**Modelos de Efeitos Aleatórios para
Análise de Dados Longitudinais
Não Balanceados em Relação ao Tempo**

Solange Andreoni

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE MESTRE
EM
ESTATÍSTICA

Área de Concentração: Estatística
Orientador: **Prof. Dr. Julio da Motta Singer**

–São Paulo, Agosto de 1989–

Aos meus pais, Hilário e Luiza

Agradecimentos

Ao Prof. Dr. Julio da Motta Singer, pela sua orientação dedicada e apoio em todos os sentidos para a realização deste trabalho.

Ao querido Gilberto e aos amigos Beti, Antonio e Vilma, pela paciente ajuda na composição final.

Ao Prof. Canton, pelo constante incentivo.

Aos colegas do I.P.T. e amigos que me ajudaram.

Solange

Índice

Capítulo 1	Introdução	1
Capítulo 2	Modelos Lineares para Dados Longitudinais	15
2.1	Modelos para Dados Longitudinais	15
2.1.1	Modelos para as Médias	17
2.1.2	Modelos para as Matrizes de Covariâncias	21
2.2	Modelos de Efeitos Aleatórios	27
Capítulo 3	Estimação e Testes de Hipóteses	34
3.1	Métodos de Máxima Verossimilhança	34
3.2	Processos Iterativos para Maximização da Verossimilhança	45
3.2.1	Algoritmos de Newton-Raphson e Scoring de Fisher	46
3.2.2	Algoritmo EM	48
3.2.3	Aspectos Computacionais	60
3.3	Soluções Explícitas	65
Capítulo 4	Aplicações Práticas	68
4.1	Recursos Computacionais	68
4.2	Exemplos de Ajuste de Modelos	70
4.2.1	Exemplo 1	70
4.2.2	Exemplo 2	81
4.2.3	Exemplo 3	89
4.2.4	Comentários Finais	92

Apêndice A	Derivadas	93
A.1	Alguns Resultados sobre Matrizes	93
A.2	Derivadas do Logaritmo da Função de Verossimilhança	94
A.3	Derivadas do Logaritmo da Função de Verossimilhança Restrita	98
Apêndice B	Programas	103
B.1	Organização do Arquivo de Dados	105
B.2	Dados de Entrada	106
B.3	Dados de Saída	107
B.4	Exemplos	109
B.5	Listagens dos Programas	112
Referências Bibliográficas		136

Capítulo 1

Introdução

Dados longitudinais, caracterizados pela observação repetida de uma ou mais variáveis respostas na mesma unidade experimental, têm um papel importante em pesquisas de diversas áreas, principalmente em Medicina, Biologia e Engenharia. Nesse contexto, a observação das variáveis respostas pode ocorrer em diferentes idades, diferentes datas, ou mesmo diferentes distâncias de uma certa origem, embora neste trabalho utilizemos frequentemente o “tempo” como variável representativa das ocasiões de observação. As unidades experimentais (pessoas, animais, etc.) podem estar classificadas em diferentes grupos segundo um ou mais fatores ou tratamentos, tais como: sexo, raça, marca etc.. Quanto às respostas, elas podem envolver variáveis categorizadas (presença ou ausência de um sintoma, por exemplo) ou contínuas (como pressão arterial, volume de ar expirado, etc.); no entanto, restringiremos a discussão ao caso de uma única variável resposta com a suposição adicional de que o conjunto de observações referentes a uma determinada unidade experimental tenha distribuição de probabilidade normal multivariada.

Frequentemente os estudos envolvendo dados longitudinais têm como objetivo o ajuste de modelos simples (curvas polinomiais, por exemplo) para explicar a variação das respostas ao longo das ocasiões de observação e investigar a influência de outros fatores ou covariáveis no padrão temporal das respostas. Assim, o nosso problema é estimar os

parâmetros da curva associada a cada grupo de unidades experimentais e testar hipóteses estatísticas sobre esses parâmetros. As comparações entre os perfis médios de resposta de diferentes grupos podem ser realizadas através de comparações dos parâmetros associados às respectivas curvas.

Como essa situação de dados longitudinais envolve observações na mesma unidade experimental, ela é frequentemente relacionada (ver Cook e Ware (1983), Ware (1985), Jennrich e Schluchter (1986)) aos planejamentos de medidas repetidas (Winer (1971), Cap 4 e 7). Nesse caso, a ordem das observações é geralmente aleatorizada para cada unidade experimental, enquanto que nos planejamentos longitudinais as observações de uma mesma unidade experimental são realizadas de forma sistemática, isto é, sequencialmente ao longo do tempo. Discussões mais detalhadas sobre estudos envolvendo dados longitudinais podem ser encontradas em Cook e Ware (1983), Ware (1985) e Singer e Andrade (1986). Embora a obtenção de dados através desses planejamentos exija esforços no acompanhamento das unidades experimentais durante a realização do estudo, eles têm como principais motivações os seguintes fatos:

- (i) requerem menos unidades experimentais do que planejamentos alternativos onde uma única observação é realizada em cada unidade experimental (conhecidos como estudos do tipo “cross-sectional”);
- (ii) proporcionam condições mais adequadas para o controle de fatores secundários que possam ter influência na resposta;
- (iii) melhoram, em geral, a precisão de contrastes associados às diferenças entre os valores médios da resposta de diferentes ocasiões e
- (iv) permitem o estudo da mudança de comportamento da resposta média da unidade experimental ao longo das ocasiões de observação, ou seja, descrevem mudanças associadas a crescimento ou envelhecimento, incorporando informações sobre a variação individual na análise.

Uma característica fundamental associada a esse tipo de dados é a possibilidade de correlações não nulas entre as observações realizadas na mesma unidade experimental. Na realidade, os itens (iii) e (iv) acima são válidos se os métodos de análise levarem esse fato em consideração.

Exemplo 1.1 (Potthoff e Roy, 1964): Num estudo realizado na Faculdade de Odontologia da Universidade da Carolina do Norte (E.U.A.) foram medidas as distâncias (em mm) entre o centro da pituitária e a fissura pterio-maxilar em 11 meninas e 16 meninos aos 8, 10, 12 e 14 anos. Os resultados coletados são apresentados na Tabela 1.1. O objetivo do estudo era tentar explicar a variação dessa distância em função da idade através de uma curva simples (reta ou parábola, por exemplo), além de comparar as curvas obtidas para os dois grupos.

Tabela 1.1: Distância (em mm) entre o centro da pituitária e a fissura pterio-maxilar em 11 meninas e 16 meninos em 4 idades.

Meninas					Meninos				
Idade (anos)					Idade (anos)				
Indivíduo	8	10	12	14	Indivíduo	8	10	12	14
1	21.0	20.0	21.5	23.0	12	26.0	25.0	29.0	31.0
2	21.0	21.5	24.0	25.5	13	21.5	22.5	23.0	26.5
3	20.5	24.0	24.5	26.0	14	23.0	22.5	24.0	27.5
4	23.5	24.5	25.0	26.5	15	25.5	27.5	26.5	27.0
5	21.5	23.0	22.5	23.5	16	20.0	23.5	22.5	26.0
6	20.0	21.0	21.0	22.5	17	24.5	25.5	27.0	28.5
7	21.5	22.5	23.0	25.0	18	22.0	22.0	24.5	26.5
8	23.0	23.0	23.5	24.0	19	24.0	21.5	24.5	25.5
9	20.0	21.0	22.0	21.5	20	23.0	20.5	31.0	26.0
10	16.5	19.0	19.0	19.5	21	27.5	28.0	31.0	31.5
11	24.5	25.0	28.0	28.0	22	23.0	23.0	23.5	25.0
					23	21.5	23.5	24.0	28.0
					24	17.0	24.5	26.0	29.5
					25	22.5	25.5	25.5	26.0
					26	23.0	24.5	26.0	30.0
					27	22.0	21.5	23.5	25.0
Média	21.18	22.23	23.09	24.09		22.88	23.81	25.72	27.47
Desvio-padrão	2.12	1.90	2.36	2.44		2.45	2.14	2.65	2.09

Figura 1.1: Distância média entre o centro da pituitária e a fissura pterio-maxilar de meninas e meninos em quatro idades.

Uma análise preliminar desses dados, feita a partir do gráfico das médias das distâncias em função da idade para os dois grupos (veja Figura 1.1), sugere que talvez retas possam ser ajustadas aos dados. A matriz de correlação amostral ($\tilde{\mathbf{R}}$), calculada com os dados dos dois grupos, indica que as distâncias medidas nos mesmos indivíduos devem ser correlacionadas.

Matriz de correlação amostral:

$$\tilde{\mathbf{R}} = \begin{pmatrix} 1 & 0.57 & 0.66 & 0.52 \\ 0.57 & 1 & 0.56 & 0.73 \\ 0.66 & 0.56 & 1 & 0.73 \\ 0.52 & 0.73 & 0.73 & 1 \end{pmatrix}$$

*há correlação
 $R_c \neq 0.2 \cdot I$*

No contexto de dados longitudinais, podemos especificar quatro formas de estruturas de dados:

- (I) dados provenientes de pesquisas onde todas unidades experimentais são observadas nas mesmas T ocasiões pré-estabelecidas;
- (II) dados provenientes de pesquisas planejadas para que as respostas de todas unidades experimentais sejam observadas nas mesmas T ocasiões pré-estabelecidas, mas por alguma razão há presença de observações incompletas, isto é, quando para algumas unidades experimentais não há observação da resposta em uma ou mais ocasiões;
- (III) dados provenientes de pesquisas onde planejamos observar diferentes unidades experimentais em diferentes subconjuntos das ocasiões para as quais o estudo foi globalmente dimensionado (conhecidos como planejamentos longitudinais mistos ou “linked cross-sectional”) e
- (IV) dados provenientes de pesquisas nas quais as ocasiões de observação da resposta variam irregularmente de unidade experimental para unidade experimental.

Um conjunto de dados longitudinais é dito *não balanceado em relação ao tempo* se a estrutura dos dados corresponde à situação descrita em (II), (III) ou em (IV). O Exemplo 1.1 pode ser classificado como sendo do tipo (I), e, nesse caso, dizemos que os dados são *completos* ou *balanceados em relação ao tempo*.

A situação de dados não balanceados em relação ao tempo pode ser melhor compreendida através dos seguintes exemplos.

Exemplo 1.2 (Singer e Mauro, 1987): Os dados apresentados na Tabela 1.2 são provenientes de um estudo na área de Fisiologia Animal, e correspondem às observações da amplitude do potencial de ação composto (PAC) do nervo conectivo cérebro-visceral de 18 mexilhões em 4 condições de salinidade da água. O objetivo do estudo era avaliar o comportamento médio do PAC em função da variação da salinidade da água (Δs_j) medida

em relação a salinidade de aclimação original dos mexilhões. Embora o estudo tenha sido planejado para que os mexilhões fossem observados nas 4 salinidades, por motivos técnicos não relacionados ao estudo, 25% das observações não foram coletadas. Observando a Figura 1.2, vemos que os valores médios do PAC diminuem conforme aumenta a variação de salinidade. A matriz de correlação amostral ($\tilde{\mathbf{R}}$), calculada considerando todos os pares do PAC observados, indica que as observações podem ser correlacionadas.

Tabela 1.2: Amplitude do potencial de ação composto (PAC) de 18 mexilhões em 4 variações de salinidade.

Mexilhão	Variação de salinidade (Δs_j)			
	-1.7%	-1.0%	-0.5%	0%
1	6.0	—	—	14.0
2	8.0	—	—	15.0
3	5.1	—	—	9.6
4	1.0	11.0	12.5	15.0
5	—	5.4	7.5	11.0
6	—	8.4	10.6	12.4
7	—	6.0	13.0	17.0
8	5.8	4.4	—	9.2
9	6.8	6.8	8.0	18.0
10	3.0	5.4	6.6	6.4
11	—	—	6.0	9.4
12	5.8	6.0	8.0	12.0
13	—	6.4	9.6	10.4
14	8.8	10.4	10.8	13.0
15	—	9.6	—	11.2
16	7.6	11.2	—	12.0
17	3.2	6.0	10.4	12.0
18	—	—	9.6	14.0
Média	5.55	7.46	9.38	12.31
Desvio-padrão	2.36	2.35	2.21	2.89

Figura 1.2: PAC médio de mexilhões em quatro variações de salinidade de água.

Matriz de correlação amostral:

$$\tilde{\mathbf{R}} = \begin{pmatrix} 1 & 0.16 & -0.22 & 0.27 \\ 0.16 & 1 & 0.58 & 0.29 \\ -0.22 & 0.58 & 1 & 0.61 \\ 0.27 & 0.29 & 0.61 & 1 \end{pmatrix}$$

Exemplo 1.3: Foram encaminhados ao NTIA/EMBRAPA pelo Centro Nacional de Pesquisa de Suínos e Aves (CNPISA) dados referentes a 13 suínos, para que se estimasse a taxa média de gordura subcutânea (ETM) dos suínos em função do peso. Várias observações foram feitas no mesmo animal, e o número de observações varia de animal para animal. Como o objetivo do estudo era explicar a ETM em função do peso (considerado como definidor das ocasiões de observação) e este varia de animal para animal, o experimento é não balanceado em relação ao “tempo”. Os dados estão apresentados na Tabela 1.3.

Tabela 1.3: Taxa de gordura (ETM) de 13 suínos em função do peso.

Animal	Peso	ETM	Animal	Peso	ETM
1	92	20.00	8	101	21.33
1	68	14.33	8	85	18.67
1	101	23.00	8	61	14.33
1	86	18.00	8	95	20.33
2	64	15.33	9	100	20.33
2	89	21.67	9	82	17.33
2	98	23.00	9	60	13.67
2	78	18.33	9	91	18.67
3	63	16.00	10	90	20.00
3	100	27.33	10	102	21.67
3	80	22.67	10	80	18.67
4	63	13.67	10	62	15.33
4	82	19.00	11	90	18.33
5	92	23.00	11	100	18.67
5	66	17.33	11	80	16.67
5	98	24.33	11	64	16.33
5	83	22.67	12	93	21.67
6	81	18.33	12	109	24.00
6	61	16.33	12	86	20.00
6	105	23.67	12	62	12.33
6	91	21.33	13	93	20.33
7	102	23.00	13	104	23.67
7	80	18.67	13	86	19.00
7	62	16.33	13	62	15.00
7	93	20.67

Figura 1.3: Gráfico de dispersão da ETM de treze suínos em função do peso.

O gráfico da ETM em função do peso (Figura 1.3) mostra que talvez uma reta possa representar essa relação. Observemos que a aplicação de técnicas usuais de Regressão para o ajuste do modelo pode não ser a forma mais adequada de análise, tendo em vista uma possível correlação entre as medidas tomadas no mesmo animal. Nesse exemplo não podemos calcular a matriz de correlação amostral como nos Exemplos 1.1 e 2.2, já que as ocasiões de observação são distintas para cada animal.

Os Exemplos 1.2 e 1.3 são casos típicos das situações (II) e (IV) respectivamente. Um exemplo da situação (III), sugerido por Schluchter (1988b), corresponde a um ensaio clínico com duração de três anos cujo objetivo é estudar uma certa variável resposta em crianças com idades de 6 a 12 anos. Cada criança é medida em 4 idades consecutivas começando com a idade da criança no início do estudo. Um outro exemplo da situação (III) corresponderia a um experimento com animais de laboratório cujo objetivo é avaliar o efeito de uma determinada droga ao longo de um período de 6 semanas após sua administração; alguns animais poderiam ser observados após 1, 2 e 3 semanas, outros após 3, 4 e 5 semanas e outros após 4, 5 e 6 semanas.

Embora situações com dados não balanceados em relação ao tempo sejam frequentes em problemas práticos, a maioria das técnicas estatísticas clássicas para análise de dados longitudinais é dirigida para conjuntos de dados completos (ou balanceados em relação ao tempo) como no Exemplo 1.1. Por esse motivo, recentemente tem havido maior interesse no desenvolvimento de métodos estatísticos que possam ser utilizados em casos mais gerais. Deve-se ressaltar, também, que esses métodos não são aplicáveis a todos os mecanismos geradores de dados não balanceados. Eles são mais apropriados para as situações (III) ou (IV), onde a natureza do desbalanceamento das observações é imposta pelo esquema amostral. Para que sejam aplicáveis na situação (II), a presença de observações incompletas deve ser de natureza completamente aleatória. Isso ocorre, por exemplo, quando um paciente falta a um exame médico onde seria avaliado por motivos não relacionados ao tratamento que está recebendo, de maneira que a probabilidade de a resposta estar faltando não dependa de seu valor e nem dos outros valores observados no conjunto de dados. Dados com essas características são ditos “incompletos por razões totalmente aleatórias” (MCAR, “missing completely at random”). Essa suposição às vezes pode ser relaxada, permitindo que a probabilidade de a resposta estar incompleta possa depender dos valores observados, mas não dos valores não observados — ditos “incompletos por razões aleatórias” (MAR, “missing at random”) — se o padrão dos dados incompletos for *ig-*

norável, isto é, se os parâmetros que desejamos estimar não dependerem dos parâmetros associados aos mecanismos geradores de dados incompletos. Os leitores interessados em definições formais sobre os conceitos de MCAR, MAR e *ignorabilidade* devem consultar Rubin (1976b), Laird (1988) e Little (1988).

A análise de dados longitudinais envolve a representação da estrutura geradora dos dados através de modelos estruturais simples (com um número pequeno de parâmetros), que reflitam os aspectos fundamentais da variação da resposta. Essencialmente esses modelos envolvem:

- (1) os valores médios da variável resposta em cada ocasião para cada tratamento;
- (2) as variâncias e covariâncias entre as respostas observadas na mesma unidade experimental.

Embora existam diversos tipos de propostas para representar a função que descreve o comportamento médio da variável resposta, serão considerados apenas modelos lineares neste trabalho. Em geral, os parâmetros desses modelos são de fácil interpretação e também fornecem, na maioria dos casos, um bom ajuste para os dados observados. Apesar de o maior interesse estar geralmente concentrado nessa função, é necessário levar em consideração a estrutura de covariância das observações. É essa estrutura que contém a informação essencial sobre a correlação entre as observações realizadas na mesma unidade experimental. Ela tem influência direta nas estimativas da variabilidade dos parâmetros associados às médias e em alguns casos pode inclusive afetar as próprias estimativas desses parâmetros.

As técnicas clássicas de análise de dados longitudinais são geralmente dirigidas para o caso de dados balanceados em relação ao tempo. Entre elas, destacamos:

- Análise Multivariada de Perfis e Análise Univariada de Perfis através de modelos mistos (ver Singer e Andrade (1986), Cap 2, por exemplo);

- Análise de Curvas de Crescimento que também pode ser realizada de modelos uni ou multivariados (ver Singer e Andrade (1986), Cap 3, por exemplo).

