)} \\ \Theta^{(p-1)} \end{bmatrix} + [H(\beta^{(p-1)}, \Theta^{(p-1)})]^{-1} s(\beta^{(p-1)}, \Theta^{(p-1)}). \quad (5.3.2.2)$$

O valor da verossimilhança é calculado em cada passo, sendo necessário que $I^{(p+1)} \leq I^{(p)}$. O cálculo direto da verossimilhança em cada passo requer matrizes de ordem $T_1 \times T_1$. Um método alternativo para o cálculo da verossimilhança é o filtro de Kalman. Como será discutido na seção 5.3.4, esse método utiliza cálculos recursivos no tempo, i.e., os cálculos são feitos à medida que se avança no tempo. Por isso, as matrizes não dependerão do número de observações individuais e em geral serão de dimensões pequenas.

Para descrever o algoritmo NR, considere o seguinte:

Exemplo 5.3.2 estrutura com σ^2 comum aos efeitos aleatórios

Considere que no modelo (5.2.1.1) haja uma componente de variação σ^2 comum aos efeitos aleatórios b_i e ε_i , i.e., $\Omega_i = \sigma^2 W_i$ e $\Gamma = \sigma^2 B$. Assim, $\Sigma_i = \sigma^2 (Z_i B Z_i' + W_i) = \sigma^2 G_i$. A função de verossimilhança para EMV é dada por (5.2.1.11) e para EMVRe, é dada por (5.2.2.13). Suprimindo os termos constantes e tomando $\Sigma_i = \sigma^2 G_i$, as funções de verossimilhança são dadas, respectivamente, por

$$l_F(\beta, \sigma, \Theta | y) = \sum_{i=1}^n [\ln |\sigma^2 G_i| + \sigma^{-2} r_i' G_i^{-1} r_i] \quad (5.3.2.3)$$

e

$$l_R(\beta, \sigma, \Theta | y) = \ln \left| \sigma^2 \sum_{i=1}^n X_i' G_i^{-1} X_i \right| + l_F(\beta, \sigma, \Theta | y), \quad (5.3.2.4)$$

onde $r_i = y_i - X_i \beta$, e os índices F e R referem-se, respectivamente, à verossimilhança completa (EMV) e parcial (EMVRe). Deve-se observar que a matriz G_i depende dos parâmetros Θ_i .

As EMV e EMVRe de σ^2 são dadas, respectivamente, por (5.2.1.12) e (5.2.2.10), i.e.,

$$\hat{\sigma}_{MV}^2(\beta, \Theta) = \frac{\sum_{i=1}^n r_i' G_i^{-1} r_i}{T} \quad (5.3.2.5)$$

e

$$\hat{\sigma}_{MVRc}^2(\beta, \Theta) = \frac{\sum_{i=1}^n r_i' G_i^{-1} r_i}{T - a} \quad (5.3.2.6)$$

Lindstrom e Bates (1988) observaram que, no caso $\Sigma_i = \sigma^2 G_i$, é possível remover o termo σ^2 dos cálculos. Substituindo-se as expressões (5.3.2.5) e (5.3.2.6), respectivamente, em (5.3.2.3) e (5.3.2.4), obtemos as funções concentradas com respeito a σ^2 , i.e.,

$$L_F^C(\beta, \Theta|y) = \sum_{i=1}^n \ln|G_i| + T \ln \left| \sum_{i=1}^n r_i' G_i^{-1} r_i \right| + \text{constante} \quad (5.3.2.7)$$

e

$$L_R^C(\beta, \Theta|y) = \sum_{i=1}^n \ln|G_i| + \ln \left| \sum_{i=1}^n X_i' G_i^{-1} X_i \right| + (T-a) \ln \left| \sum_{i=1}^n r_i' G_i^{-1} r_i \right| + \text{cte.} \quad (5.3.2.8)$$

□

Lindstrom e Bates (1988) comentam que a otimização NR pode ser feita usando tanto as funções (5.3.2.3) e (5.3.2.4) como (5.3.2.7) e (5.3.2.8). O uso dessas duas últimas formas é mais recomendada devido às vantagens práticas como a redução do número de iterações e sensível melhora no comportamento da convergência.

Note que, em (5.3.2.2), a estimação não é feita em dois passos, mas é feita simultaneamente. Lindstrom e Bates (1988) que, por causa da linearidade condicional de β , a substituição do seu valor no passo p da iteração, $\beta^{(p)}$, pela sua estimativa, $\hat{\beta}(\Theta^{(p)})$ [eq. (5.2.1.28)] aumenta sensivelmente a velocidade de convergência e facilita os cálculos da iteração seguinte. Então a formulação do NR em dois passos é feita da seguinte forma: no passo (p) da iteração, os novos valores dos parâmetros ($\beta^{(p)}$ e $\Theta^{(p)}$) serão dados por

$$\Theta^{(p)} = \Theta^{(p-1)} + \left[H_{\Theta\Theta}^{(p-1)} \right]^{-1} s_{\Theta}^{(p-1)} \quad (5.3.2.9)$$

e

$$\hat{\beta}^{(p)} = \left[\sum_{i=1}^n X_i' (G_i^{(p)})^{-1} X_i \right]^{-1} \sum_{i=1}^n X_i' (G_i^{(p)})^{-1} y_i \quad (5.3.2.10)$$

O uso de decomposições de matrizes produzem fórmulas numericamente eficientes e estáveis. Em geral, a decomposição torna o algoritmo menos suscetível a erros de arredondamento e mais rápido pela redução dos cálculos. No modelo de independência condicional, $\Omega_i = \sigma^2 I$, Lindstrom e Bates (1988) desenvolveram a decomposição para as matrizes X_i e Z_i que são reduzidas à forma ortogonal – triangular (forma QR):

$$\begin{bmatrix} Z_i \\ X_i \end{bmatrix} = Q_i R_i = \begin{bmatrix} Q_{i(1)} & Q_{i(2)} & Q_{i(3)} \end{bmatrix} \begin{bmatrix} R_{i(11)} & R_{i(12)} \\ 0 & R_{i(22)} \\ 0 & 0 \end{bmatrix}$$

onde Q é matriz ortogonal; $R_{r(11)}$ e $R_{r(22)}$ são triangulares e têm dimensões, respectivamente, $m \times m$ e $a \times a$; a matriz $R_{r(12)}$ tem dimensão $m \times a$. Utiliza-se então a transformação $y^* = Q'y$, para desenvolver a função de verossimilhança (completa ou parcial).

Lindstrom e Bates (1988) também comentam que a decomposição de Cholesky de B (i.e., determinar a matriz triangular superior F de modo que $B = FF'$) para otimizar a verossimilhança com respeito a F é importante para garantir estimativas positivas definidas para B .

Na convergência, a inversa da Hessiana de I_F^C ou I_R^C é uma estimativa para a matriz de covariância marginal para β e Θ , e pode ser utilizado para achar intervalos de confiança aproximados para β . Caso necessário, a Hessiana de I_F ou I_R poderá ser calculada após a convergência obtida para I_F^C ou I_R^C .

5.3.3. O Método do *Scoring* de Fisher

O *Scoring* de Fisher (SF) é uma variação do método NR. O método SF substitui a matriz Hessiana pela sua esperança na equação (5.3.2.2). A esperança é dada por

$$E[H(\beta, \Theta)] = \begin{bmatrix} E(H_{\beta\beta}) & 0 \\ 0 & E(H_{\Theta\Theta}) \end{bmatrix},$$

onde $E(H_{\beta\beta}) = 2 \sum_{i=1}^n X_i' \Sigma_i^{-1} X_i$, $E(H_{\beta\Theta}) = 0$; e $[E(H_{\Theta\Theta})]_{rs}$, o (rs) -ésimo elemento da matriz $E(H_{\Theta\Theta})$ é $2 \text{tr}(\Sigma_i^{-1} \Sigma_{ir} \Sigma_i^{-1} \Sigma_{is})$, para $r, s = 1, \dots, T$.

Como $E(H_{\beta\Theta}) = 0$, as atualizações $\beta^{(p)}$ e $\Theta^{(p)}$ são obtidas, resolvendo-se equações separadamente. As novas estimativas $\Theta^{(p)}$ são calculadas no passo de *scoring*:

$$\Theta^{(p)} = \Theta^{(p-1)} + [E(H_{\Theta\Theta}^{(p-1)})]^{-1} s_{\Theta}^{(p-1)}, \quad (5.3.3.1)$$

onde s_{Θ} é calculado por (5.3.2.1) usando os resíduos atualizados $r_i^{(p)} = y_i - X_i \hat{\beta}^{(p)}$. Para calcular $\beta^{(p)}$, utiliza-se o estimador $\hat{\beta}(\Theta^{(p)})$ [eq. (5.2.1.28)]:

$$\hat{\beta}^{(p)} = \left[\sum_{i=1}^n X_i' (G_i^{(p)})^{-1} X_i \right]^{-1} \sum_{i=1}^n X_i' (G_i^{(p)})^{-1} y_i. \quad (5.3.3.2)$$

Como exemplo de aplicação desse método, considere o seguinte:

Exemplo 5.3.3 modelo de efeitos aleatórios com erros AR(1)

Chi e Reinsel (1989) consideraram o método SF na estimação de MV no modelo de efeitos aleatórios com erros AR(1). A matriz de covariância individual tem a forma (5.2.1.3) com $\Omega_i = \sigma^2 W_i$, onde W_i tem a forma da matriz (3.2.1.4): estrutura de covariância do processo AR(1) contínuo. A matriz W_i é função do parâmetro de autoregressão ϕ (seção 3.2). Chi e Reinsel (1989) expressaram $\hat{\Omega}_i = \hat{\sigma}^2 \hat{W}_i$ na forma (3.2.1.5), i.e., $\hat{W}_i = \hat{R}_i^{-1} \hat{V}_i \hat{R}_i^{-1}$. Para que $\hat{\Gamma}$ seja positiva definida a cada iteração, utiliza-se a decomposição de Cholesky proposta por Lindstrom e Bates (1988). A reparametrização é feita fazendo-se $\Gamma = F'F$, onde F é uma matriz triangular superior. Assim, o conjunto de parâmetros considerado é $\Theta = \{F, \sigma^2, \phi\}$.

A função de verossimilhança completa é dada por (5.2.1.7) e a parcial por (5.2.2.13). Para aplicar o algoritmo do *scoring*, para cada iteração calcula-se o vetor dos escores s_{Θ} e a matriz de informação de Fisher $E(H_{\Theta\Theta})$. As EMV de β e Θ são obtidas por iteração de (5.3.3.1) e (5.3.3.2). Para inicializar o algoritmo, toma-se $\phi = 0$, i.e., $\Omega_i = \sigma^2 I$, e as EMV de Γ e σ^2 sob esse modelo, i.e., $\Theta^{(0)} = \{\hat{F}, \hat{\sigma}^2, 0\}$.

Nota: o desenvolvimento das derivadas da função de verossimilhança está descrito por Chi e Reinsel (1989); detalhes sobre o método do *scoring* em dados longitudinais com erros AR(1) estão desenvolvidos na tese de doutorado de Chi (1988, cap. 2).

□

O método NR requer o cálculo da matriz das derivadas segundas (matriz Hessiana). O *scoring* substitui a matriz Hessiana pelo seu valor esperado (matriz de informação). Em muitas situações, o cálculo da matriz de informação é mais simples do que o cálculo da matriz Hessiana (Searle *et alii.*, 1992, pp. 295). Jennrich e Sampson (1986) observam que o SF é mais robusto aos valores iniciais do que o NR. Eles recomendam o uso do SF nas primeiras iterações e depois a mudança para NR (que é mais rápido do que o SF).

5.3.4. O Filtro de Kalman

Em alguns dos algoritmos de estimação apresentados, como os algoritmos de otimização não linear de NR e SF, para obtermos as EMV ou EMVRe dos parâmetros das matrizes B e W_i é necessário calcular a verossimilhança completa ou parcial. O cálculo direto da verossimilhança feito em cada passo da iteração requer manipulação de matrizes $T_i \times T_i$ (por MV) e $(T_i - a) \times (T_i - a)$ (por MVRe), onde T_i é o número de observações do i -

ésimo indivíduo e a é o número de efeitos fixos. Quando T_i for pequeno, não haverá problemas computacionais e o uso da fatorização de Cholesky da matriz de covariância torna o cálculo da verossimilhança estável do ponto de vista numérico, pois garante que a estimativa da matriz de covariância seja positiva definida.

O filtro de Kalman descrito no apêndice B é um auxiliar para o cálculo da verossimilhança. Ele pode ser usado em análise de dados longitudinais quando o problema puder ser formulado na forma em espaço de estados. Uma vez modelado nessa forma, é possível calcular a verossimilhança recursivamente sem utilizar matrizes de dimensões que dependem do número das observações individuais (Harvey, 1978; Jones, 1993, pp. 78). Jones (1987) usou esse método para calcular a verossimilhança exata, suportando dados com estrutura não-balanceada e assumindo-se variáveis aleatórias gaussianas.

Nessa seção, considere o modelo de efeitos mistos de modo que $\Sigma_i = \sigma^2 (Z_i B Z_i' + W_i) = \sigma^2 G_i$, com W_i tendo a forma da estrutura de covariância do processo AR(1) contínuo (3.2.1.4). Como feito na seção 3.2, para acomodar dados desigualmente espaçados, considere que as observações para o i -ésimo indivíduo sejam tomadas nos instantes t_{ij} , para $j=1, \dots, T_i$. A defasagem de tempo entre observações consecutivas é dada por $\delta_{ij} = t_{i,j} - t_{i,j-1}$, $j = 2, \dots, T_i$.

Seja X_i a matriz de delineamento da forma polinomial:

$$X_i = \begin{bmatrix} 1 & t_{i1} & \dots & t_{i1}^{a-1} \\ 1 & t_{i2} & \dots & t_{i2}^{a-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_{iT_i} & \dots & t_{iT_i}^{a-1} \end{bmatrix} \quad (5.3.4.1)$$

Cada linha de X_i será denotada por $x_i'(t_{ij})$; cada elemento do vetor de observações y_i e do vetor de erros ε_i será representado, respectivamente, por $y_i(t_{ij})$ e $\varepsilon_i(t_{ij})$.

5.3.4.a. Utilizando Estimativas de β e Θ em 2 Passos

Quando o algoritmo de otimização é formulado em dois passos (e.g., eq. [5.3.2.9] e [5.3.2.10]), as componentes aleatórias do modelo de efeitos mistos podem ser escritas na forma em espaço de estados (Jones e Boateng, 1991). Para o i -ésimo indivíduo, a representação em espaço de estados pode ser dada por:

$$\begin{bmatrix} \varepsilon_i(t_{ij}) \\ b_i(t_{ij}) \end{bmatrix} = \begin{bmatrix} \phi(t_{ij} - t_{i,j-1}) & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \varepsilon_i(t_{i,j-1}) \\ b_i(t_{i,j-1}) \end{bmatrix} + \eta_i(t_{ij}), \quad (5.3.4.2)$$

e

$$v_i(t_{ij}) = \begin{bmatrix} 1 & z'_i(t_{ij}) \end{bmatrix} \begin{bmatrix} \varepsilon_i(t_{ij}) \\ b_i(t_{ij}) \end{bmatrix} + \zeta_i(t_{ij}) \quad (5.3.4.3)$$

onde $b_i(t_{ij})$ é o vetor dos efeitos aleatórios no tempo t_{ij} . A equação (5.3.4.2) indica que os efeitos aleatórios para o i -ésimo indivíduo são constantes no tempo, i.e., $b_i(t_{ij}) = b_i(t_{ij-1})$; $z'_i(t_{ij})$ é a i -ésima linha de Z_i no tempo t_{ij} , $\eta_i(t_{ij})$ representa os choques aleatórios do processo AR(1) contínuo e $\zeta_i(t_{ij})$ é o erro observacional com distribuição $N(0, \sigma^2 \sigma_0^2)$. Na equação (5.3.4.3), $v_i(t_{ij})$ representa a diferença $y_i(t_{ij}) - x'_i(t_{ij})\beta$. Nessa representação, o parâmetro desconhecido β é tratado como um termo fixo, sendo necessário estimá-lo juntamente com os outros parâmetros σ_0^2 , σ^2 , ϕ e B .

5.3.4.b. Utilizando a Verossimilhança Concentrada

Para evitar a estimação direta de β e de σ^2 que requer manipulação de matrizes $T_i \times T_i$ (por MV) e $(T_i - a) \times (T_i - a)$ (por MVRe), Jones e Boateng (1991) e Jones (1993, pp. 108) formulam as equações de estado de modo o filtro gere as equações normais para estimar β e σ^2 . Para isso, o parâmetro β é concentrado para fora da verossimilhança (5.3.4.4). Uma rotina de otimização não linear deve ser utilizada para obter as EMV dos parâmetros não lineares, que são: o parâmetro de autoregressão ϕ , os elementos de B e a variância do erro observacional σ_0^2 . No procedimento, todas as matrizes de covariância podem ser divididas por σ^2 , i.e., as equações podem ser expressas em função de G_i . A função de verossimilhança concentrada é obtida substituindo-se (5.2.1.13) e (5.2.1.12) em (5.2.1.11), resultando em

$$l = T \ln 2\pi \hat{\sigma}^2 + \Delta + T, \quad (5.3.4.4)$$

onde $\Delta = \sum_{i=1}^n \ln |G_i|$. O parâmetro β é estimado por (5.2.1.12) e σ^2 por (5.2.1.12). A predição de b_i é dada por (5.2.1.29) que nesse caso tem a forma $\hat{b}_i = BZ'_i \hat{G}_i^{-1} (y_i - X_i \hat{\beta})$. A fatorização de Cholesky da matriz G_i^{-1} auxilia nos cálculos de $\hat{\beta}$, $\hat{\sigma}^2$, \hat{b}_i e $\ln |G_i|$. Assim, com a decomposição $G_i^{-1} = L'L$, onde L é uma matriz triangular superior, fazemos

$$Ly_i = LX_i \beta + LZ_i b_i + L\varepsilon_i,$$

i.e.,

$$\bar{y}_i = \bar{X}_i \beta + \bar{Z}_i b_i + \bar{\varepsilon}_i. \quad (5.3.4.5)$$

Os estimadores dos parâmetros $\hat{\beta}$, $\hat{\sigma}^2$, \hat{b}_i e $\ln|G_i|$ do modelo transformado são os mesmos do modelo original, i.e., $\hat{\beta} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}$, $\hat{b}_i = B\bar{Z}'_i(\bar{y}_i - \bar{X}_i\hat{\beta})$ e $\hat{\sigma}^2 = RSS/T$, onde $RSS = \bar{y}'\bar{y} - \bar{y}'\bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}$. A recursão de Kalman é feita de modo que a matriz e o vetor das equações normais necessários para estimar β (os efeitos b_i serão preditos após a estimação dos parâmetros β e Θ , juntamente com os resíduos que serão descritos na seção 6.3), a soma dos quadrados ponderados para estimar σ^2 e o termo determinante da verossimilhança sejam acumulados durante um passo do algoritmo. Em outras palavras, o filtro deverá produzir os termos $\bar{X}'\bar{X}$, $\bar{X}'\bar{y}$ e $\bar{y}'\bar{y}$. Esses valores são utilizados quando os valores dos parâmetros obtidos através de otimização não linear forem considerados soluções que maximizam a verossimilhança. Assim, a verossimilhança dada por (5.3.4.4) é calculada no final da recursão, produzindo também as estimativas de β e σ^2 . Além disso, o filtro de Kalman fornece as predições dos efeitos aleatórios individuais e suas matrizes de covariância para cada indivíduo, e os resíduos transformados (que serão descritos na seção 6.3) e a variância do processo $\hat{\sigma}^2/(1-\hat{\phi}^2)$. Para isso, defina a matriz

$$M = \begin{bmatrix} \bar{X}'\bar{X} & \bar{X}'\bar{y} \\ \bar{y}'\bar{X} & \bar{y}'\bar{y} \end{bmatrix} \quad (5.3.4.6)$$

Para obtermos M_i é preciso que o filtro de Kalman seja executado separadamente em cada coluna das matrizes X_i e para y_i , onde cada coluna é tratada como vetor de dados. As partes da recursão que envolvem matrizes de covariância devem ser calculadas apenas uma vez. Como existe um vetor de estado para cada coluna de X_i e y_i (ver passo 3 dos passos de iteração da página seguinte), o estado será uma matriz dimensão $(m+1) \times (a+1)$. A matriz de predição dos vetores dos estados no tempo t dada a informação anterior será denotada por $S_i(t_j|t_j - \delta t)$ e a matriz de covariância dessa predição por

$$P_i(t_j|t_j - \delta t) = \sigma^2 \begin{bmatrix} P_{11}(t_j|t_j - \delta t) & P_{12}(t_j|t_j - \delta t) \\ P'_{12}(t_j|t_j - \delta t) & P_{22}(t_j|t_j - \delta t) \end{bmatrix}, \quad (5.3.4.7)$$

onde P_{11} é escalar, P'_{12} é um vetor $m \times 1$ e P_{22} é uma matriz $m \times m$. A inicialização é feita fazendo-se

$$S_i(t_{i1}|0) = 0 \text{ e } P_i(t_{i1}|0) = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & B \end{bmatrix}, \quad (5.3.4.8)$$

onde a notação $S_i(t_{i1}|0)$ indica a estimativa no tempo t_{i1} dado que não há observações prévias no indivíduo, e $\sigma^2 B$ é a matriz de covariância a priori dos efeitos aleatórios b_i . Na

inicialização, a matriz B é um chute inicial; depois, ela é a estimativa corrente produzida por um método de otimização não linear (ver Jones e Boateng, 1991; Jones, 1993, pp. 33). Como foi discutido por Lindstrom e Bates (1988) e Jones e Boateng (1991), é conveniente utilizar a decomposição de Cholesky $B = FF'$ para garantir estimativas positivas definidas de B .

Os passos da iteração são:

1. Calcular a predição do estado:

$$S_i(t_j|t_j - \delta t) = \begin{bmatrix} \phi(t_j; t_j - \delta t) & 0 \\ 0 & I \end{bmatrix} S_i(t_j - \delta t|t_j - \delta t).$$

2. Calcular a matriz de covariância da predição:

$$P_{11}(t_j|t_j - \delta t) = P_{11}(t_j - \delta t|t_j - \delta t)\phi^2(t_j; t_j - \delta t) + Q(\delta t),$$

$$P_{12}(t_j|t_j - \delta t) = P_{12}(t_j - \delta t|t_j - \delta t)\phi(t_j; t_j - \delta t),$$

$$P_{22}(t_j|t_j - \delta t) = P_{22}(t_j - \delta t|t_j - \delta t),$$

onde $Q(\delta t) = [1 - \phi^2(\delta t)]$ denota a variância do erro da predição.

3. Calcular a predição da próxima observação:

$$\begin{bmatrix} x(t_j|t_j - \delta t) & y(t_j|t_j - \delta t) \end{bmatrix} = \begin{bmatrix} 1 & z_i(t_j) \end{bmatrix} S(t_j|t_j - \delta t).$$

4. Calcular a inovação (diferença entre o observado e o predito):

$$I(t_j) = \begin{bmatrix} x(t_j) & y(t_j) \end{bmatrix} - \begin{bmatrix} x(t_j|t_j - \delta t) & y(t_j|t_j - \delta t) \end{bmatrix}$$

5. Calcular a matriz de covariância da inovação:

$$V(t_j) = \begin{bmatrix} 1 & z_i(t_j) \end{bmatrix} P(t_j|t_j - \delta t) \begin{bmatrix} 1 & z_i(t_j) \end{bmatrix}' + \sigma_0^2.$$

6. Acumular as quantidades necessárias para calcular I no final da recursão, i.e., no passo (p):

$$M^{(p)} = M^{(p-1)} + I'(t_j)I(t_j)/V(t_j)$$

e

$$\Delta^{(p)} = \Delta^{(p-1)} + \ln V(t_j),$$

onde $M^{(p)}$ é a parte triangular superior da matriz $(a+1) \times (a+1)$ necessária para calcular as estimativas de β e σ^2 , e $\Delta^{(p)}$ é o termo determinante acumulado sobre todas as observações para todos os indivíduos. Inicialmente, $M^{(0)} = 0$ e $\Delta^{(0)} = 0$.

7. Atualizar a estimativa do vetor de estado:

$$K_i(t_j) = [1 \quad z_i(t_j)]P(t_j|t_j - \delta t),$$

então

$$S_i(t_j|t_j) = S_i(t_j|t_j - \delta t) + K'(t_j)I(t_j)/V(t_j).$$

8. Calcular a matriz de covariância atualizada:

$$P(t_j|t_j) = P(t_j|t_j - \delta t) - K'(t_j)K(t_j)/V(t_j).$$

A recursão de Kalman é feita sobre todos os indivíduos e a função l é acumulada sobre os indivíduos (i.e., a matriz de estados e a matriz de covariância são reinicializadas para cada indivíduo, sendo que as quantidades M e Δ são acumuladas sobre todas as observações de todos os indivíduos). Assim, a verossimilhança dada por (5.3.4.4) é calculada no final da recursão. Como a matriz M contém as somas de quadrados, o cálculo da verossimilhança produz como subprodutos as estimativas de β e σ^2 .

Quando os valores dos parâmetros obtidos através de otimização não linear forem considerados soluções que maximizam a verossimilhança (isso é feito por critérios de convergências que não foram discutidos nesse trabalho), o filtro de Kalman fornece as predições dos efeitos aleatórios individuais e suas matrizes de covariância para cada indivíduo, e os resíduos transformados (que serão descritos na seção 6.3) e a variância do processo $\hat{\sigma}^2/(1-\hat{\phi}^2)$.

As principais vantagens do filtro de Kalman são a generalização de situações onde a verossimilhança não pode ser expressa em uma forma fechada e a rapidez do algoritmo mesmo que haja muitas observações em alguns dos indivíduos (Jones e Boateng, 1991).

A formulação em espaço de estados permite utilizar covariáveis variando no tempo, o valor mais recente da covariável pode ser usado na matriz de delineamento individual. A questão é qual matriz de delineamento deverá conter essas covariáveis: X_i ou Z_i . Se a covariável for incorporada em Z_i , o coeficiente aleatório poderá ter média diferente de zero, ao contrário do que se supõe no modelo. Se for o caso, a mesma coluna pode ser incluída em X_i como efeito fixo, e o coeficiente aleatório terá média zero. Jones (1993, pp. 120) comenta que a estrutura do erro pode ser generalizada para um ARMA(p,q) contínuo.

6

Testes de Hipóteses e Diagnóstico

6.1. Introdução

No capítulo 4 discutiu-se alguns pontos importantes na seleção do modelo mais adequado como a parcimoniosidade e a bondade de ajuste do modelo escolhido. Como extensão dos capítulos anteriores, este se propõe a tratar de algumas questões de interesse na análise e modelagem de dados longitudinais, apresentando ferramentas estatísticas como o teste de hipóteses, a análise de resíduos (diagnóstico) e predição.

Para introduzir o teste de hipótese em estudos longitudinais considere a situação descrita no exemplo 2.2.1, onde n indivíduos são alocados aleatoriamente em dois grupos. Como discutido no capítulo 4, a escolha por um modelo deve levar em consideração os objetivos de análise. Por exemplo, ajustando-se o modelo univariado (2.4.1), uma possível questão de interesse é verificar se os tratamentos administrados em cada grupo produzem os mesmos efeitos na média populacional da variável de resposta. Em outro caso, se a questão de interesse fosse testar se o efeito temporal é constante para todas as ocasiões, então o modelo (2.4.4) seria adequado.

Os testes de hipóteses sobre os parâmetros são aplicados na seleção do modelo e na sua interpretação. Na seção 6.2.1 discutiremos sobre o teste da razão de verossimilhança para a comparação entre dois modelos hierárquicos. Na seção 6.2.4 apresenta-se um teste de escores como critério de seleção entre o modelo de independência condicional e o modelo com erros AR(1). Os testes de hipóteses podem ser utilizados como uma opção aos critérios de seleção como o AIC (critério de informação de Akaike, 1973) apresentado no capítulo 4. Como foi discutido no capítulo 4, o AIC é um critério útil para responder a questão, por exemplo, sobre a existência de correlação serial ou efeitos aleatórios.

Na análise de variância, a questão de interesse está na identificação das fontes de variação que compõem o erro experimental, ou seja, estudar a variação das componentes da variância. Quando a matriz de covariância individual satisfaz a condição de circularidade, pode-se aplicar normalmente os testes F da ANOVA. Entretanto, quando essa condição não for satisfeita, é necessário tomar certos cuidados para obter as estatísticas F apropriadas.

Algumas formas de diagnósticos estão descritos na seção 6.3, onde define-se o resíduo ordinário como a diferença entre o valor observado e o ajustado, e os resíduos transformados como uma combinação dos resíduos ordinários. Como os métodos e modelos discutidos nesse trabalho assumiram em sua grande maioria normalidade, é preciso considerar meios para verificar essa suposição. A falha na suposição de normalidade não invalida as estimativas dos parâmetros de locação. Entretanto, pode invalidar os testes usuais e os intervalos de confiança que são baseados na teoria da distribuição Normal. Também é possível verificar se o modelo escolhido descreve razoavelmente bem o conjunto de dados, por exemplo, através de gráficos dos resíduos e do teste de bondade de ajuste. Outros elementos podem se tornar claros como: *outliers*, correlação residual (comportamento não aleatório dos resíduos) e pontos influentes (valores ou indivíduos). Na seção 6.3.3 apresentaremos o método do semivariograma empírico para avaliar a estrutura de correlação dos resíduos. Essa avaliação pode ser utilizada para a seleção inicial das estruturas de covariância do modelo, indicando, por exemplo, existência de correlação temporal nos erros.

6.2. Testes de Hipóteses

Nessa seção serão apresentados alguns testes de hipóteses que podem ser usados na seleção do modelo. Os testes de hipóteses aplicados ao modelo de efeitos mistos essencialmente são os mesmos aplicados, por exemplo, aos modelos de regressão em geral. Como já são amplamente conhecidos eles não serão discutidos detalhadamente. Os testes descritos são: o teste da razão de verossimilhança para a comparação de dois modelos hierárquicos (*nested*), o teste de Wald, o teste dos escores e o teste F da ANOVA.

6.2.1. Teste da razão de verossimilhança

O teste da razão de verossimilhança pode ser usado para comparar dois modelos hierárquicos. Dois modelos são hierárquicos (*nested*) quando um modelo é uma versão restrita do outro, i.e., o segundo modelo tem todos os parâmetros do primeiro mais r adicionais. Nesse caso, o modelo reduzido (primeiro modelo) é chamado de versão restrita do segundo pois os r parâmetros adicionais são nulos; outra forma de reduzir o modelo é restringir dois ou mais parâmetros a serem iguais como ilustrado a seguir:

Exemplo 6.2.1 modelo univariado

Tome por exemplo o modelo (2.4.1) com dois grupos

$$y_{i(1)j} = \beta_1 + \pi_{i(1)} + \varepsilon_{i(1)j} \quad (6.2.1.1)$$

e

$$y_{i(2)j} = \beta_1 + \beta_2 + \pi_{i(2)} + \varepsilon_{i(2)j}, \quad (6.2.1.2)$$

onde β_1 é a média populacional do grupo 1 e β_2 é a diferença da média entre os grupos. Uma possível questão de interesse é testar a hipótese nula (H_0)

$$H_0: \beta_2 = 0, \quad (6.2.1.3)$$

que produz a versão restrita de (6.2.1.2) $y_{i(2)j} = \beta_1 + \pi_{i(2)} + \varepsilon_{i(2)j}$.

Em outro exemplo, tome o modelo (2.4.4)

$$y_{i(k)j} = \mu_k + \beta_j + \pi_{i(k)} + \varepsilon_{i(k)j}, \quad (6.2.1.4)$$

onde μ_k é a média do grupo k , e o parâmetro β_j é o efeito da j -ésima ocasião. A questão de interesse é testar se o efeito temporal é constante, ou seja, com quatro ocasiões teríamos por exemplo,

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4. \quad (6.2.1.5)$$

A hipótese acima produz um modelo reduzido onde o efeito temporal é constante. □

Seja l_1 o valor de $-2 \ln$ verossimilhança do modelo reduzido (sob H_0) e l_2 o valor do outro com r parâmetros adicionais (sob a hipótese alternativa H_1). A estatística do teste é definida como

$$LR = l_1 - l_2 \quad (6.2.1.6)$$

que, sob a hipótese de que os dois modelos são iguais (H_0), a diferença LR entre l_1 e l_2 é assintoticamente distribuída como qui-quadrado com r graus de liberdade. Quanto maior for essa diferença, maiores serão as evidências contra H_0 . Em outras palavras, a não rejeição da hipótese nula (H_0) significa que o modelo com mais parâmetros não produz

significativamente melhor ajuste do que o outro e, analogamente, a rejeição de H_0 indica que há uma melhora significativa no ajuste quando parâmetros adicionais são introduzidos.

Por ser um teste assintótico, o tamanho da amostra deve ser grande, mas ele frequentemente produz resultados razoáveis para tamanhos moderados como 20 indivíduos com 6 observações cada (Jones, 1993, pp. 44). Quanto maior for o número de observações individuais maior será a informação proporcionada sobre $\sigma^2\Omega$, que é a componente intra-individual da variância; e para um número grande de indivíduos, haverá mais informação sobre $\sigma^2\Gamma$, a componente entre-indivíduos da variância.

A razão de verossimilhança também pode ser aplicada para ajustar um modelo com parâmetros não-lineares; por exemplo, é possível testar hipóteses do tipo

$$H_0: \phi = 0, \quad (6.2.1.7)$$

onde ϕ é o parâmetro (não linear) de autoregressão AR(1) dos erros (capítulo 3). Como foi discutido no capítulo 4, Jones (1990) observou que na prática a correlação serial pode ser parcialmente confundida com os efeitos aleatórios individuais. A inferência baseada na função de verossimilhança nem sempre identificará corretamente o modelo, especialmente em pequenas amostras. Além disso, ao ajustarmos vários modelos para um conjunto de dados, o uso do teste da razão de verossimilhança pode tornar-se confuso e também poderão haver modelos que não sejam hierárquicos (*nested*), de modo que esse teste não possa ser utilizado na comparação entre eles. Um procedimento alternativo é utilizar o critério de seleção de modelo como o AIC (Capítulo 4). O AIC é baseado na teoria de decisão e penaliza $l = -2 \ln \text{verossimilhança}$ do número de parâmetros ajustados aos dados (r_n) para evitar excesso de parametrização. Na seção 6.2.4 apresentaremos uma aplicação do teste do *scoring* de Fisher para testar (6.2.1.7)

O valor minimizado de l_2 deve ser menor que l_1 uma vez que algumas restrições foram removidas. Em problemas de otimização não-linear, às vezes, l_2 aumenta. Isso teoricamente não deve ocorrer, mas o programa converge ao mínimo local que é maior do que o mínimo do primeiro modelo (l_1). Uma forma de evitar convergência para um mínimo local diferente é começar o segundo modelo a partir de valores dos parâmetros que minimizam l_1 com os r parâmetros iguais a zero. Isso inicia a otimização a partir de l_1 e um bom algoritmo irá apenas na direção do declive. Por causa do risco de terminar num mínimo local ao invés do mínimo global, é uma boa prática tentar vários valores iniciais dos parâmetros para ver se eles convergem ao mesmo ponto.

6.2.2. Teste de Wald

Quando o interesse é testar os parâmetros lineares β , o ajuste de dois modelos para comparação pelo teste de razão de verossimilhança pode ser evitado, ajustando-se apenas um modelo mais completo (com mais parâmetros) e testando-se apenas os contrastes entre os elementos do vetor de efeitos β .

Como foi ilustrado na seção anterior, um exemplo simples é o teste da hipótese de que um dos coeficientes é zero, i.e.,

$$H_0: \beta_k = 0, \quad (6.2.2.1)$$

onde k denota o elemento do vetor de efeitos β que está sendo testado. O teste da razão de verossimilhança ajustaria dois modelos: um modelo com a presença desse parâmetro e o outro sem ele. Depois, as verossimilhanças associadas a cada modelo seriam comparadas para verificar se a mudança encontrada é significativa. Essa comparação seria feita por (6.2.1.6) considerando-se a distribuição Qui-quadrada com um grau de liberdade.

Outro meio para se testar essa hipótese é usar a matriz de covariância estimada dos efeitos estimados. O k -ésimo elemento dessa matriz é a variância estimada de $\hat{\beta}_k$, $\text{vâr}(\hat{\beta}_k)$. A estatística do teste, conhecido com estatística de Wald, é

$$t_{gl} = \frac{\hat{\beta}_k}{\sqrt{\text{vâr}(\hat{\beta}_k)}}, \quad (6.2.2.2)$$

onde gl é o número de graus de liberdade, $gl = T - a$.

Equivalentemente,

$$F_{1, gl} = \frac{\hat{\beta}_k^2}{\text{vâr}(\hat{\beta}_k)}. \quad (6.2.2.3)$$

Na regressão linear, as estatísticas (6.2.2.2) e (6.2.2.3) são, respectivamente, o teste t exato e o teste F exato para as hipóteses, assumindo-se erros gaussianos independentes com variância constante (homocedasticia). Convém observar que o quadrado da estatística t equivale à estatística F com 1 grau de liberdade no numerador apenas quando um único coeficiente está sendo testado. Quando mais de um coeficiente estão sendo testados simultaneamente, o teste F é mais apropriado.

O teste de Wald é assintoticamente equivalente ao teste da razão de verossimilhança (para tamanhos amostrais grandes) o que significa que o grau de liberdade é grande. Nesse caso, quando os graus de liberdade crescem, as distribuições das estatísticas t e F se aproximam respectivamente à distribuição Normal padrão z e Qui-quadrado com 1 grau de liberdade, χ_1^2 :

$$t_{gl} \sim z, \quad (6.2.2.4)$$

e

$$F_{gl} \sim \chi_1^2, \quad (6.2.2.5)$$

Para um número moderado de graus de liberdade, os testes t e F são mais conservativos que os testes z e χ_1^2 , ou seja, possuem valores críticos mais altos para um nível de significância.

Considere agora o caso em que se deseja comparar coeficientes, como por exemplo:

$$H_0: \beta_1 = \beta_2. \quad (6.2.2.6)$$

Isso representa um contraste, que pode ser escrito na forma matricial:

$$[1 \quad -1] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = 0,$$

onde $C = [1 \quad -1]$ é a matriz de contraste. A hipótese (6.2.2.6) pode ser escrita matricialmente como

$$H_0: C\beta = 0. \quad (6.2.2.7)$$

A estatística do teste é

$$F_{1, gl} = \frac{(C\hat{\beta})^2}{\text{var}(C\hat{\beta})}, \quad (6.2.2.8)$$

que assintoticamente tem distribuição χ_1^2 .

Suponha agora que seja de interesse testar o seguinte:

$$H_0: \beta_1 = \beta_2 \text{ e } \beta_2 = \beta_3,$$

i.e.,

$$H_0: \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (6.2.2.9)$$

A estatística dos teste é dada por

$$(C\hat{\beta})' [C \text{ var}(\hat{\beta}) C']^{-1} C\hat{\beta}, \quad (6.2.2.10)$$

que sob H_0 é assintoticamente distribuída como χ_2^2 .

6.2.4. Teste do Escore de Fisher

Considere o modelo de efeitos aleatórios com erros AR(1) do exemplo 5.3.3.1. Nessa seção descreveremos a aplicação do teste dos escores proposto por Chi e Reinsel (1989) para testar a existência de correlação serial AR(1), i.e:

$$H_0: \phi = 0 \quad (6.2.4.1)$$

A hipótese H_0 equivale à hipótese de independência condicional $\Omega_1 = \sigma^2 I$. A rejeição da hipótese nula não significa que o modelo especificado na hipótese alternativa seja apropriado, mas deve sugerir a existência de correlação nos erros e que o AR(1) é uma estrutura possível para os erros.

Considere o vetor de parâmetros $\vartheta' = \{\beta', \Theta, \phi\}$, onde $\Theta = \{\gamma \in \Gamma \text{ e } \sigma^2\}$, e seja $\vartheta'_0 = \{\hat{\beta}', \hat{\Theta}, 0\}$, onde $\hat{\beta}$ e $\hat{\Theta}$ são EMV de β e Θ no modelo sob a hipótese nula (6.2.4.1). Quando calculado em ϑ_0 o vetor de escores definido em (5.3.2.4 b) é

$$s(\vartheta)|_{\vartheta_0} = \frac{\partial l}{\partial \vartheta} \Big|_{\vartheta_0} = \begin{bmatrix} 0 \\ 0 \\ s_\phi|_{\vartheta_0} \end{bmatrix},$$

onde $s_\phi = \frac{\partial l}{\partial \phi}$. A matriz de informação (5.3.3.1), denotada por

$$J = E\left[\frac{1}{2} \partial^2 l / \partial \vartheta \partial \vartheta'\right],$$

assume uma forma particionada em bloco:

$$J = \begin{bmatrix} J_{11} & 0 & 0 \\ 0 & J_{22} & J_{23} \\ 0 & J'_{23} & J_{33} \end{bmatrix},$$

onde os elementos da matriz J são dadas em (5.3.3.1), e.g., $J_{11} = E(H_{\beta\beta})$.

A estatística do teste de score é

$$\lambda = \left(s(\vartheta)|_{\vartheta_0} \right)' \left(J|_{\vartheta_0} \right)^{-1} \left(s(\vartheta)|_{\vartheta_0} \right) = \frac{s_{\phi}^2|_{\vartheta_0}}{J_*}, \quad (6.2.4.2)$$

onde $J_* = J_{33}^0 - (J_{23}^0)'(J_{22}^0)^{-1}(J_{23}^0)$ e o índice 0 indica que os elementos de J são calculados em ϑ_0 . Sob condições de regularidade consideradas por Chi e Reinsel (1989), o vetor de escores λ é assintoticamente χ_1^2 sob H_0 .

Nota: o desenvolvimento das derivadas da função de verossimilhança está descrito por Chi (1988) e Chi e Reinsel (1989).

Um valor grande de λ , comparada com a distribuição χ_1^2 , irá sugerir a existência de autocorrelação nos resíduos. Como observado por Chi e Reinsel (1989), embora a estatística dos escores seja válida para padrões irregulares nos tempos de observação (dados desigualmente espaçados), o poder do teste tenderá a crescer se as observações forem tomadas consecutivamente (dados igualmente espaçados).

6.2.5. Teste F da ANOVA

No modelo univariado ou no modelo componente de variâncias, quando a matriz de covariância individual satisfaz a condição de circularidade, pode-se aplicar os testes F da ANOVA (Andrade e Singer, 1986). Por exemplo, o padrão de uniformidade (2.4.3) é uma estrutura circular utilizada em medidas repetidas. Quando se espera a existência de uma estrutura de correlação que decai com o tempo, é preciso tomar certos cuidados. Pela simplicidade na obtenção das estatísticas F da ANOVA, autores com Box (1954) e Geisser e Greenhouse (1958) propuseram correções no número dos graus de liberdade das estatísticas F. Essas correções são utilizadas para acomodar os desvios do padrão circular de Σ .

Nessa seção ilustraremos dois casos simples para construção dos testes F.

Exemplo 6.2.5.1 ANOVA dos modelos dos exemplos 4.3.1 e 4.3.2

Considere o modelo univariado de efeitos fixos descrito no exemplo 4.3.1; a tabela de ANOVA para esse caso é dada pela Tabela 6.2.5.1.

Tabela 6.2.5.1 - ANOVA para o modelo $y_{ij} = \mu + \gamma_i + \epsilon_{ij}$

Fonte de variação	g.l.	Soma de Quadrados	Quadrado Médio	Estatística F
Entre Grupos	$a-1$	SSA	$MSA = SSA/(a-1)$	$F = MSA/MSE$
Dentro do Grupo	$a(n-1)$	SSE	$MSE = SSE/a(n-1)$	
Total	$an-1$	SST		

onde, $SSA = n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2$; $SSE = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2$; $SST = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$

e o uso do ponto em lugar do índice (i, j) refere-se à soma com respeito ao índice; por exemplo:

$$\bar{y}_{..} = \frac{\sum_{i=1}^a \sum_{j=1}^n y_{ij}}{an}$$

Sob suposição de normalidade dos dados, para se testar hipóteses do tipo

$$H_0: \gamma_1 = \dots = \gamma_a,$$

pode-se utilizar a estatística F da Tabela 6.2.5.1 que possui distribuição F com $a-1$ e $n(a-1)$ graus de liberdade, respectivamente, no numerador e no denominador.

Uma questão que surge é: como utilizar a Tabela 6.2.5.1 para o modelo que trata os efeitos como aleatórios? Em primeiro lugar, é preciso levar em consideração a diferença entre os efeitos fixos e aleatórios. Como já se discutiu, no modelo fixo os efeitos correspondem a níveis específicos, cuidadosamente escolhidos, dos fatores de interesse (e.g., variedade de tomates). No modelo aleatório (exemplo 4.3.2), os efeitos correspondem a uma amostra aleatória dos níveis de um fator dentro de uma população. Em outras palavras, no modelo de efeitos fixos, o interesse está em um conjunto particular de efeitos que ocorrem nos dados, enquanto que no modelo de efeitos aleatórios, o interesse está em se fazer inferências sobre a população, em particular, sobre sua variância. Logo, a preocupação no modelo fixo se restringe às médias e no aleatório sobre as variâncias; i.e., no modelo do exemplo 4.3.2, a estatística F deve testar a seguinte hipótese:

$$H_0: \sigma_y^2 = 0.$$

Para casos mais gerais, entretanto, deve-se tomar o cuidado de não utilizar inadequadamente a estatística F, pois a construção de testes para modelos mistos ou aleatórios mais gerais pode não ser tão imediata, principalmente no caso não balanceado (Searle *et alii.*, 1992, pp. 22).

□

Em segundo lugar, deve-se considerar a questão: que parte da tabela ANOVA será utilizada para estimar as componentes de variância? A resposta para essa pergunta é extensiva e foi tratada, e.g., por Searle *et alii.* (1992, cap. 3 a 5). Como ilustração de um modelo misto, considere o seguinte exemplo:

Exemplo 6.2.5.2 ANOVA do modelo 2.4.4

Considere o modelo (2.4.4) do capítulo 2; as estatísticas da ANOVA estão na Tabela 6.2.5.2 e algumas das interpretações na Tabela 6.2.5.3:

A condição necessária e suficiente para que as estatísticas F_2 e F_3 da Tabela 6.2.5.2 tenham distribuição exata F central é: os elementos de Σ , $\sigma_{jj'}$, devem satisfazer

$$\sigma_{jj'} = \begin{cases} d_j + d_{j'} + \varpi, & \text{para } j = j' \\ d_j + d_{j'}, & \text{para } j \neq j' \end{cases} \quad (6.2.5.1)$$

onde d_j , $d_{j'}$ e $\varpi > 0$ são constantes. A forma (6.2.5.1) é conhecido como *condição de circularidade*. O padrão de uniformidade é um caso particular dela, com $d_j = d_{j'} = \sigma^2 \rho/2$ e $\varpi = \sigma^2(1-\rho)$.

Tabela 6.2 5.2 - Análise de variância para o modelo 2.4.4
(assumindo-se que $n_k = r$, $k = 1, \dots, g$)

Fonte de Variação	graus de liberdade	Quadrado Médio	Estatística F
Grupo	$g-1$	QM_1	$F_1 = \frac{(n-g) QM_1}{(g-1) QM_2}$
Indivíduo dentro do Grupo	$N-g$	QM_2	
Ocasião	$T-1$	QM_3	$F_2 = (n-g) \frac{QM_3}{QM_5}$
Grupo \times Ocasião	$(g-1) \times (T-1)$	QM_4	$F_3 = \frac{(g-1) QM_4}{(n-g) QM_5}$
Ocasião \times Indivíduo dentro do Grupo	$(n-g) \times (T-1)$	QM_5	
Total	$nT-1$	QM_6	

onde $QM_1 = \frac{1}{rT} \sum_{k=1}^g y_{..k}^2 - \frac{1}{nT} y_{...}^2$, $QM_2 = \frac{1}{T} \sum_{i=1}^r \sum_{k=1}^g y_{i(k)}^2 - \frac{1}{rT} \sum_{k=1}^g y_{..k}^2$,

$$QM_3 = \frac{1}{n} \sum_{j=1}^T y_{.j}^2 - \frac{1}{nT} y_{...}^2,$$

$$QM_4 = \frac{1}{r} \sum_{k=1}^g \sum_{j=1}^T y_{.kj}^2 - \frac{1}{nT} y_{...}^2 - QM_1 - QM_3,$$

$$QM_5 = QM_6 - \sum_{F=1}^4 QM_F, \quad QM_6 = \frac{1}{rT} \sum_{i=1}^r \sum_{k=1}^g \sum_{j=1}^T y_{i(k)j}^2 - \frac{1}{nT} y_{...}^2$$

O uso do ponto em lugar do índice (i, j, k) refere-se à soma com respeito ao índice, por exemplo,

$$y_{...} = \sum_{i=1}^r \sum_{k=1}^g \sum_{j=1}^T y_{i(k)j}$$

Tabela 6.2.5.3 - Exemplo de algumas interpretações da ANOVA

Efeito	Resultado significativo pode indicar que:
Grupo	há diferenças verticais entre as curvas
Tempo	as curvas não são constantes no tempo
Grupo \times Tempo	a forma das curvas diferem entre os grupos

A correção de Geisser e Greenhouse κ é um fator multiplicativo que pode ser estimado por:

$$\hat{\kappa} = \frac{T^2(\bar{\sigma}_d - \bar{\sigma}_\cdot)^2}{(T-1)(\sum_j \sum_k \sigma_{jk}^2 - 2T \sum_j \bar{\sigma}_j^2 + T^2 \bar{\sigma}_\cdot^2)}, \quad (6.2.5.2)$$

onde T é o número de observações individuais; o termo σ_{jk} é o jk -ésimo elemento da matriz de covariância Σ ; o termo $\bar{\sigma}_j = \sum_k \sigma_{jk} / T$; $\bar{\sigma}_\cdot = \sum_j \sum_k \sigma_{jk} / T^2$; e $\bar{\sigma}_d = \sum_j \sigma_{jj} / T$.

Schwertman e Carter (1993) estudaram o efeito da correlação AR(1) na análise dos dados sob correção de Geisser e Greenhouse e mostraram que é um método prático e efetivo de ajuste. A magnitude dessa correção pode ser utilizada como uma medida da importância da estrutura de correlação na análise dos dados.

Para o caso onde Σ assume uma estrutura geral, Box (1954) e Geisser e Greenhouse (1958) mostraram que as estatísticas F_2 e F_3 da Tabela 6.2.5.2 têm aproximadamente distribuição F central com, respectivamente, $(T-1)\kappa$ e $(n-g)(T-1)\kappa$; e $(g-1)(T-1)\kappa$ e $(n-g)(T-1)\kappa$ graus de liberdade.

Box (1954) mostrou que $1/(T-1) \leq \kappa \leq 1$. O valor $\kappa = 1$ corresponde à condição de circularidade e à medida que Σ se afasta desse padrão, o valor de κ tende para $1/(T-1)$, que é a redução máxima no número dos graus de liberdade. Quando se utiliza $\kappa = 1$ ou $1/(T-1)$, o resultado pode ser um teste, respectivamente, muito liberal ou muito conservativo. Um procedimento em três estágios foi proposto por Geisser e Greenhouse (1958):

- i) usa-se o teste F exato com $\kappa = 1$; caso o teste seja significativo (rejeição da hipótese nula) procede-se o próximo estágio;
- ii) usa-se o teste F exato com $\kappa = 1/(T-1)$; caso o teste seja não significativo passa-se para o próximo estágio;
- iii) usa-se o teste F aproximado com κ estimado a partir de (6.2.5.2).

Quando a matriz Σ assume o padrão AR(1) na forma (3.2.1.2), o fator de correção tem o limite inferior $5(T+1)/(2T^2+7)$ que é maior do que $1/(T-1)$ para $T > 2$ (Andrade e Singer, 1986, pp. 41). Nesse caso, o segundo estágio do procedimento anterior é feito com $\kappa = 5(T+1)/(2T^2+7)$.

6.3. Diagnóstico

Após o ajuste do modelo é necessário checar se as suposições e ajuste são adequados através, por exemplo, da análise dos resíduos. Como os métodos e modelos discutidos nesse trabalho assumiram em sua grande maioria normalidade, é preciso considerar meios para verificar essa suposição. A falha na suposição de normalidade não invalida as estimativas dos parâmetros de locação. Entretanto, pode invalidar os testes usuais e os intervalos de confiança que são baseados na teoria da distribuição Normal. De modo geral, tanto a análise de resíduos como a bondade de ajuste podem ser realizadas a partir de técnicas e testes conhecidos na literatura, como por exemplo, a validação cruzada (*cross-validation*) e o teste Qui-quadrado. Como referência podemos citar Draper e Smith (1981, cap. 3, pp. 141) e Searle (1971, pp. 129) que tratam do diagnóstico. Nesta fase a análise (principalmente gráfica) dos resíduos tem importância fundamental. Todas as técnicas tradicionais de análise utilizadas em análise de regressão podem ser consideradas aqui; entretanto, como já são amplamente conhecidas (e.g., Draper e Smith, 1981, cap. 3, pp. 141; e Searle, 1971, pp. 129) elas não serão discutidas aqui. Nessa seção discutiremos apenas algumas técnicas mais específicas como o uso dos resíduos transformados, o gráfico de probabilidade normal ponderado e o método do semivariograma empírico para o diagnóstico do modelo de efeitos mistos.

O resíduo ordinário, que essencialmente é a diferença do valor observado com o ajustado, possui deficiências como a existência de correlação, i.e., a matriz de covariância dos resíduos ordinários não é necessariamente diagonal (são heterocedásticos). Os resíduos transformados (seção 6.3.1), por outro lado, são independentes e identicamente distribuídos (iid). Os resíduos transformados podem ser usados para checar as suposições de normalidade e homocedasticidade do modelo através, por exemplo, do gráfico de probabilidade normal (seção 6.3.2). Galpin e Hawkins (1984) observam que em comparação aos resíduos ordinários, os recursivos são mais sensíveis às mudanças de padrões e de valores aberrantes (*outliers*).

No modelo de efeitos mistos, para se checar a suposição de normalidade dos efeitos aleatórios, o gráfico de probabilidade normal usual (GPN) atribui pesos iguais para todas as observações, i.e., considera-se que as quantidades de informações disponíveis para cada indivíduo são as mesmas, o que não ocorre no caso de dados não balanceados. O GPN ponderado a ser apresentado na seção 6.3.2 atribui pesos às observações e são mais sensíveis aos desvios em relação à distribuição assumida aos efeitos aleatórios.

Na seção 6.3.3 apresentaremos o método do semivariograma empírico para avaliar a estrutura de correlação dos resíduos. Essa avaliação pode ser utilizada para a seleção inicial

das estruturas de covariância do modelo, indicando, por exemplo, existência de correlação temporal nos erros.

6.3.1. Resíduos

Após o ajuste do modelo de efeitos fixos, os valores ajustados ou preditos para a curva média populacional, dado um vetor de covariáveis x , são dados por

$$\hat{y} = x\hat{\beta}, \quad (6.3.1.1)$$

e a variância estimada é

$$\text{vâr}(\hat{y}) = x\text{vâr}(\hat{\beta})x', \quad (6.3.1.2)$$

onde $\text{vâr}(\hat{\beta})$ pode ser obtida substituindo-se Σ pela sua estimativa em (5.2.1.32).

No modelo de efeitos mistos, permite-se fazer predições individuais. Considere a predição para um novo indivíduo introduzido no conjunto de dados e que não tenha ainda nenhum dado coletado ou com observações perdidas. Como as estimativas a priori dos efeitos aleatórios são nulas, a predição para esse indivíduo será da mesma forma que (6.3.1.1), mas a sua variância estimada será dada por

$$\text{vâr}(\hat{y}) = x\text{vâr}(\hat{\beta})x' + z\Gamma z' + \hat{\sigma}_e^2, \quad (6.3.1.3)$$

onde z é uma linha da matriz de delineamento dos efeitos aleatórios e os dois últimos termos representam, respectivamente, a contribuição das componentes de variância entre indivíduos e dentro do indivíduo.