O objetivo principal da técnica de Análise de Perfis é testar hipóteses sobre os valores médios da variável resposta nas ocasiões de observação e compará-los entre os diferentes grupos de unidades experimentais. O perfil de cada grupo é dado pelo vetor de valores médios em cada ocasião de observação. Essencialmente esse tipo de análise visa responder a três perguntas:

1. Os perfis médios dos diferentes tratamentos são paralelos? (i.e., não existe interação entre tratamento e tempo?)
2. Se os perfis são paralelos, eles são coincidentes? (i.e., não existe efeito de tratamento?)
3. Se os perfis são paralelos, eles são horizontais? (i.e., não existe efeito de “tempo”?)

A Análise Multivariada de Perfis (ver, por exemplo, Morrison (1976, Cap 4 e 5) e Timm (1980)) admite que as variâncias da resposta em cada ocasião de observação e as covariâncias das respostas entre ocasiões distintas possam ser diferentes. A Análise de Variância através de Modelos Mistos (ver Winer (1971), Cap 4 e 7, por exemplo) geralmente requer que a variância da resposta seja constante nas ocasiões de observação, assim como as covariâncias entre as respostas de diferentes ocasiões sejam iguais. A adequabilidade desta estrutura de covariância pode ser verificada através do teste da razão de verossimilhança generalizada como indicado, por exemplo, em Rao (1965) e Winer (1971, Cap 7). Mesmo quando essa estrutura de covariância não é razoável, podemos considerar testes aproximados através da correção dos graus de liberdade dos testes F usuais (ver Greenhouse e Geisser (1959), Singer e Andrade (1986)). Esses métodos também podem ser utilizados para estimar diferenças, tendências, e outras combinações lineares das médias nas ocasiões de observação (Bock (1975)).

Para o método de Curvas de Crescimento, o objetivo principal é descrever o comportamento médio da variável resposta nas diversas ocasiões de observação através de uma curva polinomial. Nesse contexto, Potthoff e Roy (1964), Khatri (1966) e Grizzle e Allen (1969) desenvolveram técnicas de estimação e testes de hipóteses sobre os parâmetros das curvas envolvidas considerando que os dados são balanceados em relação ao tempo e sem impor uma estrutura na matriz de covariância das respostas. Rao (1965, 1975) considera a possibilidade de uma matriz de covariância mais restritiva, induzida através da especificação de modelos de efeitos aleatórios, isto é, assumindo que cada unidade experimental tenha sua própria curva com desvios aleatórios de uma curva populacional. Os modelos de Curvas de Crescimento para dados balanceados em relação ao tempo também estão descritos em Singer (1977).

Quando os dados não são balanceados em relação ao tempo, as técnicas de análise clássicas citadas podem não ser eficientes. Se o desbalanceamento for causado pela presença de poucas observações incompletas, uma solução seria realizar a análise omitindo essas unidades experimentais ou substituindo os valores incompletos por estimativas. Algumas técnicas para lidar com esse tipo de dados longitudinais são revistas por Berk (1987). Entre elas temos a utilização de mínimos quadrados não balanceados (Bartlett (1937), Rubin (1976a), Schwertman (1978)) e mínimos quadrados generalizados (Kleinbaum (1973a,b)). Contudo, essas soluções não podem ser empregadas quando há muitas observações incompletas.

Uma abordagem sistemática para análise de dados não balanceados em relação ao tempo envolve a especificação de um modelo para as médias e de outro para a estrutura de covariância e a estimação dos parâmetros através dos métodos de máxima verossimilhança ou máxima verossimilhança restrita. Nesse sentido, a maioria dos trabalhos era anteriormente dirigida para extensões da Análise Multivariada de Perfis (Hartley e Hocking (1971), Orchard e Woodbury (1972), Beale e Little (1975)) e Análise de Perfis através de modelos mistos (Hartley e Rao (1967), Hemmerle e Hartley (1973), Jennrich e Sampson (1976)).

Neste trabalho nos concentraremos em propostas recentes que permitem a definição de modelos lineares para a função que descreve o comportamento médio da resposta e de outras formas para a estrutura de covariância, generalizando as técnicas de Análise de Perfis e Curvas de Crescimento para dados não balanceados em relação ao tempo. Assim, na Seção 2.1 do Capítulo 2 apresentamos vários modelos para a estrutura de médias e de covariâncias que podem ser utilizados nessas situações. Dentre os modelos apresentados, nos deteremos na Seção 2.2 nos modelos de efeitos aleatórios (Laird e Ware (1982); Laird, Lange e Stram (1987); Fairclough e Helms (1986)), pois permitem tratar situações em que os dados não são balanceados em relação ao tempo e têm algumas interpretações interessantes. No Capítulo 3 apresentamos a estimação dos parâmetros do modelo de efeitos aleatórios através do método de máxima verossimilhança e máxima verossimilhança restrita. Como será visto, frequentemente necessitaremos de processos iterativos para encontrar as soluções que maximizam a verossimilhança.

O Capítulo 4 é dedicado a aplicações práticas utilizando os modelos de efeitos aleatórios para análise de dados longitudinais. Como, até o momento, não temos programas disponíveis nos pacotes estatísticos para obter as estimativas dos parâmetros desses modelos, apresentamos alguns programas com essa finalidade. Esses programas constam do Apêndice B e utilizam o módulo CM do pacote estatístico SOC (EMBRAPA) que pode ser implantado em microcomputadores de 16 bits ou em qualquer máquina para a qual um compilador C esteja disponível.

Capítulo 2

Modelos Lineares para Dados Longitudinais

Neste capítulo consideramos modelos lineares para dados longitudinais balanceados ou não em relação ao tempo. Na Seção 2.1 apresentamos diversos modelos para as estruturas de médias e de covariâncias sugeridos por Ware (1985), Jennrich e Schluchter (1986) e Schluchter (1988a), discutindo em quais situações eles são aplicáveis. Dentre eles, os modelos de efeitos aleatórios, estudados mais detalhadamente por Laird e Ware (1982), Fairclough e Helms (1986) e Laird, Lange e Stram (1987), são os que se mostram mais flexíveis para dados não balanceados em relação ao tempo e têm interpretações interessantes; por esse motivo eles são discutidos em detalhes na Seção 2.2.

2.1 Modelos para Dados Longitudinais

Suponhamos um conjunto de n unidades experimentais onde em cada unidade i , $i = 1, \dots, n$, é observada uma variável resposta ao longo de t_i ocasiões conhecidas. Seja $\mathbf{y}_i = (y_{i1}, \dots, y_{it_i})'$ o vetor $(t_i \times 1)$ de respostas associado à i -ésima unidade experimental. Um modelo linear que associa a resposta da i -ésima unidade experimental a um conjunto de

variáveis explicativas é dado por:

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

onde $\mathbf{X}_i$ ($t_i \times p$) é a matriz (conhecida) de especificação do modelo, $\boldsymbol{\alpha}$ ($p \times 1$) é o vetor de parâmetros populacionais desconhecidos e $\mathbf{e}_i$ ($t_i \times 1$) é um vetor de erros aleatórios. Além disso, supomos que $\mathbf{e}_1, \dots, \mathbf{e}_n$ são independentes com $\mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$, onde cada $\boldsymbol{\Sigma}_i$ é uma matriz simétrica positiva definida. Consequentemente

$$\mathbf{y}_i \sim N(\mathbf{X}_i \boldsymbol{\alpha}, \boldsymbol{\Sigma}_i) \quad \text{para } i = 1, \dots, n, \quad (2.2)$$

e $\mathbf{y}_1, \dots, \mathbf{y}_n$ são independentes ($\text{Cov}(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{0}$ se $i \neq j$).

Assumimos, também, que a matriz $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)'$ ($\sum_{i=1}^n t_i \times p$) tem posto $p \leq \sum_{i=1}^n t_i$ e que os elementos de $\boldsymbol{\Sigma}_i$, $i = 1, \dots, n$, são funções conhecidas de um vetor $\boldsymbol{\theta}$ ($k \times 1$) de parâmetros desconhecidos não redundantes, cujos elementos *não dependem* de $\boldsymbol{\alpha}$, isto é, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ para $i = 1, \dots, n$.

O modelo (2.1) pode também ser escrito na forma vetorial utilizada por Harville (1974, 1976, 1977):

$$\mathbf{y} = \mathbf{X} \boldsymbol{\alpha} + \mathbf{e}, \quad (2.3)$$

com $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$ ($\sum_{i=1}^n t_i \times 1$), $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)'$ ($\sum_{i=1}^n t_i \times p$) e $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_n)'$ ($\sum_{i=1}^n t_i \times 1$). Dessa forma,

$$E(\mathbf{y}) = \mathbf{X} \boldsymbol{\alpha} \quad \text{e} \quad \text{Var}(\mathbf{y}) = \mathbf{V}(\boldsymbol{\theta}) = \text{diag}(\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \dots, \boldsymbol{\Sigma}_n(\boldsymbol{\theta})),$$

onde $\text{diag}(\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \dots, \boldsymbol{\Sigma}_n(\boldsymbol{\theta}))$ representa uma matriz diagonal em blocos.

Essa maneira de escrever o modelo para cada unidade experimental é bastante flexível, pois não requer que as unidades experimentais sejam observadas nas mesmas ocasiões ou tenham o mesmo número de observações.

Sejam y_{it} a resposta da unidade experimental i numa ocasião t e $\mathbf{x}'_{it}$ a linha da matriz $\mathbf{X}_i$ associada a essa resposta. A partir da distribuição de $\mathbf{y}_i$ dada por (2.2) o

modelo condicional de y_{it} dadas as observações anteriores $y_{i1}, \dots, y_{it-1}$ é

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\alpha} + \sum_{s=1}^{t-1} \gamma_{is}(y_{is} - \mathbf{x}'_{is}\boldsymbol{\alpha}) + w_{it},$$

onde $w_{it} \sim N(0, \sigma_{it}^2)$ e $\gamma_{i1}, \dots, \gamma_{it-1}$ e σ_{it}^2 são determinados pela estrutura de $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$. No caso em que $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \sigma^2 \mathbf{I}_{t_i}$ temos $\gamma_{is} = 0$ e $\sigma_{it}^2 = \sigma^2$ que é o modelo de Regressão usual. A natureza longitudinal dos dados associada às possíveis correlações entre as medidas realizadas na mesma unidade experimental pode ser incorporada no modelo através da consideração de outras estruturas para $\text{Var}(\mathbf{e}_i) = \text{Var}(\mathbf{y}_i) = \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$.

Como os parâmetros $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$ variam independentemente, o modelo (2.1) pode ser encarado como sendo composto por dois componentes separados: um para os valores médios envolvendo os parâmetros em $\boldsymbol{\alpha}$ e outro para a estrutura de covariância envolvendo $\boldsymbol{\theta}$. A seguir discutiremos algumas alternativas para esses dois componentes.

2.1.1 Modelos para as Médias

O modelo (2.1) permite, por exemplo, a expressão da resposta média num tempo t em termos de variáveis explicativas medidas no mesmo tempo t ou em tempos anteriores. A forma de especificação das matrizes $\mathbf{X}_i$'s (ou $\mathbf{X}$) é similar àquela utilizada em modelos de Regressão. Dessa maneira, as colunas de $\mathbf{X}_i$ podem estar associadas:

- (i) aos fatores que definem a estrutura das subpopulações;
- (ii) ao fator Tempo, identificando, por exemplo, a forma da curva a ser ajustada e
- (iii) a possíveis covariáveis, cujo efeito na resposta média desejamos investigar.

Como ilustração consideremos os exemplos apresentados no Capítulo 1.

No Exemplo 1.1, expressando os vetores de respostas observadas na forma do modelo (2.1), temos

$$y_1 = \begin{pmatrix} 21.0 \\ 20.0 \\ 21.5 \\ 23.0 \end{pmatrix}, y_2 = \begin{pmatrix} 21.0 \\ 21.5 \\ 24.0 \\ 25.5 \end{pmatrix}, \dots, y_{27} = \begin{pmatrix} 22.0 \\ 21.5 \\ 23.5 \\ 25.0 \end{pmatrix}.$$

Suponhamos primeiramente que queremos estudar a resposta média em cada idade (8, 10, 12 e 14 anos) para cada grupo (meninas e meninos). Isso pode ser realizado, por exemplo, através da parametrização de médias de caselas, onde as matrizes X_i 's são da forma:

$$X_i = a'_i \otimes I_4,$$

$$I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

com $a'_i = (1, 0)$ se o i -ésimo indivíduo pertence ao grupo das meninas ou $a'_i = (0, 1)$ se o i -ésimo indivíduo pertence ao grupo dos meninos; I_4 é a matriz identidade de ordem 4 e $\otimes$ denota o produto direto (ou de Kronecker) definido de tal forma que se A é uma matriz $m \times n$ com elementos a_{ij} e B é uma matriz $q \times p$, então, $A \otimes B$ ($mq \times np$) é uma matriz formada por mn blocos com o bloco de ordem (i, j) igual a $a_{ij}B$. Assim, temos:

$$X_i = \begin{pmatrix} I_4 & 0_{4 \times 4} \end{pmatrix} \quad \text{para } i = 1, \dots, 11 \text{ (meninas)}$$

e

$$X_i = \begin{pmatrix} 0_{4 \times 4} & I_4 \end{pmatrix} \quad \text{para } i = 12, \dots, 27 \text{ (meninos),}$$

com $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_8)'$, onde $\alpha_1, \alpha_2, \alpha_3$ e α_4 representam respectivamente as respostas médias aos 8, 10, 12 e 14 anos para o grupo das meninas e $\alpha_5, \alpha_6, \alpha_7$ e α_8 para o grupo dos meninos.

Se desejamos ajustar retas diferentes para cada grupo a fim de descrevermos o comportamento da resposta média em função da idade, as matrizes X_i 's podem ser da forma:

$$X_i = \begin{pmatrix} 1 & 8 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 1 & 12 & 0 & 0 \\ 1 & 14 & 0 & 0 \end{pmatrix} \quad \text{para } i = 1, \dots, 11 \text{ (meninas)}$$

e

$$\mathbf{X}_i = \begin{pmatrix} 0 & 0 & 1 & 8 \\ 0 & 0 & 1 & 10 \\ 0 & 0 & 1 & 12 \\ 0 & 0 & 1 & 14 \end{pmatrix} \quad \text{para } i = 12, \dots, 27 \text{ (meninos).}$$

Dessa maneira, $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$, onde α_1 e α_2 são respectivamente os coeficientes linear e angular do grupo das meninas, e α_3 e α_4 os coeficientes linear e angular do grupo dos meninos.

O modelo (2.1) permite ainda que as curvas associadas aos diferentes grupos sejam diferentes. Por exemplo, se queremos ajustar uma reta para o grupo das meninas e uma parábola para o grupo dos meninos, temos:

$$\mathbf{X}_i = \begin{pmatrix} 1 & 8 & 0 & 0 & 0 \\ 1 & 10 & 0 & 0 & 0 \\ 1 & 12 & 0 & 0 & 0 \\ 1 & 14 & 0 & 0 & 0 \end{pmatrix} \quad \text{para } i = 1, \dots, 11 \text{ (meninas)}$$

e

$$\mathbf{X}_i = \begin{pmatrix} 0 & 0 & 1 & 8 & 64 \\ 0 & 0 & 1 & 10 & 100 \\ 0 & 0 & 1 & 12 & 144 \\ 0 & 0 & 1 & 14 & 196 \end{pmatrix} \quad \text{para } i = 12, \dots, 27 \text{ (meninos),}$$

com $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)'$, onde $\alpha_1, \alpha_2, \alpha_3$ e α_4 têm a mesma interpretação que no modelo anterior e α_5 é coeficiente do termo de 2º grau do grupo dos meninos.

No Exemplo 1.2, se expressarmos o vetor de respostas na forma do modelo (2.1) e quisermos utilizar um modelo de médias de caselas, teremos:

$$\mathbf{y}_1 = \begin{pmatrix} 6.0 \\ 14.0 \end{pmatrix}, \dots, \mathbf{y}_9 = \begin{pmatrix} 6.8 \\ 6.8 \\ 8.0 \\ 18.0 \end{pmatrix}, \dots, \mathbf{y}_{18} = \begin{pmatrix} 9.6 \\ 14.0 \end{pmatrix},$$

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \dots, \mathbf{X}_9 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \dots, \mathbf{X}_{18} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

e $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$, onde α_j é o valor esperado do PAC na j -ésima variação de salinidade ($j = 1, \dots, 4$).

Uma outra possibilidade para modelagem das médias do PAC nesse exemplo é através de uma parametrização de casela de referência. Por exemplo, tomando como referência a salinidade de aclimação (0%), podemos fazer:

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \dots, \mathbf{X}_9 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \dots, \mathbf{X}_{18} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

com $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$, onde α_1 , α_2 e α_3 são as variações da resposta média em relação à resposta média na salinidade de aclimação (α_4) correspondentes às variações de salinidade -1.7%, -1.0% e -0.5%.

Se queremos descrever a relação entre o PAC médio e a variação de salinidade através de uma reta, podemos ter:

$$\mathbf{X}_1 = \begin{pmatrix} 1 & -1.7 \\ 1 & 0.0 \end{pmatrix}, \dots, \mathbf{X}_9 = \begin{pmatrix} 1 & -1.7 \\ 1 & -1.0 \\ 1 & -0.5 \\ 1 & 0.0 \end{pmatrix}, \dots, \mathbf{X}_{18} = \begin{pmatrix} 1 & -0.5 \\ 1 & 0.0 \end{pmatrix},$$

com $\alpha = (\alpha_1, \alpha_2)'$, onde α_1 e α_2 são respectivamente o coeficiente linear e angular.

Finalmente, no Exemplo 1.3, supondo que uma reta seja razoável para descrever o comportamento médio da ETM em função do peso dos suínos, temos que:

$$\mathbf{y}_1 = \begin{pmatrix} 20.00 \\ 14.33 \\ 23.00 \\ 18.00 \end{pmatrix}, \dots, \mathbf{y}_4 = \begin{pmatrix} 13.67 \\ 19.00 \end{pmatrix}, \dots, \mathbf{y}_{13} = \begin{pmatrix} 20.33 \\ 23.67 \\ 19.00 \\ 15.00 \end{pmatrix},$$

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 92 \\ 1 & 68 \\ 1 & 101 \\ 1 & 86 \end{pmatrix}, \dots, \mathbf{X}_4 = \begin{pmatrix} 1 & 63 \\ 1 & 82 \end{pmatrix}, \dots, \mathbf{X}_{13} = \begin{pmatrix} 1 & 93 \\ 1 & 104 \\ 1 & 86 \\ 1 & 62 \end{pmatrix},$$

com $\alpha = (\alpha_1, \alpha_2)'$, onde α_1 é o coeficiente linear e α_2 o coeficiente angular populacional.

Notemos que nesse exemplo não faz sentido parametrizar os valores médios da ETM com modelos de médias de caselas. Contudo, um modelo de médias possível seria

$$\mathbf{X}_i \alpha = \mathbf{1}_{t_i} \alpha,$$

onde $\mathbf{1}_{t_i}$ ($t_i \times 1$) é um vetor com todos elementos iguais a 1 e α seria a média geral da ETM.