Agora considere a predição para um indivíduo que têm observações disponíveis. Isso significa que há informação sobre os efeitos aleatórios que pode ser utilizada na predição, reduzindo-se a variância estimada (6.3.1.5) quando comparada com (6.3.1.3). A predição será dada por

$$\hat{y} = x\hat{\beta} + z\hat{b}_1, \quad (6.3.1.4)$$

e a variância estimada dessa predição é

$$\text{vâr}(\hat{y}) = x\text{vâr}(\hat{\beta})x' + z\text{vâr}(\hat{b}_1 - b_1)z' + \hat{\sigma}_e^2, \quad (6.3.1.5)$$

No caso do BLUP, quando Θ é conhecido, um intervalo de predição $100(1-\alpha)\%$ será

$$\hat{y} - t_{\alpha_2(T-a)}\sqrt{\text{vâr}(\hat{y})} \leq y \leq \hat{y} + t_{\alpha_1(T-a)}\sqrt{\text{vâr}(\hat{y})}, \quad (6.3.1.6)$$

onde $\alpha = \alpha_1 + \alpha_2$, com α_1 e α_2 positivos e $t_{\alpha_1(T-a)}$ denota o $(1-\alpha_1)$ -ésimo quantil da distribuição t de Student com $T-a$ graus de liberdade.

Quando utilizamos o BLUP empírico, denotado por $\hat{y}_{\hat{\Theta}}$, a distribuição de $(\hat{y}_{\hat{\Theta}} - y) / \sqrt{\text{vâr}(\hat{y}_{\hat{\Theta}})}$ poderá ter caudas mais pesadas do que a distribuição t com $T-a$ graus de liberdade. Isso ocorre, segundo Harville e Carriquiry (1992) por dois motivos: a variância de $\hat{y}_{\hat{\Theta}}$ tende a subestimar $\text{var}(\hat{y})$ e os graus de liberdade d da distribuição Qui-quadrado que melhor aproxima a distribuição de $d\text{vâr}(\hat{y}_{\hat{\Theta}}) / E[\text{vâr}(\hat{y}_{\hat{\Theta}})]$ pode ser maior do que $T-a$. Conseqüentemente, a probabilidade real de cobertura do intervalo de predição do BLUP empírico tenderá a ser menor do que $1-\alpha$.

O resíduo ordinário pode ser definido como a diferença entre o valor observado e o valor ajustado (ou predito), i.e.,

$$r_{it} = y_{it} - \hat{y}_{it}. \quad (6.3.1.7.a)$$

No modelo de efeitos mistos, o valor ajustado pode envolver a predição dos efeitos aleatórios. Assim,

$$r_{it} = y_{it} - x_{it}\hat{\beta} - z_{it}\hat{b}_1. \quad (6.3.1.7.b)$$

Quando a predição não é considerada, o ajuste do modelo envolve apenas os efeitos fixos estimados:

$$r_{it} = y_{it} - x_{it}\hat{\beta}. \quad (6.3.1.7.c)$$

Similarmente, os resíduos longitudinais podem ser definidos como os desvios das observações sucessivas de seus valores esperados, dadas as respostas prévias e os valores das covariáveis incluídas no modelo linear. Por exemplo, no caso considerado por Ware (1985),

$$y_{it} = X_{it}\beta + \varepsilon_{it}, \text{ com } \varepsilon_{it} = \phi\varepsilon_{it-1} + a_{it},$$

que pode ser reescrito como

$$y_{it} = \phi y_{it-1} + (X_{it} - \phi X_{it-1})\beta + a_{it};$$

assim, os resíduos longitudinais têm a forma

$$r_{it} = y_{it} - \phi y_{it-1} - (X_{it} - \phi X_{it-1})\beta, \text{ para } t = 2, \dots, T_i. \quad (6.1.3.8)$$

Embora os resíduos longitudinais possam ser descritos analogamente em modelos multivariados ou modelos de efeitos aleatórios, suas formas variam de ocasião para ocasião, e possivelmente de indivíduo para indivíduo, fazendo com que a análise de resíduos seja mais difícil de ser interpretada.

Os resíduos ordinários dados por (6.3.1.7) têm deficiências como existência de correlação e, em geral, nem todos têm a mesma variância (i.e., não são iid). Por exemplo, no modelo de regressão $y = X\beta + \varepsilon$, se $\varepsilon \sim N(0, \sigma^2 I)$, então, assumindo-se que o modelo está correto, a distribuição do vetor de resíduos cujos elementos são definidos por (6.3.1.7) será $N[0, \sigma^2 (I - X(X'X)^{-1}X')]$, i.e., a distribuição dos resíduos depende do padrão da matriz de delineamento X . Como o conjunto dos resíduos ordinários é resultado da mistura de distribuições normais, é possível que, por exemplo, o gráfico de probabilidade normal acuse não normalidade. Os resíduos ordinários também podem mascarar a existência de valores aberrantes (Galpin e Hawkins, 1984).

Para se evitar as deficiências dos resíduos ordinários, os resíduos transformados podem ser usados para testar o ajuste e as suposições do modelo. Essencialmente, os resíduos transformados, que são iid, resultam de uma transformação linear dos resíduos ordinários, i.e., são definidos por

$$w_i = L r_i, \quad (6.3.1.9)$$

onde r_i é o vetor de resíduos ordinários do indivíduo i definido por (6.3.1.7.c) e L é a matriz triangular superior da decomposição de Cholesky da inversa da matriz de covariância individual (Σ_i^{-1}), i.e., $\Sigma_i^{-1} = L'L$. Assim,

$$E(w_i w_i') = E(L r_i r_i' L') = L E(r_i r_i') L' = L' \Sigma_i L = I,$$

$$\text{pois } \Sigma_i^{-1} = L'L \Rightarrow \Sigma_i = L^{-1}(L^{-1})' \Rightarrow L' \Sigma_i L = I.$$

Os resíduos transformados podem ser obtidos através do filtro de Kalman (seção 5.3). (Na realidade, como Jones e Boateng (1991) observaram, o filtro de Kalman é uma implementação recursiva do cálculo da decomposição de Cholesky).

Como ilustração dos resíduos transformados, considere o seguinte exemplo:

Exemplo 6.3.1 resíduos transformados

Considere o modelo $y_i = 1\beta + \varepsilon_i$ com os erros seguindo um processo AR(1), i.e., $\Sigma_i = \sigma^2 W_i$, onde W_i é a matriz dada por (3.1.2). Tome as estimativas de MV dos parâmetros do modelo $\hat{\beta}$, $\hat{\sigma}^2$ e $\hat{\phi}$, onde $\hat{\phi}$ é o parâmetro de autoregressão estimado. A inversa de W_i é uma matriz tridiagonal; i.e.,

$$\Sigma_i^{-1} = \sigma^{-2} W_i^{-1} = \frac{1}{\sigma^2 (1 - \hat{\phi}^2)} \begin{bmatrix} 1 & -\hat{\phi} & 0 & \dots & 0 & 0 \\ -\hat{\phi} & 1 + \hat{\phi}^2 & -\hat{\phi} & \dots & 0 & 0 \\ 0 & -\hat{\phi} & 1 + \hat{\phi}^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\hat{\phi} & 1 \end{bmatrix}$$

Assim, a decomposição de Cholesky de $\Sigma_i^{-1} = \sigma^{-2} W_i^{-1}$ resulta na matriz triangular superior

$$L = \frac{1}{\sigma \sqrt{1 - \hat{\phi}^2}} \begin{bmatrix} \sqrt{1 - \hat{\phi}^2} & 0 & 0 & \dots & 0 & 0 \\ -\hat{\phi} & 1 & 0 & \dots & 0 & 0 \\ 0 & -\hat{\phi} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\hat{\phi} & 1 \end{bmatrix}$$

Multiplicando-se a matriz L com o vetor dos resíduos ordinários ($w_i = L r_i$) obtemos os resíduos transformados; como os elementos da matriz L é estimada por $\hat{\sigma}^2$ e $\hat{\phi}$, então:

$$w_{i1} = \frac{r_{i1}}{\hat{\sigma}}, w_{i2} = \frac{r_{i2} - \hat{\phi} r_{i1}}{\hat{\sigma} \sqrt{1 - \hat{\phi}^2}}, w_{i3} = \frac{r_{i3} - \hat{\phi} r_{i2}}{\hat{\sigma} \sqrt{1 - \hat{\phi}^2}}, \dots, w_{i,t-1} = \frac{r_{i,t-1} - \hat{\phi} r_{i,t-2}}{\hat{\sigma} \sqrt{1 - \hat{\phi}^2}} \text{ e } w_{i,t} = \frac{r_{i,t} - \hat{\phi} r_{i,t-1}}{\hat{\sigma} \sqrt{1 - \hat{\phi}^2}}.$$

□

6.3.2. O Gráfico de Probabilidade Normal

O gráfico de probabilidade normal (GPN) é um método gráfico para checar a suposição de normalidade do modelo misto. Para introduzir o GPN, considere o modelo

$$y_i = b_i + \varepsilon_i, \quad (6.3.2.1)$$

onde b_i e ε_i são efeitos aleatórios independentes com distribuições, respectivamente, $N(\beta, \sigma^2)$ e $N(0, \sigma_i^2)$.

Sob o modelo (6.3.2.1), as variáveis y_i são iid com média β e variância $\sigma^2 + \sigma_i^2$. Logo, quando o modelo especificado é correto, as variáveis padronizadas

$$z_i = \frac{y_i - \beta}{\sqrt{\sigma^2 + \sigma_i^2}}$$

formam uma amostra independente de uma população normal padrão, e a bondade de ajuste pode ser verificada construindo-se o GPN de z_1, \dots, z_n ; o gráfico pode ser feito plotando-se os valores de z_i contra $\Phi^{-1}[F_n(z_i)]$, onde $\Phi[x]$ é o valor da função de distribuição acumulada da normal padrão no ponto x e $F_n(x)$ é a função de distribuição empírica acumulada dada por

$$F_n(x) = \frac{\sum_{i=1}^n I(x - z_i)}{n}, \quad (6.3.2.2)$$

onde $I(x) = 1$ se $x \geq 0$ e 0 caso contrário.

Como observado por Dempster e Ryan (1985) e Lange e Ryan (1989) e estudado por Guimarães (1994, pp. 60), a desvantagem da formulação de (6.3.2.2) é que atribui-se para cada observação os mesmos pesos. Como a quantidade de informações sobre os efeitos aleatórios pode diferir entre os indivíduos, e.g. no caso de dados não balanceados, a atribuição de pesos para cada informação torna-se importante. Assim, a função de distribuição empírica ponderada acumulada é dada por

$$F_n^*(x) = \frac{\sum_{i=1}^n I(x - z_i) P_i}{\sum_{i=1}^n P_i}, \quad (6.3.2.3)$$

onde P_i são os pesos que são funções da variância. Por exemplo, para o modelo (6.3.2.1), uma escolha simples para os pesos é

$$P_i = \frac{1}{\sigma^2 + \sigma_i^2},$$

que satisfaz

$$\frac{\sum_{i=1}^n P_i z_i^2}{\sum_{i=1}^n P_i} = 1.$$

Considere o agora o modelo de efeitos mistos; se o objetivo for avaliar a normalidade dos efeitos aleatórios (ou da combinação linear dos efeitos aleatórios), as variáveis padronizadas podem ser construídas fazendo-se

$$z_i = \frac{c' \hat{b}_i}{\sqrt{c' \text{var}(\hat{b}_i) c}},$$

onde c é um vetor de constantes. A função empírica é dada por (6.3.2.3), e os pesos podem ser dados por

$$P_i = c' \text{var}(\hat{b}_i) c.$$

As respectivas regiões críticas de $(1-\alpha)\%$ do GPN e do GPNP para o ponto $\Phi^{-1}[F_n(x)]$ e $\Phi^{-1}[F_n^*(x)]$, são aproximadamente

$$x \pm \frac{\Phi^{-1}(1-\alpha) \{ \Phi(x) [1 - \Phi(x)] / n \}^{\frac{1}{2}}}{\phi(x)}$$

$$x \pm \frac{\Phi^{-1}(1-\alpha) \{ (1 + v/\eta^2) \Phi(x) [1 - \Phi(x)] / n \}^{\frac{1}{2}}}{\phi(x)},$$

onde $\eta = \frac{\sum_{i=1}^n P_i}{n}$, $v = \frac{\sum_{i=1}^n P_i^2}{n} - \eta^2$ e $\phi(x)$ é o valor da função de densidade da normal padrão no ponto x . Nota-se que o fator $(1 + v/\eta^2)$ aumenta a largura da região de confiança.

Como Dempster e Ryan (1985) observaram, o GPNP deve ser ajustado para os pontos finais, i.e., $\Phi^{-1}[F_n(x)]$ não pode ser calculado para valores grandes de z_i . Lange e Ryan (1989) também sugerem correções para a região crítica pelo uso de estimativas ao

invés dos parâmetros verdadeiros. Nesse trabalho, entretanto, essas correções não serão implementadas.

Para ilustrar algumas interpretações do GPN ou do GPNP, considere a Figura 6.3.2.1. Se as suposições de normalidade são satisfeitas, a forma do gráfico deve ser uma linha reta como mostra (a). Se os valores padronizados têm uma distribuição com caudas menos pesadas que as da distribuição normal, então a curva terá uma forma de "S" como mostra (b). Se, por outro lado, a distribuição tiver caudas mais pesadas que as da distribuição normal, então a forma de "S" da curva será invertida (c). No caso de assimetria, a curva será a combinação de (b) e (c), resultando em (d). Em (e) há uma reta e alguns pontos na extremidade fora dela, o que deve indicar presença de valores aberrantes (*outliers*). Em (f) a reta está deslocada do ponto central (0,0), isso pode indicar que a média das variáveis padronizadas é diferente de zero.

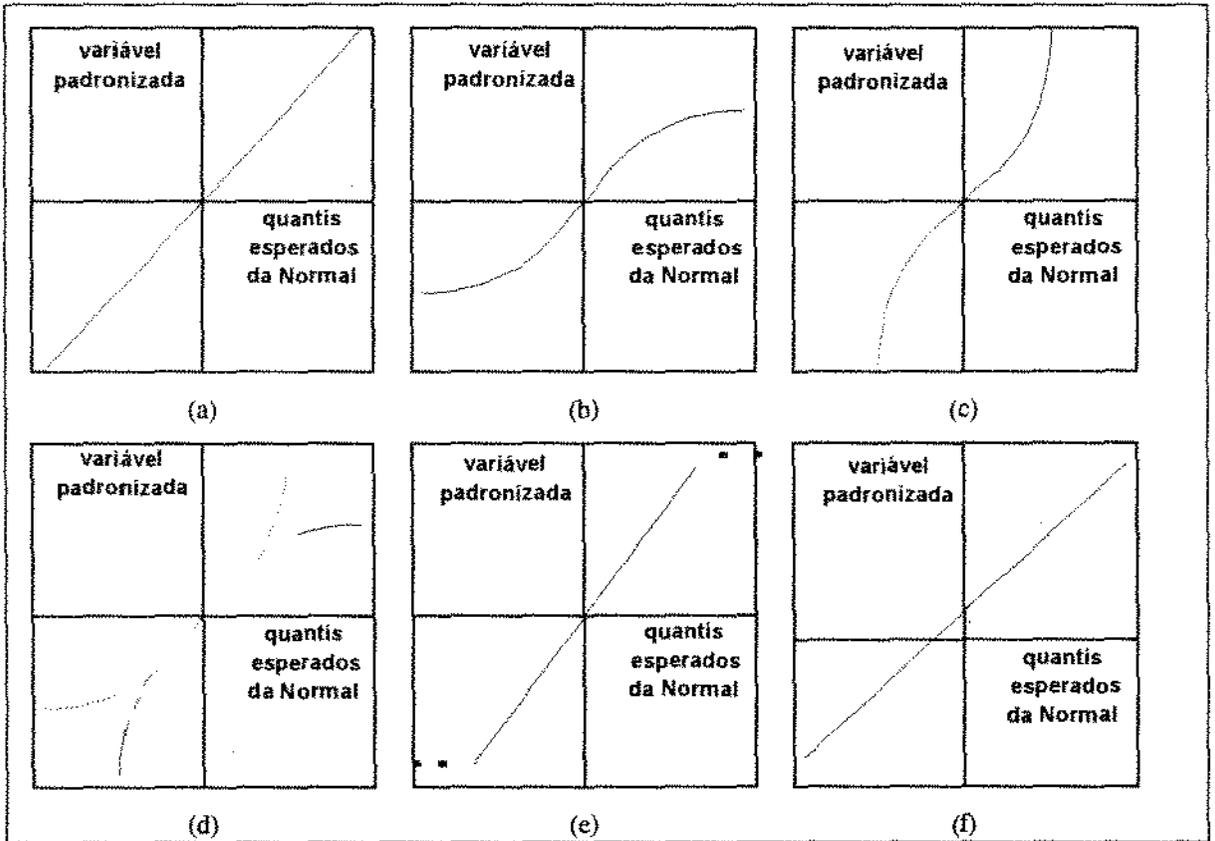


Fig. 6.3.2.1 - exemplos de GPN

A Figura 7.2.9 do capítulo 7 é um exemplo da situação (e) da Figura 6.3.2.1, onde os pontos da extremidade indicam pontos aberrantes. A Figura 7.2.12 ilustra a situação (a), onde os pontos estão dentro dos limites de confiança de 95%.

6.3.3. O Semivariograma

O semivariograma pode ser utilizado para avaliar a estrutura de correlação dos resíduos transformados. Essa avaliação pode ser utilizada para a seleção inicial das estruturas de covariância do modelo (Diggle, 1988), indicando, por exemplo, existência de correlação temporal nos erros, ou também pode ser usado para verificar se os resíduos longitudinais do modelo final são "ruídos brancos" (*white noise*).

O semivariograma de um processo aleatório $\{Y(t): t \in \mathfrak{R}\}$ é a função

$$\gamma(u) = E\left[\frac{1}{2}\{Y(t) - Y(t-u)\}^2\right],$$

onde $u \geq 0$ representa a defasagem (lag). Note que se o processo for estacionário com a função de covariância dada por $C(u) = \text{cov}\{Y(t), Y(t-u)\}$, então $\gamma(u) = C(0) - C(u)$.

O semivariograma empírico de uma série temporal $\{\hat{y}(t_j): j=1, \dots, T\}$ é um gráfico das diferenças quadradas $d_{jk}^2 = \frac{1}{2}[y(t_j) - y(t_k)]^2$ contra as respectivas quantidades $u_{jk} = t_j - t_k$. O semivariograma empírico é dado por:

$$\hat{\gamma}(u_{jk}) = \frac{\sum_{(j,k)} d_{jk}^2}{n_{jk}},$$

onde n_{jk} é o número de pares de pontos distanciados de u_{jk} (intervalo de tempo). Assim, por exemplo (Uzumaki, 1994, pp. 17), se $j = 0, 1, 2, 3$, então o primeiro *lag* usará os pares de distância 1: (0,1), (1,2) e (2,3); o segundo *lag* os pares de distância 2: (0,2) e (1,3); e o terceiro o par de distância 3: (0,3). Se o processo for gaussiano, a distribuição amostral de cada d_{jk}^2 é proporcional ao χ_1^2 (Diggle, 1988).

Exemplo 6.3.3.1 algumas formas de covariância utilizadas em geoestatística

Seja um processo estocástico de covariância estacionário (estacionário de 2ª ordem), com média $m = E[Y(t)]$, então a função de covariância, $C(u)$, é dada por

$$C(u) = E\{Y(t+u)Y(t)\} - E^2\{Y(t)\}, \quad \forall t \in \mathfrak{R}.$$

A estacionariedade de 2ª ordem implica a estacionariedade da variância e do semivariograma, e $\sigma^2 = \text{Var}[Y(t)] = E\{[Y(t) - m]^2\} = C(0)$, que é chamado de *patamar*.

Como exemplo de estruturas de covariância estacionários (os nomes das estruturas desse exemplo são utilizados em geoestatística) podemos citar (Journel e Huijbregts, 1978):

i) Modelo esférico:

$$C(r) = \begin{cases} \sigma^2 \left[1 - \frac{3}{2a}r + \frac{1}{2a^3}r^3 \right] & 0 \leq r \leq a \\ 0 & r > a \end{cases}$$

onde: a é chamado de *amplitude* (lag onde se tem o patamar); e $\sigma^2 = \text{Var}[Y(t)] = C(0)$ é chamado de patamar. A Figura 6.3.3.1 ilustra o gráfico do semivariograma correspondente à função de covariância do modelo esférico com patamar igual a 1, segundo a relação $\gamma(r) = C(0) - C(r)$.

ii) Modelo exponencial (equivale ao AR(1)):

$$C(r) = \sigma^2 \exp\left(-\frac{1}{a}r\right), \quad a > 0;$$

está ilustrado na Figura 6.3.3.2. Nesse modelo o "*a efetivo*" ocorre em $3a$.

iii) Modelo gaussiano:

$$C(r) = \sigma^2 \exp\left(-\frac{1}{a^2}r^2\right), \quad a > 0$$

está ilustrado na Figura 6.3.3.3. Nesse caso o "*a efetivo*" ocorre em $a\sqrt{3}$.

iv) Ausência de correlação:

$$C_0(r) = \begin{cases} \sigma^2 & r = 0 \\ 0 & \forall |r| > \varepsilon \end{cases}$$

onde ε é um valor tão pequeno quanto se queira. Essa estrutura é equivalente ao "ruído branco" (Fig. 6.3.3.4).

□

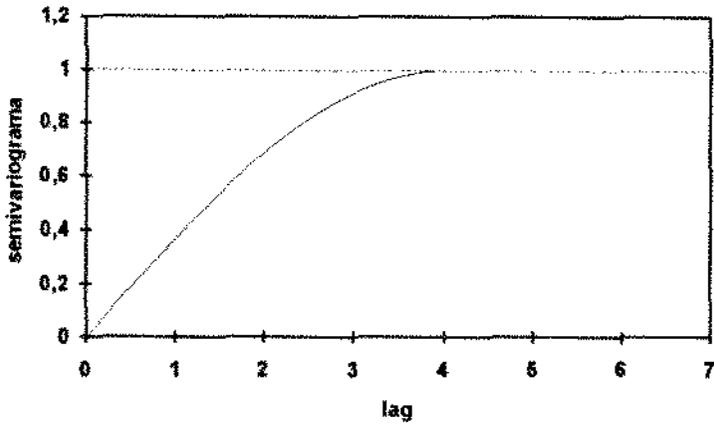


Fig. 6.3.3.1 - gráfico do semivariograma: modelo esférico

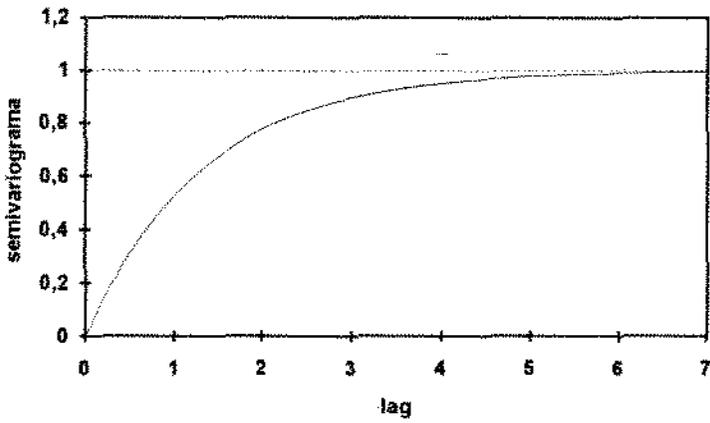


Fig. 6.3.3.2 - gráfico do semivariograma: modelo exponencial

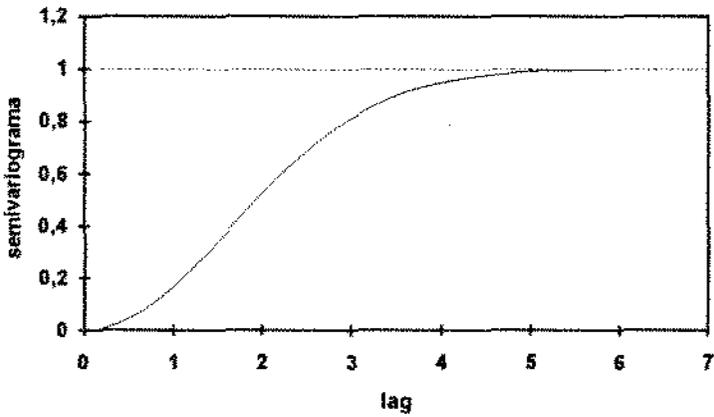


Fig. 6.3.3.3 - gráfico do semivariograma: modelo gaussiano

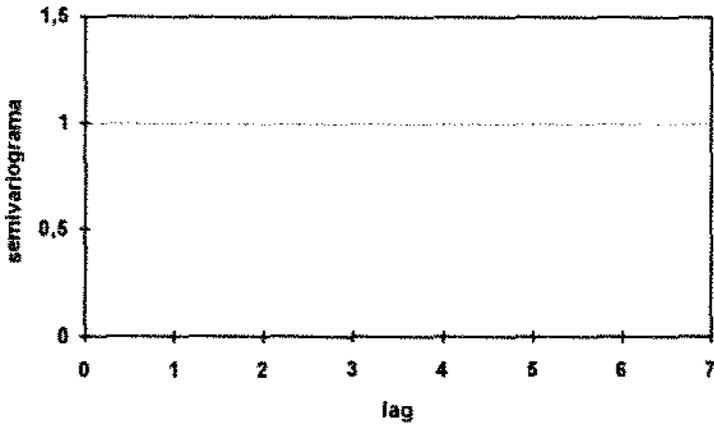


Fig. 6.3.3.4 - gráfico do semivariograma: ruído branco

Em dados longitudinais o semivariograma é utilizado, em geral, como forma de diagnóstico da estrutura de correlação (Diggle, 1988). O uso do mesmo para identificação da estrutura é mais complexo porque a existência de uma estrutura de correlação pode estar associada aos efeitos aleatórios, aos erros correlacionados ou ambos. Como exemplo de aplicação, o semivariograma dos resíduos transformados da Figura 7.2.7 do capítulo 7 indica a existência de uma estrutura de correlação. Ao modelarmos os erros como um processo AR(1), o semivariograma dos resíduos transformados da Figura 7.2.15 indica um "patamar" ao longo dos lags.

7

Aplicação

7.1. Introdução

Nesse capítulo aplicaremos ilustrativamente algumas das técnicas de análise estudadas, tomando-se como exemplo o conjunto de dados experimentais apresentado na seção 1.2.

O conjunto de dados é formado por dois grupos de ratos: ratos treinados (grupo 1) e ratos sedentários (grupo 2). Os ratos são sujeitos à atividade física durante 1 hora, observando-se o nível individual de ácido láctico. As observações foram feitas de 10 em 10 minutos (incluindo o instante inicial) e estão registradas nas Tabelas 1.2.1 e 1.2.2. Cada i -ésima linha dessas tabelas representa as observações do i -ésimo rato tomadas no tempo ($j = 1, \dots, 7$) de forma igualmente espaçada (a cada 10 minutos).

Entre os objetivos na análise podemos considerar:

- a) modelagem e comparação das curvas de resposta (nível de ácido láctico) dos 2 grupos. O modelo de efeitos mistos permite estudar tanto o padrão de resposta dos grupos e compará-los, como permite estudar o padrão individual de resposta e compará-lo com os outros indivíduos (Análise I, seção 7.2). Nessa parte ajustou-se várias formas do modelo misto descrito no exemplo 2.5.6, considerando-se no máximo 3 efeitos aleatórios com uma matriz de covariância não estruturada ou diagonal (efeitos não correlacionados) com erros independentes ou AR(1). A estratégia da modelagem está descrita na seção 7.2.
- b) estudo e comparação das várias estruturas de correlação possíveis para as componentes aleatórias individuais (Γ e Ω_i). Entre as estruturas possíveis incluem-se: estrutura de independência ($\sigma^2 I$), simetria composta, não estruturada, AR(1) e Toeplitz (Análise II, seção 7.3). A formulação do modelo nessa parte difere um pouco daquela adotada na análise I (modelo misto descrito no exemplo 3.3.3, considerando-se os efeitos aleatórios dependentes do tempo com uma matriz de covariância estruturada). Como o interesse nessa seção é estudar o ajuste utilizando as formas citadas para as matrizes de covariância, consideraremos apenas um modelo mais geral (saturado) sem considerar os modelos reduzidos.