2.1.2 Modelos para as Matrizes de Covariâncias

Uma consideração importante para a escolha de uma estrutura de covariância se refere a maneira pela qual os dados foram coletados. Essencialmente podemos distinguir dois casos:

- (i) situações nas quais planejamos fazer observações nas mesmas ocasiões para todas unidades experimentais (essencialmente aquelas correspondentes às estruturas (I) e (II) descritas na Introdução).
- (ii) situações em que as ocasiões de observação variam irregularmente de unidade experimental para unidade experimental (basicamente aquelas correspondentes às estruturas (III) e (IV) descritas na Introdução; também podem ser classificadas aqui situações correspondentes à estrutura (II), quando o número de observações incompletas é muito grande).

No primeiro caso, o vetor y_i ($t_i \times 1$) de respostas de uma unidade experimental i é considerado um subvetor do conjunto de respostas y_i^* ($T \times 1$) que seriam potencialmente medidas, onde T é o número de ocasiões distintas em que planejamos avaliar a resposta. Dessa maneira, podemos supor que o modelo para y_i^* é dado por:

$$y_i^* = X_i^* \alpha + e_i^* \quad \text{para } i = 1, \dots, n,$$

onde X_i^* ($T \times p$) é a matriz de especificação do modelo para os dados “completos” e e_i^* ($T \times 1$) o vetor de erros aleatórios. Nessa situação, frequentemente supomos que $\text{Var}(e_i^*) = \text{Var}(y_i^*) = \Sigma(\theta) = \Sigma$ para $i = 1, \dots, n$, assim, cada $\Sigma_i = \text{Var}(y_i)$ é obtida de Σ ($T \times T$) selecionando-se linhas e colunas associadas às t_i ocasiões de observação efetiva da i -ésima unidade experimental. Isso pode ser realizado fazendo $\Sigma_i = E_i \Sigma E_i'$, onde E_i ($t_i \times T$) é uma matriz cujos elementos são zeros e uns convenientemente colocados para escolher as variâncias e covariâncias apropriadas de Σ . É claro que se os dados são balanceados em relação ao tempo, como no Exemplo 1.1, isso não é necessário, pois

$y_i = y_i^*$ e $\Sigma = \Sigma_i$ ($E_i = I_T$). Por outro lado, para o Exemplo 1.2, que é um caso de dados incompletos com Σ (4×4), a matriz de covariância associada ao mexilhão 15 (observado somente na segunda e na terceira variação de salinidade) é

$$\Sigma_{15} = E_{15} \Sigma E_{15}', \text{ onde } E_{15} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

e a do mexilhão 18 (observado na terceira e quarta variação de salinidade) é

$$\Sigma_{18} = E_{18} \Sigma E_{18}', \text{ onde } E_{18} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Nesse contexto de *dados incompletos* muitos autores, como Jennrich e Schluchter (1986) e Schluchter (1988a) por exemplo, têm sugerido várias estruturas para Σ , que são discutidas a seguir.

A estrutura de covariância mais geral é dada por:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1T} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2T} \\ \vdots & \vdots & & \vdots \\ \sigma_{1T} & \sigma_{2T} & \dots & \sigma_T^2 \end{pmatrix} \quad (2.4)$$

Os modelos que assumem a estrutura de covariância (2.4) são usualmente denominados “modelos multivariados gerais” e Σ é dita “não estruturada”. Nesse caso não é feita nenhuma restrição sobre as variâncias e as covariâncias entre os elementos de y_i . Assim, essa estrutura é aplicável quando:

- (i) não tivermos bases empíricas ou teóricas para estabelecer algum padrão para a matriz de covariância e
- (ii) não houver necessidade de extrapolar o modelo além das ocasiões de observação consideradas.

Como o número de ocasiões distintas de avaliação das respostas é T , a matriz Σ tem $T(T+1)/2$ parâmetros contidos em $\theta = (\sigma_1^2, \sigma_{12}, \dots, \sigma_{1T}, \sigma_2^2, \dots, \sigma_T^2)'$. Quando o número n de unidades experimentais é pequeno em relação a T , a estimação dos $T(T+1)/2$ parâmetros

da matriz de covariância (2.4) pode levar a estimativas ineficientes dos parâmetros α associados à função que descreve o comportamento das respostas médias; em outros casos pode ser computacionalmente impossível estimar os parâmetros dessa matriz de covariância. Suponhamos, por exemplo, um conjunto de dados longitudinais incompletos da forma:

Unidade Experimental	Tempo			
	1	2	3	4
1	y_{11}	y_{12}	—	—
2	—	—	y_{23}	y_{24}
3	y_{31}	y_{32}	—	—
4	—	—	y_{43}	y_{44}

onde “—” denota a falta de uma observação. Se Σ for dada por (2.4), os parâmetros σ_{13} , σ_{14} , σ_{23} e σ_{24} não poderão ser estimados. Portanto quando n é pequeno em relação a T , ou há presença de muitas observações incompletas, estruturas mais restritivas para a matriz de covariância devem ser consideradas.

Outras formas para a estrutura de covariância com um número menor de parâmetros podem ser obtidas quando cada vetor de erros e_i^* é considerado como uma série temporal obedecendo a processos auto-regressivos estacionários ou de médias móveis (Box e Jenkins (1976)). Eles são especialmente úteis quando há poucas unidades experimentais (n pequeno) e um número grande de observações por unidade (T ou t_i 's grandes), mesmo com a presença de observações incompletas. Dessa maneira, supondo que os e_i^* 's sigam um processo auto-regressivo de primeira ordem (AR(1)), o padrão de correlação serial pode ser representado por:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 \end{pmatrix} \quad (2.5)$$

com $|\rho| < 1$ e $\sigma^2 > 0$. Nesse caso, com apenas dois parâmetros ($\theta = (\sigma^2, \rho)'$), conseguimos modelar uma situação comum na prática, onde a correlação entre as observações realizadas na mesma unidade experimental decresce à medida que a distância entre elas aumenta.

Uma outra classe de modelos associados aos de séries temporais é obtida quando os e_i^* 's seguem um processo de médias móveis de primeira ordem (MA(1)), cuja estrutura de

covariância é dada por:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & 0 & \dots & 0 \\ \rho & 1 & \rho & \dots & 0 \\ 0 & \rho & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \quad (2.6)$$

com $|\rho| < 1$ e $\sigma^2 > 0$. Com a estrutura (2.6) supomos que somente as respostas de ocasiões adjacentes são correlacionadas. Nesse caso $\theta = (\sigma^2, \rho)'$.

Todos os modelos que assumem que os e_i^* 's são gerados por processos estacionários de séries temporais são casos especiais da estrutura de covariância denominada "modelo auto-regressivo geral" que é caracterizada por:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_2 & \sigma_3 & \dots & \sigma_T \\ \sigma_2 & \sigma_1^2 & \sigma_2 & \dots & \sigma_{T-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_T & \sigma_{T-1} & \sigma_{T-2} & \dots & \sigma_1^2 \end{pmatrix} \quad (2.7)$$

Esse modelo supõe que as variâncias das respostas são iguais em todas ocasiões de observação e que as covariâncias entre as respostas de diferentes ocasiões separadas pelo mesmo intervalo de tempo também são iguais. Essa estrutura tem T parâmetros contidos em $\theta = (\sigma_1^2, \sigma_2, \sigma_3, \dots, \sigma_T)'$.

Os leitores interessados na análise de dados longitudinais através desses tipos de modelos podem consultar, por exemplo, Lavange (1984), Louis e Spiro (1986), Wilson (1988) e Rochon e Helms (1989) para maiores detalhes. Vale lembrar que os métodos de estimação usuais só podem ser aplicados quando as observações são igualmente espaçadas.

Uma outra classe importante de estruturas de covariâncias é aquela induzida através da especificação de modelos de efeitos aleatórios. Nessa situação supomos que o modelo para os dados "completos" (y_i^* 's) possa ser escrito como:

$$y_i^* = X_i^* \alpha + Z b_i + u_i^* \quad \text{para } i = 1, \dots, n, \quad (2.8)$$

onde Z ($T \times q$) é uma matriz conhecida, b_i ($q \times 1$) e u_i^* ($T \times 1$) são vetores aleatórios independentes com $b_i \sim N(0, D)$ e $u_i^* \sim N(0, \sigma^2 I_T)$. Assim, Σ é dada por

$$\Sigma = Z D Z' + \sigma^2 I_T. \quad (2.9)$$

Esse modelo implica o seguinte modelo para y_i :

$$y_i = X_i \alpha + Z_i b_i + u_i \quad \text{para } i = 1, \dots, n, \quad (2.10)$$

onde X_i , Z_i e u_i contêm as linhas de X_i^* , Z e u_i^* respectivamente, que correspondem aos elementos de y_i^* efetivamente observados.

Um caso especial do modelo de efeitos aleatórios é obtido de (2.8) e (2.9) fazendo $Z = 1_T$ e $D = \sigma_b^2$. Assim, $\Sigma = \sigma^2 I_T + \sigma_b^2 11'$, que é um caso particular da estrutura de covariância uniforme:

$$\Sigma = \sigma_*^2 \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix} \quad (2.11)$$

com $\sigma_*^2 = \sigma^2 + \sigma_b^2$ e $\rho = \sigma_b^2 / (\sigma^2 + \sigma_b^2)$. A estrutura (2.11) supõe que as variâncias e as correlações da variável resposta são constantes nas ocasiões de observação.

Outras classes de estruturas de covariância podem ser consideradas. Por exemplo, notemos que com exceção do modelo auto-regressivo de primeira ordem, todas as outras estruturas de covariância apresentadas podem ser escritas na forma linear:

$$\Sigma(\theta) = \sum_{g=1}^k \theta_g G_g, \quad (2.12)$$

onde as matrizes G_g 's são conhecidas e os θ_g 's são os parâmetros desconhecidos da matriz de covariância Σ . Além de Jennrich e Schluchter (1986), Szatrowski (1983) também apresenta técnicas de estimação para esses modelos.

Quando as ocasiões de observação da resposta variam irregularmente de unidade experimental para unidade experimental (como no Exemplo 1.3) é necessário especificar diretamente a estrutura das matrizes $\Sigma_i(\theta)$'s. Nesse sentido, Louis (1988) e Wilson (1988) generalizam o modelo auto-regressivo de primeira ordem para essa situação supondo que a covariância entre as observações sucessivas em uma mesma unidade experimental i é $\sigma^2 \rho^{|t_{ij} - t_{ik}|}$, onde ρ agora é restrito a $0 \leq \rho < 1$ e $|t_{ij} - t_{ik}|$ é a diferença entre os tempos de observação j e k da resposta da unidade experimental i .

Ainda dentro desse contexto, um modelo de efeitos aleatórios mais geral que (2.8) é obtido quando especificamos o modelo (2.10) diretamente para y_i . Assim, Z_i pode depender de covariáveis ou dos tempos de observação associados unicamente a i -ésima unidade experimental. Essa situação é ainda um caso particular do modelo (2.1) fazendo $e_i = Z_i b_i + u_i$. Assim, a estrutura de covariância da i -ésima unidade experimental é dada por:

$$\Sigma_i = Z_i D Z_i' + \sigma^2 I_{t_i} \quad (2.13)$$

Essa estrutura de covariância permite que diferentes unidades experimentais tenham matrizes de covariância diferentes, porém dependendo de no máximo $q(q+1)/2 + 1$ parâmetros desconhecidos. Esse modelo é estudado por Laird e Ware (1982), Fairclough e Helms (1986), Laird, Lange e Stram (1987) e Lindstrom e Bates (1988). Com esta flexibilidade para a modelagem de dados não balanceados em relação ao tempo, os modelos de efeitos aleatórios generalizam os modelos de Curvas de Crescimento e de Análise de Perfis utilizando modelos mistos em dados longitudinais, o que será visto na próxima seção.

Andrade e Helms (1984b) apresentam técnicas de análise e estimação em modelos cuja estrutura de covariância de cada unidade experimental i é da forma linear:

$$\Sigma_i(\theta) = \sum_{g=1}^k \theta_g G_{ig}, \quad (2.14)$$

onde as matrizes G_{ig} 's são conhecidas e os θ_g 's são os parâmetros desconhecidos das matrizes de covariância Σ_i 's.

Finalmente, vale lembrar que a estrutura mais simples para Σ ou Σ_i ocorre quando as observações na mesma unidade experimental são consideradas independentes e com variância constante, ou seja:

$$\Sigma_i = \sigma^2 I_{t_i}. \quad (2.15)$$

Os modelos que se enquadram nessa estrutura podem ser analisados por métodos amplamente divulgados na literatura como Análise de Variância ou Análise de Regressão.

2.2 Modelos de Efeitos Aleatórios

Os modelos de efeitos aleatórios oferecem boas alternativas para análise de dados longitudinais, pois além de permitirem a modelagem de matrizes de covariância com um número menor de parâmetros que a de modelos multivariados gerais (matriz de covariância não estruturada), sua formulação pode ser naturalmente estendida para dados não balanceados em relação ao tempo.

Com base no trabalho de Harville (1977), Laird e Ware (1982) propuseram o seguinte modelo de efeitos aleatórios:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i \quad \text{para } i = 1, \dots, n, \quad (2.16)$$

onde $\mathbf{y}_i$ ($t_i \times 1$) é o vetor de respostas nas t_i ocasiões, $\mathbf{X}_i$ ($t_i \times p$) é a matriz (conhecida) de especificação do modelo populacional, $\boldsymbol{\alpha}$ ($p \times 1$) é o vetor de parâmetros populacionais desconhecidos, $\mathbf{b}_i$ ($q \times 1$) é o vetor de efeitos aleatórios associados a i -ésima unidade experimental, $\mathbf{Z}_i$ ($t_i \times q$) é a matriz de especificação conhecida ligando $\mathbf{b}_i$ a $\mathbf{y}_i$ e $\mathbf{e}_i$ ($t_i \times 1$) é um vetor de erros aleatórios. Além disso, supomos que $\mathbf{e}_1, \dots, \mathbf{e}_n, \mathbf{b}_1, \dots, \mathbf{b}_n$ são independentes com $\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_{t_i})$ e $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$ para $i = 1, \dots, n$, onde $\mathbf{D}$ ($q \times q$) e $\mathbf{R}_{t_i}$ ($t_i \times t_i$) são matrizes de covariância; $\mathbf{R}_{t_i}$ depende de i somente através de sua dimensão t_i (em geral $\mathbf{R}_{t_i} = \sigma^2 \mathbf{I}_{t_i}$).

Uma forma alternativa interessante do ponto de vista de interpretação é aquela considerada por alguns autores (Laird e Ware (1982), por exemplo) que apresentam esse modelo através de um procedimento em *dois estágios*, onde a distribuição de probabilidade das respostas de cada unidade experimental tem a mesma forma, mas com parâmetros que variam de unidade experimental para unidade experimental. A distribuição desses “parâmetros” ou “efeitos aleatórios” na população constituem o segundo estágio do modelo. Assim, podemos escrever (2.16) na forma:

Estágio 1: Para cada unidade experimental $i, i = 1, \dots, n$, consideramos o seguinte modelo condicional:

$$y_i | b_i = X_i \alpha + Z_i b_i + e_i.$$

Nesse estágio, α e b_i são considerados fixos. Os e_i 's são independentes e cada $e_i \sim N(0, R_{t_i})$.

Estágio 2: Supomos que os b_i 's são independentes com $b_i \sim N(0, D)$ e que são independentes dos e_i 's.

Consequentemente, temos que a distribuição marginal (não condicional) de cada vetor de respostas y_i é dada por:

$$y_i \sim N(X_i \alpha, \Sigma_i), \quad (2.17)$$

com $\Sigma_i = R_{t_i} + Z_i D Z_i'$ e $\text{Cov}(y_i, y_j) = 0$ se $i \neq j$. É comum assumir que $R_{t_i} = \sigma^2 I_{t_i}$, embora outras alternativas possam ser consideradas. Lindstrom e Bates (1988) consideram um modelo de efeitos aleatórios similar, mas com $R_{t_i} = \sigma^2 R_{t_i}^*$ e $D = \sigma^2 D^*$. Fairclough e Helms (1986) supõem que $R_{t_i} = \sigma^2 V_{t_i}$, onde V_{t_i} é uma matriz simétrica positiva definida conhecida. Prosseguiremos supondo que $R_{t_i} = \sigma^2 I_{t_i}$ para $i = 1, \dots, n$; neste caso, o modelo é denominado "*modelo de independência condicional*", já que $R_{t_i} = \sigma^2 I_{t_i}$ implica que as respostas da i -ésima unidade experimental são independentes dados α e b_i . Assumiremos que D possa ser escrita na forma linear $D = \sum_{g=1}^{k-1} \theta_g A_g$, onde as matrizes A_g 's ($q \times q$) são conhecidas e os θ_g 's ($g = 1, \dots, k-1$) representam parâmetros desconhecidos. Por exemplo, no caso em que $q = 2$ e D é não estruturada teremos:

$$D = \theta_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \theta_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \theta_3 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Assim, cada $\Sigma_i(\theta)$ pode também ser escrita na forma linear (2.14), isto é,

$$\begin{aligned} \Sigma_i(\theta) &= Z_i D Z_i' + \sigma^2 I_{t_i} \\ &= Z_i \left(\sum_{g=1}^{k-1} \theta_g A_g \right) Z_i' + \sigma^2 I_{t_i} \\ &= \sum_{g=1}^{k-1} \theta_g Z_i A_g Z_i' + \sigma^2 I_{t_i} \\ &= \sum_{g=1}^k \theta_g G_{ig} \end{aligned}$$

com $\theta = (\theta_1, \dots, \theta_{k-1}, \theta_k = \sigma^2)$ e

$$G_{ig} = \begin{cases} Z_i A_g Z_i' & \text{para } g = 1, \dots, k-1 \\ I_{t_i} & \text{para } g = k. \end{cases} \quad (2.18)$$

De (2.16) temos que para $i = 1, \dots, n$:

$$E(y_i) = X_i \alpha$$

e

$$\text{Var}(y_i) = \Sigma_i = Z_i D Z_i' + \sigma^2 I_{t_i}.$$

Nesse caso, os parâmetros desconhecidos do modelo (2.16) são α e θ , onde θ é o vetor $(k \times 1)$ constituído por σ^2 e os elementos que definem D . Sob essa formulação podemos considerar vários casos particulares que apresentaremos a seguir.