Os cálculos desse capítulo foram realizados através do PROC MIXED (SAS Institute), utilizado na análise II, e do programa CARI (Jones e Boateng, 1991), utilizado na análise I. O CARI, apesar de não ser tão sofisticado quanto o SAS, é um programa útil que permite ajustar modelos polinomiais do exemplo 2.5.6 com observações desigualmente espaçadas (diferentes instantes de observações para diferentes indivíduos). Os erros individuais são assumidos independentes ou seguem um processo AR(1), com a opção de se incluir erros observacionais. Esse programa suporta apenas a matriz de covariância não estruturada ou diagonal para os coeficientes aleatórios. Os métodos de estimação disponíveis nesse programa são a máxima verossimilhança (MV) e a máxima verossimilhança restringida (MVRe). O programa utiliza o filtro de Kalman para calcular a verossimilhança. As estimativas dos parâmetros não lineares são calculadas por uma rotina de otimização não linear.

O PROC MIXED (SAS Institute) implementa de forma abrangente a metodologia do modelo misto, permitindo análises de medidas repetidas, delineamentos *split-plot*, componentes de variância e coeficientes aleatórios. O MIXED permite várias estruturas de correlação para as matrizes Γ e Ω . Para ambas, dentre as estruturas possíveis estão incluídas: estrutura de independência ($\sigma^2 I$), simetria composta, não estruturada, AR(1) e Toeplitz. O MIXED permite trabalhar com dados perdidos de forma mais apropriada que o PROC GLM, por exemplo. Os métodos de estimação disponíveis são a MV, MVRe e o MIVQUE, implementados através do algoritmo de Newton-Raphson. Nas análises I e II restringiremo-nos aos métodos de EMV e EMVRe.

7.2. Análise I

Como foi descrito no exemplo 2.5.6, o modelo polinomial de efeitos mistos pode ser representado individualmente como:

$$y_{i(1)t} = (\beta_{01} + \beta_{11}t + \dots + \beta_{c-1,1}t^{c-1}) + (b_{0i} + b_{1i}t + \dots + b_{m-1,i}t^{m-1}) + \varepsilon_{i(1)t},$$

$$y_{i(2)t} = (\beta_{01} + \beta_{11}t + \dots + \beta_{c-1,1}t^{c-1}) + (\beta_{02} + \beta_{12}t + \dots + \beta_{d-1,2}t^{d-1}) + (b_{0i} + \dots + b_{m-1,i}t^{m-1}) + \varepsilon_{i(2)t},$$

onde $y_{i(1)t}$ e $y_{i(2)t}$ denotam os níveis de ácido láctico dos i -ésimos ratos, respectivamente, do grupo 1 e do grupo 2 no instante t . Os parâmetros lineares β_{01} a $\beta_{c-1,1}$ descrevem a curva do grupo de referência (ratos treinados) e os restantes, β_{02} a $\beta_{d-1,2}$, descrevem os efeitos diferenciais do grupo 2 ao grupo 1. As componentes aleatórias $(b_{0i}, \dots, b_{m-1,i})$ descrevem a variação entre os indivíduos, assumindo-se que tenham distribuição $N(0, \sigma^2 B)$ e os erros $(\varepsilon_{i(1)0}, \dots, \varepsilon_{i(1)60})$ descrevem a variação dentro do indivíduo, assumindo-se que tenham distribuição $N(0, \sigma^2 W_i)$. A componente de variação σ^2 é chamada de componente de variação dentro do indivíduo e os elementos de B são as componentes de variação entre os indivíduos.

Nessa seção, utilizaremos até 3 efeitos aleatórios (uma possível interpretação desses efeitos será dada na seção 7.2.5). O estudo será restrito no caso onde a matriz B não é estruturada com $m(m+1)/2$ parâmetros ($m = 0, 1, 2, 3$) ou B é diagonal com m parâmetros; e os erros ou são independentes ($W_i = I$) ou seguem um processo AR(1). Assim, por exemplo, considerando-se 3 efeitos aleatórios com uma matriz de covariância não estruturada teremos:

$$z b_i = \begin{bmatrix} 1 & t & t^2 \end{bmatrix} \begin{bmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \end{bmatrix},$$

com Γ não estruturada:
$$\Gamma = \text{var}(b_i) = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{12} & \gamma_{22} & \gamma_{23} \\ \gamma_{13} & \gamma_{23} & \gamma_{33} \end{bmatrix},$$

ou diagonal:
$$\Gamma = \text{var}(b_i) = \begin{bmatrix} \gamma_{11} & 0 & 0 \\ 0 & \gamma_{22} & 0 \\ 0 & 0 & \gamma_{33} \end{bmatrix}.$$

A estratégia de seleção do modelo consiste basicamente de 3 partes (Diggle, 1988):

- i) Estimar um modelo preliminar (modelo saturado) com erros independentes e com um efeito aleatório b_{i0} . Esse estágio envolve uma transformação apropriada dos dados para se obter um padrão homocedástico nos resíduos. Pelos gráficos dos resíduos padronizados individuais do modelo inicial (Fig. 7.2.1 e Fig. 7.2.2), nota-se que as respostas de ácido são mais estáveis no início do experimento, tornando-se cada vez mais variáveis ao longo do experimento. Uma forma de se contornar isso é transformar os dados, por exemplo, tomando-se o logaritmo dos níveis de ácido láctico. A transformação dos dados possivelmente produz um padrão mais satisfatório nos resíduos (Fig. 7.2.5 e Fig. 7.2.6), atenuando o efeito da heterocedasticia temporal. Mesmo assim, ainda se observa alguma heterogeneidade temporal. Entretanto, como será discutido na seção 7.2.5, essa heterogeneidade ("aparente", segundo Jones, 1990) pode ser em parte explicada pelos efeitos aleatórios. Uma alternativa à transformação é incorporar a heterocedasticia ao modelo, introduzindo-o como uma covariável dependente do tempo como será estudado na análise II.
- ii) Seleção inicial das estruturas de covariância para Γ e Ω . Na análise I considera-se somente três casos: Γ não estruturada com erros independentes ($\Omega = \sigma^2 I$), Γ não estruturada com erros AR(1) e Γ diagonal com erros AR(1).
- iii) Esse estágio utiliza os critérios de seleção de modelo para a escolha do modelo e a análise de diagnóstico e resíduos.

A Tabela 7.2.1 explica as notações adotadas para representar as estruturas de covariância de Γ e Ω :

Tabela 7.2.1 - Notações para representar as estruturas de covariância

matriz	nome da estrutura	notação
Γ	não estruturada (<i>unstructured</i>)	UN
Γ	diagonal	DIAG
Γ ou Ω	independente (<i>simple</i>)	SIM
Ω	AR(1)	AR1

7.2.1. Investigação Inicial

Inicialmente ajustou-se o modelo saturado para os efeitos fixos ($c = d = 7$) com um coeficiente aleatório ($m = 0$) pelos métodos de estimação MV e MVRe. Plotando-se os resíduos transformados individuais de MV e MVRe, respectivamente, nas Figuras 7.2.1 e 7.2.2, observa-se que a variação residual no início (0 min) é pequena, diferindo bastante do padrão de variação dos demais instantes (10 a 60 min). Uma alternativa para tornar o padrão de variação mais homogêneo é transformar os dados, tomando-se o logaritmo natural dos níveis de ácido ($\ln y_{i(k)}$).

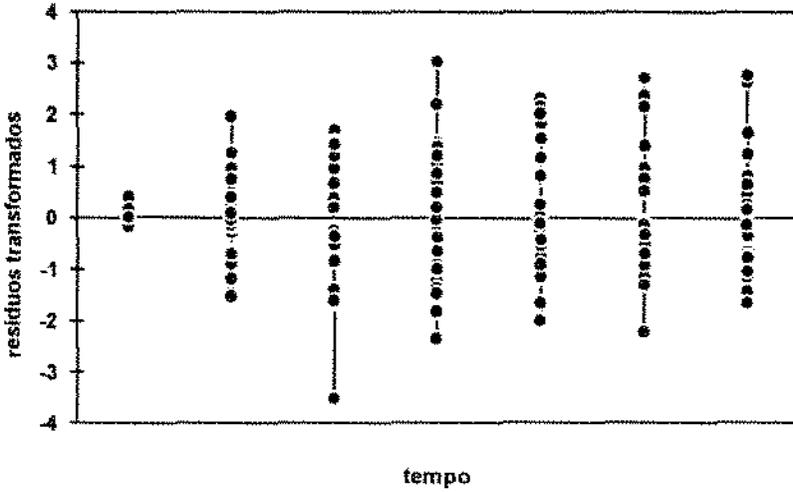


Fig. 7.2.1 - gráfico dos resíduos transformados dos 36 ratos do modelo inicial ajustado por MV

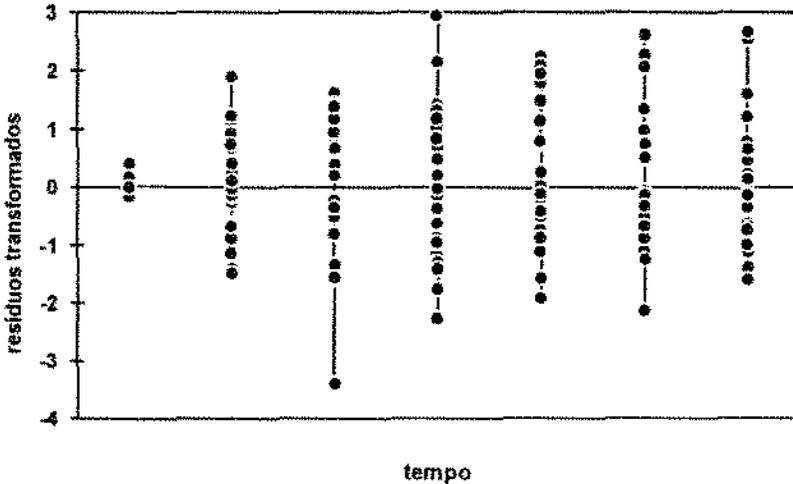


Fig. 7.2.2 - gráfico dos resíduos transformados dos 36 ratos do modelo inicial ajustado por MVRe

7.2.2. Transformação dos Dados

Os dados transformados (logaritmo natural dos níveis de ácido) estão representados graficamente na Figura 7.2.3 e os perfis médios transformados estão representados na Figura 7.2.5.

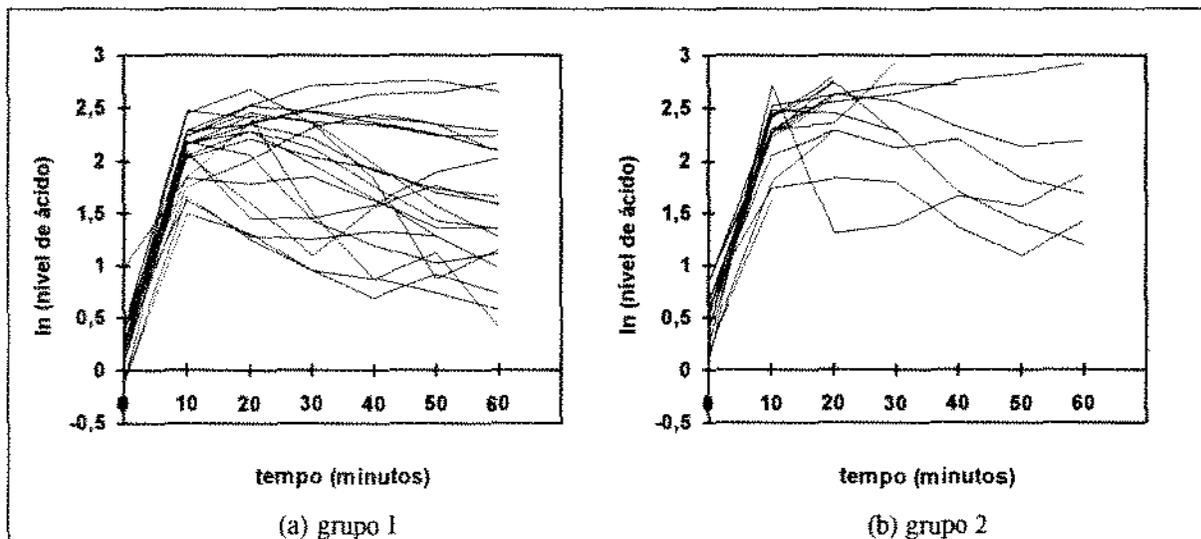


Fig. 7.2.3 - gráfico dos dados transformados: \ln (nível de ácido láctico)

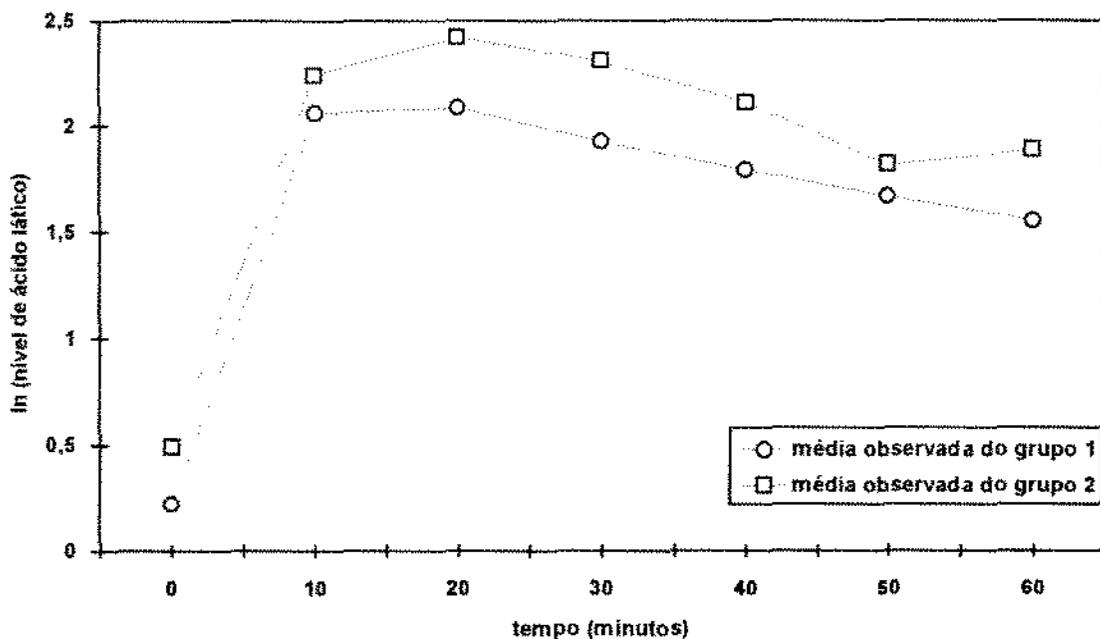


Fig. 7.2.4 - gráfico dos perfis médios observados para dados transformados - \ln (dados)

Novamente, agora para os dados transformados, ajustou-se o modelo saturado para os efeitos fixos ($c = d = 7$) com um efeito aleatório ($m = 0$) pelos métodos de estimação MV e MVRe. Esse modelo será chamado de **modelo 1**. Pelas Figuras 7.2.5 e 7.2.6, observa-se que o padrão residual está melhor distribuído em todos os instantes se comparado ao padrão apresentado nas Figuras 7.2.1 e 7.2.2; e, portanto, serão utilizados os dados transformados para o ajuste de modelos. Mesmo assim, ainda se pode observar alguma heterogeneidade temporal que, como será discutido na seção 7.2.5, pode ser parcialmente explicada através dos efeitos aleatórios. Entretanto, a simples visualização dos resíduos a partir desses gráficos não é suficiente para avaliar as possíveis estruturas existentes nos resíduos. Nos gráficos dos semivariogramas dos resíduos transformados obtidos por MV (Fig. 7.2.7) e MVRe (Fig. 7.2.8), há indícios de estruturas de correlação nos resíduos. A adoção de outras estruturas para Γ e Ω , será feita na seção 7.2.4.

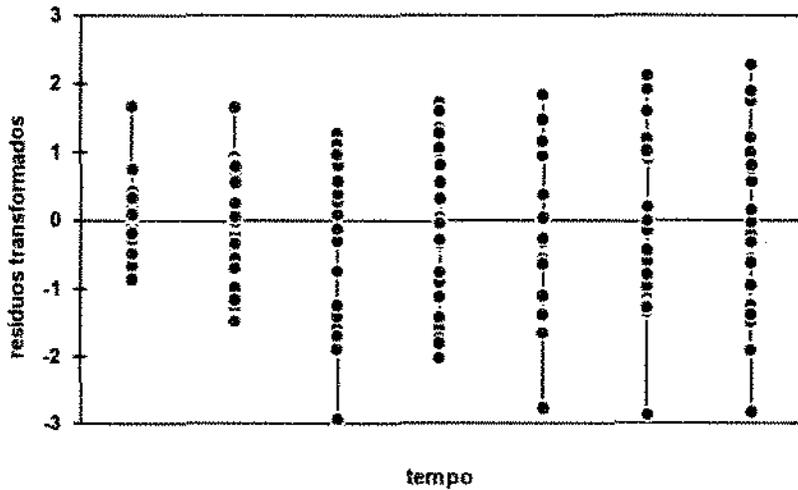


Fig. 7.2.5 - gráfico dos resíduos transformados dos 36 ratos do modelo 1 ajustado por MV para \ln (nível de ácido láctico)

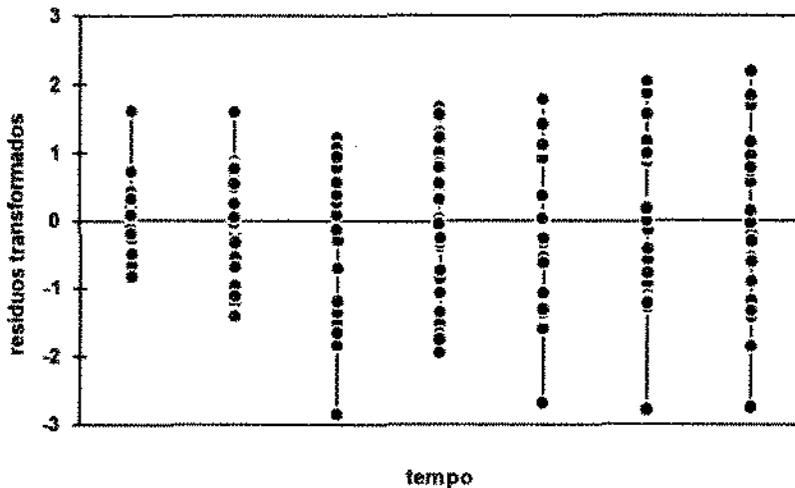


Fig. 7.2.6 - gráfico dos resíduos transformados dos 36 ratos do modelo 1 ajustado por MVRe para \ln (nível de ácido láctico)

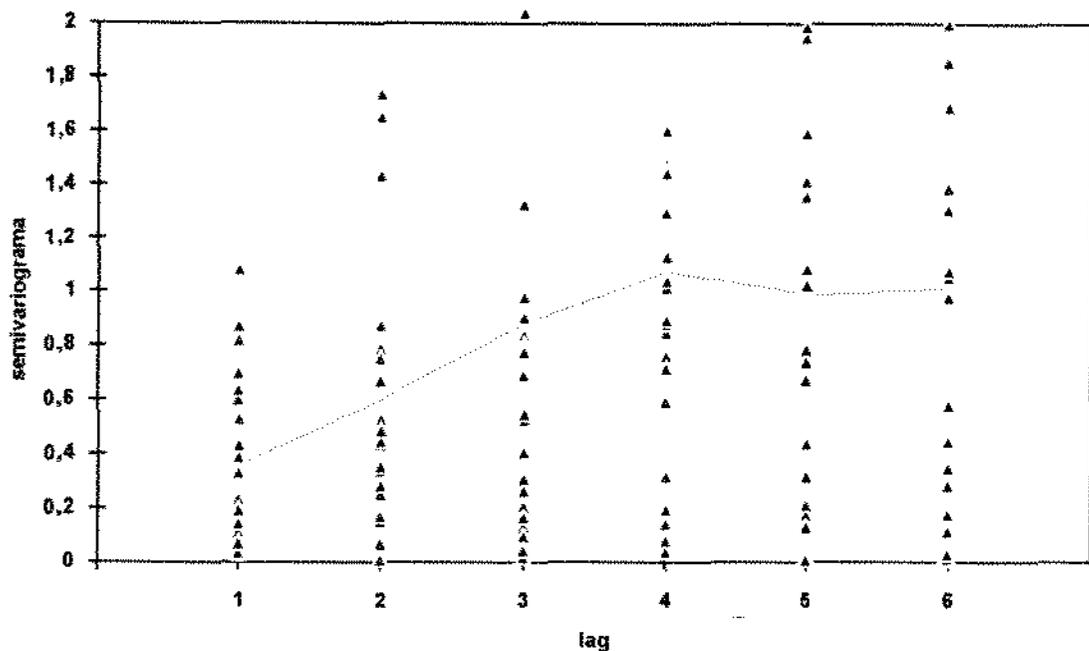


Fig. 7.2.7 - gráfico do semivariograma empírico dos resíduos transformados do modelo 1 ajustado por MV (dados transformados):
 -- representa o variograma médio
 (15 pontos com semivariograma acima de 2 foram omitidos)

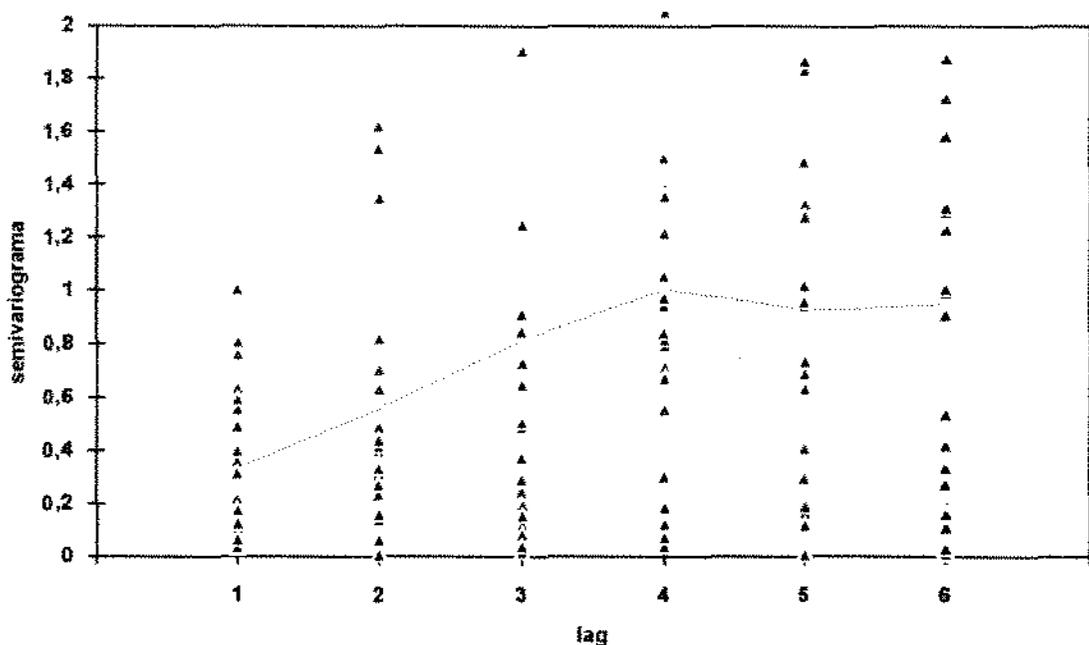
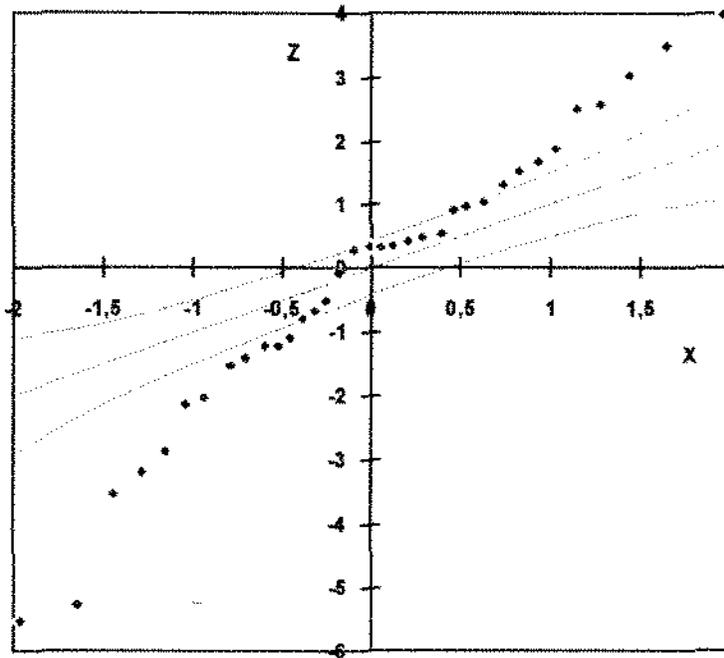
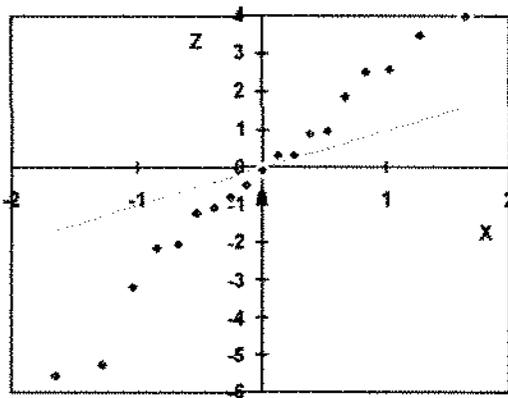


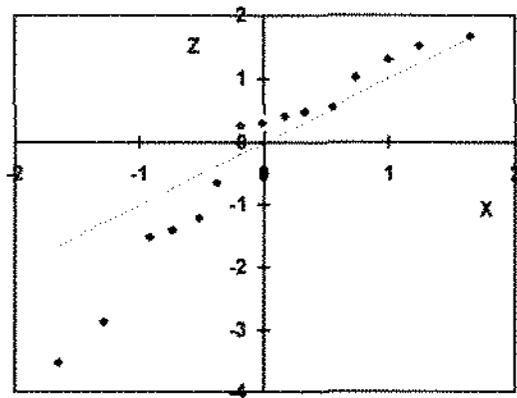
Fig. 7.2.8 - gráfico do semivariograma empírico dos resíduos transformados do modelo 1 ajustado por MVRe (dados transformados):
 -- representa o variograma médio
 (12 pontos com semivariograma acima de 2 foram omitidos)



(a) grupos 1 e 2



(b) grupo 1

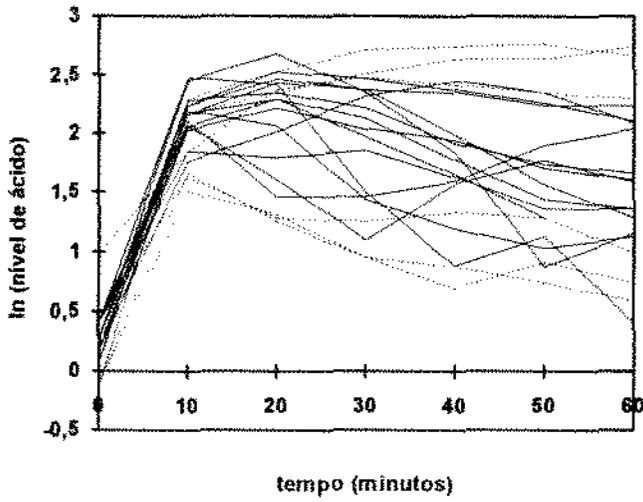


(c) grupo 2

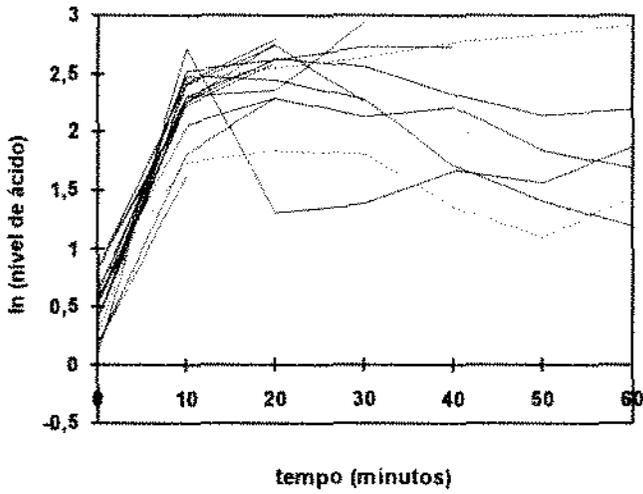
Fig. 7.2.9 - gráficos normais ponderados das previsões dos efeitos aleatórios \hat{b}_{10} do modelo 1 ajustado por MVRe; (limites de confiança de 95 %); z é a variável padronizada e x é o quantil da Normal (0,1)

No gráfico de probabilidade normal ponderado das previsões dos efeitos aleatórios conjuntamente para os dois grupos (Fig. 7.2.9 (a)), é possível observar ratos que podem ser considerados como valores aberrantes. Pelas Figuras 7.2.9 (b) e (c), nota-se que, para os dois grupos, não há normalidade no padrão das previsões dos efeitos aleatórios.