Consideremos primeiramente o Exemplo 1.3. A Figura 1.3 sugere que o valor médio da ETM em função do peso possa ser representado através de uma reta. Supondo também que a cada suíno esteja associada uma curva individual com desvios aleatórios em relação a curva populacional, um possível modelo de efeitos aleatórios para representar essa relação é:

$$y_{ij} = \alpha_1 + \alpha_2 x_{ij} + b_{i1} + b_{i2} x_{ij} + e_{ij}$$

para $i = 1, \dots, 13$ e $j = 1, \dots, t_i$, onde y_{ij} denota a resposta (ETM) do i -ésimo suíno para o peso x_{ij} , α_1 e α_2 são respectivamente o coeficiente linear e angular da reta populacional e b_{i1} e b_{i2} são os desvios aleatórios dos parâmetros das curvas individuais em relação aos parâmetros populacionais. Esse modelo pode ser escrito na forma (2.16) fazendo $y_i = (y_{i1}, \dots, y_{it_i})'$,

$$X_1 = \begin{pmatrix} 1 & 92 \\ 1 & 68 \\ 1 & 101 \\ 1 & 86 \end{pmatrix}, \dots, X_4 = \begin{pmatrix} 1 & 63 \\ 1 & 82 \end{pmatrix}, \dots, X_{13} = \begin{pmatrix} 1 & 93 \\ 1 & 104 \\ 1 & 86 \\ 1 & 62 \end{pmatrix},$$

com $\alpha = (\alpha_1, \alpha_2)'$, $Z_i = X_i$, $b_i = (b_{i1}, b_{i2})'$ e $e_i = (e_{i1}, \dots, e_{it_i})'$ para $i = 1, \dots, 13$. Se $\text{Var}(e_i) = \sigma^2 I_{t_i}$ e $\text{Var}(b_i) = D$, com D não estruturada, a matriz de covariância associada

ao i -ésimo suíno é dada por $\Sigma_i = \mathbf{X}_i \mathbf{D} \mathbf{X}_i' + \sigma^2 \mathbf{I}_{t_i}$ (note que os elementos de Σ_i dependem do peso através de $\mathbf{X}_i$).

Se considerarmos as matrizes $\mathbf{X}_i$'s como anteriormente e retirarmos b_{i2} do modelo teremos $\mathbf{Z}_i = \mathbf{1}_{t_i}$ para $i = 1, \dots, 13$, e estaremos supondo que as retas que representam as respostas médias individuais são paralelas. Esse fato implica uma estrutura de covariância uniforme, isto é, que as variâncias e as covariâncias entre a ETM nos diversos pesos são constantes, já que nesse caso $\Sigma_i = \sigma^2 \mathbf{I}_{t_i} + \sigma_b^2 \mathbf{1}_{t_i} \mathbf{1}_{t_i}'$.

Alternativamente, podemos supor que o grau da curva polinomial associada a cada suíno é maior que o da curva populacional. Por exemplo, considerando as matrizes $\mathbf{X}_i$'s como anteriormente e fazendo

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & 92 & 8464 \\ 1 & 68 & 4664 \\ 1 & 101 & 10201 \\ 1 & 86 & 7696 \end{pmatrix}, \dots, \mathbf{Z}_4 = \begin{pmatrix} 1 & 63 & 3969 \\ 1 & 82 & 6724 \end{pmatrix}, \dots, \mathbf{Z}_{13} = \begin{pmatrix} 1 & 93 & 8694 \\ 1 & 104 & 10816 \\ 1 & 86 & 7396 \\ 1 & 62 & 3844 \end{pmatrix},$$

teremos uma reta para representar a variação média populacional da resposta ($\mathbf{X}_i \boldsymbol{\alpha}$) e parábolas para representar a variação média individual da resposta ($\mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i$).

Modelos de efeitos aleatórios similares podem ser propostos para os Exemplos 1.1 e 1.2. Ainda mais, nesses exemplos que se enquadram nos casos de dados balanceados em relação ao tempo ou dados incompletos respectivamente, podemos assumir Σ_i não estruturada ($i = 1, \dots, n$). Isso pode ser obtido do modelo (2.16) fazendo $\sigma^2 = 0$ e definindo $\mathbf{Z}_i = \mathbf{E}_i$ como foi discutido na Seção 2.1. Assim,

$$\text{Var}(\mathbf{y}_i) = \Sigma_i = \begin{cases} \mathbf{D} & \text{se } t_i = T \\ \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' = \mathbf{E}_i \mathbf{D} \mathbf{E}_i' = \mathbf{D}_i & \text{se } t_i < T. \end{cases} \quad (2.19)$$

Nos modelos apresentados, nenhuma restrição foi colocada em $\mathbf{D}$ ($\mathbf{D}$ é não estruturada). Outros modelos poderiam ser considerados, por exemplo, quando $\mathbf{D}$ tem estrutura de processos auto-regressivos, médias móveis ou outra estrutura linear, como na seção anterior. Por exemplo, Elston e Grizzle (1962) consideram o caso em que os efeitos aleatórios são independentes, como acontece quando supomos $\mathbf{D}$ uma matriz diagonal.

Note que se fizermos $\mathbf{Z}_i = \mathbf{0}$ ($t_i \times 1$) para $i = 1, \dots, n$ no modelo (2.16), obteremos o modelo de regressão usual, onde $\Sigma_i = \sigma^2 \mathbf{I}_{t_i}$.

Em alguns casos, a estrutura de covariância pode diferir entre subpopulações. Suponhamos que no Exemplo 1.1 desejássemos ajustar retas diferentes para cada grupo (meninas e meninos), mas com a suposição adicional que a cada indivíduo esteja associada sua própria reta que difere aleatoriamente da reta populacional. Sejam $\mathbf{a}'_i = (1, 0)$ e $\mathbf{a}'_i = (0, 1)$ se o i -ésimo indivíduo pertence ao grupo de meninas ou meninos respectivamente. Um modelo de efeitos aleatórios possível nesse caso é:

$$y_i = (\mathbf{a}'_i \otimes \mathbf{Z}_i) \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i, \quad (2.20)$$

onde

$$\mathbf{Z}_i = \begin{pmatrix} 1 & 8 \\ 1 & 12 \\ 1 & 10 \\ 1 & 14 \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} \quad \text{e} \quad \mathbf{b}_i = \begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix}.$$

Isso resulta em $\mathbf{X}_i = \mathbf{a}'_i \otimes \mathbf{Z}_i$, levando ao modelo desejado para as curvas populacionais (retas diferentes para cada grupo), mas com a estrutura de covariância independente do grupo ($\Sigma_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i + \sigma^2 \mathbf{I}_{t_i}$). Alternativamente, podemos fazer com que a estrutura de covariância dependa do grupo, escrevendo:

$$y_i = (\mathbf{a}'_i \otimes \mathbf{Z}_i) \boldsymbol{\alpha} + (\mathbf{a}'_i \otimes \mathbf{Z}_i) \mathbf{b}_i^* + \mathbf{e}_i, \quad (2.21)$$

onde $\mathbf{b}_i^* = (b_{i1}^*, b_{i2}^*, b_{i3}^*, b_{i4}^*)' \sim N(\mathbf{0}, \mathbf{D}^*)$, com

$$\mathbf{D}^* = \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix}$$

e $\mathbf{D}_1$ (2×2) e $\mathbf{D}_2$ (2×2) são matrizes simétricas positivas definidas. Assim, para o grupo de meninas, $\mathbf{a}'_i = (1, 0)$, temos:

$$y_i = \mathbf{Z}_i \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \mathbf{Z}_i \begin{pmatrix} b_{i1}^* \\ b_{i2}^* \end{pmatrix} + \mathbf{e}_i,$$

e para o grupo de meninos, $\mathbf{a}'_i = (0, 1)$, temos:

$$y_i = \mathbf{Z}_i \begin{pmatrix} \alpha_3 \\ \alpha_4 \end{pmatrix} + \mathbf{Z}_i \begin{pmatrix} b_{i3}^* \\ b_{i4}^* \end{pmatrix} + \mathbf{e}_i.$$

Nesse caso, segue-se que a matriz de covariância associada ao grupo das meninas é dada por:

$$\Sigma_i = \mathbf{Z}_i \mathbf{D}_1 \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{t_i},$$

enquanto que no grupo dos meninos é:

$$\Sigma_i = \mathbf{Z}_i \mathbf{D}_2 \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{t_i}.$$

O vetor $\mathbf{b}_i^* = (b_{i1}^*, b_{i2}^*, b_{i3}^*, b_{i4}^*)'$ é artificialmente construído já que metade de seus componentes nunca entram realmente no modelo (2.16).

Um caso particular do modelo de efeitos aleatórios na formulação em dois estágios tem origens na literatura sobre Curvas de Crescimento. Rao (1965, 1975) considerou o problema do ajuste de curvas de crescimento polinomiais para um único grupo de unidades experimentais observadas nas mesmas T ocasiões (balanceado em relação ao tempo). Generalizando o modelo proposto por Rao para dados não balanceados em relação ao tempo e para diversas subpopulações de unidades experimentais, Ware (1985) e Laird et al. (1987) propuseram uma abordagem alternativa para especificar o modelo, caracterizando primeiramente a curva de crescimento de cada unidade experimental e depois modelando os parâmetros das curvas individuais como funções lineares das características das diversas subpopulações de unidades experimentais. Especificamente, o modelo associado à i -ésima unidade experimental é dado por:

$$\mathbf{y}_i = \mathbf{T}_i \boldsymbol{\beta}_i + \mathbf{e}_i, \quad (2.22)$$

onde $\mathbf{y}_i$, $\mathbf{T}_i$ e $\mathbf{e}_i$ são, respectivamente, equivalentes a $\mathbf{y}_i$, $\mathbf{Z}_i$ e $\mathbf{e}_i$ do modelo (2.16). O componente $\mathbf{T}_i \boldsymbol{\beta}_i$ define a curva da i -ésima unidade experimental. Os $\boldsymbol{\beta}_i$'s são os "parâmetros" aleatórios independentes associados à curva da i -ésima unidade experimental; supomos que $\boldsymbol{\beta}_i \sim N(\mathbf{A}_i \boldsymbol{\alpha}^*, \mathbf{D}^*)$, onde $\boldsymbol{\alpha}^*$ e $\mathbf{D}^*$ são como $\boldsymbol{\alpha}$ e $\mathbf{D}$ em (2.16) e $\mathbf{A}_i$ é uma matriz ($q \times p$) conhecida de especificação do modelo para $\boldsymbol{\beta}_i$. Nesse caso, o modelo populacional é dado por:

$$E(\mathbf{y}_i) = \mathbf{T}_i \mathbf{A}_i \boldsymbol{\alpha}^*$$

e

$$\text{Var}(\mathbf{y}_i) = \mathbf{T}_i \mathbf{D}^* \mathbf{T}_i' + \sigma^2 \mathbf{I}_{t_i}.$$

Vemos, assim, que (2.22) é um caso particular do modelo (2.16) com $\mathbf{X}_i = \mathbf{T}_i \mathbf{A}_i$, $\mathbf{Z}_i = \mathbf{T}_i$, $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ e $\mathbf{D} = \mathbf{D}^*$. Então, $\boldsymbol{\beta}_i = \mathbf{A}_i \boldsymbol{\alpha}^* + \mathbf{b}_i$ e $\mathbf{b}_i$ pode ser visto como um *resíduo generalizado* referente aos desvios entre os parâmetros associados às curvas individuais e os parâmetros da curva de crescimento populacional. Vale notar a diferença entre esse resíduo e o resíduo comum definido por:

$$\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}.$$

De uma forma geral, é mais natural pensar nesse contexto de modelos de curvas de crescimento e muitos investigadores acham mais fácil utilizar essa abordagem. Essa formulação de curva de crescimento parece impor algumas limitações no modelo, pois a dimensão de $\mathbf{Z}_i$ determina não somente a estrutura da matriz de covariância $\boldsymbol{\Sigma}_i$, mas também o modelo para o vetor de médias $\mathbf{Z}_i \mathbf{A}_i \boldsymbol{\alpha}$. Assim, para modelar o vetor de médias adequadamente, podemos estar forçando uma estrutura de covariância por demais complicada. No entanto, através de um artifício, é possível escrever qualquer modelo de efeitos aleatórios da forma (2.16) como uma curva de crescimento generalizada da forma (2.22); para isso, basta especificar os elementos de $\boldsymbol{\alpha}^*$ e $\mathbf{D}^*$ convenientemente. Por exemplo, fazendo $\mathbf{T}_i = (\mathbf{X}_i, \mathbf{Z}_i)$, $\boldsymbol{\alpha}^{*'} = (\boldsymbol{\alpha}', \mathbf{0})$, $\mathbf{A}_i = \mathbf{I}_{p+q}$ e

$$\mathbf{D}^* = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}$$

em (2.22), temos

$$\boldsymbol{\beta}_i = \mathbf{A}_i \boldsymbol{\alpha}^* + \mathbf{b}_i = \mathbf{I}_{p+q} \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{b}_i \end{pmatrix}.$$

Então,

$$\mathbf{y}_i = \mathbf{T}_i \boldsymbol{\beta}_i + \mathbf{e}_i = (\mathbf{X}_i, \mathbf{Z}_i) \left[\begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{b}_i \end{pmatrix} \right] + \mathbf{e}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i$$

que corresponde a (2.16).

Capítulo 3

Estimação e Testes de Hipóteses

Neste capítulo tratamos do processo de estimação dos parâmetros $\alpha = (\alpha_1, \dots, \alpha_p)'$ e $\theta = (\theta_1, \dots, \theta_k)'$ do modelo de efeitos aleatórios (2.16) com base na suposição de normalidade do vetor de respostas de cada unidade experimental. Dentro dessa abordagem, apresentamos os métodos de máxima verossimilhança (MV) e o de máxima verossimilhança restrita (MVR) na Seção 3.1, segundo o enfoque de Harville (1977). Geralmente a resolução das equações normais resultantes da aplicação desses métodos requer a utilização de processos iterativos. Assim, na Seção 3.2 são apresentados os processos iterativos mais comuns para obtenção das estimativas de máxima verossimilhança dos parâmetros envolvidos, discutindo alguns aspectos computacionais. Na Seção 3.3 consideramos casos para os quais é possível encontrar formas explícitas para os estimadores de máxima verossimilhança e máxima verossimilhança restrita.

3.1 Métodos de Máxima Verossimilhança

O método de máxima verossimilhança usual envolve a estimação de α e θ simultaneamente com base na distribuição conjunta de $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$. Como, sob a suposição de normalidade, a distribuição marginal dos $\mathbf{y}_i$'s é dada por (2.17), a função de verossimilhança

$$y_i \sim N(x_i \beta, \Sigma_i)$$

associada é

$$L(\mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = (2\pi)^{-\frac{N}{2}} \prod_{i=1}^n |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \right], \quad (3.1)$$

onde $N = \sum_{i=1}^n t_i$ e $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{t_i}$.

Para obtermos estimativas de máxima verossimilhança de $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$ precisamos encontrar $\hat{\boldsymbol{\alpha}}$ e $\hat{\boldsymbol{\theta}}$ que maximizam (3.1). Isso é equivalente a maximizar o logaritmo da função de verossimilhança:

$$l(\mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = K - \frac{1}{2} \sum_{i=1}^n \ln |\boldsymbol{\Sigma}_i| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}), \quad (3.2)$$

onde $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{t_i}$ e K é uma constante que não depende de $\boldsymbol{\alpha}$ nem de $\boldsymbol{\theta}$. Nesse sentido, necessitamos calcular as primeiras e segundas derivadas parciais de $l(\mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta})$ em relação a $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)'$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$. Essas derivadas são dadas por:

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_{\boldsymbol{\alpha}} \\ \mathbf{s}_{\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \partial l / \partial \boldsymbol{\alpha} \\ \partial l / \partial \boldsymbol{\theta} \end{bmatrix}$$

e

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & \mathbf{H}_{\boldsymbol{\alpha}\boldsymbol{\theta}} \\ \mathbf{H}_{\boldsymbol{\theta}\boldsymbol{\alpha}} & \mathbf{H}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \partial^2 l / \partial \boldsymbol{\alpha}' \partial \boldsymbol{\alpha} & \partial^2 l / \partial \boldsymbol{\alpha}' \partial \boldsymbol{\theta} \\ \partial^2 l / \partial \boldsymbol{\theta}' \partial \boldsymbol{\alpha} & \partial^2 l / \partial \boldsymbol{\theta}' \partial \boldsymbol{\theta} \end{bmatrix}.$$

Os cálculos de $\mathbf{s}$ e $\mathbf{H}$ são facilitados com a utilização de resultados de álgebra e diferenciação matricial e estão detalhados no Apêndice A. As expressões dos componentes de $\mathbf{s}$ e $\mathbf{H}$ são:

$$\begin{matrix} \mathbf{s}_{\boldsymbol{\alpha}} \\ (p \times 1) \end{matrix} = \sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}), \quad (3.3)$$

$$\begin{matrix} \mathbf{s}_{\boldsymbol{\theta}} \\ (k \times 1) \end{matrix} = \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \boldsymbol{\Sigma}_i^{-1} [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \boldsymbol{\Sigma}_i] \boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ig} \right\rangle_g \right], \quad (3.4)$$

$$\begin{matrix} \mathbf{H}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \\ (p \times p) \end{matrix} = - \sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i, \quad (3.5)$$

$$\begin{matrix} \mathbf{H}_{\boldsymbol{\alpha}\boldsymbol{\theta}} \\ (p \times k) \end{matrix} = \mathbf{H}_{\boldsymbol{\theta}\boldsymbol{\alpha}}' = \left[\left\langle - \sum_{i=1}^n \mathbf{x}_{ij}' \boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \right\rangle_{jg} \right] \quad (3.6)$$

e

$$\begin{matrix} \mathbf{H}_{\boldsymbol{\theta}\boldsymbol{\theta}} \\ (k \times k) \end{matrix} = \left[\left\langle - \frac{1}{2} \sum_{i=1}^n \text{tr} \boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} [2(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \boldsymbol{\Sigma}_i] \boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ih} \right\rangle_{gh} \right], \quad (3.7)$$

onde x_{ij} é a j -ésima coluna de X_i , as matrizes G_{ig} e G_{ih} são dadas por (2.18), expressões do tipo $[< >_{ab}]$ denotam matrizes cujo elemento de ordem (a, b) é dado pela expressão entre $< >$ e expressões do tipo $[< >_a]$ denotam vetores cujo a -ésimo elemento é dado pela expressão entre $< >$.

Os estimadores de máxima verossimilhança de α e θ correspondem à solução das seguintes $p + k$ equações normais:

$$s = \begin{bmatrix} \partial l / \partial \alpha \\ \partial l / \partial \theta \end{bmatrix} = 0_{p+k}$$

Mais especificamente, eles são as soluções $\hat{\alpha}$ e $\hat{\theta}$ de

$$\sum_{i=1}^n X_i' \hat{\Sigma}_i^{-1} (y_i - X_i \hat{\alpha}) = 0_p \quad (3.8)$$

e

$$\frac{1}{2} \sum_{i=1}^n \text{tr} \hat{\Sigma}_i^{-1} [(y_i - X_i \hat{\alpha})(y_i - X_i \hat{\alpha})' - \hat{\Sigma}_i] \hat{\Sigma}_i^{-1} G_{ig} = 0 \text{ para } g = 1, \dots, k, \quad (3.9)$$

se elas estiverem contidas no espaço paramétrico definido pelo modelo (isto é, se $\Sigma_i(\hat{\theta})$ for positiva definida para $i = 1, \dots, n$) e se $H(\hat{\alpha}, \hat{\theta})$ for negativa definida (para garantir que $\hat{\alpha}$ e $\hat{\theta}$ são pontos de máximo).

De (3.8) e (3.9) obtemos:

$$\hat{\alpha} = \left(\sum_{i=1}^n X_i' \hat{\Sigma}_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^n X_i' \hat{\Sigma}_i^{-1} y_i \right) \quad (3.10)$$

e

$$\sum_{i=1}^n \text{tr} \hat{\Sigma}_i^{-1} G_{ig} = \sum_{i=1}^n \text{tr} \hat{\Sigma}_i^{-1} G_{ig} \hat{\Sigma}_i^{-1} (y_i - X_i \hat{\alpha})(y_i - X_i \hat{\alpha})' \text{ para } g = 1, \dots, k. \quad (3.11)$$

Podemos observar de (3.10) que o estimador de máxima verossimilhança de α quando as matrizes Σ_i 's (ou θ) são conhecidas é dado por:

$$\hat{\alpha} = \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^n X_i' \Sigma_i^{-1} y_i \right). \quad (3.12)$$

Além disso, convém notar a dificuldade de obtermos soluções explícitas para $\hat{\alpha}$ e $\hat{\theta}$ a partir de (3.10) e (3.11). A equação (3.11) pode ser ligeiramente simplificada em função da forma linear (2.14) para Σ_i . Nesse caso, podemos mostrar que (ver, por exemplo, Andrade e Helms (1984a)):

$$\hat{\theta} = \left[\left\langle \sum_{i=1}^n \text{tr} \hat{\Sigma}_i^{-1} \mathbf{G}_{ig} \hat{\Sigma}_i^{-1} \mathbf{G}_{ih} \right\rangle_{gh} \right]^{-1} \left[\left\langle \sum_{i=1}^n \text{tr} \hat{\Sigma}_i^{-1} \mathbf{G}_{ig} \hat{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\alpha}) (\mathbf{y}_i - \mathbf{X}_i \hat{\alpha})' \right\rangle_g \right] \quad (3.13)$$

As equações (3.8) e (3.9) ou (3.10) e (3.11) sugerem que em muitas situações necessitaremos de processos iterativos para encontrar estimativas de máxima verossimilhança de α e θ .

A matriz de informação empírica associada a função de verossimilhança (3.1) é $-\mathbf{H}(\alpha, \theta)$ e a correspondente matriz de informação de Fisher é $\mathbf{I}(\alpha, \theta) = -\mathbf{E}[\mathbf{H}(\alpha, \theta)]$. Mais especificamente, temos:

$$\mathbf{I}(\alpha, \theta) = \begin{bmatrix} \mathbf{I}_{\alpha\alpha} & \mathbf{I}_{\alpha\theta} \\ \mathbf{I}_{\theta\alpha} & \mathbf{I}_{\theta\theta} \end{bmatrix},$$

onde

$$\mathbf{I}_{\alpha\alpha} = -\mathbf{E}[\mathbf{H}_{\alpha\alpha}] = \sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} \mathbf{X}_i \quad (3.14)$$

$$\begin{aligned} \mathbf{I}_{\alpha\theta} &= \mathbf{I}_{\theta\alpha}' = -\mathbf{E}[\mathbf{H}_{\alpha\theta}] \\ &= \mathbf{E} \left[\left\langle \sum_{i=1}^n \mathbf{x}_{ij}' \Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \alpha) \right\rangle_{jg} \right] \\ &= \left[\left\langle \sum_{i=1}^n \mathbf{x}_{ij}' \Sigma_i^{-1} \mathbf{G}_{ig} (\mathbf{E}(\mathbf{y}_i) - \mathbf{X}_i \alpha) \right\rangle_{jg} \right] \\ &= \left[\left\langle \sum_{i=1}^n \mathbf{x}_{ij}' \Sigma_i^{-1} \mathbf{G}_{ig} (\mathbf{X}_i \alpha - \mathbf{X}_i \alpha) \right\rangle_{jg} \right] \\ &= \mathbf{0} \end{aligned} \quad (3.15)$$

$$\begin{aligned} \mathbf{I}_{\theta\theta} &= -\mathbf{E}[\mathbf{H}_{\theta\theta}] \\ &= \mathbf{E} \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} [2(\mathbf{y}_i - \mathbf{X}_i \alpha)(\mathbf{y}_i - \mathbf{X}_i \alpha)' - \Sigma_i] \Sigma_i^{-1} \mathbf{G}_{ih} \right\rangle_{gh} \right] \\ &= \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} [2\text{Var}(\mathbf{y}_i) - \Sigma_i] \Sigma_i^{-1} \mathbf{G}_{ih} \right\rangle_{gh} \right] \end{aligned}$$

$$\begin{aligned}
&= \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} \Sigma_i \Sigma_i^{-1} \mathbf{G}_{ih} \right\rangle_{gh} \right] \\
&= \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} \mathbf{G}_{ih} \right\rangle_{gh} \right] \quad (3.16)
\end{aligned}$$

Como os estimadores de máxima verossimilhança são consistentes e assintoticamente eficientes, as seguintes expressões, baseadas na matriz de informação de Fisher observada, podem ser utilizadas como estimativas da matriz de covariância assintótica dos estimadores dos parâmetros (ver Searle (1970)):

$$\widehat{\text{Var}}_A(\hat{\alpha}) = \left(\sum_{i=1}^n \mathbf{X}_i' \hat{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \quad (3.17)$$

$$\widehat{\text{Var}}_A(\hat{\theta}) = 2 \left[\left\langle \sum_{i=1}^n \text{tr} \hat{\Sigma}_i^{-1} \mathbf{G}_{ig} \hat{\Sigma}_i^{-1} \mathbf{G}_{ih} \right\rangle_{gh} \right]^{-1} \quad (3.18)$$

$$\widehat{\text{Cov}}_A(\hat{\alpha}, \hat{\theta}) = \mathbf{0}. \quad (3.19)$$

Alguns autores (ver, por exemplo, Jennrich e Schluchter (1986) ou Little (1988)) sugerem que no caso de dados não balanceados em relação ao tempo a estimativa da matriz de covariância assintótica dos estimadores dos parâmetros seja baseada na matriz de informação empírica $-\mathbf{H}(\hat{\alpha}, \hat{\theta})$, devido à presença de observações incompletas. Uma justificativa para esse fato é que as variâncias assintóticas dos estimadores dos parâmetros baseadas na matriz de informação de Fisher só são válidas se os dados incompletos são MCAR. As variâncias assintóticas baseadas na matriz de informação empírica são preferíveis quando os dados não são MCAR, pois permanecem válidas mesmo quando os dados são MAR e ignoráveis (ver Laird (1988), por exemplo).

O ajuste dos modelos hierárquicos para as estruturas de médias e covariâncias pode ser avaliado através da comparação dos modelos mais simples com modelos mais gerais. No caso do método de máxima verossimilhança, a estatística utilizada para testar o ajuste é a da Razão de Verossimilhança Generalizada (ver Jennrich e Schluchter (1986)), dada por

$$Q_1 = 2 [l(\hat{\alpha}^G, \hat{\theta}^G) - l(\hat{\alpha}^S, \hat{\theta}^S)], \quad (3.20)$$

onde $\hat{\alpha}^G$ e $\hat{\theta}^G$ são os valores de α e θ que maximizam a verossimilhança sob o modelo mais geral e $\hat{\alpha}^S$ e $\hat{\theta}^S$ sob o modelo mais simples. Sob a hipótese H_0 de que o modelo mais simples é adequado, Q_1 tem distribuição assintótica Qui-Quadrado com $(p^G + k^G - p^S - k^S)$ graus de liberdade, onde $(p^G$ e $k^G)$ e $(p^S$ e $k^S)$ correspondem ao número de parâmetros de α e θ sob os modelos mais geral e mais simples, respectivamente.

Uma crítica que é geralmente feita aos estimadores de máxima verossimilhança dos elementos de θ é que eles não levam em consideração a perda de graus de liberdade resultantes da estimação de α (ver, por exemplo, Harville (1977)), sendo, em geral, viciados (subestimam os elementos de θ). Esse fato pode ser verificado, por exemplo, quando os dados são balanceados em relação ao tempo e as técnicas de Análise de Variância para modelos mistos citadas no Capítulo 1 são aplicáveis. Nessa situação, podemos observar que alguns componentes de $\hat{\theta}$ são menores que os correspondentes obtidos por Análise de Variância, que são os melhores estimadores não viciados quando os dados são normais e balanceados (ver Graybill e Hultquist (1961) e Albert (1976), por exemplo). As estimativas de máxima verossimilhança de α , coincidem com as obtidas através da ANOVA; entretanto, as estimativas de $\text{Var}(\hat{\alpha})$ são afetadas pela subestimação de θ . O vício (para baixo) da estimativa de máxima verossimilhança de $\hat{\theta}$ pode ser facilmente observado através de um exemplo de Regressão linear, obtido fazendo no modelo (2.16) $Z_i = 0_{t_i}$, de modo que $\Sigma_i = \sigma^2 I_{t_i}$ ($\theta = \sigma^2$). Nesse exemplo, o estimador de máxima verossimilhança de σ^2 tem esperança $\sigma^2(N - p)/N$, onde $N = \sum_{i=1}^n t_i$ e p é a dimensão de α (um exemplo com $p = 2$ está descrito em Mood, Graybill e Boes (1974), p. 487-490). Assim, o estimador de máxima verossimilhança de σ^2 é viciado para baixo por uma quantia igual a $\sigma^2 p/N$, que pode ser considerável se $N - p$ for pequeno. Em outras estruturas de covariância esse vício para menos também ocorre, mas métodos gerais e simples de correção podem não existir, principalmente quando os dados não são balanceados em relação ao tempo.

Uma possível solução para a eliminação desse vício está no método de máxima verossimilhança restrita (ver Patterson e Thompson (1971) ou Harville (1974, 1977)). Esse

método consiste em maximizar a função de verossimilhança associada a um conjunto específico de $N - p$ contrastes de erro linearmente independentes, em vez da função usual de verossimilhança (3.1). Entendemos por *contraste de erro* uma combinação linear $u'y$ das observações tal que $E(u'y) = 0$ (isto é, $u'X = 0$ para $X = (X'_1, \dots, X'_n)'$), onde u ($N \times 1$) não depende nem de α nem de θ . Harville (1974) derivou a *função de verossimilhança restrita* associada a um conjunto de $N - p$ resíduos de mínimos quadrados independentes para um modelo mais geral que contém (2.1) e (2.16) como casos particulares. Assim, escolhendo $N - p$ combinações lineares de $\bar{e}_i = y_i - X_i\bar{\alpha}$, para $i = 1, \dots, n$, onde $\bar{\alpha} = (\sum_{i=1}^n X'_i X_i)^{-1} (\sum_{i=1}^n X'_i y_i)$, a função de verossimilhança restrita resultante é:

$$l_R(y; \theta) = (2\pi)^{-\frac{(N-p)}{2}} \left| \sum_{i=1}^n X'_i X_i \right|^{\frac{1}{2}} \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \left| \sum_{i=1}^n X'_i \Sigma_i^{-1} X_i \right|^{-\frac{1}{2}} \times \\ \times \exp \left[-\frac{1}{2} \sum_{i=1}^n (y_i - X_i \tilde{\alpha})' \Sigma_i^{-1} (y_i - X_i \tilde{\alpha}) \right], \quad (3.21)$$

onde $\tilde{\alpha} = (\sum_{i=1}^n X'_i \Sigma_i^{-1} X_i)^{-1} (\sum_{i=1}^n X'_i \Sigma_i^{-1} y_i)$. Segundo Harville (1974) qualquer que seja o conjunto de contrastes de erro linearmente independentes utilizado, a função de verossimilhança restrita será sempre proporcional a (3.21). É importante notar que a função de verossimilhança restrita não envolve α ; então, uma vez encontrado o estimador de máxima verossimilhança de θ ($\hat{\theta}_R$), α pode ser estimado através do método de mínimos quadrados generalizados, substituindo-se Σ_i por $\Sigma_i(\hat{\theta}_R)$ na equação (3.12). As estimativas de θ obtidas através da maximização de (3.21) concordam com as estimativas obtidas através da ANOVA no caso de dados balanceados em relação ao tempo ou quando ajustamos modelos de Regressão linear, sempre que esses métodos (ANOVA) produzem estimativas positivas dos componentes de θ associados às variâncias. Dessa forma, os graus de liberdade perdidos na estimação de α são levados em conta no cálculo de $\hat{\theta}_R$. Devido a esse fato, os estimadores de máxima verossimilhança restrita de θ são muitas vezes preferidos aos de máxima verossimilhança (ver Harville (1977), Dempster, Rubin e Tsutakawa (1981), Laird e Ware (1982) ou Berk (1987), por exemplo).

Para obter estimativas de máxima verossimilhança restrita de θ podemos maximizar

o logaritmo natural da função (3.21) dado por:

$$l_R(\mathbf{y}; \boldsymbol{\theta}) = K - \frac{1}{2} \sum_{i=1}^n \ln |\boldsymbol{\Sigma}_i| - \frac{1}{2} \ln \left| \sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}), \quad (3.22)$$

onde $\tilde{\boldsymbol{\alpha}} = (\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i)$ e K é uma constante que não depende de $\boldsymbol{\theta}$. Para esse fim, necessitamos calcular as primeiras e segundas derivadas parciais de $l_R(\mathbf{y}; \boldsymbol{\theta})$ em relação a $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$, que são dadas por:

$$\mathbf{s}_R = \left[\left\langle \frac{\partial l_R}{\partial \theta_g} \right\rangle_g \right]$$

e

$$\mathbf{H}_R = \left[\left\langle \frac{\partial^2 l_R}{\partial \theta_g \partial \theta_h} \right\rangle_{gh} \right].$$

Detalhes sobre os cálculos correspondentes constam do Apêndice A e as expressões para os elementos de $\mathbf{s}_R$ e $\mathbf{H}_R$ são:

$$\frac{\partial l_R}{\partial \theta_g} = \frac{1}{2} \sum_{i=1}^n \text{tr} \left[\boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}) (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' - \mathbf{P}_i \mathbf{G}_{ig} \right] \quad \text{para } g = 1, \dots, k, \quad (3.23)$$

e

$$\begin{aligned} \frac{\partial^2 l_R}{\partial \theta_g \partial \theta_h} = & - \sum_{i=1}^n \text{tr} \boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}) (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' \\ & + \frac{1}{2} \sum_{i=1}^n \text{tr} \mathbf{P}_i \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \quad \text{para } g, h = 1, \dots, k, \end{aligned} \quad (3.24)$$

onde $\mathbf{P}_i = \boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i (\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1}$.

O estimador de máxima verossimilhança restrita de $\boldsymbol{\theta}$ é a solução de

$$\mathbf{s}_R = [\partial l_R / \partial \boldsymbol{\theta}] = \mathbf{0}_k$$

que corresponde à solução $(\hat{\boldsymbol{\theta}}_R)$ das seguintes k equações:

$$\frac{1}{2} \sum_{i=1}^n \text{tr} [\hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{G}_{ig} \hat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}) (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' - \hat{\mathbf{P}}_i \mathbf{G}_{ig}] = 0 \quad \text{para } g = 1, \dots, k, \quad (3.25)$$

sob a restrição de que $\hat{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}_i(\hat{\boldsymbol{\theta}}_R)$ deve ser positiva definida para $i = 1, \dots, n$. Se $\mathbf{H}_R(\hat{\boldsymbol{\theta}}_R)$ for negativa definida podemos garantir que $\hat{\boldsymbol{\theta}}_R$ é um ponto de máximo de $l_R(\mathbf{y}; \boldsymbol{\theta})$.

Como no método de máxima verossimilhança, frequentemente necessitaremos de processos iterativos para resolver as equações dadas por (3.25).

A matriz de informação de Fisher associada à função de verossimilhança restrita é dada por:

$$\mathbf{I}_R(\boldsymbol{\theta}) = -E[\mathbf{H}_R(\boldsymbol{\theta})] = -E \left[\left\langle \frac{\partial l_R}{\partial \theta_g \partial \theta_h} \right\rangle_{gh} \right].$$

Consequentemente:

$$\begin{aligned} \mathbf{I}_R(\boldsymbol{\theta}) &= \left[\left\langle \sum_{i=1}^n \text{tr} \boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \boldsymbol{\Sigma}_i^{-1} \text{Var}(\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}) \right\rangle_{gh} \right] \\ &\quad - \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \mathbf{P}_i \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \right\rangle_{gh} \right] \end{aligned}$$

Como

$$\begin{aligned} \text{Var}(\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}) &= \text{Var}(\mathbf{y}_i) + \mathbf{X}_i \text{Var}(\tilde{\boldsymbol{\alpha}}) \mathbf{X}_i' - 2 \text{Cov}(\mathbf{y}_i, \tilde{\boldsymbol{\alpha}}) \mathbf{X}_i' \\ &= \boldsymbol{\Sigma}_i + \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}_i' - 2 \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \\ &= \boldsymbol{\Sigma}_i - \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \\ &= \boldsymbol{\Sigma}_i \left[\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \right] \boldsymbol{\Sigma}_i \\ &= \boldsymbol{\Sigma}_i \mathbf{P}_i \boldsymbol{\Sigma}_i \end{aligned}$$

temos:

$$\begin{aligned} \mathbf{I}_R(\boldsymbol{\theta}) &= \left[\left\langle \sum_{i=1}^n \text{tr} \boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Sigma}_i \mathbf{P}_i \boldsymbol{\Sigma}_i \right\rangle_{gh} \right] - \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \mathbf{P}_i \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \right\rangle_{gh} \right] \\ &= \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \mathbf{P}_i \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \right\rangle_{gh} \right]. \end{aligned} \quad (3.26)$$

Podemos obter uma estimativa da matriz de covariância assintótica do estimador $\hat{\boldsymbol{\theta}}_R$, substituindo $\boldsymbol{\theta}$ por $\hat{\boldsymbol{\theta}}_R$ em $\mathbf{I}_R(\boldsymbol{\theta})$; mais especificamente:

$$\widehat{\text{Var}}_A(\hat{\boldsymbol{\theta}}_R) = 2 \left[\left\langle \sum_{i=1}^n \text{tr} \mathbf{P}_i \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \right\rangle_{gh} \right]^{-1} \quad (3.27)$$

A matriz de covariância de $\hat{\alpha}$ pode ser estimada por:

$$\widehat{\text{Var}}(\hat{\alpha}) = \left(\sum_{i=1}^n \mathbf{X}_i' \hat{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1}, \quad (3.28)$$

onde $\hat{\Sigma}_i = \Sigma_i(\hat{\theta}_R)$.

Tanto no método da máxima verossimilhança, quanto no método de máxima verossimilhança restrita, testes de hipóteses sobre combinações lineares da forma $H_0 : \mathbf{C}\alpha = \mathbf{0}$, onde $\mathbf{C}$ ($c \times p$) é uma matriz de constantes conhecidas de posto c ($c \leq p$), podem ser realizados através da estatística de Wald:

$$Q_c = (\mathbf{C}\hat{\alpha})' \left[\mathbf{C} \left(\sum_{i=1}^n \mathbf{X}_i' \hat{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{C}' \right]^{-1} (\mathbf{C}\hat{\alpha}). \quad (3.29)$$

Sob H_0 , Q_c tem distribuição assintótica Qui-Quadrado com c graus de liberdade.