Os ratos identificados como tendo valores anormais são: 3, 5, 7, 9, 10 e 11 (do grupo 1), e 24 e 33 (do grupo 2). As observações desses ratos (destacadas em linhas pontilhadas na Fig. 7.2.10) serão excluídas da modelagem.



(a) grupo 1



(b) grupo 2

Fig. 7.2.10 - gráfico dos dados transformados: \ln (nível de ácido láctico) destacando-se os ratos com valores aberrantes (linhas pontilhadas)

7.2.3. Retirada dos Ratos com Valores Aberrantes

Os ratos com valores aberrantes (identificados na seção anterior) serão excluídos da análise. Talvez a retirada de 8 ratos é exagerada, mas isso serve aqui para verificar ilustrativamente a influência desses valores (aberrantes) nas estimativas.

Assim, os perfis médios observados do conjunto de dados transformados sem esses ratos serão modificados, como estão representados na Fig. 7.2.11. Por essa figura, nota-se que as médias observadas de ambos os grupos sem os ratos com valores aberrantes são ligeiramente modificadas em comparação às medias observadas com todos os ratos. Portanto, os perfis médios não foram muito influenciados pela retirada desses ratos.

Entretanto, como o número de observações foi reduzido de 36 ratos (214 observações no total) para 28 ratos (158 observações), a precisão das estimativas pode ser afetada. Deve-se lembrar que a falha da suposição de normalidade detectada na Figura 7.2.9 não invalida as estimativas dos parâmetros de locação do modelo, mas pode invalidar as inferências baseadas na distribuição normal.

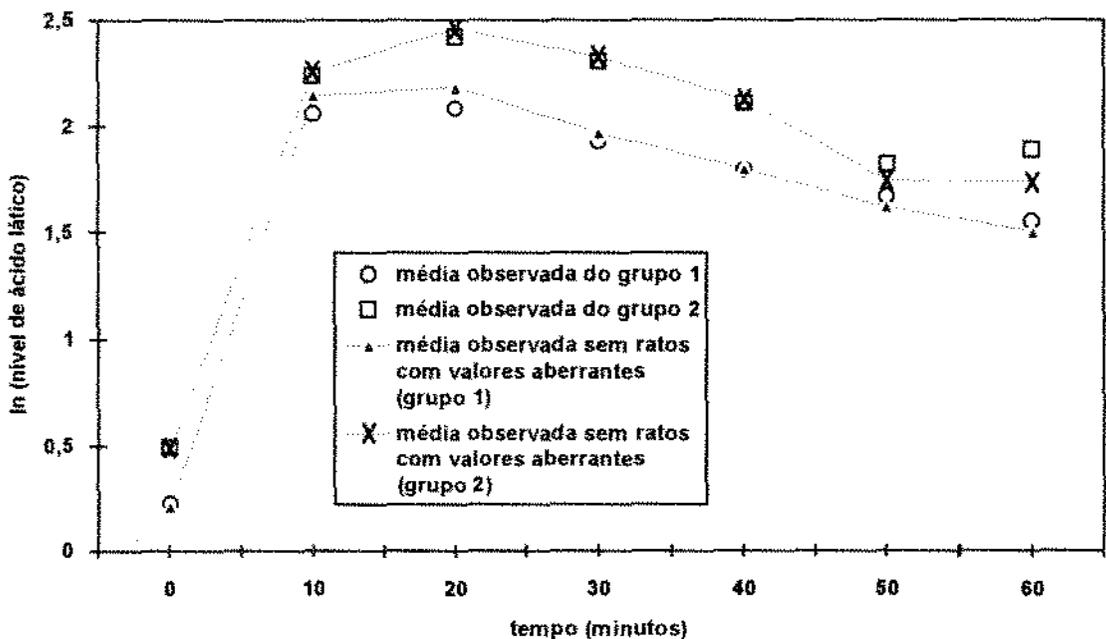
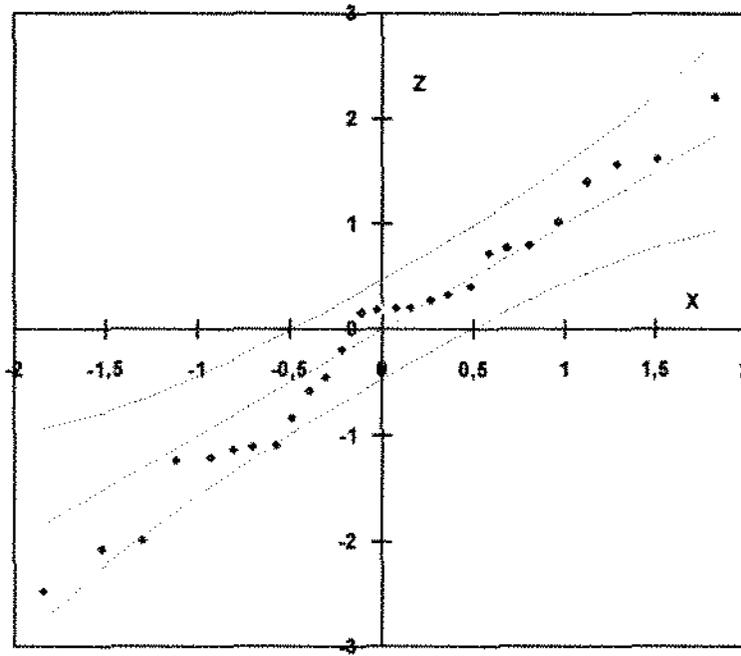


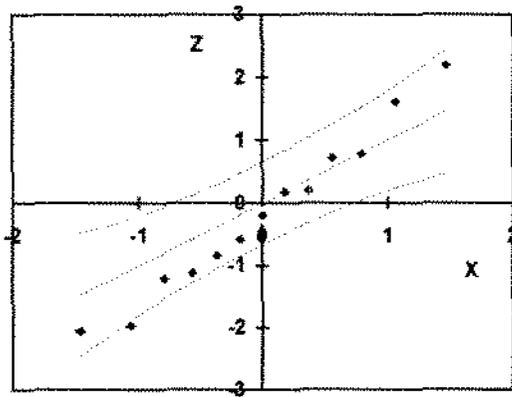
Fig. 7.2.11 - gráfico dos perfis médios observados para dados transformados - $\ln(\text{dados})$ excluindo-se os ratos considerados com valores aberrantes.

Novamente, agora para os dados transformados sem os ratos com valores aberrantes, ajustou-se o **modelo 1**. Agora no gráfico de probabilidade normal ponderado

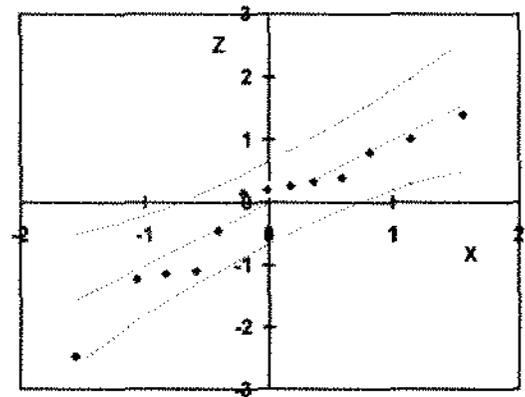
(Fig. 7.2.12 (a)) é possível observar um padrão razoável de normalidade das predições dos efeitos aleatórios. Pelas Figuras 7.2.12 (b) e (c), nota-se que o mesmo padrão das predições dos efeitos aleatórios pode ser considerado para ambos os grupos.



(a) grupos 1 e 2



(b) grupo 1



(c) grupo 2

Fig. 7.2.12 - gráficos normais ponderados das predições dos efeitos aleatórios \hat{b}_{i0} do modelo 1 (dados sem ratos com valores aberrantes) ajustado por MVRe; (limites de confiança de 95%) z é a variável padronizada e x é o quantil da Normal (0,1)

Como já foi observado na seção 7.2.2 e novamente pela Fig. 7.2.13 (gráfico do semivariograma), nota-se que existe uma estrutura de correlação nos resíduos. Na próxima seção tentaremos ajustar outros modelos, utilizando as estruturas descritas na Tabela 7.2.1 para as matrizes de covariância. Os resultados de alguns desses ajustes estão resumidos na Tabela 7.2.2.

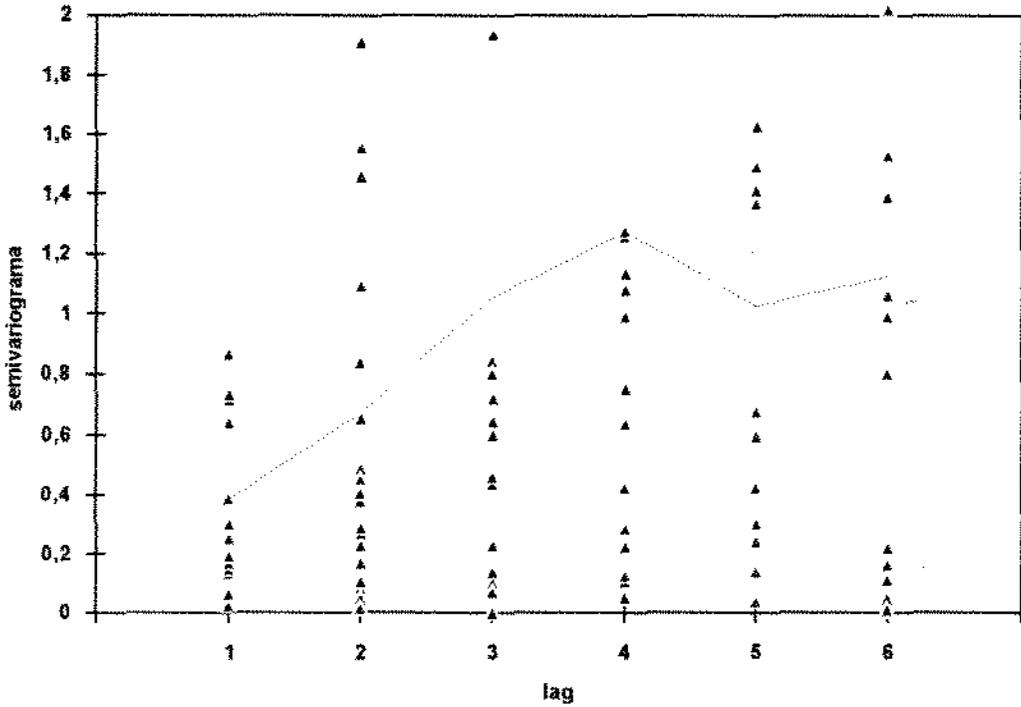


Fig. 7.2.13 - gráfico do semivariograma empírico dos resíduos transformados do modelo 1 ajustado por MVRc (dados transformados sem ratos aberrantes):
 -- representa o variograma médio
 (13 pontos com semivariograma acima de 2 foram omitidos)

7.2.4. Ajuste e Seleção dos Modelos

Nessa seção, investigamos modelos (saturados e reduzidos) com $m = 0, 1, 2, 3$, com matriz de covariância dos efeitos aleatórios B não estruturada (UN) ou diagonal (DIAG), considerando-se os erros independentes (SIM) ou AR(1). Ajustando-se por MV e MVRe, a Tabela 7.2.2 apresenta os respectivos valores obtidos de l ($-2\ln$ verossimilhança), l_R ($-2\ln$ verossimilhança restringida), com o número de parâmetros estimados (r) e valores do AIC e AICRe (AIC que utiliza o l_R).

Tabela 7.2.2 - Resumo dos ajustes

modelo	$c-1$	$d-1$	$m-1$	Γ	Ω_i	r	l	AIC	l_R	AICRe
1	6	6	0	SIM	SIM	16	111,43	143,43	434,87	466,87
2	6	6	1	DIAG	SIM	17	75,43	109,43	401,60	435,60
3	6	6	1	UN	SIM	18	71,10	107,10	397,65	433,65
4	6	6	2	DIAG	SIM	18	69,99	105,99	396,50	432,50
5	6	6	2	UN	SIM	21	40,78	82,78	370,55	412,55
6	6	5	0	SIM	SIM	15	111,48	141,48	397,22	427,22
7	6	4	0	SIM	SIM	14	111,48	139,48	363,90	391,90
8	6	3	0	SIM	SIM	13	113,10	139,10	337,10	363,10
9	6	2	0	SIM	SIM	12	113,59	137,59	314,62	338,62
10	6	1	0	SIM	SIM	11	113,60	135,60	297,21	319,21
11	6	0	0	SIM	SIM	10	113,78	133,78	285,82	305,82
12	6	0	1	DIAG	SIM	11	79,29	101,29	253,00	275,00
13	6	0	1	UN	SIM	12	75,31	99,31	249,31	273,31
14	6	0	2	DIAG	SIM	12	74,19	98,19	262,71	286,71
15	6	0	2	UN	SIM	15	58,60	88,60	225,43	255,43
16	6	0	0	SIM	AR1	11	61,33	83,33	235,51	257,51
17	5	0	0	SIM	AR1	10	61,53	81,53	246,56	266,57
18	4	0	0	SIM	AR1	9	118,34	136,34	216,09	234,09
19	3	0	0	SIM	AR1	8	145,91	161,91	213,20	229,20
20	2	0	0	SIM	AR1	7	244,23	258,23	285,87	299,87
21	5	0	1	UN	AR1	12	75,33	99,33	209,63	233,63
22	4	0	1	UN	AR1	11	81,92	103,92	180,72	202,72
23	5	0	2	UN	AR1	15	56,62	86,62	191,10	221,10
24	4	0	2	UN	AR1	14	61,87	89,87	163,22	191,22

$c-1$ é o grau do polinômio ajustado, r representa o número de parâmetros estimados, l é o valor de $-2\ln(\text{veross.})$ e AIC é o critério de informação de Akaike: $AIC = l + 2r$ e, similarmemente, l_R é o valor de $-2\ln(\text{veross. restringida})$ e $AICRe = l_R + 2r$. As notações SIM, DIAG, UN e AR1 estão descritas na Tabela 7.2.1.

Comparando-se os modelos 1, 2, 3, 4 e 5 pelos critérios do AIC e AICRe, a inclusão de 3 efeitos aleatórios diminui o valor dos critérios (estrutura UN para Γ).

Os demais modelos (6, 7, ..., 24) são reduções dos modelos saturados (com 7 coeficientes fixos). Comparando-se os modelos 6 a 11, o modelo 11 possui o menor AIC e AICRe. O modelo 11 assume que as curvas dos grupos 1 e 2 são "paralelas", i.e., a diferença entre os perfis médios é uma constante como está ilustrado na Figura 7.2.14.

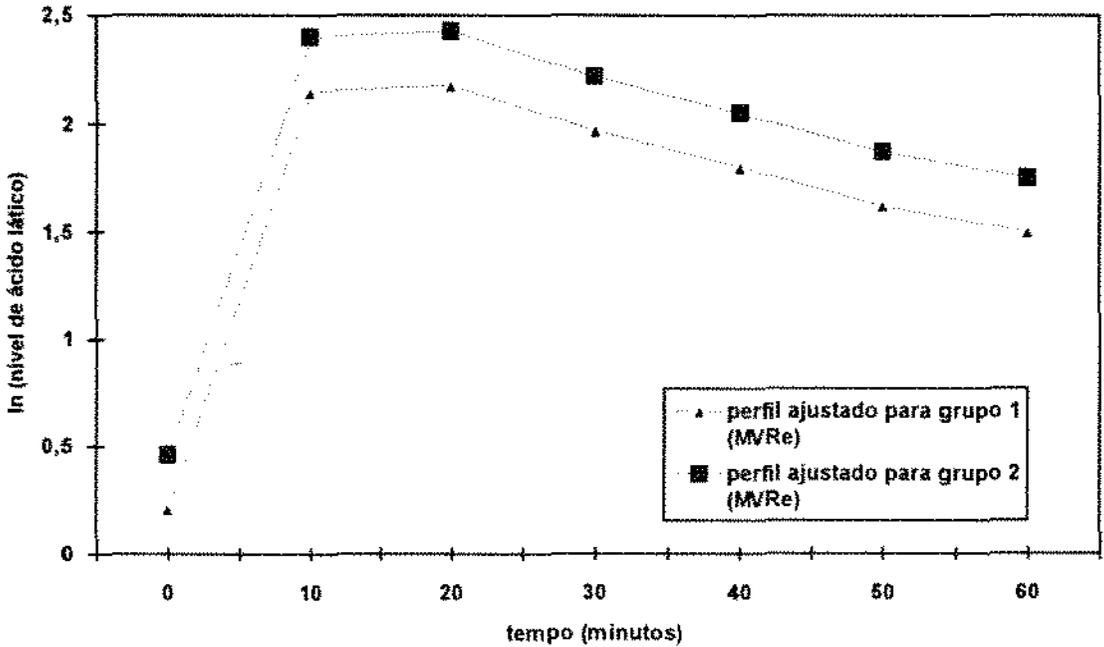


Fig. 7.2.14 - gráfico dos perfis médios ajustados pelo modelo 11 para dados transformados, - $\ln(\text{dados})$, excluindo-se os ratos considerados influentes (MVRe).

Os modelos 11 a 24 assumem curvas "paralelas". Comparando-se os modelos 12 a 15, o modelo 15 possui o menor AIC e AICRe. Assim, novamente, a inclusão de 3 efeitos aleatórios diminuiu o valor dos critérios (estrutura UN para Γ). Nos modelos 16 a 24 uma estrutura AR(1) é adotada para Ω . Ajustando-se por MV, o modelo escolhido pelo AIC é o modelo 17 (modelo polinomial de grau 5 com 1 efeito aleatório), enquanto que ajustando-se por MVRe, o modelo escolhido pelo AICRe é o modelo 24 (modelo polinomial de grau 4 com 3 efeitos aleatórios).

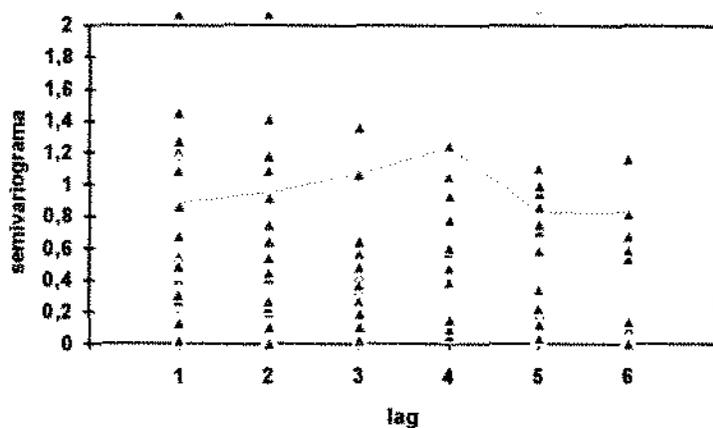


Fig. 7.2.15 - gráfico do semivariograma empírico dos resíduos transformados do modelo 17 ajustado por MV (dados transformados sem ratos influentes):
 -- representa o variograma médio
 (15 pontos com semivariograma acima de 2 foram omitidos)

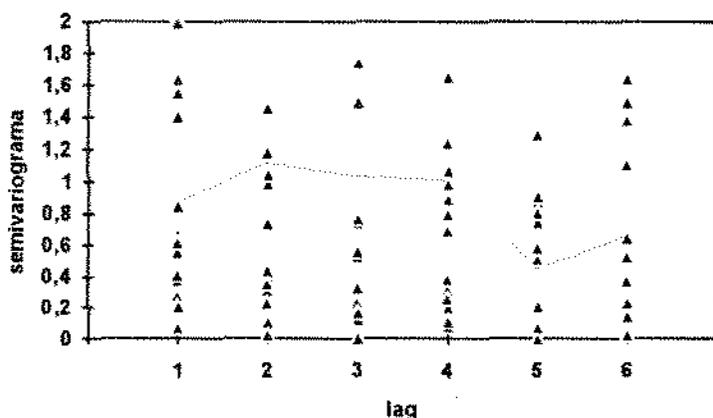


Fig. 7.2.16 - gráfico do semivariograma empírico dos resíduos transformados do modelo 24 ajustado por MVRe (dados transformados sem ratos influentes):
 -- representa o variograma médio
 (12 pontos com semivariograma acima de 2 foram omitidos)

Os variogramas dos resíduos transformados dos modelos 17 e 24 (respectivamente, Fig. 7.2.15 e 7.2.16) não apresentam padrões que possam ser identificados. Assim, não há indícios de padrões na estrutura de covariância que comprometam os modelos ajustados.

7.2.5. Interpretação dos Efeitos Aleatórios

Considere o modelo 24 (ajustado por MVRe) que possui 3 efeitos aleatórios ($m - 1 = 2$). Pelos perfis dos dados transformados (Fig. 7.2.3) ainda se pode observar (aparentemente) alguma heterocedasticia ao longo do tempo (os dados no início são mais estáveis do que os restantes). Como Jones (1990) observa, uma parte dessa heterocedasticia pode ser explicada através dos efeitos aleatórios. Como será discutido aqui, os 3 efeitos aleatórios podem descrever esse comportamento.

Os efeitos aleatórios para o i -ésimo rato podem ser escritos como

$$b_i = \sigma L' \eta_i,$$

onde L' é matriz triangular inferior da decomposição de Cholesky da matriz de covariância dos efeitos aleatórios ($B = L'L$), e $\eta_i \sim N(0, I)$; I é matriz identidade 3×3 .

Como ilustração nessa seção, os parâmetros da matriz L a serem estimados são os elementos apenas da primeira linha de L , ou seja,

$$L = \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

(O modelo 24 com a matriz L dada acima produz $AIC = 205,33$ e $I_R = 183,33$).

Então,

$$(Z_i b_i) = Z_i \sigma L' \eta_i,$$

e, substituindo-se os valores estimados de L' e σ , obtemos: $(Z_i b_i)^* = Z_i \hat{\sigma} \hat{L}' \eta_i =$

$$= \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_7 & t_7^2 \end{bmatrix} \begin{bmatrix} 0,1377 & 0 & 0 \\ 0,0138 & 0 & 0 \\ -0,0007 & 0 & 0 \end{bmatrix} \eta_i =$$

$$= \begin{bmatrix} 0,1377 + 0,0138t_1 - 0,0007t_1^2 & 0 & 0 \\ 0,1377 + 0,0138t_2 - 0,0007t_2^2 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0,1377 + 0,0138t_7 - 0,0007t_7^2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{bmatrix}$$

Assim, a curva quadrática estimada associada ao efeito aleatório η_{1i} é dada por:

$$h_1(t) = 0,1377 + 0,0138t - 0,0007t^2,$$

e está ilustrada na Figura 7.2.17:

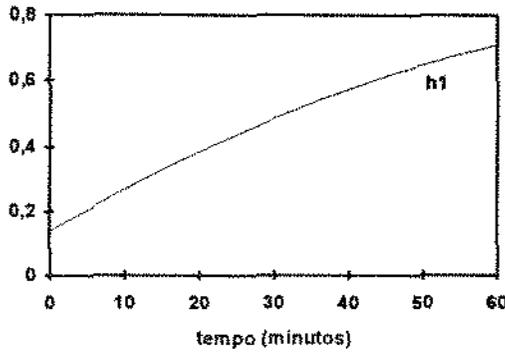


Fig. 7.2.17 - gráfico da curva quadrática estimada associada ao efeito aleatório η_{1i} do modelo 24;

A curva quando multiplicada por uma normal padrão nos instantes das observações pode ajudar a explicar a heterogeneidade da variância dos dados.

7.2.6. Perfis Individuais Estimados

Considere o modelo 24 ajustado por MVRe. Os gráficos dos perfis médios estimados para os grupos estão ilustrados na Fig. 7.2.18; e os perfis individuais na Fig. 7.2.19.

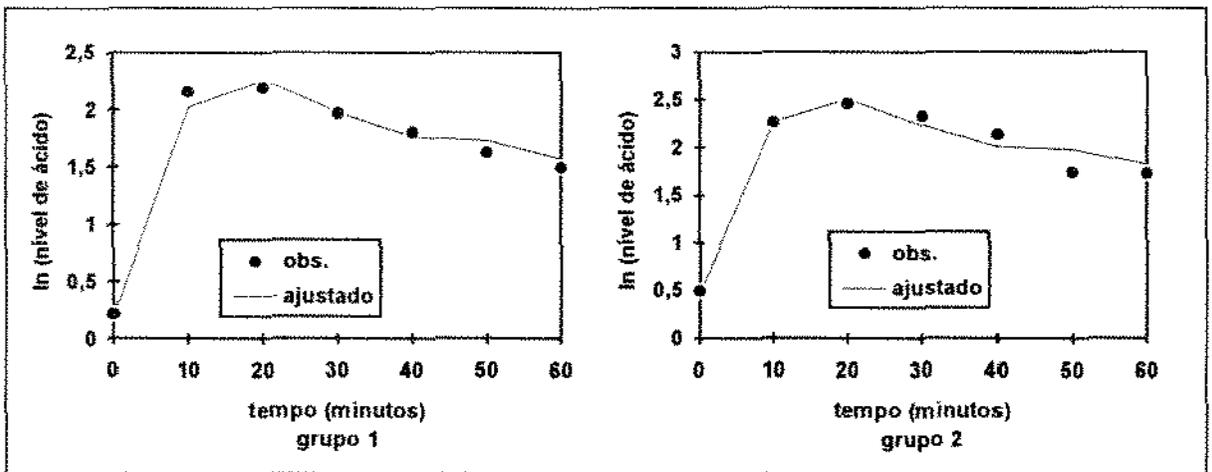


Fig. 7.2.18 - gráfico dos perfis médios ajustados pelo modelo 24 (MVRe) para dados transformados, excluindo-se os ratos com valores aberrantes.