Quando usamos o método de máxima verossimilhança restrita podemos considerar hipóteses relativas a adequabilidade da estrutura de covariância proposta através da comparação entre uma estrutura mais simples e uma mais geral (não esquecendo que os modelos a serem comparados devem ter a mesma estrutura de médias). Essas hipóteses podem ser testadas através da estatística da razão de verossimilhança generalizada, dada por:

$$Q_2 = 2 [l_R(\hat{\theta}_R^G) - l_R(\hat{\theta}_R^S)], \quad (3.30)$$

onde $\hat{\theta}_R^G$ é o valor de θ que maximiza a máxima verossimilhança restrita sob o modelo mais geral e $\hat{\theta}_R^S$ sob o modelo mais simples. Sob a hipótese de que o modelo mais simples é adequado, Q_2 tem distribuição assintótica Qui-Quadrado com $(k^G - k^S)$ graus de liberdade, onde k^G e k^S representam o número de elementos de θ sob os modelos mais geral e mais simples respectivamente.

Outros elementos que podem ser de interesse quando trabalhamos com o modelo de efeitos aleatórios (2.16) são as “estimativas” dos efeitos aleatórios $\mathbf{b}_1, \dots, \mathbf{b}_n$, já que eles podem ser interpretados como resíduos generalizados (ver Seção 2.2). Laird e Ware (1982) mostram que uma estimativa de $\mathbf{b}_i$ é:

$$\hat{\mathbf{b}}_i = \hat{\mathbf{D}}\mathbf{Z}_i' \hat{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\alpha}), \quad (3.31)$$

onde $\hat{\alpha} = (\sum_{i=1}^n \mathbf{X}_i' \hat{\Sigma}_i^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^n \mathbf{X}_i' \hat{\Sigma}_i^{-1} \mathbf{y}_i)$ e as estimativas $\hat{\Sigma}_1, \dots, \hat{\Sigma}_n$ são obtidas por máxima verossimilhança ou máxima verossimilhança restrita. Uma justificativa para a expressão (3.31) no caso de $\Sigma_1, \dots, \Sigma_n$ serem *conhecidas* é dada através do Teorema de Gauss-Markov generalizado (ver Harville (1976, 1977)), segundo o qual $\lambda_1' \hat{\alpha} + \lambda_2' \hat{\mathbf{b}}_i$ é o melhor estimador não viciado da combinação linear $\lambda_1' \alpha + \lambda_2' \mathbf{b}_i$. Quando $\Sigma_1, \dots, \Sigma_n$ são desconhecidas e são estimadas por máxima verossimilhança ou máxima verossimilhança restrita, a expressão (3.31) pode ser interpretada como sendo um estimador empírico de Bayes (ver Laird e Ware (1982)). Uma estimativa da variância de $\hat{\mathbf{b}}_i$ é dada por:

$$\widehat{\text{Var}}(\hat{\mathbf{b}}_i - \mathbf{b}_i) = \hat{\mathbf{D}}(\mathbf{I}_q - \mathbf{Z}_i' \hat{\mathbf{P}}_i \mathbf{Z}_i \hat{\mathbf{D}}), \quad (3.32)$$

onde $\hat{\mathbf{P}}_i = \hat{\Sigma}_i^{-1} - \hat{\Sigma}_i^{-1} \mathbf{X}_i (\sum_{i=1}^n \mathbf{X}_i' \hat{\Sigma}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i' \hat{\Sigma}_i^{-1}$.

É importante lembrar que as propriedades estatísticas discutidas acima são assintóticas e que não existem estudos sobre o comportamento desses estimadores para amostras pequenas. Consequentemente, os modelos considerados aqui devem ser encarados muito mais sob o ponto de vista exploratório do que sob uma ótica inferencial.

3.2 Processos Iterativos para Maximização da Verossimilhança

Nas situações onde não podemos encontrar soluções explícitas para os estimadores de máxima verossimilhança de α e θ e de máxima verossimilhança restrita de θ , precisamos empregar processos iterativos para obter as estimativas correspondentes.

Denotemos por ϕ o conjunto de m parâmetros que desejamos estimar e por $l(\phi)$ o logaritmo da função de verossimilhança. Geralmente os processos iterativos para encontrar o valor $\hat{\phi}$ que maximiza $l(\phi)$ têm a forma: $\phi^{(w+1)} = f(\phi^{(w)})$ para $w = 0, 1, 2, \dots$, onde w denota a iteração, f é uma função que depende do processo iterativo empregado e $\phi^{(0)}$ é um valor inicial que deve ser fornecido. Como critérios de parada das iterações, os mais comuns são:

- (i) $\max_{1 \leq u \leq m} |\phi_u^{(w+1)} - \phi_u^{(w)}| < \epsilon$,
- (ii) $\frac{\|\phi^{(w+1)} - \phi^{(w)}\|}{\|\phi^{(w)}\|} < \epsilon$,
- (iii) $|l(\phi^{(w+1)}) - l(\phi^{(w)})| < \epsilon$ e
- (iv) combinações de (i) e (iii) ou de (ii) e (iii),

onde ϵ é a precisão desejada.

Jennrich e Schluchter (1986) sugerem três algoritmos para obtenção das estimativas de máxima verossimilhança: Newton-Raphson, “Scoring” de Fisher e EM híbrido com “Scoring”. Laird e Ware (1982) sugerem a aplicação do algoritmo EM através da formulação de modelos de efeitos aleatórios. Discutiremos a aplicação e alguns aspectos computacionais relativos a esses processos iterativos nas próximas subseções.

3.2.1 Algoritmos de Newton-Raphson e Scoring de Fisher

O algoritmo de Newton-Raphson (ver, por exemplo, Bard (1974)) é um processo iterativo utilizado para encontrar pontos de máximo de uma função. Mais especificamente, na iteração $(w + 1)$ do algoritmo, a função $l(\phi)$ é aproximada até o termo de segunda ordem da sua expansão em séries de Taylor ao redor de pontos $\phi^{(w)}$, ou seja:

$$l(\phi) \doteq l(\phi^{(w)}) + s(\phi^{(w)})'(\phi - \phi^{(w)}) + \frac{1}{2}(\phi - \phi^{(w)})' \mathbf{H}(\phi^{(w)})(\phi - \phi^{(w)}), \quad (3.33)$$

onde $s(\phi^{(w)})$ e $\mathbf{H}(\phi^{(w)})$ são dados por:

$$\underset{(m \times 1)}{s(\phi)} = \left[\left\langle \frac{\partial l(\phi)}{\partial \phi_u} \right\rangle_u \right]$$

e

$$\underset{(m \times m)}{\mathbf{H}(\phi)} = \left[\left\langle \frac{\partial^2 l(\phi)}{\partial \phi_u \partial \phi_v} \right\rangle_{uv} \right],$$

calculados no ponto $\phi = \phi^{(w)}$.

Derivando a expressão (3.33) em relação a ϕ , temos a seguinte aproximação para $s(\phi)$:

$$s(\phi) \doteq s(\phi^{(w)}) + \mathbf{H}(\phi^{(w)})(\phi - \phi^{(w)}). \quad (3.34)$$

Como estamos interessados em encontrar a solução $\hat{\phi}$ tal que $s(\phi) = 0$, basta igualar a expressão (3.34) a zero, fazer $\phi = \phi^{(w+1)}$ e resolver esse sistema em $\phi^{(w+1)}$, obtendo assim uma aproximação $\phi^{(w+1)}$ para $\hat{\phi}$. As sequências dessas aproximações definem o processo iterativo de Newton-Raphson, dado por:

$$\phi^{(w+1)} = \phi^{(w)} - \mathbf{H}^{-1}(\phi^{(w)})s(\phi^{(w)}) \quad \text{para } w = 0, 1, 2, \dots \quad (3.35)$$

Uma modificação do algoritmo de Newton-Raphson que pode ser utilizada quando queremos encontrar estimativas de máxima verossimilhança é o algoritmo “Scoring” de Fisher (ver, por exemplo, Jennrich e Schluchter (1986)). Esse algoritmo consiste em substituir na expressão (3.35) $\mathbf{H}(\phi^{(w)})$ por $-\mathbf{I}(\phi^{(w)})$, onde $\mathbf{I}(\phi) = -E[\mathbf{H}(\phi)]$ é a matriz de

informação de Fisher. Assim, o algoritmo “Scoring” de Fisher pode ser definido a partir de:

$$\phi^{(w+1)} = \phi^{(w)} + \mathbf{I}^{-1}(\phi^{(w)})\mathbf{s}(\phi^{(w)}) \quad \text{para } w = 0, 1, 2, \dots \quad (3.36)$$

Mais especificamente, a $(w+1)$ -ésima iteração do algoritmo de Newton-Raphson para obtenção estimativas de máxima verossimilhança é da forma:

$$\begin{bmatrix} \alpha^{(w+1)} \\ \theta^{(w+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(w)} \\ \theta^{(w)} \end{bmatrix} - \begin{bmatrix} \mathbf{H}_{\alpha\alpha} & \mathbf{H}_{\alpha\theta} \\ \mathbf{H}_{\theta\alpha} & \mathbf{H}_{\theta\theta} \end{bmatrix}_{\substack{\alpha=\alpha^{(w)} \\ \theta=\theta^{(w)}}}^{-1} \begin{bmatrix} \mathbf{s}_{\alpha} \\ \mathbf{s}_{\theta} \end{bmatrix}_{\substack{\alpha=\alpha^{(w)} \\ \theta=\theta^{(w)}}} \quad (3.37)$$

onde as expressões para $\mathbf{s}_{\alpha}$, $\mathbf{s}_{\theta}$, $\mathbf{H}_{\alpha\theta}$ e $\mathbf{H}_{\theta\theta}$ são dadas por (3.3) a (3.7) respectivamente.

No caso do algoritmo de Newton-Raphson, a $(w+1)$ -ésima iteração para obtenção das estimativas de máxima verossimilhança restrita é dada por:

$$\theta^{(w+1)} = \theta^{(w)} - \mathbf{H}_R^{-1}(\theta^{(w)})\mathbf{s}_R(\theta^{(w)}), \quad (3.38)$$

onde as expressões para $\mathbf{s}_R$ e $\mathbf{H}_R$ são dadas por (3.23) e (3.24) respectivamente. Quando o critério de convergência desejado é atingido, podemos calcular $\hat{\alpha}$ usando (3.12) com a matriz Σ_i substituída por $\Sigma_i(\hat{\theta}_R)$, $i = 1, \dots, n$.

Na $(w+1)$ -ésima iteração do algoritmo “Scoring” de Fisher para obtenção das estimativas de máxima verossimilhança temos:

$$\begin{bmatrix} \alpha^{(w+1)} \\ \theta^{(w+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(w)} \\ \theta^{(w)} \end{bmatrix} + \begin{bmatrix} \mathbf{I}_{\alpha\alpha} & \mathbf{I}_{\alpha\theta} \\ \mathbf{I}_{\theta\alpha} & \mathbf{I}_{\theta\theta} \end{bmatrix}_{\substack{\alpha=\alpha^{(w)} \\ \theta=\theta^{(w)}}}^{-1} \begin{bmatrix} \mathbf{s}_{\alpha} \\ \mathbf{s}_{\theta} \end{bmatrix}_{\substack{\alpha=\alpha^{(w)} \\ \theta=\theta^{(w)}}} \quad (3.39)$$

com $\mathbf{s}_{\alpha}$ e $\mathbf{s}_{\theta}$ dados por (3.3) e (3.4) e $\mathbf{I}_{\alpha}$, $\mathbf{I}_{\alpha\theta}$ e $\mathbf{I}_{\theta\theta}$ dados pelas expressões (3.14), (3.15) e (3.16) respectivamente. No entanto, como $\mathbf{I}_{\alpha\theta} = \mathbf{I}'_{\theta\alpha} = \mathbf{0}$ é possível simplificar (3.39) de forma a calcular $\alpha^{(w)}$ e $\theta^{(w)}$ resolvendo equações separadas reduzindo, assim, a dimensão da matriz $\mathbf{I}^{(w)}$ a ser invertida. Dessa maneira, também podemos calcular os elementos $(\mathbf{y}_i - \mathbf{X}_i\alpha)$ de (3.4) usando o novo valor $\alpha^{(w+1)}$ em vez de $\alpha^{(w)}$. Consequentemente, o algoritmo “Scoring” de Fisher pode ser definido a partir de:

$$\alpha^{(w+1)} = \left(\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1(w)} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1(w)} \mathbf{y}_i \right) \quad (3.40)$$

e

$$\boldsymbol{\theta}^{(w+1)} = \boldsymbol{\theta}^{(w)} + \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\alpha}^{(w)}, \boldsymbol{\theta}^{(w)}) \mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{\alpha}^{(w+1)}, \boldsymbol{\theta}^{(w)}). \quad (3.41)$$

Utilizando a idéia de Szatrowski (1983), como $\boldsymbol{\Sigma}_i = \sum_{g=1}^n \theta_g \mathbf{G}_{ig}$ a expressão (3.41) pode ainda ser reduzida a

$$\boldsymbol{\theta}^{(w+1)} = \left[\left\langle \sum_{i=1}^n \text{tr} \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{G}_{ih} \right\rangle_{gh} \right]^{-1} \left[\left\langle \sum_{i=1}^n \text{tr} \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{r}_i^{(w+1)} \mathbf{r}_i^{(w+1)'} \right\rangle_g \right], \quad (3.42)$$

com $\mathbf{r}_i^{(w+1)} = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w+1)}$. Notemos que no algoritmo “Scoring” assim definido as equações (3.40) e (3.42) são similares a (3.10) e (3.13).

No caso de máxima verossimilhança restrita, a $(w+1)$ -ésima iteração do algoritmo “Scoring” de Fisher é dada por:

$$\boldsymbol{\theta}^{(w+1)} = \boldsymbol{\theta}^{(w)} + \mathbf{I}_R^{-1}(\boldsymbol{\theta}^{(w)}) \mathbf{s}_R(\boldsymbol{\theta}^{(w)}), \quad (3.43)$$

onde as expressões para os elementos de $\mathbf{s}_R$ e $\mathbf{I}_R$ são dadas por (3.23) e (3.26) respectivamente. Quando o critério de convergência desejado é atingido, estimamos $\boldsymbol{\alpha}$ usando (3.12) com as matrizes $\boldsymbol{\Sigma}_i$'s substituídas por $\boldsymbol{\Sigma}_i(\hat{\boldsymbol{\theta}}_R)$.

3.2.2 Algoritmo EM

O algoritmo EM, apresentado em Dempster, Laird e Rubin (1977), é um processo iterativo para calcular estimativas de máxima verossimilhança em situações que possam ser consideradas como tendo dados incompletos ou não observáveis. A idéia básica do EM é considerar os dados observáveis $\mathbf{Y}$ como uma parte de um sistema maior de dados completos $\mathbf{U} = (\mathbf{Y}, \mathbf{W})$, onde $\mathbf{W}$ são os dados incompletos ou não observáveis. Cada iteração do algoritmo EM consiste de dois passos, denominados passo E (esperança) e passo M (maximização). O passo M corresponde ao processo de estimação de máxima verossimilhança que seria realizado se os dados fossem completos; mas como os dados são incompletos a maximização no passo M só pode ser realizada numa alternativa da verossimilhança que

é calculada a partir do passo E. Quando a distribuição dos dados completos é da *família exponencial regular* o passo E consiste na obtenção da esperança das estatísticas suficientes para os parâmetros contidos em ϕ baseadas nos dados completos, condicionada aos dados observados e às estimativas em vigor de ϕ (dadas por $\phi^{(w)}$). A teoria geral do EM garante que a cada iteração a verossimilhança não diminui.

Para utilizar o algoritmo EM precisamos definir o sistema de dados completos U , encontrar as estatísticas suficientes para ϕ baseadas nos dados completos U e fornecer estimativas iniciais $\phi^{(0)}$ para ϕ . Para dados completos pertencentes a *família exponencial regular*, a $(w + 1)$ -ésima iteração do algoritmo EM é dada por:

Passo E: Estimação das estatísticas suficientes $t(U)$ a partir de

$$t^{(w)} = E[t(U) \mid Y = y, \phi = \phi^{(w)}]. \quad (3.44)$$

Passo M: Determinação de $\phi^{(w+1)}$ como sendo a solução em ϕ das equações de verossimilhança dos dados completos:

$$E[t(U) \mid \phi] = t^{(w)}. \quad (3.45)$$

No caso geral o passo M pode requerer um processo iterativo adicional para resolução das equações (3.45). Um processo iterativo geralmente empregado nesse caso é o algoritmo “Scoring”, como sugerido por Jennrich e Schluchter (1986).

Dadas essas considerações, o algoritmo EM pode ser utilizado segundo dois enfoques.

O primeiro enfoque, considerado, por exemplo, por Szatrowski (1983) e Jennrich e Schluchter (1986), é dirigido para os *modelos de dados incompletos* discutidos na Seção 2.1. Nesse contexto, o sistema de dados completos no EM corresponde aos vetores $y_i^* = (y_i, y_i^+)'$ (ou $e_i^* = (e_i, e_i^+)'$, ver Seção 2.1) onde os elementos associados a y_i são observados e aqueles associados a y_i^+ não são observados. Em algumas situações, como no Exemplo 1.3, não podemos usar o algoritmo EM na forma sugerida por esses autores pois ela envolve a estimação de uma matriz de covariância não estruturada que, nesse caso, não pode ser

efetuada; isto ocorre nos casos em que as ocasiões de observação variam irregularmente de unidade experimental para unidade experimental ou quando as matrizes Z_i 's do modelo de efeitos aleatórios contêm covariáveis específicas das unidades experimentais a que estão associadas.

Outro enfoque utilizado para a aplicação do algoritmo EM surge no contexto de modelos de efeitos aleatórios. Nesse caso, podemos considerar o sistema de dados completos como sendo composto pelas respostas efetivamente observadas $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$ mais os efeitos aleatórios, $\mathbf{b} = (\mathbf{b}'_1, \dots, \mathbf{b}'_n)'$ e os termos de erro do modelo (2.16) $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_n)'$, ambos não observáveis. Laird (1982), Laird e Ware (1982) e Laird et al. (1987) usam esse tipo de abordagem. Dempster, Rubin e Tsutakawa (1981), Andrade e Helms (1984a,b) e Fairclough e Helms (1986) também consideram abordagens similares. Concentraremos a atenção nas propostas de Laird e Ware (1982) e Laird et al. (1987) pois as expressões resultantes também podem ser utilizadas para o ajuste de modelos de matrizes de covariância não estruturadas. A aplicação do algoritmo EM proposta por Jennrich e Schluchter (1986) pode ser considerada como caso particular.

No caso de máxima verossimilhança consideramos o sistema de dados completos $(\mathbf{y}', \mathbf{b}', \mathbf{e}')$, onde $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$ são os dados observáveis, $\mathbf{b} = (\mathbf{b}'_1, \dots, \mathbf{b}'_n)'$ e $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_n)'$ são os dados não observáveis. A distribuição conjunta desse sistema de dados completos é:

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \\ \mathbf{e}_i \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{X}_i \boldsymbol{\alpha} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2 \mathbf{I}_{t_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i & \mathbf{Z}_i \mathbf{D} & \sigma^2 \mathbf{I}_{t_i} \\ \mathbf{D} \mathbf{Z}'_i & \mathbf{D} & \mathbf{0} \\ \sigma^2 \mathbf{I}_{t_i} & \mathbf{0} & \sigma^2 \mathbf{I}_{t_i} \end{pmatrix} \right] \quad (3.46)$$

para $i = 1, \dots, n$. Dado o modelo de efeitos aleatórios $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i$, onde os $\mathbf{b}_i$'s e $\mathbf{e}_i$'s são independentes, a função de verossimilhança dos dados completos pode ser escrita na forma:

$$\begin{aligned} L^*(\mathbf{y}, \mathbf{b}, \mathbf{e}; \boldsymbol{\alpha}, \boldsymbol{\theta}) &= \prod_{i=1}^n f(\mathbf{y}_i | \mathbf{e}_i, \mathbf{b}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) f(\mathbf{b}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) f(\mathbf{e}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) \\ &= \prod_{i=1}^n f(\mathbf{b}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) f(\mathbf{e}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) \end{aligned}$$

$$= K(\sigma^2)^{-\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i \right] |\mathbf{D}|^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \text{tr} \mathbf{D}^{-1} \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i' \right],$$

onde $N = \sum_{i=1}^n t_i$ e K é uma constante que não depende de $\mathbf{D}$, σ^2 e $\boldsymbol{\alpha}$. Notemos que as estatísticas suficientes para σ^2 e $\mathbf{D}$ são $\mathbf{t}_1 = \sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i$ e $\mathbf{T}_2 = \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i'$ respectivamente. Assim, considerando o sistema de dados completos, os estimadores de máxima verossimilhança de σ^2 e $\mathbf{D}$ com $\mathbf{D}$ não estruturada¹ (ver Mardia et al. (1979), Cap 4) são:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i}{N} = \frac{\mathbf{t}_1}{N} \quad (3.47)$$

e

$$\hat{\mathbf{D}} = \frac{\sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i'}{n} = \frac{\mathbf{T}_2}{n}. \quad (3.48)$$

O estimador de máxima verossimilhança de $\boldsymbol{\alpha}$ pode ser obtido considerando a equação (3.10). Assim, temos

$$\hat{\boldsymbol{\alpha}} = \left(\sum_{i=1}^n \mathbf{X}_i' \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i' \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{y}_i \right). \quad (3.49)$$

As equações (3.47), (3.48) e (3.49) definem o passo M do algoritmo EM nessa aplicação.

No passo E necessitamos calcular as esperanças condicionais de $\mathbf{t}_1$ e $\mathbf{T}_2$ dado $\mathbf{y}$ para $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$ fixados. Assim, para $\mathbf{t}_1$ temos que

$$\begin{aligned} E[\mathbf{t}_1 | \mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta}] &= E \left[\sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i | \mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta} \right] \\ &= \sum_{i=1}^n E[\mathbf{e}_i' \mathbf{e}_i | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}]. \end{aligned}$$

Usando o fato de que a esperança de uma forma quadrática $\mathbf{u}' \mathbf{Q} \mathbf{u}$ é

$$E(\mathbf{u}' \mathbf{Q} \mathbf{u}) = E(\mathbf{u}') \mathbf{Q} E(\mathbf{u}) + \text{tr}[\mathbf{Q} \text{Var}(\mathbf{u})],$$

temos:

$$E[\mathbf{t}_1 | \mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta}] = \sum_{i=1}^n E[\mathbf{e}_i' | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}] E[\mathbf{e}_i | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}] + \sum_{i=1}^n \text{tr} \text{Var}[\mathbf{e}_i | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}].$$

¹veremos mais adiante como relaxar essa suposição.

Para T_2 , temos:

$$\begin{aligned} E[T_2 | y; \alpha, \theta] &= E \left[\sum_{i=1}^n b_i b_i' | y; \alpha, \theta \right] \\ &= \sum_{i=1}^n E[b_i b_i' | y_i; \alpha, \theta]. \end{aligned}$$

Lembrando que para um vetor aleatório u , $E(uu') = E(u)E(u') + \text{Var}(u)$, temos:

$$E[T_2 | y; \alpha, \theta] = \sum_{i=1}^n E[b_i | y_i; \alpha, \theta] E[b_i' | y_i; \alpha, \theta] + \sum_{i=1}^n \text{Var}[b_i | y_i; \alpha, \theta].$$

As esperanças e variâncias condicionais de e_i e b_i dado y_i são mais facilmente calculadas a partir da distribuição conjunta de y_i , b_i e e_i dada por (3.46), embora através da formulação do modelo de efeitos aleatórios em 2 estágios vista na Seção 2.2, elas também possam ser calculadas através do Teorema de Bayes, como indicado em Lindley e Smith (1972). Por outro lado, lembrando que, se

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \right],$$

então, $x_2 | x_1 \sim N[\mu_2 + V_{21}V_{11}^{-1}(x_1 - \mu_1), V_{22} - V_{21}V_{11}^{-1}V_{12}]$, podemos concluir a partir de (3.46) que:

- (i) $E[b_i | y_i; \alpha, \theta] = DZ_i'(\sigma^2 I_{t_i} + Z_i D Z_i')^{-1}(y_i - X_i \alpha)$
 $= DZ_i' \Sigma_i^{-1}(y_i - X_i \alpha),$
- (ii) $\text{Var}[b_i | y_i; \alpha, \theta] = D - DZ_i'(\sigma^2 I_{t_i} + Z_i D Z_i')^{-1} Z_i D$
 $= D - DZ_i' \Sigma_i^{-1} Z_i D,$
- (iii) $E[e_i | y_i; \alpha, \theta] = \sigma^2 I_{t_i}(\sigma^2 I_{t_i} + Z_i D Z_i')^{-1}(y_i - X_i \alpha)$
 $= \sigma^2 \Sigma_i^{-1}(y_i - X_i \alpha) e$
- (iv) $\text{Var}[e_i | y_i; \alpha, \theta] = \sigma^2 I_{t_i} - \sigma^2 I_{t_i}(\sigma^2 I_{t_i} + Z_i D Z_i')^{-1} I_{t_i} \sigma^2$
 $= \sigma^2 I_{t_i} - (\sigma^2)^2 \Sigma_i^{-1}.$

Alternativamente, como $e_i = y_i - X_i \alpha - Z_i b_i$, temos:

$$\begin{aligned} E[e_i | y_i; \alpha, \theta] &= E[y_i - X_i \alpha - Z_i b_i | y_i; \alpha, \theta] \\ &= y_i - X_i \alpha - Z_i E[b_i | y_i; \alpha, \theta]. \end{aligned}$$

Então, no caso em que $\mathbf{D}$ é não estruturada, a $(w + 1)$ -ésima iteração do algoritmo EM para calcular estimativas de máxima verossimilhança é dada por:

$$\begin{aligned} \sigma^{2(w+1)} &= \frac{1}{N} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)} - \mathbf{Z}_i \mathbf{b}_i^{(w)})' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)} - \mathbf{Z}_i \mathbf{b}_i^{(w)}) \\ &\quad + \frac{\sigma^{2(w)}}{N} \sum_{i=1}^n \text{tr} [\mathbf{I}_{t_i} - \sigma^{2(w)} \boldsymbol{\Sigma}_i^{-1(w)}] \end{aligned} \quad (3.50)$$

$$\mathbf{D}^{(w+1)} = \mathbf{D}^{(w)} + \frac{1}{n} \sum_{i=1}^n \mathbf{b}_i^{(w)} \mathbf{b}_i^{(w)'} - \frac{\mathbf{D}^{(w)}}{n} \left(\sum_{i=1}^n \mathbf{Z}_i' \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{Z}_i \right) \mathbf{D}^{(w)} \quad (3.51)$$

$$\boldsymbol{\alpha}^{(w+1)} = \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{y}_i \right) \quad (3.52)$$

onde $N = \sum_{i=1}^n t_i$, $\boldsymbol{\Sigma}_i^{-1(w)} = [\sigma^{2(w)} \mathbf{I}_{t_i} + \mathbf{Z}_i \mathbf{D}^{(w)} \mathbf{Z}_i']^{-1}$ e $\mathbf{b}_i^{(w)} = E[\mathbf{b}_i \mid \mathbf{y}_i; \boldsymbol{\alpha}^{(w)}, \boldsymbol{\theta}^{(w)}] = \mathbf{D}^{(w)} \mathbf{Z}_i' \boldsymbol{\Sigma}_i^{-1(w)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)})$.

A aplicação do algoritmo EM no caso de máxima verossimilhança restrita pode ser mais facilmente compreendida considerando a seguinte formulação “Bayesiana” do modelo de efeitos aleatórios em dois estágios apresentado na Seção 2.2. O Estágio 1 permanece inalterado, de modo que condicionalmente a $\boldsymbol{\alpha}$ e $\mathbf{b}_i$, o vetor $\mathbf{y}_i$ é Normal com média $\mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i$ e variância $\sigma^2 \mathbf{I}_{t_i}$, para $i = 1, \dots, n$. No Estágio 2, $\boldsymbol{\alpha}, \mathbf{b}_1, \dots, \mathbf{b}_n$ são considerados independentes e normalmente distribuídos com vetores de médias iguais a $\mathbf{0}$ e $\text{Var}(\boldsymbol{\alpha}) = \boldsymbol{\Gamma}$, $\text{Var}(\mathbf{b}_i) = \mathbf{D}$ e $\text{Cov}(\boldsymbol{\alpha}, \mathbf{b}_i) = \mathbf{0}$ para $i = 1, \dots, n$. Dessa maneira, a distribuição marginal de $\mathbf{y}_i$, para $i = 1, \dots, n$, nessa formulação “Bayesiana” é dada por:

$$\mathbf{y}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{t_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \mathbf{X}_i \boldsymbol{\Gamma} \mathbf{X}_i') \quad (3.53)$$

Denotando ainda por $\boldsymbol{\theta}$ o vetor formado pelos parâmetros desconhecidos de $\mathbf{D}$ e σ^2 , Laird e Ware (1982) salientam que essa distribuição marginal de $\mathbf{y}_i$ nos dá informação somente sobre $\boldsymbol{\theta}$ e não sobre $\boldsymbol{\Gamma}$. Portanto, uma estratégia razoável é fazer $\boldsymbol{\Gamma}^{-1} \rightarrow \mathbf{0}$, indicando uma vaga informação a priori sobre $\boldsymbol{\alpha}$ e estimar $\boldsymbol{\theta}$ através da maximização da verossimilhança baseada na distribuição marginal limite de $\mathbf{y}_i, i = 1, \dots, n$ dada por (3.53) quando $\boldsymbol{\Gamma}^{-1} \rightarrow \mathbf{0}$. Harville (1974) demonstrou que essa função de verossimilhança limite é equivalente à função de verossimilhança restrita dada por (3.21). Assim, as estimativas de $\boldsymbol{\theta}$ obtidas nesse contexto são as estimativas de máxima verossimilhança restrita.

Dessa maneira, para obter estimativas de máxima verossimilhança restrita utilizando o algoritmo EM, podemos considerar que o sistema de dados completos é $(\mathbf{y}, \boldsymbol{\alpha}, \mathbf{b}, \mathbf{e})$ cuja distribuição conjunta é dada por:

$$\begin{pmatrix} \mathbf{y} \\ \boldsymbol{\alpha} \\ \mathbf{b} \\ \mathbf{e} \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2 \mathbf{I}_N + \mathbf{ZHZ}' + \mathbf{XGX}' & \mathbf{XG} & \mathbf{XH} & \sigma^2 \mathbf{I}_N \\ \mathbf{GX}' & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{HZ}' & \mathbf{0} & \mathbf{H} & \mathbf{0} \\ \sigma^2 \mathbf{I}_N & \mathbf{0} & \mathbf{0} & \sigma^2 \mathbf{I}_N \end{pmatrix} \right], \quad (3.54)$$

onde $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)'$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$, $\mathbf{H} = \mathbf{I}_n \otimes \mathbf{D}$ e $\mathbf{G}^{-1} \rightarrow \mathbf{0}$.

Lembrando que dado $\boldsymbol{\alpha}, \mathbf{b}$ e $\mathbf{e}$, $\mathbf{y}$ está completamente especificado e $\boldsymbol{\alpha}, \mathbf{b}$ e $\mathbf{e}$ são independentes, a função de verossimilhança associada aos dados completos pode ser escrita como:

$$\begin{aligned} L_R^*(\mathbf{y}, \boldsymbol{\alpha}, \mathbf{b}, \mathbf{e}; \boldsymbol{\theta}) &= f(\mathbf{y} | \boldsymbol{\alpha}, \mathbf{b}, \mathbf{e}; \boldsymbol{\theta}) f(\boldsymbol{\alpha}; \boldsymbol{\theta}) f(\mathbf{b}; \boldsymbol{\theta}) f(\mathbf{e}; \boldsymbol{\theta}) \\ &= K |\mathbf{G}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \boldsymbol{\alpha}' \mathbf{G}^{-1} \boldsymbol{\alpha} \right] |\mathbf{D}|^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \text{tr} \mathbf{D}^{-1} \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i' \right] \times \\ &\quad \times (\sigma^2)^{-\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i \right]. \end{aligned}$$

Como $\mathbf{G}^{-1} \rightarrow \mathbf{0}$ e supondo que $|\mathbf{G}|^{-1} \neq \mathbf{0}$ (Dempster, Rubin e Tsutakawa (1981)), temos:

$$L_R^*(\mathbf{y}, \boldsymbol{\alpha}, \mathbf{b}, \mathbf{e}; \boldsymbol{\theta}) = K (\sigma^2)^{-\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i \right] |\mathbf{D}|^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \text{tr} \mathbf{D}^{-1} \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i' \right], \quad (3.55)$$

onde K é uma constante que não depende de $\boldsymbol{\theta}$.

Analogamente ao método de máxima verossimilhança, as estatísticas suficientes para σ^2 e $\mathbf{D}$ baseadas nos dados completos são respectivamente $\mathbf{t}_1 = \sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i$ e $\mathbf{T}_2 = \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i'$. Também neste caso, os estimadores de máxima verossimilhança de σ^2 e $\mathbf{D}$ são dados por:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i}{N} = \frac{\mathbf{t}_1}{N} \quad (3.56)$$

e

$$\hat{\mathbf{D}} = \frac{\sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i'}{n} = \frac{\mathbf{T}_2}{n}. \quad (3.57)$$

As equações (3.56) e (3.57) definem o passo M do algoritmo EM.

No Passo E necessitamos calcular a esperança condicional de t_1 e T_2 dado y com θ fixado. Essas esperanças não envolvem o condicionamento em α já que a função que vamos maximizar não depende de α . Assim,

$$\begin{aligned} E[t_1 | y; \theta] &= E \left[\sum_{i=1}^n e_i' e_i | y; \theta \right] \\ &= \sum_{i=1}^n E[e_i' e_i | y; \theta] \\ &= \sum_{i=1}^n E[e_i' | y; \theta] E[e_i | y; \theta] + \sum_{i=1}^n \text{tr} \text{Var}[e_i | y; \theta] \end{aligned}$$

e

$$\begin{aligned} E[T_2 | y; \theta] &= E \left[\sum_{i=1}^n b_i b_i' | y; \theta \right] \\ &= \sum_{i=1}^n E[b_i b_i' | y; \theta] \\ &= \sum_{i=1}^n E[b_i | y; \theta] E[b_i' | y; \theta] + \sum_{i=1}^n \text{Var}[b_i | y; \theta]. \end{aligned}$$

Para calcular as esperanças e variâncias condicionais de e_i e b_i dado y , vamos considerar a distribuição conjunta de (y, α, b, e) dada por (3.54) e obter primeiramente as distribuições condicionais de α , b e e dado y com θ fixado. Uma outra maneira de calcular essas esperanças e variâncias condicionais é através da formulação em dois estágios e usar o Teorema de Bayes, como indicado em Lindley e Smith (1972).

Usando resultados conhecidos sobre a distribuição condicional de variáveis aleatórias normais, temos:

- (i) $E[\alpha | y; \theta] = \Gamma X' (\sigma^2 I_N + Z H Z' + X \Gamma X')^{-1} y$,
- (ii) $\text{Var}[\alpha | y; \theta] = \Gamma - \Gamma X' (\sigma^2 I_N + Z H Z' + X \Gamma X')^{-1} X \Gamma$,
- (iii) $E[b | y; \theta] = H Z' (\sigma^2 I_N + Z H Z' + X \Gamma X')^{-1} y$,
- (iv) $\text{Var}[b | y; \theta] = H - H Z' (\sigma^2 I_N + Z H Z' + X \Gamma X')^{-1} Z H$,
- (v) $E[e | y; \theta] = \sigma^2 (\sigma^2 I_N + Z H Z' + X \Gamma X')^{-1} y$ e

$$(vi) \text{Var}[\mathbf{e} | \mathbf{y}; \boldsymbol{\theta}] = \sigma^2 \mathbf{I}_N - (\sigma^2)^2 (\sigma^2 \mathbf{I}_N + \mathbf{ZHZ}' + \mathbf{X}\boldsymbol{\Gamma}\mathbf{X}')^{-1}.$$

Para incorporar no procedimento o fato de que $\boldsymbol{\Gamma}^{-1} \rightarrow \mathbf{0}$ precisamos das seguintes identidades matriciais:

$$(a) (\mathbf{C}_1 + \mathbf{A}_1 \mathbf{C}_2 \mathbf{A}_1')^{-1} = \mathbf{C}_1^{-1} - \mathbf{C}_1^{-1} \mathbf{A}_1 (\mathbf{A}_1' \mathbf{C}_1^{-1} \mathbf{A}_1 + \mathbf{C}_2^{-1})^{-1} \mathbf{A}_1' \mathbf{C}_1^{-1} \text{ e}$$

$$(b) \mathbf{C}_2 \mathbf{A}_1' (\mathbf{C}_1 + \mathbf{A}_1 \mathbf{C}_2 \mathbf{A}_1')^{-1} = (\mathbf{A}_1' \mathbf{C}_1^{-1} \mathbf{A}_1 + \mathbf{C}_2^{-1})^{-1} \mathbf{A}_1' \mathbf{C}_1^{-1},$$

onde $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{A}_1$ e $\mathbf{A}_2$ são matrizes de ordens compatíveis para as operações consideradas. O resultado (a) está provado em Henderson et al. (1959), por exemplo. O resultado (b) pode ser obtido de (a), pré-multiplicando-se (a) por $\mathbf{C}_2 \mathbf{A}_1'$.