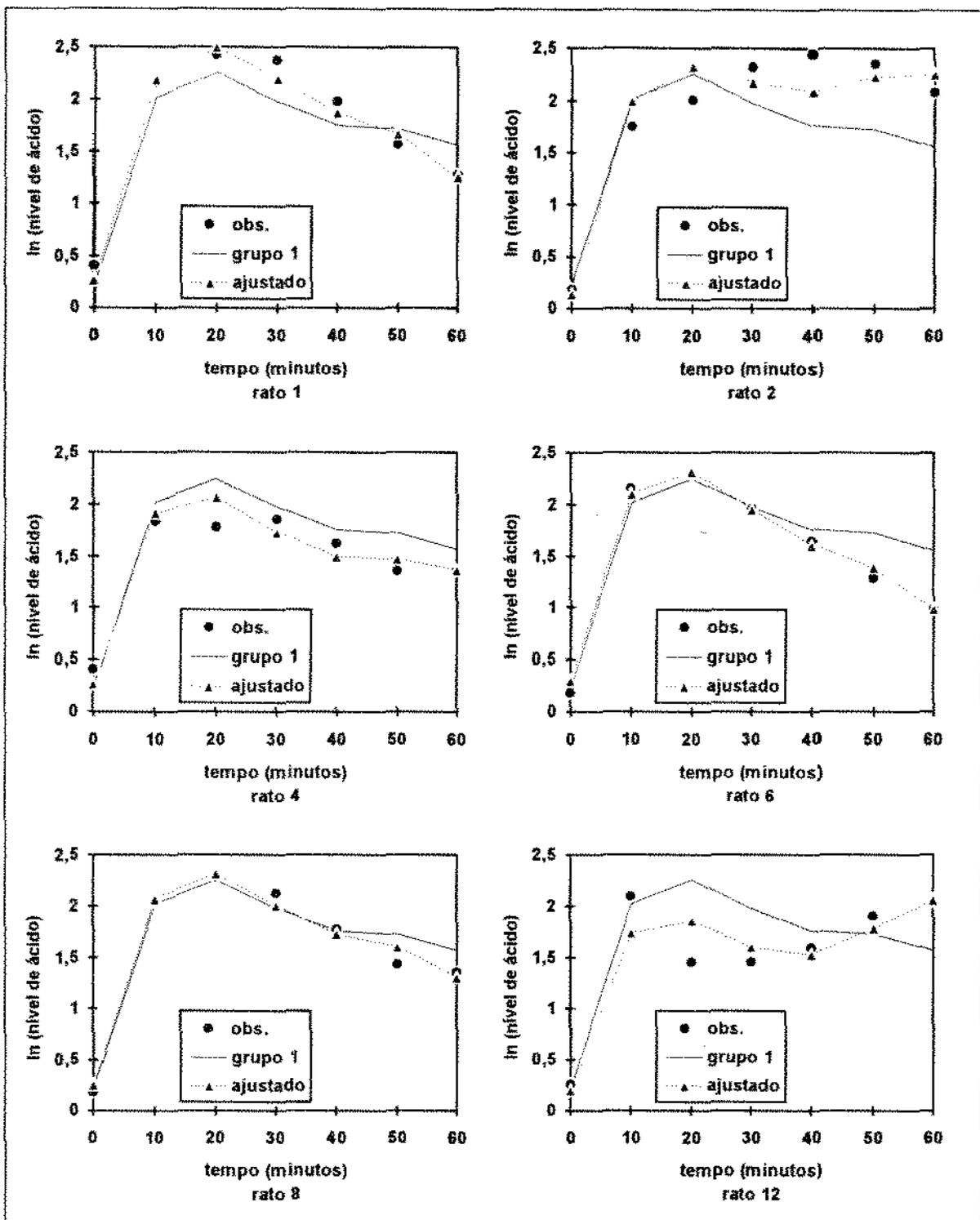


Fig. 7.2.19 - gráfico dos perfis individuais ajustados pelo modelo 24 (MVRe) para dados transformados, excluindo-se os ratos com valores aberrantes (continua)

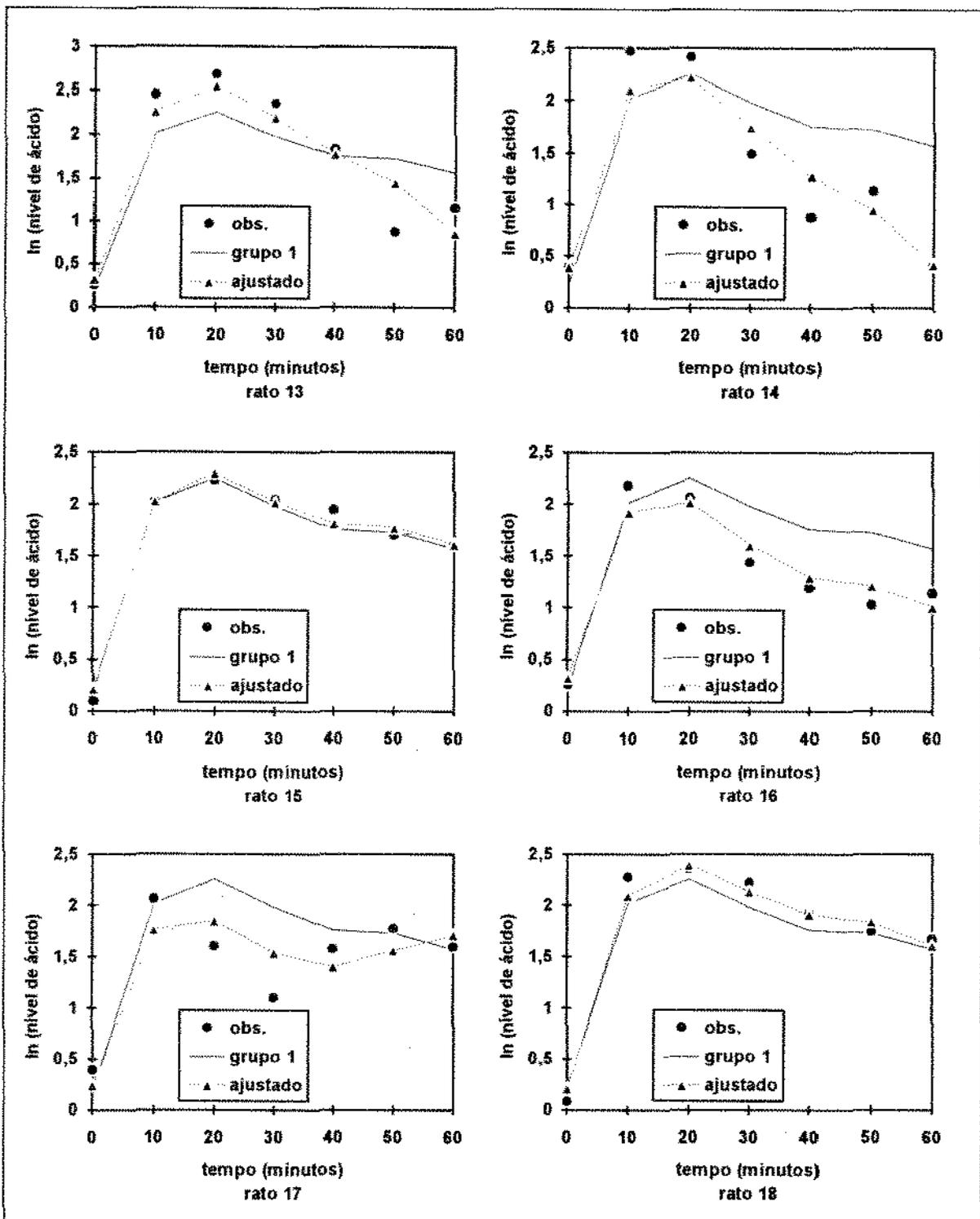


Fig. 7.2.19 - gráfico dos perfis individuais ajustados pelo modelo 24 (MVRe) para dados transformados, excluindo-se os ratos com valores aberrantes (continuação)

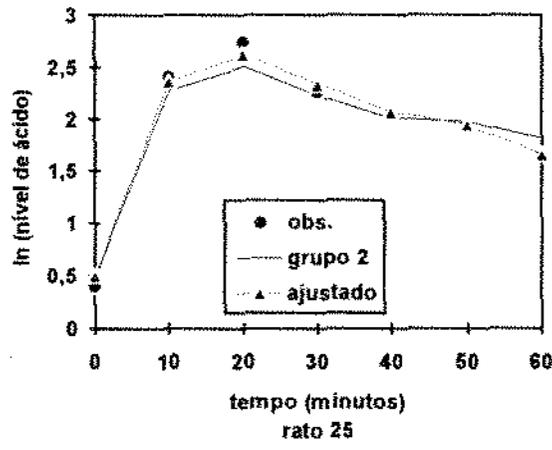
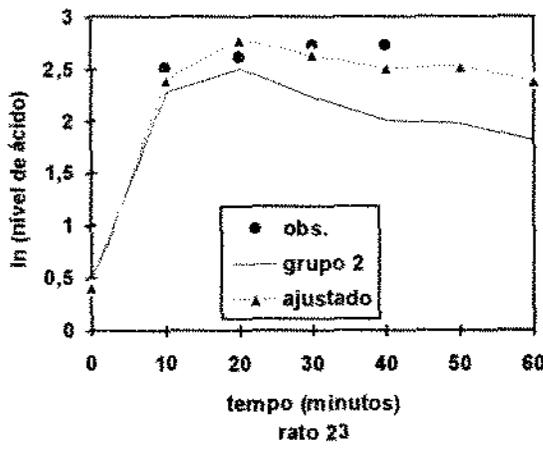
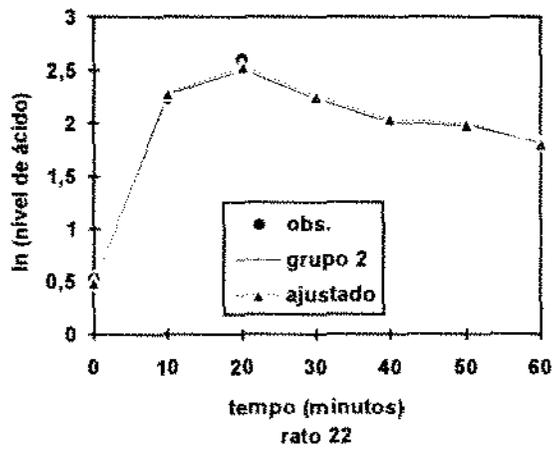
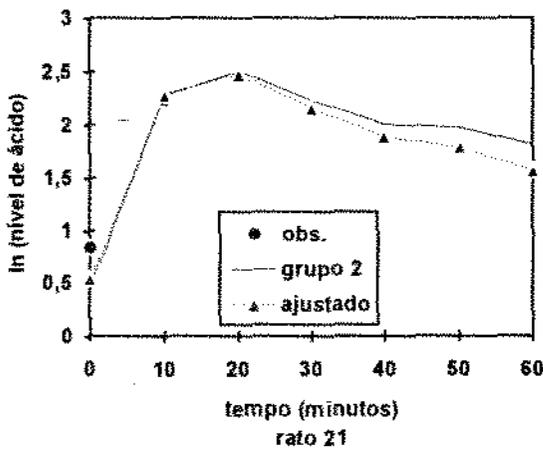
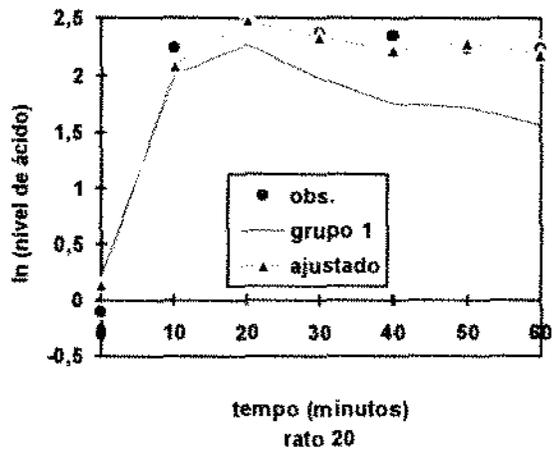
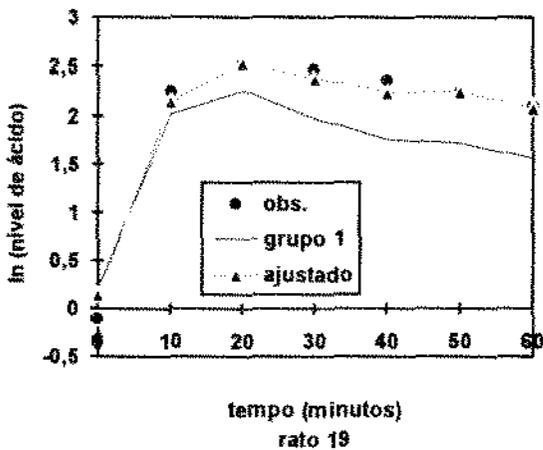


Fig. 7.2.19 - gráfico dos perfis individuais ajustados pelo modelo 24 (MVRe) para dados transformados, excluindo-se os ratos com valores aberrantes (continuação)

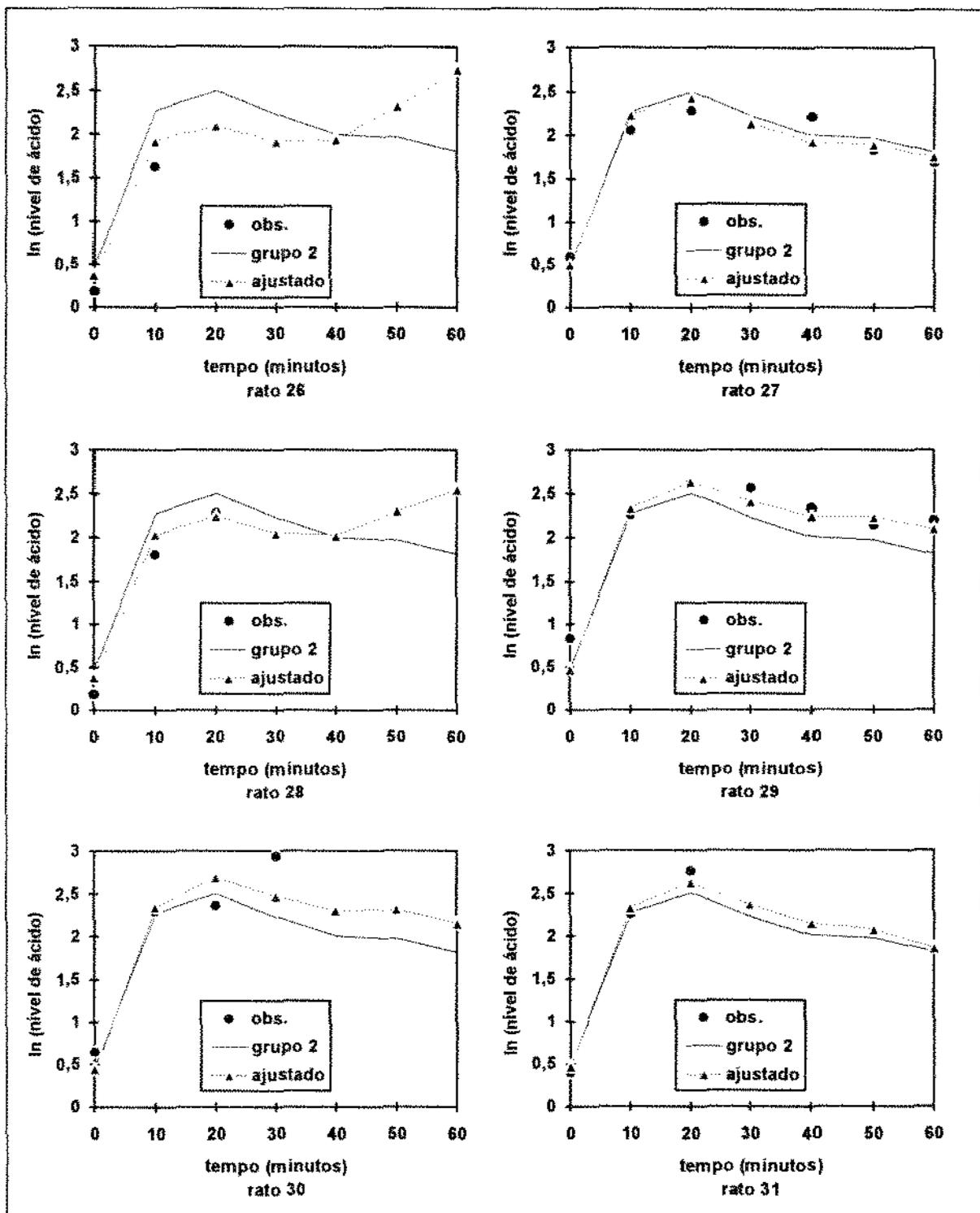


Fig. 7.2.19 - gráfico dos perfis individuais ajustados pelo modelo 24 (MVRe) para dados transformados, excluindo-se os ratos com valores aberrantes (continuação)

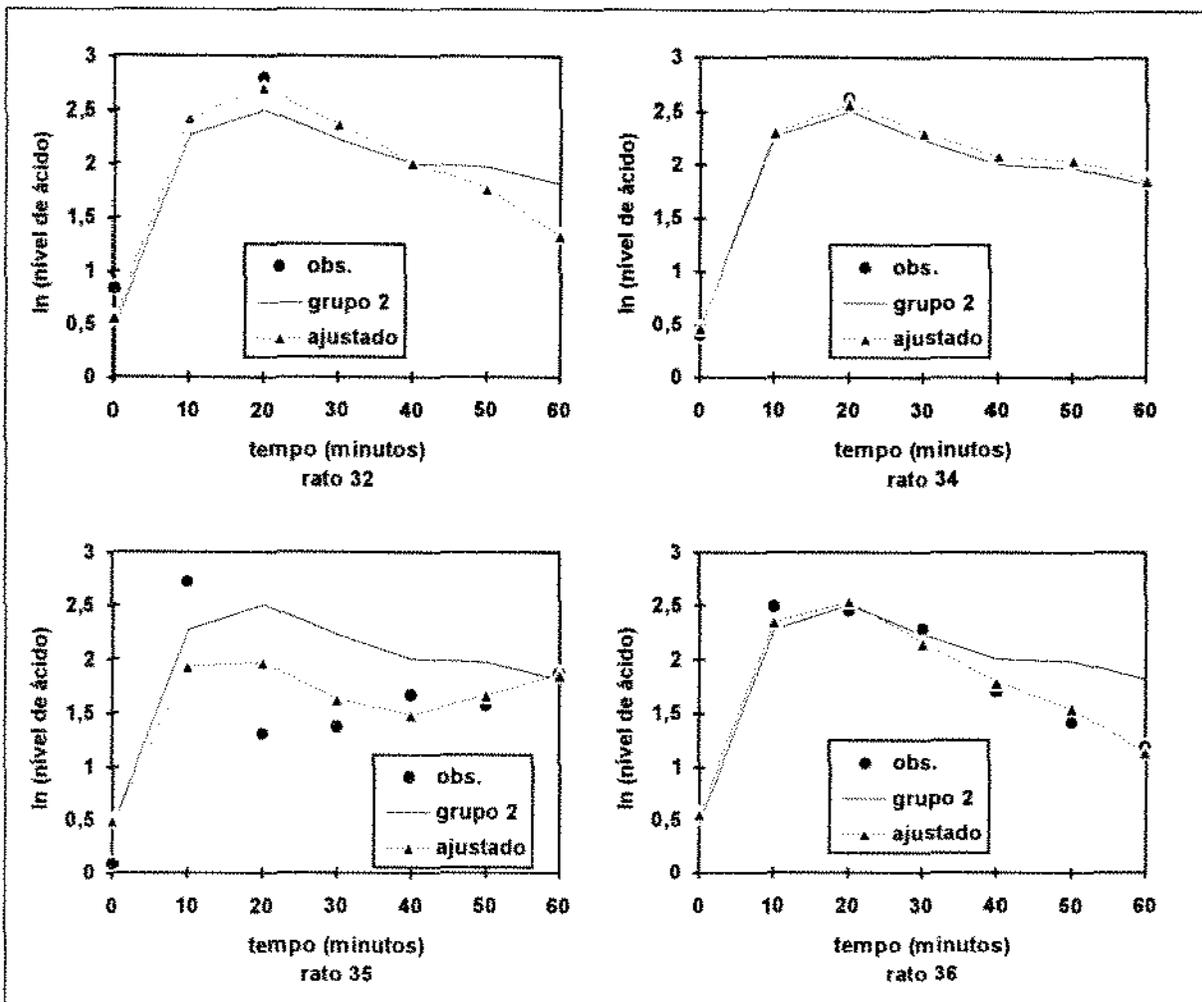


Fig. 7.2.19 - gráfico dos perfis individuais ajustados pelo modelo 24 (MVRe) para dados transformados, excluindo-se os ratos com valores aberrantes (continuação)

7.3. Análise II

Na seção anterior, foram investigados os modelos descritos no exemplo 2.5.6. Nessa seção, toma-se novamente o conjunto original dos dados (sem transformações nem retirada de ratos do conjunto). O modelo dessa seção será escrito em uma forma semelhante ao descrito no exemplo 3.3.3:

$$y_{i(1)t} = \beta_{1t} + z_t b_{i(1)t} + \varepsilon_{i(1)t},$$

$$y_{i(2)t} = \beta_{2t} + z_t b_{i(2)t} + \varepsilon_{i(2)t},$$

onde $y_{i(1)t}$ e $y_{i(2)t}$ denotam os níveis de ácido láctico dos i -ésimos ratos, respectivamente, do grupo 1 e do grupo 2 no instante t . Os efeitos fixos, β_{1t} e β_{2t} , descrevem o perfil médio do grupo 1 e 2, respectivamente, em cada instante (modelo saturado). A componente aleatória b_{it} descreve a variação de cada indivíduos no instante t , assumindo-se o vetor das componentes aleatórias $b_i = (b_{i1}, \dots, b_{i7})'$ que tenha distribuição $N(0, \Gamma)$. As formas da matriz Γ consideradas aqui serão: independência ($\sigma_b^2 I$), forma de simetria composta (SC) e AR(1). Os erros $(\varepsilon_{i(k)0}, \dots, \varepsilon_{i(k)60})$ descrevem a variação dentro do indivíduo, assumindo-se que tenham distribuição $N(0, \Omega_i)$. As formas da matriz Ω_i consideradas serão: independência ($\sigma_\varepsilon^2 I$), forma de simetria composta e AR(1).

As covariáveis z_t introduzem a heterocedasticia temporal ao modelo cujos valores estão indicados na Tabela 7.3.1. Esses valores de z_t são valores aproximados dos erros padrões observados indicados na Tabela 1.2.3 (pp. 10). O modelo descrito no exemplo 3.3.3, onde assume-se $z_t^* = \hat{\sigma}_t / \hat{\sigma}_b$ em lugar de z_t , não foi ajustado pela limitação do programa utilizado. Uma alternativa seria o ajuste em dois (ou mais) estágios: o primeiro com z_t e o segundo com $z_t^* = \hat{\sigma}_t / \hat{\sigma}_b$, utilizando-se as estimativas do primeiro estágio. Nessa seção, entretanto, assumiremos simplesmente os valores da Tabela 7.3.1 para descrever z_t .

Tabela 7.3.1 - Valores assumidos para z_t

t	0	10	20	30	40	50	60
z_t	0,4	2,4	3,5	4,3	3,3	4,6	4,9

Através da estimação via MVRe, considerando-se algumas formas diferentes para as matrizes B e W_1 , obtém-se os resultados da Tabela 7.3.2.

Tabela 7.3.2 - Resumo dos resultados dos modelos ajustados via MVRe

modelo	Γ	Ω_1	r	l_R	MC	g/l	χ^2	P-value	AICRe
1	-	$\sigma_e^2 I$	15	1100,161	-	-	-	-	1130,161
2	$\sigma_b^2 I$	$\sigma_e^2 I$	16	968,722	1	1	131,439	0,000	1000,72
3	SC	$\sigma_e^2 I$	17	907,713	2	1	61,009	0,000	941,713
4	$\sigma_b^2 I$	$AR1_e$	17	968,485	2	1	0,237	0,626	1002,485
5	SC	$AR1_e$	18	907,653	3	1	0,060	0,807	943,653
6	$AR1_b$	$\sigma_e^2 I$	17	865,252	2	1	103,470	0,000	899,252
7	$AR1_b$	CS	18	875,128	6	1	9,876	0,002	911,128
8	$AR1_b$	$AR1_e$	18	865,251	6	1	0,001	0,975	901,252

SC representa a estrutura simetria composta, r é o número de parâmetros estimados, MC é o modelo comparado no teste de razão de verossimilhança com g/l graus de liberdade, χ^2 é o valor da estatística do teste (aproximado) de razão de verossimilhança, e AICRe é o critério de escolha para EMVRe.

Da Tabela 7.3.2, aplicando-se sequencialmente o teste aproximado da razão de verossimilhança, um possível candidato é o modelo 7, enquanto que o critério AIC escolhe o modelo 6. Pelo *portfolio* (seção 4.4, pp. 66) o modelo 8 também é competitivo, pois seu AICRe é menor do que $899,25 + 2,3 = 901,55$.

Os resultados das estimativas MVRe dos parâmetros de covariância dos modelos 6 e 7 estão na Tabela 7.3.3.

Tabela 7.3.3 - Resultados das estimativas MVRe dos parâmetros de covariância

modelo 6			
parâmetros	estimativa	erro padrão	razão
variância em Γ (σ_b^2)	0,828	0,141	3,347
autoregressão em Γ (ϕ_b)	0,775	0,048	3,134
variância em Ω_1 (σ_e^2)	0,247	0,087	1,000
modelo 7			
parâmetros	estimativa	erro padrão	razão
variância em Γ (σ_b^2)	0,804	0,125	5,554
autoregressão em Γ (ϕ_b)	0,750	0,050	5,181
variância em Ω_1 (σ_e^2)	0,145	0,045	1,000
elementos SC em Ω_1	0,220	0,069	1,519

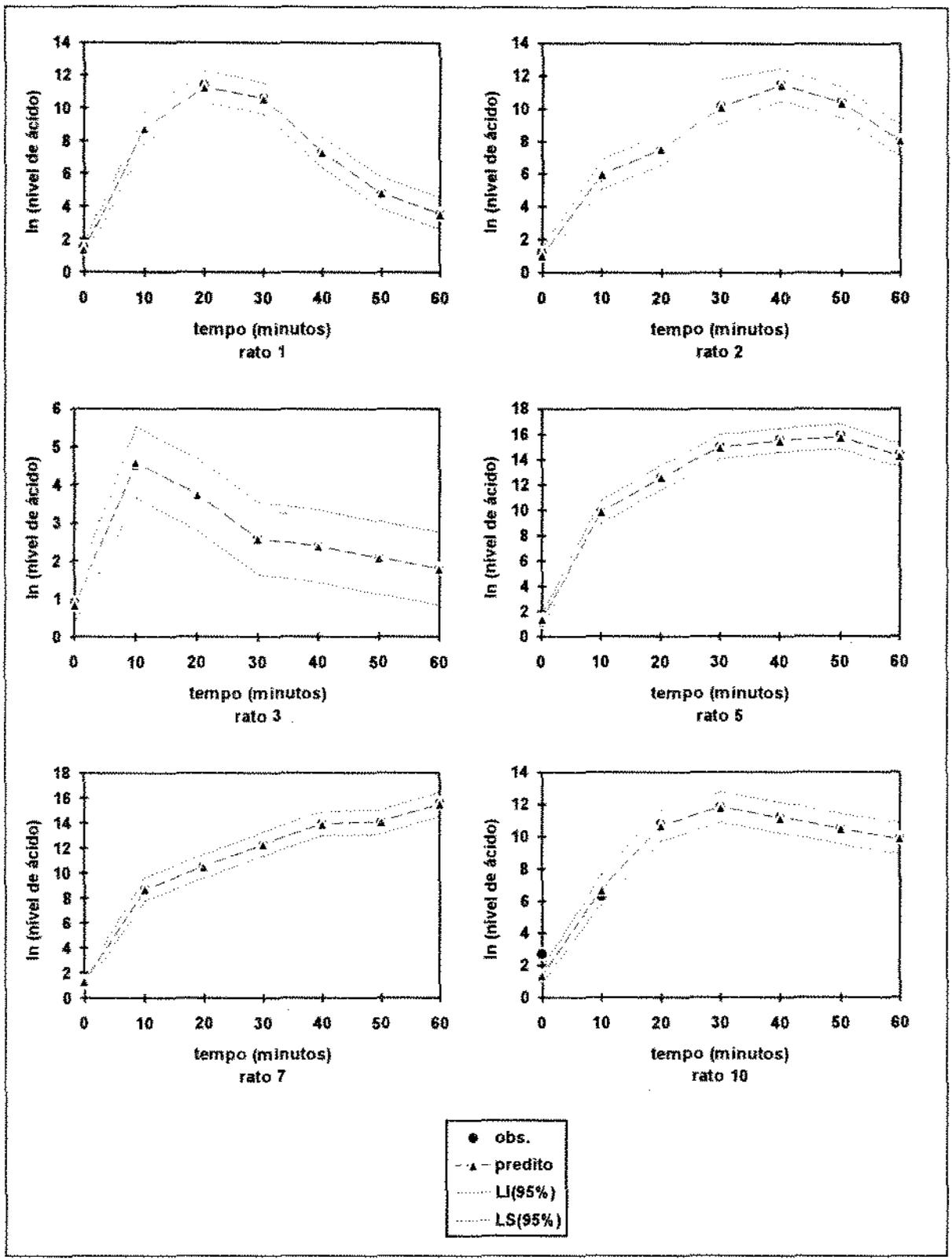


Fig. 7.3.1 - gráfico dos perfis ajustados pelo modelo 6 (MVRe) de alguns ratos (grupo 1)

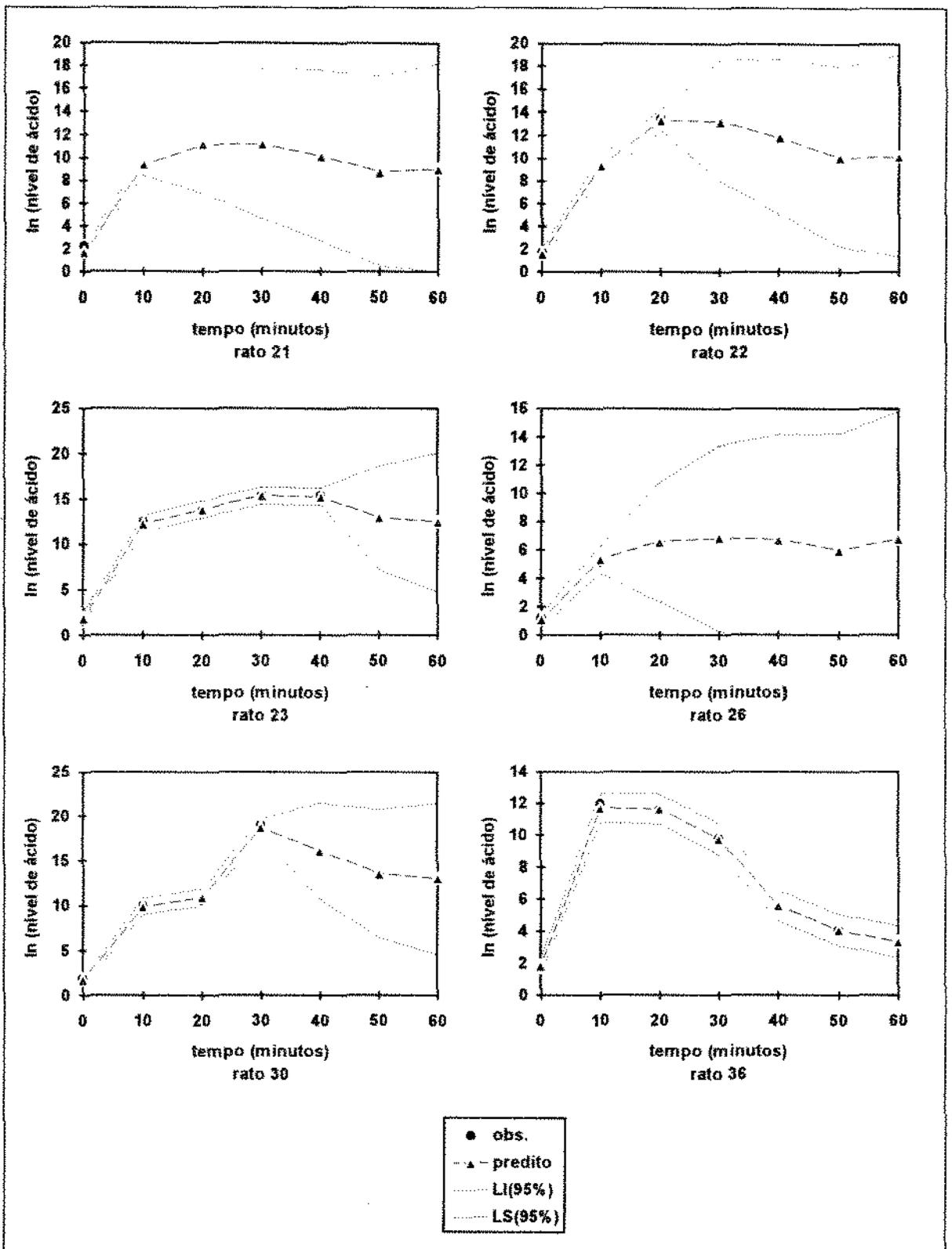


Fig. 7.3.2 - gráfico dos perfis ajustados pelo modelo 6 (MVRe) de alguns ratos (grupo 2)

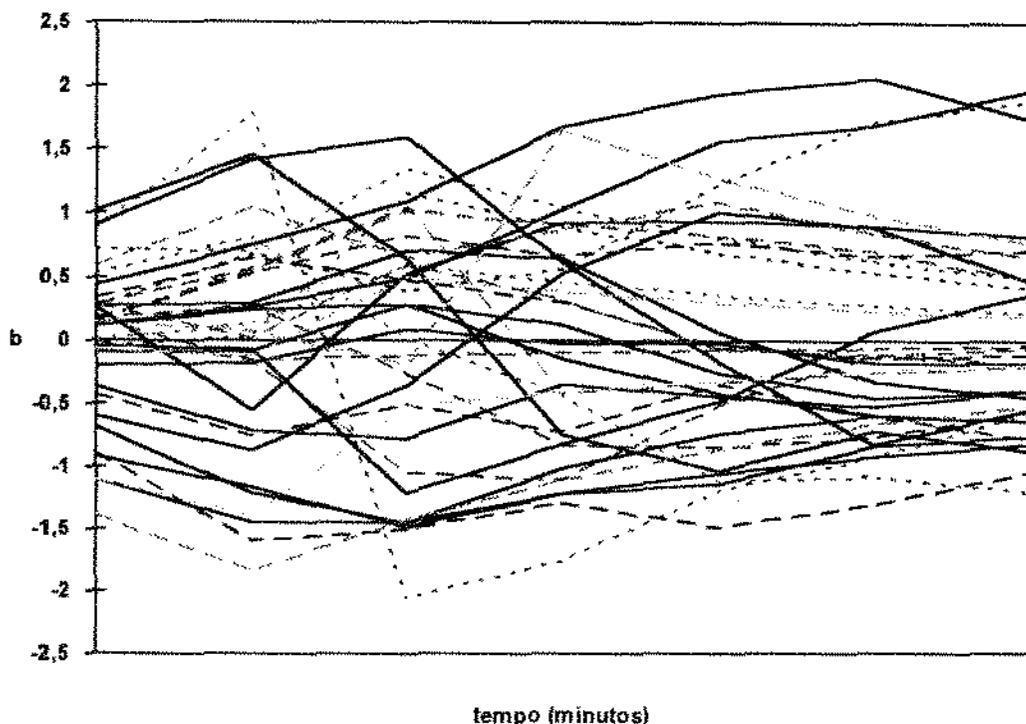


Fig. 7.3.3 - gráfico das previsões dos efeitos aleatórios individuais do modelo 6 (MVRe): $AR(1)$ com $\hat{\phi}_b = 0,775$.

As Figuras 7.3.1 e 7.3.2 ilustram os perfis de alguns ratos com os respectivos valores preditos e os limites de confiança de 95%. Os limites de confiança das previsões dos níveis de ácido láctico dos ratos com observações perdidas (grupo 2) se tornam mais largos à medida que se propaga a falta de informação.

A Fig. 7.3.3 ilustra as previsões dos efeitos aleatórios individuais que, segundo o modelo 6, seguem um processo $AR(1)$. O valor estimado do coeficiente de autoregressão é 0,775.

7.4. Algumas Considerações

Nesse capítulo aplicamos como ilustração algumas das técnicas de análise estudadas, tomando-se o conjunto de dados experimentais apresentado na seção 1.2.

Na análise I (seção 7.2), foram ajustados vários modelos para as curvas de resposta do logaritmo do nível de ácido láctico dos 2 grupos. O interesse restringiu-se à comparação entre os grupos sem preocuparmos em fazer interpretações sobre os modelos. Assim, foi preciso escolher um modelo dentre várias formas do modelo de efeitos mistos descritas na Tabela 7.2.2. Considerando-se no máximo 3 efeitos aleatórios com uma matriz de

covariância não estruturada ou diagonal (efeitos não correlacionados) com erros independentes ou AR(1), o critério do AIC "escolheu" o modelo 17 ajustado via MV (vide Tabela 7.2.2) e o AICRe "escolheu" o modelo 24 ajustado via MVRe. Em ambos modelos, as curvas ajustadas para os dois grupos são "paralelas". Embora haja muitas críticas ao AIC (por exemplo, não é consistente), ele é um critério utilizado por autores como Jones (1993) e Wolfinger (1993). O programa CAR1 (descrito na seção 7.1) e o PROC MIXED do SAS fornecem o valor calculado do AIC (o SAS fornece ainda o valor do critério de Schwarz). O uso do AIC é justificado por ser ainda o mais popular, sendo utilizado apenas como indicação para a escolha do modelo (não é um critério absoluto). Como foi discutido na seção 4.4, Jones (1993) sugere que os modelos que produzem AIC dentro de 2 unidades são modelos competitivos; ou então pelo *Portfolio* estudado por Poskitt e Tremayne (1987) que sugerem 2,3 unidades.

Na análise II (seção 7.3) ajustou-se algumas estruturas de correlação possíveis para as componentes aleatórias individuais (Γ e Ω). Entre as estruturas (Tabela 7.3.2) estão: estrutura de independência ($\sigma^2 I$), simetria composta, não estruturada, AR(1) e Toeplitz. A formulação do modelo nessa parte diferiu daquela adotada na análise I (modelo misto semelhante ao descrito no exemplo 3.3.3, considerando-se os efeitos aleatórios dependentes do tempo com uma matriz de covariância estruturada). Como o interesse nessa seção foi estudar formas para as matrizes de covariância, ilustrando a variedade de estruturas para as matrizes de covariância, consideramos apenas um modelo mais geral (saturado) sem considerar os modelos reduzidos e sem fazer comparações entre os grupos.

A

Algoritmo EM

O algoritmo EM é um procedimento iterativo, onde cada iteração consiste de uma etapa de Esperança (passo E) seguida de uma etapa de Maximização (passo M).

A idéia básica do EM, formalizada por Dempster *et alii.* (1977), é considerar os dados observados \mathbf{y} como um conjunto "incompleto" de dados que veio de um conjunto mais amplo de quantidades ou, simplesmente, um conjunto "completo" de dados, denotado por \mathbf{x} ; este conjunto inclui os parâmetros do modelo, por exemplo, \mathbf{x} e β .

Temos então dois espaços amostrais X e Y e uma aplicação não injetora de X para Y . Ao invés de observarmos os dados "completos" \mathbf{x} em X , observamos os dados "incompletos" $\mathbf{y} = \mathbf{y}(\mathbf{x})$ em Y , ou seja, os dados observados são uma realização de Y e o correspondente \mathbf{x} em X não é observado diretamente, mas indiretamente através de \mathbf{y} .

Seja $f(\mathbf{x}|\omega)$ a função de densidade de \mathbf{x} com parâmetros $\omega \in \Omega$ e seja a densidade de \mathbf{y} dada por

$$g(\mathbf{y}|\omega) = \int_{X(\mathbf{y})} f(\mathbf{x}|\omega) d\mathbf{x}, \quad (\text{A.1})$$

onde $X(\mathbf{y}) = \{ \mathbf{x} : \mathbf{y} = \mathbf{y}(\mathbf{x}) \}$ e $\omega \in \Omega$ são os parâmetros desconhecidos. Especificamente, assume-se que há uma aplicação $\mathbf{x} \rightarrow \mathbf{y}$ de X para Y e que $X(\mathbf{y}) = \{ \mathbf{x} : \mathbf{y} = \mathbf{y}(\mathbf{x}) \}$.

Os parâmetros ω serão estimados por Máxima Verossimilhança (MV), i.e., maximizando-se $g(\mathbf{y}|\omega)$ sobre $\omega \in \Omega$. Em muitos problemas estatísticos, a maximização quando se especifica completamente os dados, $f(\mathbf{x}|\omega)$, é mais simples do que quando se especifica dados incompletos, $g(\mathbf{y}|\omega)$.

O passo M é a maximização de $f(\mathbf{x}|\omega)$ sobre $\omega \in \Omega$. O passo M simula a estimação de MV que poderia ser realizada se os dados fossem completos; mas como os dados são incompletos, pois \mathbf{x} não é observável, substitui-se o $\ln f(\mathbf{x}|\omega)$ pela sua esperança condicional dado \mathbf{y} (que, no caso da família exponencial regular descrita em (A.6), é equivalente à esperança a posteriori das estatísticas suficientes baseadas nos dados completos como veremos adiante) e a estimativa corrente dos parâmetros $\omega^{(p)}$ (passo E).

Assim, o passo M pode ser executado através das estatísticas suficientes produzidas no passo E para a verossimilhança. Para famílias de distribuições gerais que não pertencem à família exponencial regular, o método é utilizado heurísticamente, não mais em função de

estatísticas suficientes, mas em função de uma partição da função de verossimilhança quando se especifica dados incompletos (A.3 e A.4). Para esse fim, seja

$$k(\mathbf{x}|\mathbf{y}, \omega) = \frac{f(\mathbf{x}|\omega)}{g(\mathbf{y}|\omega)} \quad (\text{A.2})$$

a densidade condicional de \mathbf{x} dados \mathbf{y} e ω . Então, a função $\ln(\text{verossimilhança})$ pode ser escrita como

$$L(\tilde{\omega}) = \ln g(\mathbf{y}|\tilde{\omega}) = Q(\tilde{\omega}|\omega) - H(\tilde{\omega}|\omega), \quad (\text{A.3.})$$

onde assume-se que as funções $Q(\tilde{\omega}|\omega) = E[\ln f(\mathbf{x}|\tilde{\omega})|\mathbf{y}, \omega]$ e $H(\tilde{\omega}|\omega) = E[\ln k(\mathbf{x}|\mathbf{y}, \tilde{\omega})|\mathbf{y}, \omega]$ existem para todos os pares $(\tilde{\omega}, \omega)$. Define-se a iteração EM $\omega^{(p)} \rightarrow \omega^{(p+1)} \in M(\omega^{(p)})$ como:

passo E: Determinar $Q(\omega|\omega^{(p)})$. (A.4)

passo M: Escolher $\omega^{(p+1)}$ como um valor de $\omega \in \Omega$ que maximiza $Q(\omega|\omega^{(p)})$. (A.5)

Note que $M(\omega^{(p)})$ é um conjunto de valores ω que maximiza $Q(\omega|\omega^{(p)})$ sobre $\omega \in \Omega$

Muitas aplicações do EM consideram densidades que pertencem à família exponencial regular, para o qual os passos E e M tem formas especiais. É dentro dessa família exponencial que o algoritmo EM está fundamentada estatisticamente em sua formulação. Na situação geral, o algoritmo é formulado heurísticamente (Dempster *et alii.*, 1977). Considere que $f(\mathbf{x}|\omega)$ pertença à família exponencial regular da forma

$$f(\mathbf{x}|\omega) = \frac{b(\mathbf{x}) \exp[\omega' t(\mathbf{x})]}{a(\omega)}, \quad (\text{A.6})$$

onde ω denota o vetor $r \times 1$ de parâmetros, $t(\mathbf{x})$ é o vetor $r \times 1$ das estatísticas suficientes dos dados completos e $0 < a(\omega) < \infty$. A família exponencial regular possui características como:

- a) o conjunto dos parâmetros ω restringe-se a um conjunto convexo r -dimensional Ω (subconjunto aberto do \mathbf{R}^n) de modo que defina uma densidade para todo ω em Ω ,

- b) $\omega' t(\mathbf{x}) = \sum_{i=1}^r \omega_i t_i(\mathbf{x})$ representa o produto interno entre as variáveis canônicas ω e $t(\mathbf{x})$;
- c) o suporte de $f(\mathbf{x}|\omega)$ é o mesmo para $\forall \omega \in \Omega$, permitindo a redução dos dados através de suficiência;
- d) $t(\mathbf{x})$ é uma estatística suficiente e completa.

Os passos da iteração na família exponencial regular podem ser descritos como:

passo E: Estimar as estatísticas suficientes $t(\mathbf{x})$ através de

$$t^{(p)} = E[t(\mathbf{x}) | \mathbf{y}, \omega^{(p)}]. \quad (\text{A.7})$$

passo M: Determinar $\omega^{(p+1)}$ como solução das equações

$$E[t(\mathbf{x}) | \omega] = t^{(p)} \quad (\text{A.8})$$

Às vezes, não é possível realizar numericamente o passo M. Dempster *et alii.* (1977) definiram um algoritmo EM generalizado (GEM) como o sistema iterativo $\omega^{(p)} \rightarrow \omega^{(p+1)} \in M(\omega^{(p)})$, onde $\omega \rightarrow M(\omega)$ é tal que

$$Q(\ddot{\omega} | \omega) \geq Q(\omega | \omega), \text{ para todo } \ddot{\omega} \in M(\omega).$$

O EM é uma caso particular do GEM. Para qualquer caso $\{\omega^{(p)}\}$ de um algoritmo GEM (Dempster *et alii.*, 1977):

$$L(\omega^{(p+1)}) \geq L(\omega^{(p)})$$

e

$$H(\omega | \omega) \geq H(\ddot{\omega} | \omega), \text{ para qualquer } \ddot{\omega} \in \Omega.$$

Detalhes sobre convergência do método são discutidos por Dempster *et alii.* (1977) e Wu (1983).

B

Filtro de Kalman

Kalman (1960) desenvolveu um método recursivo que pode ser utilizado na estimação e predição de séries temporais baseado no modelo espaço de estados. Esse modelo foi originalmente desenvolvido para aplicações em engenharia de controle.

Para descrever o modelo espaço de estados considere um conjunto de m variáveis de estado que mudam com o tempo, representadas pelo vetor de estado α_t . As variáveis de estado podem ser, por exemplo, um sinal que denota a posição e velocidade de um foguete. Na maioria dos casos, o sinal não é diretamente observável, sendo sujeito a distorções sistemáticas e contaminação por ruídos. As N variáveis contidas no vetor y_t ($N \times 1$), observadas em tempos igualmente espaçados, i.e., $t = 1, \dots, T$, são associadas às variáveis de estado através das equações de medida ou observação, que podem ser representadas por:

$$y_t = Z_t \alpha_t + f_t + S_t \zeta_t, \quad (\text{B.1})$$

onde y_t é o vetor $N \times 1$ de observações; Z_t e S_t são matrizes fixadas de dimensões, respectivamente, $N \times m$ e $N \times n$; α_t é o vetor de estados $m \times 1$; f_t é o vetor $N \times 1$ de efeitos não aleatórios e ζ_t é o vetor $n \times 1$ de ruídos aleatórios com média 0 e matriz de covariância H_t .

Embora α_t não seja diretamente observável, assume-se que ele seja governado por um processo definido pelas equações de transição,

$$\alpha_t = \Phi_t \alpha_{t-1} + f_{\alpha_t} + R_t \eta_t, \quad (\text{B.2})$$

onde Φ_t e R_t são matrizes fixadas de dimensões $m \times m$ e $m \times g$, respectivamente; a matriz Φ_t é chamada de matriz de transição de estado do tempo $t-1$ para t ; f_{α_t} é o vetor que representa a entrada não aleatória ao estado no tempo t e η_t é o vetor $g \times 1$ de ruídos aleatórios com média 0 e matriz de covariância Q_t .

Assume-se que os ruídos ζ_t e η_t são não correlacionados, ou seja,

$$\begin{pmatrix} \zeta_t \\ \eta_t \end{pmatrix} \sim \text{WN} \left[0, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right], \quad (\text{B.3})$$

e que $E(\alpha_0 \eta_t') = 0$ e $E(\alpha_0 \zeta_t') = 0$, onde WN significa *White Noise* (ruído branco) e α_0 é o vetor inicial de estado.

A estimativa (ou predição) de α_t condicionada às informações até o tempo t , denotada por $\alpha_{t|t-1}$ é dada por

$$\alpha_{t|t-1} = E(\alpha_t | y_1, \dots, y_{t-1}),$$

com matriz de covariância dada por

$$P_{t|t-1} = E[(\alpha_t - \alpha_{t|t-1})(\alpha_t - \alpha_{t|t-1})' | y_1, \dots, y_{t-1}].$$

Como os dados podem ser desigualmente espaçados no tempo, ou podendo haver dados perdidos, é conveniente modificar a notação. Assim, as equações de medidas e de transição podem ser representadas, respectivamente, por

$$y(t) = Z(t)\alpha(t) + f(t) + S(t)\zeta(t), \quad (\text{B.4})$$

e

$$\alpha(t) = \Phi(t, t - \delta t)\alpha(t - \delta t) + f_\alpha(t) + R(t)\eta(t), \quad (\text{B.5})$$

onde δt é o intervalo de tempo entre as observações. Nesse caso, a matriz de transição, $\Phi(t, t - \delta t)$, e a matriz de covariância de $\eta(t)$, $Q(\delta t)$, dependem do espaçamento entre as observações.

A representação do modelo linear dinâmico descrito pelas equações (B.1) e (B.2), ou (B.4) e (B.5), é conhecido como forma em espaço de estados. Qualquer modelo ARMA (apêndice C) pode ser colocado na forma em espaço de estado (Harvey, 1981, pp. 103). Também é possível representar nessa forma, por exemplo, o modelo de efeitos aleatórios com erros correlacionados serialmente ou modelar sistemas com parâmetros variando no tempo.

O filtro de Kalman é um conjunto de equações que permite atualizar as estimativas a cada nova observação disponível. Por exemplo, à medida que o foguete envia seus dados, a estimativa de sua posição é imediatamente atualizada. O termo filtro é usado porque a estimação dos estados do sistema (e.g., posição e velocidade) é feita no instante em que se dispõe de um conjunto mais recente de observações.

O filtro de Kalman produz uma solução ótima aos problemas de predição e atualização (Harvey, 1981, pp. 101). Se as observações forem normalmente distribuídas, e se o estimador do vetor de estado for o melhor disponível, então a predição e a atualização serão as melhores disponíveis. Schweppe (1965) mostrou que o filtro de Kalman pode ser usado para calcular verossimilhanças quando os erros têm distribuição gaussiana. Na ausência de normalidade, a predição e atualização serão as melhores disponíveis dentro de uma classe de estimadores e preditores lineares nas observações (Harvey, 1981, pp. 102). Como o vetor de estado é estocástico, o termo "melhor" deve ser interpretado cuidadosamente nesse contexto. Atenção deve ser dada ao erro de estimação, que envolve o conceito de *Estimador de Mínima Média Quadrática* (EMMQ), descrito por Harvey (1981, pp. 104).

À medida que as equações de filtragem são aplicadas, é produzida uma série de erros de predição. Harvey (1981, pp. 102) comenta que a função de verossimilhança para um conjunto de observações dependentes deve ser decomposta em termo desses erros. O filtro de Kalman fornece um mecanismo natural para lidar com os erros de predição.

O filtro de Kalman é conduzido basicamente em duas etapas: a primeira consiste em produzir um preditor ótimo da próxima observação, dada todas as informações passadas, através das equações de predição. Quando uma nova observação é obtida, ela é incorporada à estimativa do vetor de estado através das equações de atualização (segundo passo).

A recursão inicia-se com a estimativa do estado no tempo $t - \delta t$, $\alpha(t - \delta t | t - \delta t)$, e sua matriz de covariância $P(t - \delta t | t - \delta t)$.

Os passos do filtro são:

1. Calcular a predição do vetor de estado:

$$\alpha(t | t - \delta t) = \Phi(t, t - \delta t) \alpha(t - \delta t | t - \delta t) + f_u.$$

2. Calcular a matriz de covariância da predição:

$$P(t | t - \delta t) = \Phi(t, t - \delta t) P(t - \delta t | t - \delta t) \Phi'(t, t - \delta t) + R(t) Q(\delta t) R'(t).$$

3. Calcular a predição da próxima observação:

$$\tilde{y}(t | t - \delta t) = Z(t) \alpha(t | t - \delta t) + f.$$

4. Calcular a inovação (diferença entre o observado e o predito):

$$I(t) = y(t) - \tilde{y}(t|t - \delta t).$$

5. Calcular a matriz de covariância da inovação:

$$V(t) = Z(t)P(t|t - \delta t)Z'(t) + S(t)H(t)S'(t).$$

6. Atualizar a estimativa do vetor de estado:

$$\alpha(t|t) = \alpha(t|t - \delta t) + P(t|t - \delta t)Z(t)V^{-1}(t)I(t).$$

7. Calcular a matriz de covariância atualizada:

$$P(t|t) = P(t|t - \delta t) - P(t|t - \delta t)Z'(t)V^{-1}(t)Z(t)P(t|t - \delta t).$$

A inovação (ou erro de predição) $I(t)$ é um vetor $N \times 1$ com média zero e matriz de covariância $V(t)$, i.e.,

$$E(I(t)) = 0 \text{ e } E(I(t)I'(t)) = V(t);$$

e, como será verificado na representação da função de verossimilhança das inovações, as inovações $I(t)$, $t = 1, \dots, T$, são independentes.

As inovações desempenham o papel chave na atualização do vetor de estados. A "correção" do estado $\alpha(t|t - \delta t)$ para $\alpha(t|t)$ no passo 6 da recursão é feita somando-se o ganho do filtro de Kalman multiplicado por $I(t)$:

$$G(t|t - \delta t)I(t),$$

onde $G(t|t - \delta t) = P(t|t - \delta t)Z(t)V^{-1}(t)$ é o ganho do filtro de Kalman.

Para descrever a função de verossimilhança dos erros de predição (ou inovação), considere por simplicidade que $N = 1$ (uma variável é observada a cada instante) e considere que $\delta t = 1$ (as observações são igualmente espaçadas).

Considere um conjunto de T observações dependentes com média μ e matriz de covariância Σ retirada de uma distribuição normal multivariada, i.e., $y \sim N(\mu, \Sigma)$. Por exemplo, nas equações (B.4) e (B.5) suponha que $\zeta(t) \sim N(0, H(t))$, $\eta(t) \sim N(0, Q(t))$ e $\alpha(0) \sim N(\alpha(0|0), P(0|0))$, onde $\alpha(0|0)$ e $P(0|0)$ são conhecidos. Os valores iniciais $\alpha(0|0)$ e $P(0|0)$ juntamente com os parâmetros nas equações de transição e de medida determinam as especificações de μ e Σ . Em geral, entretanto, não é necessário calcular $\theta = \{\mu, \Sigma\}$. A aplicação do filtro de Kalman produz o EMMQ de $y(t)$, ($t = 1, \dots, T$), dado as observações anteriores, e produz também as inovações $I(t)$. Assim, a função de verossimilhança deve ser escrita de modo que se possa estimar $\theta^* = \{y(t)|y(t-1), \dots, y(1)$ e $\text{var}(y(t)|y(t-1), \dots, y(1)) = v(t)\}$.

A função $l = -2 \ln(\text{verossimilhança})$ é dada por

$$l_{\theta}(y) = T \ln 2\pi + \ln |\Sigma| + (y - \mu)' \Sigma^{-1} (y - \mu) \quad (\text{B.6})$$

que, pela propriedade de probabilidade condicional, pode ser representada como

$$l_{\theta, \theta^*}(y(1), \dots, y(T)) = l_{\theta}(y(1), \dots, y(T-1)) + l_{\theta^*}(y(T)|y(1), \dots, y(T-1)). \quad (\text{B.7})$$

Considere o problema de estimar $y(T)$ dado que $y(T-1), \dots, y(1)$ são conhecidos. Se $\hat{y}(T|T-1)$ é um estimador de $y(T)$ dada as observações anteriores, o erro de predição pode ser escrito como $y(T) - y(T|T-1) =$

$$= [y(T) - E(y(T)|y(T-1), \dots, y(1))] + [E(y(T)|y(T-1), \dots, y(1)) - \hat{y}(T|T-1)],$$

onde $E(y(T)|y(T-1), \dots, y(1))$ é a média da distribuição de $y(T)$ condicionada a $y(T-1), \dots, y(1)$. O Erro Quadrático Médio (EQM) de $y(T)$ condicionado a $y(T-1), \dots, y(1)$ é:

$$E(y(T) - y(T|T-1))^2 =$$

$$= \text{Var}(y(T)|y(T-1), \dots, y(1)) + E[\hat{y}(T|T-1) - E(y(T)|y(T-1), \dots, y(1))]^2.$$

Assim, o EMMQ de $y(T)$ condicionado a $y(T-1), \dots, y(1)$ é dado por

$$\hat{y}(T|T-1) = E(y(T)|y(T-1), \dots, y(1)),$$

e a variância do erro de predição é

$$E(y(T) - \tilde{y}(T|T-1))^2 = \text{Var}(y(T)|y(T-1), \dots, y(1)) = v(T).$$

Logo, o termo $l_{\theta} (y(T)|y(1), \dots, y(T-1))$ de (B.7) pode ser escrito como:

$$l_{\theta} (y(T)|y(1), \dots, y(T-1)) = \ln 2\pi + \ln v(T) + \frac{(y(T) - \tilde{y}(T|T-1))^2}{v(T)},$$

e pode ser interpretado como a função de verossimilhança do erro de predição $I(T) = y(T) - \tilde{y}(T|T-1)$.

Aplicando-se novamente a propriedade de probabilidade condicional ao termo $l_{\theta} (y(1), \dots, y(T-1))$ de (B.7), e repetindo-se sucessivamente, obtém-se a seguinte expressão:

$$l_{\theta} (y(1), \dots, y(T)) = \sum_{t=2}^T l_{\theta} (y(t)|y(1), \dots, y(t-1)) + l_{\theta} (y(1)). \quad (\text{B.8})$$

Para $t = 2, \dots, T$, a média de $y(t)$ condicionada a $y(t-1), \dots, y(1)$ é igual ao EMMQ de $y(t)$ dada as observações passadas:

$$\tilde{y}(t|t-1) = E(y(t)|y(t-1), \dots, y(1)).$$

Assim, cada uma das distribuições condicionais é a distribuição da inovação associada ao preditor ótimo, enquanto que $l_{\theta} (y(1))$ é a distribuição não condicional de $y(1)$. Entretanto, se considerarmos $\mu(1)$ como o EMMQ de $y(1)$, dado que não há informações prévias, então o termo $y(1) - \mu(1)$ pode ser considerado como a inovação $I(1)$. Assim, sua variância pode ser denotada como $v(1)$.

A expressão (B.8) é a forma (B.6) decomposta em T distribuições independentes das inovações $I(1), \dots, I(T)$, onde assume-se que $\tilde{y}(t|0) = \mu(1)$. Portanto, para $t = 1, \dots, T$, cada inovação $I(t)$ tem média 0 e variância $v(t)$; a verossimilhança (B.8) pode ser expressa como

$$l_{\theta} (y(1), \dots, y(T)) = T \ln 2\pi + \sum_{t=1}^T \ln v(t) + \sum_{t=1}^T \frac{I^2(t)}{v(t)}. \quad (\text{B.9})$$

A decomposição do erro de predição pode ser interpretada em termos da decomposição de Cholesky de Σ^{-1} . Se L é uma matriz triangular inferior cuja diagonal

principal é formada por 1 (uns), então a matriz Σ^{-1} pode ser decomposta na forma $\Sigma^{-1} = L'DL$, onde $D = \text{diag}(v^{-1}(1), \dots, v^{-1}(T))$. Essa fatorização é única (Harvey, 1981, pp. 14) e os erros de predição são dados pela transformação $I = L y$, onde I é o vetor $T \times 1$ das inovações. Como o Jacobiano dessa transformação é $|L| = 1$, a função da verossimilhança conjunta dos elementos do vetor das inovações é dada por (B.9):

note que $|\Sigma^{-1}| = |L'| |D| |L| = |D|$ e $\ln|\Sigma| = \sum_{t=1}^T \ln v(t)$.

Para o caso com N variáveis o argumento é o mesmo utilizado anteriormente, com $I(t)$ sendo um vetor $N \times 1$. Nesse caso, a função de verossimilhança pode ser escrita como (Harvey, 1981, pp. 16):

$$l_{\theta^*}(y(1), \dots, y(T)) = TN \ln 2\pi + \sum_{t=1}^T \ln |V(t)| + \sum_{t=1}^T I'(t) V^{-1}(t) I(t), \quad (\text{B.10})$$

onde $y(1) \sim N(\mu(1), V(1))$, $I(1) = y(1) - \mu(1)$ e $\theta^* = \{I(t), V(t), \text{ para } t \geq 1\}$.

C

Modelos ARMA

O modelo autoregressivo e média móvel de ordem p e q com média zero, denotado por ARMA(p,q), pode ser representado por (Box e Jenkins, 1976, pp. 74):

$$\phi(B)\varepsilon_t = \theta(B)a_t, \quad (C.1)$$

onde $t = 0, 1, \dots, T$; $\phi(B) = (1 - B\phi_1 - B^2\phi_2 - \dots - B^p\phi_p)$ e $\theta(B) = (1 - B\theta_1 - B^2\theta_2 - \dots - B^q\theta_q)$, com $p+q < T$. Os termos $\phi(B)$ e $\theta(B)$ são funções polinomiais de B (operador *atraso*, tal que $B\varepsilon_t = \varepsilon_{t-1}$ e $Ba_t = a_{t-1}$) que representam, respectivamente, os componentes autoregressivo (AR(p)) e média móvel (MA(q)). Assume-se que o processo seja estacionário e invertível, i.e. que as raízes de $\phi(B)$ e $\theta(B)$ caiam fora do círculo unitário no plano Complexo (sem raízes em comum); e o termo a_t é o ruído branco com média 0, variância 1 e $E(a_t a_{t'}) = 0$ para $t \neq t'$.

Aplicando-se os operadores B em (C.1), o ARMA(p,q) resulta em:

$$\varepsilon_t - \phi_1\varepsilon_{t-1} - \phi_2\varepsilon_{t-2} - \dots - \phi_p\varepsilon_{t-p} = a_t - \theta_1a_{t-1} - \theta_2a_{t-2} - \dots - \theta_qa_{t-p}, \quad (C.2)$$

onde os coeficientes ϕ_1 a ϕ_p e θ_1 a θ_q representam, respectivamente, as componentes autoregressivas (AR(p)) e média-móveis (MA(q))

Para o modelo ser estacionário e invertível, as raízes complexas de

$$1 - \sum_{k=1}^p \phi_k z^k = 0,$$

e

$$1 - \sum_{k=1}^q \theta_k z^k = 0.$$

devem cair fora do círculo unitário. Assim, dizemos que o processo é estacionário quando a série formada pelos pesos $\phi_k z^k$ for convergente; e dizemos que o processo é invertível quando a série formada pelos pesos $\theta_k z^k$ for convergente. O AR(p) (ARMA($p,0$)) é sempre invertível e MA(q) (ARMA($0,q$)) sempre é estacionário (Box e Jenkins, 1976, cap. 3).

A função de autocorrelação (fac) do ARMA(p,q) pode ser obtida tomando-se a esperança de $\varepsilon_t \varepsilon_{t-k}$. Denotando-se a covariância como $E(\varepsilon_t \varepsilon_{t-k}) = \gamma_k$, obtemos:

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + \gamma_{za(k)} - \theta_1 \gamma_{za(k-1)} - \dots - \theta_q \gamma_{za(k-q)}, \quad (C.3)$$

onde $\gamma_{za(k)} = E(\varepsilon_{t-k} \alpha_t)$ é a covariância cruzada entre ε e α .

Como ε_{t-k} depende apenas dos choques ocorridos até o tempo $t - k$, então segue que

$$\gamma_{za(k)} = 0, \text{ se } k > 0 \text{ e } \gamma_{za(k)} \neq 0 \text{ se } k \leq 0. \quad (C.4)$$

Assim, para $k \geq q + 1$, a covariância é dada por

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}. \quad (C.5)$$

Dividindo-se a expressão anterior (C.5) por $\gamma_0 = E(\varepsilon_t^2) = \text{var}(\varepsilon_t)$ obtemos a fac para $k \geq q + 1$:

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}. \quad (C.6)$$

A Tabela C.1 ilustra a forma das funções de autocorrelação para o AR(1), AR(2), ARMA(1,1) e ARMA(2,1) obtidas por (C.3):

Tabela C.1 - Exemplos de fac

Modelo:	fac
AR(1)	$\rho_0 = 1$ $\rho_k = \phi^k, k \geq 1$
AR(2)	$\rho_0 = 1$ $\rho_1 = \frac{\phi_1}{1 - \phi_2}$ $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, k \geq 1$
ARMA(1,1)	$\rho_0 = 1$ $\rho_1 = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta}$ $\rho_k = \phi^{k-1} \rho_1, k \geq 1$
ARMA(2,1)	$\rho_0 = 1$ $\rho_1 = \frac{(1 - \phi_1\theta)(\phi_1 - \theta) + \theta\phi_2^2}{(1 - \phi_2)(1 + \theta^2) - 2\phi_1\theta}$ $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, k \geq 2$

D

Tempo Contínuo

D.1. AR(1) Contínuo

Um processo autoregressivo de primeira ordem em tempo contínuo com média 0, A(1), pode ser representado por uma equação diferencial de primeira ordem, (Pandit e Wu, 1983, pp. 220),

$$\partial x(t) + \alpha x(t) \partial t = \partial w(t), \quad (D.1)$$

onde $\alpha > 0$, para que seja estacionário, e $\partial w(t)$ é "ruído branco" em tempo contínuo com média zero. "Ruído branco" em tempo contínuo quer dizer que sua integral $w(t)$, o ruído branco integrado, é um processo de Wiener, ou passeio aleatório em tempo contínuo, ou moção Browniana.

O processo de Wiener é o limite de um passeio aleatório discreto quando o intervalo de tempo diminui, isto é,

$$w(t + \delta t) = w(t) + \eta(t), \quad (D.2)$$

onde $\delta t > 0$ e $\eta(t)$ são variáveis aleatórias independentes com médias 0. Sobre um intervalo de tempo limitado, a variância é proporcional ao tamanho do intervalo, isto é, para $\delta t > 0$

$$\text{var}(\eta(t)) = \text{var}(w(t + \delta t) - w(t)) = \quad (D.3)$$

$$= \text{var}\left(\int_t^{t+\delta t} \delta w(t)\right) = \delta t Q, \quad (D.4)$$

onde Q é chamada de variância do ruído branco.

Se o processo for observado em um dado momento t , então essa observação será somada às informações passadas para predizer o futuro. A predição do processo no tempo $t + \delta t$ (D.8) pode ser calculada solucionando-se a equação diferencial sem o ruído aleatório $\partial w(t)$ (D.5) e depois integrando-o de t a $t + \delta t$, isto é, uma solução para

$$\partial x(t) + \alpha x(t) \partial t = 0 \quad (D.5)$$

é

$$x(t) = \exp[-\alpha \delta t]. \quad (D.6)$$

Integrando-se (D.5) de t a $t + \delta t$ obtemos,

$$x(t + \delta t) = \exp[-\alpha \delta t] \exp[-\alpha t] = \quad (D.7)$$

$$= \exp[-\alpha \delta t] x(t) \quad (D.8)$$

Com hipótese de estacionariedade, $\alpha > 0$,

$$\phi(\delta t) = \exp[-\alpha \delta t] \in (0, 1).$$

O processo de Markov em tempo contínuo (D.1.), tem função de covariância dada por

$$\gamma(\tau) = \frac{Q \times \exp(-\alpha|\tau|)}{2\alpha},$$

onde τ é a defasagem.

A entrada aleatória sobre um intervalo de tempo limitado é

$$\eta(\delta t) = \int_t^{t+\delta t} \exp[-\alpha(t + \delta t - \tau)] \delta w(\tau)$$

e tem variância

$$\frac{Q\{1 - \exp[-2\alpha(\delta t)]\}}{2\alpha}.$$

Quando o intervalo δt cresce, a variância se aproxima da variância do processo.

Se o processo $A(1)$ for observado em tempos igualmente espaçados (h), o processo resultante é um $AR(1)$ com coeficiente $\phi = \exp(-\alpha h)$ e $\sigma^2 = Q\{1 - \exp[-2\alpha h]\}/2\alpha$.

D.2. ARMA(p,q) Contínuo

Os modelos ARMA em tempo contínuo são mais complicados. Pandit e Wu (1983, pp. 229) e Jones (1993, pp. 128) discutem que os processos discretos podem ser vistos como processos contínuos observados em tempos igualmente espaçados. Se um processo ARMA(p,q) contínuo com $q < p$ for amostrado em tempos igualmente espaçados, o resultado é um processo discreto ARMA(p,p-1). Assim, o AR(1) contínuo quando amostrado em tempos igualmente espaçados produz o AR(1) discreto (ARMA(1,0)). Se um processo ARMA(p,q) contínuo com erro observacional (i.e., variância é dada por

$\sigma^2(1 + \sigma_0^2)$) e com $q < p$ for amostrado em tempos igualmente espaçados, então o processo resultante é o ARMA(p,p).

O ARMA(p,q) contínuo é definido pela equação diferencial linear (estocástica):

$$\frac{\partial^p \varepsilon(t)}{\partial t^p} + \alpha_{p-1} \frac{\partial^{p-1} \varepsilon(t)}{\partial t^{p-1}} + \dots + \alpha_0 \varepsilon(t) = \delta_q \frac{\partial^q \eta(t)}{\partial t^q} + \delta_{q-1} \frac{\partial^{q-1} \eta(t)}{\partial t^{q-1}} + \dots + \eta(t), \quad (D.9)$$

onde $\eta(t)$ é o "ruído branco" em tempo contínuo, como no caso do AR(1) contínuo e os coeficientes α e δ representam, respectivamente, os coeficientes de autoregressão e média móvel.

Para o processo ser estacionário é necessário que $q < p$ e que as raízes complexas de

$$1 + \sum_{k=0}^{p-1} \alpha_k z^k = 0 \quad (D.10)$$

tenham partes reais negativas. As equações diferenciais lineares têm soluções da forma exponencial, e se as raízes complexas têm partes reais negativas, os distúrbios do sistema diminuem com o tempo. A equação (D.10) é chamada de *equação do sistema*. O sistema tem um atraso mínimo (*minimum phase* ou *minimum delay*) se as raízes de

$$1 + \sum_{k=0}^{q-1} \delta_k z^k = 0 \quad (D.10)$$

tiverem partes reais não positivas.

Referências Bibliográficas

- AKAIKE, H. (1973) Information theory and an extension of the maximum likelihood principle. *Second international symposium on information theory*, Budapest: Akademia Kaido, p. 267-281.
- ANDERSON, R.L. & BANCROFT, T.A. (1952) *Statistical Theory in Research*. New York, McGraw-Hill.
- ANDRADE, D.F. & SINGER, J.M. (1986) *Análise de Dados Longitudinais*. Campinas, VII Simpósio Nacional de Probabilidade e Estatística.
- BOX, G.E.P. (1954) Some theorems on quadratic forms applied to the study of analysis of variance problems. *Annals of Mathematical Statistics*. 25: 290-302.
- BOX, G.E.P. & JENKINS, G.M. (1976) *Time Series Analysis, Forecasting and Control*. San Francisco, Holden-Day.
- BOZDOGAN, H. (1987) Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*. 52: 345-370.
- CHI, E.M. (1988) *Analysis of Longitudinal Data with Random Effects and Autocorrelated Errors*. Unpublished thesis manuscript, University of Wisconsin, Madison.
- CHI, E.M. & REINSEL, G.C. (1989) Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association*. 84: 452-459.
- COLE, J.W.L. & GRIZZLE, J.E. (1966) Applications of multivariate analysis of variance to repeated measurements experiments. *Biometrics*. 22: 810-828.
- CORBEIL, R.R. & SEARLE, S.R. (1976) Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics*. 18: 31-38.
- CROWDER, M.J. & HAND, D.J. (1990) *Analysis of Repeated Measures*. London, Chapman and Hall.
- DEMPSTER, A.P.; LAIRD, N.M. and RUBIN, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological*. 39: 1-38.
- DEMPSTER A.P. & RYAN, L.M. (1985) Weighted normal plots. *Journal of the American Statistical Association*. 80: 845-850.
- DIELMAN, T.E. (1989) *Pooled Cross-sectional and Time Series Data Analysis*. New York, Marcel Dekker.
- DIGGLE, P.J. (1988) An approach to the analysis of repeated measurements. *Biometrics*. 44: 959-971.

- DRAPER, N.R. & SMITH, H. (1981) *Applied Regression Analysis*. New York, John Wiley.
- EFRON, B. & MORRIS, C. (1973) Stein's estimation rule and its competitors - an empirical Bayes approach. *Journal of the American Statistical Association*. 68: 117-130.
- ELSTON, R.C. & GRIZZLE, J.E. (1962) Estimation of time-response curves and their confidence bands. *Biometrics*. 18: 148-159.
- GALPIN, J.S. & HAWKINS, D.M. (1984) The use of recursive residuals in checking model fit in linear regression. *The American Statistician*. 38: 94-105.
- GEISSER, S. & GREENHOUSE, S.W. (1958) An extension of Box's results on the use of the F-distribution in multivariate analysis. *Annals of Mathematical Statistics*. 29: 885-891.
- GRIZZLE, J.E. & ALLEN, D.M. (1969) Analysis of growth and dose response curves. *Biometrics*. 25: 357-381.
- GUIMARÃES, P.R.B. (1994) *Modelo Linear Misto de Laird-Ware: Predição de Efeitos Aleatórios e Estimação de Parâmetros via Filtro de Kalman*. Dissertação de Mestrado (IMECC), Universidade de Campinas, SP.
- HARTIGAN, J.A. (1969) Linear Bayesian methods. *Journal of the Royal Statistical Society, Series B, Methodological*. 31: 446-454.
- HARVEY, A.C. (1978) The estimation of time-varying parameters from panel data. IN: Annales de L'insée, 30-31 *The Econometrics of Panel Data*. Paris, Institut national de la statistique et des études économiques, p. 203-226.
- HARVEY, A.C. (1981) *Time Series Models*. Oxford, Philip Allan.
- HARVILLE, D.A. (1974) Bayesian inference for variance components using only error contrasts. *Biometrika*. 61: 383-385.
- HARVILLE, D.A. (1976) Extension of the Gauss-Markov theorem to include the estimation of random effects. *The Annals of Statistics*. 4: 384-395.
- HARVILLE, D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*. 72: 320-340.
- HARVILLE, D.A. & CARRIQUIRY, A.L. (1992) Classical and bayesian prediction as applied to an unbalanced mixed linear model. *Biometrics*. 48: 987-1003.
- HASTIE, T. & TIBSHIRANI, R. (1993) Varying-coefficient models. *Journal of the Royal Statistical Society, Series B, Methodological*. 55: 757-796.
- HAUSMAN, J.A. (1978) Specification tests in econometrics. *Econometrica*. 46: 1251-1271.
- HENDERSON, C.R. (1950) Estimation of genetic parameters (abstract). *Annals of Mathematical Statistics*. 21: 309-310.

- HENDERSON, C.R. (1963) Selection index and expected genetic advance. IN: Hanson, W.D. and Robinson, H.F. *Statistical Genetics and Plant Breeding*. Washington, D.C. National Academy of Sciences and National Research Council Publication, No. 982, p. 141-163.
- HENDERSON, C.R. (1973) *Maximum Likelihood Estimation of Variance Components*. Unpublished manuscript, Department of Animal Science, Cornell University, Ithaca, New York.
- HENDERSON, C.R. (1984) *Applications of Linear Models in Animal Breeding*. Ontario, University of Guelph.
- HSIAO, C. (1986) *Analysis of Panel Data*. Cambridge, Cambridge University Press.
- HUYNH, H. & FELDT, L.S. (1970) Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association*, 65: 1582-1589.
- JENNRICH, R.I. & SAMPSON, P.F. (1976) Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, 18: 11-17.
- JENNRICH, R.I. & SCHLUCHTER, M.D. (1986) Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42: 805-820.
- JONES, R.H. (1985) Time series analysis with unequally spaced data. IN: HANNAN, E.J., KRISHNAIAH, P.R. and RAO, M.M. *Handbook of Statistics, vol. 5: Time Series in the Time Domain*. New York, North-Holland, p. 157-177.
- JONES, R.H. (1987) Serial correlation in unbalanced mixed models. IN: *Bulletin of the International Statistical Institute, Proceedings of the 46th session*, book 4. Tokyo: International Statistical Institute, p. 105-122.
- JONES, R.H. (1990) Serial correlation or random subject effects? *Communications in Statistics B, Simulation and Computation*, 19: 1105-1123.
- JONES, R.H. (1993) *Longitudinal Data with Serial Correlation: A State-Space Approach*. London, Chapman and Hall.
- JONES, R.H. & BOADI-BOATENG, F. (1991) Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics*, 47: 161-175.
- JOURNEL, A. & HUIJBREGTS, C. J. (1978) *Mining Geostatistics*. London, Academic Press.
- KALMAN, R.E. (1960) A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* 82D: 35-45.
- KHATRI, C.G. (1966) A note on MANOVA model applied to problems in growth curves. *Annals of the Institute of Statistical Mathematics*, 18: 75 - 86.
- KLEINBAUM, D.G. (1973) A generalization of the growth curve model which allows missing data. *Journal of Multivariate Analysis*, 3: 117 - 124.

- KRAFT, J. & RODEKOH, M. (1978) Regional demand for gasoline: A temporal cross-section specification. *Journal of Regional Sciences*. 18: 45-55.
- LAIRD, N.M. (1982) Computation of variance components using the EM algorithm. *Journal of Statistical Computation and Simulation*. 14: 295-303.
- LAIRD, N.M.; LANGE, N. and STRAM, D. O. (1987) Maximum likelihood computations with repeated measures: application of the EM algorithm. *Journal of the American Statistical Association*. 82: 97-105.
- LAIRD, N.M. & WARE, J.H. (1982) Random-effects models for longitudinal data. *Biometrics*. 38: 963-974.
- LANGE, L. & RYAN, L. (1989) Assessing normality in random effects models. *The Annals of Statistics*. 17: 624-642.
- LIANG, K.Y. & ZEGER, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*. 73: 13-22.
- LIN, C.Y. & McALLISTER, A.J. (1984) Monte Carlo comparison of four methods for estimation of genetic parameters in the univariate case. *Journal of Dairy Science*. 67: 2389-2398.
- LINDSTROM, M.J. & BATES, D.M. (1988) Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*. 83: 1014-1022.
- LITTLE, R.J.A. & RUBIN, D.B. (1987) *Statistical Analysis with Missing Data*. New York, John Wiley.
- LOUIS, T.A. (1988) Discussion: analysis of data with missing values. *Statistics in Medicine*. 7: 357-360.
- LOUIS, T.A. (1988b) General methods for analysing repeated measures. *Statistics in Medicine*. 7: 29-45.
- MARDIA, K.V., KENT, J.T. and BIBBY, J.M. (1979) *Multivariate analysis*. London, Academic Press.
- MARITZ, J.S. & LWIN, T. (1989) *Empirical Bayes Methods*. 2nd.. London, Chapman and Hall.
- McCULLAGH, P. & NELDER, J.A. (1983) *Generalized linear models*. London, Chapman and Hall.
- MEHTA, J.S., NARASIMHAM, G.V.L. and SWAMY, P.A.V.B. (1978) Estimation of a dynamic demand function for gasoline with different schemes of parameter variation. *Journal of Econometrics*. 7: 263 - 279.
- MUÑOZ, A.; CAREY, V.; SCHOUTEN, J.P.; SEGAL, M. and ROSNER, B. (1992) A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*. 48: 733-742.

- NELDER, J.A. & WEDDERBURN, R.W. (1972) Generalized linear models. *Journal of the Royal Statistical Society, A*. 135: 370-384.
- PANDIT, S.M. & WU, S.M. (1983) *Time Series and System Analysis with Applications*. New York, John Wiley.
- PANTULLA, S.G. & POLLOCK, K.H. (1985) Nested analysis of variance with autocorrelated errors. *Biometrics*. 41: 909-920.
- PARK, T. & WOOLSON, R.F. (1992) Generalized multivariate models for longitudinal data. *Communications in Statistics B, Simulation and Computation*. 21: 925-946.
- PARK, T. (1993) A comparison of the generalized estimating equation approach with the maximum likelihood approach for repeated measurements. *Statistics in Medicine*. 12: 1723-1732.
- PATTERSON, H.D. & THOMPSON, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*. 58: 545-554.
- POSKITT, D.S. & TREMAYNE, A.R. (1987) Determining a portfolio of linear time series models. *Biometrika*. 74: 125-137.
- POTTHOFF, R.F. & ROY, S.N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*. 51: 313-326.
- RAO, C.R. (1959) Some problems involving linear hypotheses in multivariate analysis. *Biometrika*. 46: 49-58.
- RAO, C.R. (1965) The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*. 52: 447-458.
- RAO, C.R. (1966) Covariance adjustment and related problems in multivariate analysis. IN: KRISHNAIAH, P.R. *Multivariate Analysis, vol. 2*. New York, Academic Press, p. 87-103.
- RAO, C.R. (1973) *Linear Statistical Inference and its applications*. 2nd. New York, John Wiley.
- RAO, C.R. (1975) Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics*. 31: 545-554.
- ROBINSON, D. L. (1987) Estimation and use of variance components. *The Statistician*. 36: 3-14.
- ROBINSON, G.K. (1991) That BLUP is a good thing: the estimation of random effects. *Statistical Science*. 6: 15-51.
- ROCHON, J. (1992) ARMA covariance structures with time heteroscedasticity for repeated measures experiments. *Journal of the American Statistical Association*. 87: 777-784.
- ROCHON, J. & HELMS, R.W. (1989) Maximum likelihood estimation for incomplete repeated-measures experiments under an ARMA covariance structure. *Biometrics*. 45: 207-218.

- ROSENBERG, B. (1973) Linear regression with randomly dispersed parameters. *Biometrika*. 60: 65-72.
- RUSSEL, T.S. & BRADLEY, R.A. (1958) One-way variances in the two-way classification. *Biometrika*. 45: 111-129.
- SCHWARZ, G. (1978) Estimating the dimension of a model. *Annals of Statistics*. 6: 461-464.
- SCHWEPPE, F.C. (1965) Evaluation of likelihood functions for gaussian signals. *IEEE Trans. Inform. Theory*. 11: 61-70.
- SCHWERTMAN, N.C. & CARTER, N.J. (1993) The effect of first order autocorrelation on data analysis as measured by the Geisser-Greenhouse adjustment. *Communications in Statistics B, Simulation and Computation*. 22: 791-796.
- SEARLE, S.R. (1971) *Linear Models*. New York, John Wiley.
- SEARLE, S.R., CASELLA, G. and McCULLOCH, C.E. (1992) *Variance Components*. New York, John Wiley.
- STANEK III, E.J. & KOCH, G.G. (1985) The equivalence of parameter estimates from growth curve models and seemingly unrelated regression models. *The American Statistician*. 39: 149-152.
- THOMPSON, W.A. (1962) The problem of negative estimates of variance components. *Annals of Mathematical Statistics*. 33: 273-289.
- THOMPSON, R. (1977) The estimation of heritability with unbalanced data. I. Observations available on parents and offspring. *Biometrics*. 33: 485-495.
- TIMM, M.H. (1980) Multivariate analysis of variance of repeated measures. IN: KRISHNAIAH, P.R. *Handbook of Statistics, vol. 1*. North Holland, Amsterdam, pp. 42-47.
- UZUMAKI, E.T. (1994) *Geoestatística Multivariada: Estudo de Métodos de Predição*. Dissertação de Mestrado (IMECC), Universidade de Campinas, SP.
- WARE, J.H. (1983) Groth curves. IN: KOTZ S. & JOHNSON, N.L. *Encyclopedia of Statistical Sciences, vol. 3*. New York, John Wiley, p. 539-542.
- WARE, J.H. (1985) Linear models for the analysis of longitudinal studies. *The American Statistician*. 39: 95-101.
- WEDDERBURN, R.W.M. (1974) Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*. 61: 439-447.
- WILSON, P. D. (1988) Autoregressive growth curves and Kalman filtering. *Statistics in Medicine*. 7: 73-86.
- WOLFINGER, R. (1993) Covariance structure selection in general mixed models. *Communications in Statistics B, Simulation and Computation*. 22: 1079-1106.
- WU, C.F.J. (1983) On the convergence properties of the EM algorithm. *The Annals of Statistics*. 11: 95-103.

- ZEGER, S.L. & LIANG, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 42: 121-130.
- ZEGER, S.L., LIANG, K.Y. and ALBERT, P.S. (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 44: 1049-1060.
- ZELLNER, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*. 57: 348 - 368.