Fazendo $\mathbf{C}_1 = \sigma^2 \mathbf{I}_N + \mathbf{ZHZ}'$, $\mathbf{C}_2 = \boldsymbol{\Gamma}$ e $\mathbf{A}_1 = \mathbf{X}$ em (i), (iii), (iv), (v) e (vi); e $\mathbf{C}_1^{-1} = \boldsymbol{\Gamma}$, $\mathbf{A}_1 = \mathbf{X}'$ e $\mathbf{C}_2^{-1} = \sigma^2 \mathbf{I}_N + \mathbf{ZHZ}'$ em (ii), com $\boldsymbol{\Gamma}^{-1} \rightarrow \mathbf{0}$ em todos os casos, obtemos:

$$(i) \begin{aligned} E[\boldsymbol{\alpha} | \mathbf{y}; \boldsymbol{\theta}] &= [\mathbf{X}'(\sigma^2 \mathbf{I}_N + \mathbf{ZHZ}')^{-1} \mathbf{X}]^{-1} \mathbf{X}'(\sigma^2 \mathbf{I}_N + \mathbf{ZHZ}')^{-1} \mathbf{y} \\ &= \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i \right) \\ &= \hat{\boldsymbol{\alpha}} \end{aligned}$$

$$(ii) \begin{aligned} \text{Var}[\boldsymbol{\alpha} | \mathbf{y}; \boldsymbol{\theta}] &= [\mathbf{X}'(\sigma^2 \mathbf{I}_N + \mathbf{ZHZ}')^{-1} \mathbf{X}]^{-1} \\ &= \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \end{aligned}$$

$$(iii) \begin{aligned} E[\mathbf{b} | \mathbf{y}; \boldsymbol{\theta}] &= \mathbf{HZ}'(\sigma^2 \mathbf{I}_N + \mathbf{ZHZ}')^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) \\ &= \begin{bmatrix} \mathbf{DZ}_1' \boldsymbol{\Sigma}_1^{-1}(\mathbf{y}_1 - \mathbf{X}_1 \hat{\boldsymbol{\alpha}}) \\ \vdots \\ \mathbf{DZ}_n' \boldsymbol{\Sigma}_n^{-1}(\mathbf{y}_n - \mathbf{X}_n \hat{\boldsymbol{\alpha}}) \end{bmatrix} = \begin{bmatrix} E[\mathbf{b}_1 | \mathbf{y}; \boldsymbol{\theta}] \\ \vdots \\ E[\mathbf{b}_n | \mathbf{y}; \boldsymbol{\theta}] \end{bmatrix} \end{aligned}$$

Fazendo $\mathbf{V} = \sigma^2 \mathbf{I}_N + \mathbf{ZHZ}'$, temos:

$$(iv) \text{Var}[\mathbf{b} | \mathbf{y}; \boldsymbol{\theta}] = \mathbf{H} - \mathbf{HZ}'[\mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}] \mathbf{ZH}.$$

Notemos que de $\text{Var}[\mathbf{b} | \mathbf{y}; \boldsymbol{\theta}]$ só nos interessam as matrizes de sua “diagonal”, dadas por:

$$\begin{aligned} \text{Var}[\mathbf{b}_i | \mathbf{y}; \boldsymbol{\theta}] &= \mathbf{D} - \mathbf{DZ}_i' [\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i (\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1}] \mathbf{Z}_i \mathbf{D} \\ &= \mathbf{D} - \mathbf{DZ}_i' \mathbf{P}_i \mathbf{Z}_i \mathbf{D}, \quad i = 1, \dots, n \end{aligned}$$

$$\begin{aligned}
 \text{(v)} \quad E[\mathbf{e} \mid \mathbf{y}; \boldsymbol{\theta}] &= \sigma^2 [\mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}] \mathbf{y} \\
 &= \begin{bmatrix} \sigma^2 \Sigma_1^{-1} (\mathbf{y}_1 - \mathbf{X}_1 \hat{\boldsymbol{\alpha}}) \\ \vdots \\ \sigma^2 \Sigma_n^{-1} (\mathbf{y}_n - \mathbf{X}_n \hat{\boldsymbol{\alpha}}) \end{bmatrix}
 \end{aligned}$$

$$\text{(vi)} \quad \text{Var}[\mathbf{e} \mid \mathbf{y}; \boldsymbol{\theta}] = \sigma^2 \mathbf{I}_N - (\sigma^2)^2 [\mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}],$$

cujas “diagonal” é formada pelas matrizes:

$$\begin{aligned}
 \text{Var}[\mathbf{e}_i \mid \mathbf{y}; \boldsymbol{\theta}] &= \sigma^2 \mathbf{I}_{t_i} - (\sigma^2)^2 [\Sigma_i^{-1} - \Sigma_i^{-1} \mathbf{X}_i (\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i' \Sigma_i^{-1}] \\
 &= \sigma^2 \mathbf{I}_{t_i} - (\sigma^2)^2 \mathbf{P}_i, \quad i = 1, \dots, n.
 \end{aligned}$$

Alternativamente, como $\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha} - \mathbf{Z}_i \mathbf{b}_i$, podemos usar:

$$E[\mathbf{e}_i \mid \mathbf{y}; \boldsymbol{\theta}] = \mathbf{y}_i - \mathbf{X}_i E[\boldsymbol{\alpha} \mid \mathbf{y}; \boldsymbol{\theta}] - \mathbf{Z}_i E[\mathbf{b}_i \mid \mathbf{y}; \boldsymbol{\theta}].$$

Assim, se $\mathbf{D}$ for uma matriz não estruturada, podemos combinar os passos E e M do algoritmo resultando que a $(w+1)$ -ésima iteração para calcular estimativas de máxima verossimilhança restrita é da forma:

$$\begin{aligned}
 \sigma^{2(w+1)} &= \frac{1}{N} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)} - \mathbf{Z}_i \mathbf{b}_i^{(w)})' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)} - \mathbf{Z}_i \mathbf{b}_i^{(w)}) \\
 &\quad + \frac{\sigma^{2(w)}}{N} \sum_{i=1}^n \text{tr}[\mathbf{I}_{t_i} - \sigma^{2(w)} \mathbf{P}_i^{(w)}]
 \end{aligned} \tag{3.58}$$

$$\mathbf{D}^{(w+1)} = \mathbf{D}^{(w)} + \frac{1}{n} \sum_{i=1}^n \mathbf{b}_i^{(w)} \mathbf{b}_i^{(w)'} - \frac{\mathbf{D}^{(w)}}{n} \left(\sum_{i=1}^n \mathbf{Z}_i' \mathbf{P}_i^{(w)} \mathbf{Z}_i \right) \mathbf{D}^{(w)} \tag{3.59}$$

onde

$$N = \sum_{i=1}^n t_i,$$

$$\boldsymbol{\alpha}^{(w)} = (\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1(w)} \mathbf{X}_i)^{-1} (\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1(w)} \mathbf{y}_i),$$

$$\Sigma_i^{-1(w)} = [\sigma^{2(w)} \mathbf{I}_{t_i} + \mathbf{Z}_i \mathbf{D}^{(w)} \mathbf{Z}_i']^{-1},$$

$$\mathbf{b}_i^{(w)} = E[\mathbf{b}_i \mid \mathbf{y}; \boldsymbol{\theta}^{(w)}] = \mathbf{D}^{(w)} \mathbf{Z}_i' \Sigma_i^{-1(w)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)})$$

e

$$\mathbf{P}_i^{(w)} = \Sigma_i^{-1(w)} - \Sigma_i^{-1(w)} \mathbf{X}_i (\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1(w)} \mathbf{X}_i)^{-1} \mathbf{X}_i' \Sigma_i^{-1(w)}.$$

Notemos que em cada iteração temos também o valor de $\boldsymbol{\alpha}^{(w)}$.

Um aspecto interessante relativo a aplicação dos algoritmos EM propostos por Laird et al. (1987) é a flexibilidade de ajustar outros modelos, além do modelo de efeitos aleatórios com $\mathbf{D}$ não estruturada. Por exemplo, podemos ajustar modelos de efeitos aleatórios com $\mathbf{D}$ dependendo do grupo de unidades experimentais, como em (2.21). Para tanto, basta fornecermos $\mathbf{D}^{(0)}$ diagonal em blocos e construir $\mathbf{b}_i^*$ como em (2.21); ficando garantido que os elementos de $\mathbf{D}^{(w)}$ fora das matrizes de sua diagonal permanecerão igual a 0 nas iterações subsequentes de (3.51) ou (3.59).

Quando especificamos $\mathbf{Z}_i = \mathbf{0}_{t_i}$ para $i = 1, \dots, n$ e um valor positivo qualquer para $\mathbf{D}$, podemos encontrar a estimativa de máxima verossimilhança de σ^2 ou de máxima verossimilhança restrita de σ^2 (que é equivalente a obtida em Regressão Linear).

É importante notarmos a forma das equações desses algoritmos EM, quando $\sigma^2 = 0$, $\mathbf{Z}_i = \mathbf{E}_i$ e $\mathbf{D}$ representa a matriz de covariância entre as observações realizadas na mesma unidade experimental, como em (2.19). Fazendo $\sigma^{2(0)} = 0$, temos $\mathbf{Z}_i \mathbf{b}_i^{(0)} = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(0)}$, e assim de (3.50) ou (3.58), $\sigma^{2(1)} = 0$. Vemos, então, que $\sigma^{2(w+1)} = 0$ para $w = 0, 1, 2, \dots$. Nesse caso, quando estimamos $\mathbf{D}$ por máxima verossimilhança, a equação (3.51) fica:

$$\mathbf{D}^{(w+1)} = \frac{1}{n} \sum_{i=1}^n \left(\mathbf{b}_i^{(w)} \mathbf{b}_i^{(w)'} + \mathbf{R}_i^{(w)} \right), \quad (3.60)$$

onde

$$\mathbf{b}_i^{(w)} = \begin{cases} \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)} & \text{se } t_i = T \\ \mathbf{D}^{(w)} \mathbf{Z}_i' \mathbf{D}_i^{-1(w)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)}) & \text{se } t_i < T \end{cases}$$

$$\mathbf{R}_i^{(w)} = \begin{cases} \mathbf{0} & \text{se } t_i = T \\ \mathbf{D}^{(w)} (\mathbf{I} - \mathbf{Z}_i' \mathbf{D}_i^{-1(w)} \mathbf{Z}_i \mathbf{D}^{(w)}) & \text{se } t_i < T \end{cases}$$

e

$$\mathbf{D}_i^{(w)} = \mathbf{Z}_i \mathbf{D}^{(w)} \mathbf{Z}_i'.$$

Se $t_i = T$ para todo i (nenhuma observação incompleta), o algoritmo se reduz a IRLS (mínimos quadrados ponderados iterativamente):

$$\mathbf{D}^{(w+1)} = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)})'. \quad (3.61)$$

Quando estimamos $\mathbf{D}$ por máxima verossimilhança restrita, temos:

$$\mathbf{D}^{(w+1)} = \frac{1}{n} \sum_{i=1}^n \left(\mathbf{b}_i^{(w)} \mathbf{b}_i^{(w)'} + \mathbf{R}_i^{(w)} + \mathbf{T}_i^{(w)} \right), \quad (3.62)$$

onde $\mathbf{b}_i^{(w)}$ e $\mathbf{R}_i^{(w)}$ são como antes, e

$$\mathbf{T}_i^{(w)} = \mathbf{D}^{(w)} \mathbf{Z}_i' \mathbf{D}_i^{-1(w)} \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i' \mathbf{D}_i^{-1(w)} \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{D}_i^{-1(w)} \mathbf{Z}_i \mathbf{D}^{(w)}.$$

Note que se $t_i = T$ para todo i , então, $\mathbf{Z}_i = \mathbf{I}$, $\mathbf{D}_i^{(w)} = \mathbf{D}$, $\mathbf{R}_i^{(w)} = \mathbf{0}$ e

$$\mathbf{T}_i^{(w)} = \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i' \mathbf{D}^{-1(w)} \mathbf{X}_i \right)^{-1} \mathbf{X}_i'.$$

Dessa maneira, a versão para máxima verossimilhança restrita de IRLS fica:

$$\mathbf{D}^{(w+1)} = \frac{1}{n} \sum_{i=1}^n \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)})' + \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i' \mathbf{D}^{-1(w)} \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \right]. \quad (3.63)$$

Fórmulas análogas a (3.60) e (3.62) são encontradas nos algoritmos EM propostos por Jennrich e Schluchter (1986), que são aplicáveis quando $\mathbf{D}$ é não estruturada. Se $\mathbf{D}$ for estruturada ($\mathbf{D} = \mathbf{D}(\boldsymbol{\theta})$), a forma geral do passo M deve ser modificada, pois (3.60) e (3.62) não valem mais. Nessa situação, Jennrich e Schluchter (1986) sugerem a realização de um passo "Scoring" no passo M para calcular os elementos contidos em $\boldsymbol{\theta}^{(w+1)}$ de $\mathbf{D}^{(w+1)}$.

3.2.3 Aspectos Computacionais

Se o logaritmo da função de verossimilhança for razoavelmente bem aproximado por uma função quadrática do tipo de (3.33) na vizinhança do estimador de máxima verossimilhança, e os valores iniciais fornecidos dos parâmetros estiverem suficientemente próximos das estimativas procuradas, então as sequências de aproximações geradas pelos algoritmos de Newton-Raphson e “Scoring” de Fisher convergirão para as estimativas de máxima verossimilhança. Se a função de verossimilhança satisfizer as condições de regularidade para a teoria Normal assintótica (ver, por exemplo, Mood, Graybill e Boes (1974), p. 315-316), então há convergência, pelo menos em grandes amostras (Lawless (1982), Apêndice F). Notemos que uma condição necessária para que esses dois processos iterativos convirjam para um ponto de máximo é que as matrizes $\mathbf{H}$ ou $-\mathbf{I}$, em cada iteração, sejam invertíveis e negativas definidas. Nessas condições, os algoritmos de Newton-Raphson e “Scoring” de Fisher convergem rapidamente (isto é, têm convergência quadrática) para as estimativas procuradas. Em situações menos regulares, alguns problemas podem surgir e Jennrich e Schluchter (1986) fazem algumas sugestões para melhorar a convergência. Se os valores iniciais dos parâmetros estiverem longe das estimativas procuradas, as matrizes $\mathbf{H}$ ou $-\mathbf{I}$ na próxima iteração podem não ser negativas definidas; se elas forem invertíveis esse fato pode gerar novos valores dos parâmetros que impliquem um valor menor da função de verossimilhança. Nesse caso, é sugerido que durante a inversão de $\mathbf{H}$ ou $-\mathbf{I}$ suas diagonais sejam aumentadas de modo a garantir que elas se tornem negativas definidas e dirijam os algoritmos para um ponto de máximo. Em outras situações é possível obter, em alguma iteração, valores dos parâmetros fora do espaço paramétrico definido pelo modelo (elementos de $\boldsymbol{\theta}^{(w)}$ associados a variâncias negativos ou $\Sigma_i^{(w)}$ que não seja positiva definida). Às vezes, esses problemas são solucionados através da escolha de um melhor valor inicial e da monitoração do tamanho do passo desses processos iterativos. Nesse sentido, pode-se considerar o algoritmo na forma:

$$\boldsymbol{\phi}^{(w+1)} = \boldsymbol{\phi}^{(w)} - d \mathbf{Q}^{-1} \mathbf{s}(\boldsymbol{\phi}^{(w)}) \text{ para } w = 0, 1, 2, \dots, \quad (3.64)$$

onde $\mathbf{Q} = \mathbf{H}(\boldsymbol{\phi}^{(w)})$ (Newton-Raphson) ou $\mathbf{Q} = -\mathbf{I}(\boldsymbol{\phi}^{(w)})$ (“Scoring” de Fisher) e d é um escalar cuja função é regular o tamanho do passo. Uma estratégia conveniente é começar a iteração com $d = 1$ e verificar se $\boldsymbol{\theta}^{(w)}$ assim obtido está dentro do espaço paramétrico; caso contrário, fazer $d = d/2$ e recalcular $\boldsymbol{\theta}^{(w)}$ até que esteja dentro do espaço paramétrico e aumente a verossimilhança em relação à iteração anterior. Jennrich e Schluchter (1986) sugerem, também, um procedimento híbrido, envolvendo a realização de algumas iterações pelo algoritmo “Scoring” e uma mudança para o algoritmo de Newton-Raphson quando a variação no logaritmo da função de verossimilhança for menor que 1. Isso se deve ao fato de que o algoritmo “Scoring” tem melhor comportamento que o de Newton-Raphson “longe” das estimativas procuradas, enquanto que o algoritmo de Newton-Raphson tende a se comportar melhor que o “Scoring” “perto” do ponto de máximo.

As iterações dos algoritmos EM para modelos de efeitos aleatórios propostos por Laird e Ware (1982) e Laird et al. (1987) são bem mais rápidas do que a dos algoritmos de Newton-Raphson ou “Scoring” de Fisher. Por outro lado, o algoritmo EM geralmente requer um número substancialmente maior de iterações do que os outros dois algoritmos (as taxas de convergência dos algoritmos EM são lineares, exceto nas equações (3.61) e (3.63), onde as taxas de convergência são quadráticas). Algumas sugestões para acelerar a convergência dos algoritmos EM podem ser encontradas em Laird et al. (1987).

Uma vantagem dos algoritmos EM apresentados sobre os outros algoritmos é que eles podem ser usados para estimar os parâmetros de vários tipos de modelos como indicado na seção anterior. Os algoritmos de Newton-Raphson e “Scoring” de Fisher necessitam uma reprogramação para o ajuste de outros tipos de modelos, já que eles envolvem as primeiras e segundas derivadas da função de verossimilhança em relação a cada parâmetro contido em $\boldsymbol{\theta}$. Os algoritmos EM devem ser preferíveis aos de Newton-Raphson e “Scoring” sempre que o número de parâmetros envolvidos for grande, já que estes últimos requerem a inversão das matrizes $\mathbf{H}$ ou $\mathbf{I}$ (cujas dimensões dependem do número de parâmetros).

Lindstrom e Bates (1988) sugerem algumas modificações no algoritmo de Newton-

Raphson além das sugeridas por Jennrich e Schluchter (1986). Elas envolvem:

- (i) uma reparametrização do modelo (2.16), que facilita os cálculos através de um procedimento independente para a estimação de σ^2 ;
- (ii) o uso de decomposições para algumas matrizes envolvidas com a finalidade de estabilizar o processo do ponto de vista numérico e
- (iii) otimização da verossimilhança resultante como função dos elementos não nulos da decomposição de Cholesky de $\mathbf{D}$.

Esses autores mostram que essas modificações aceleram a convergência, e asseguram que as estimativas de $\mathbf{D}$ (obtidas tanto por máxima verossimilhança quanto por máxima verossimilhança restrita) sejam positivas definidas quando calculadas através do algoritmo de Newton-Raphson. Eles também utilizam as decomposições matriciais para os algoritmos EM discutidos anteriormente e concluem que o algoritmo de Newton-Raphson modificado pode ser uma boa alternativa (em termos do tempo total para convergência) em relação ao algoritmo EM, mesmo quando utilizamos os métodos de aceleração da convergência propostos por Laird et al. (1987).

A experiência prática mostra que em todos os casos é importante calcular o valor da função de verossimilhança para verificar se ela realmente está aumentando e como isso está acontecendo. Embora a utilização do algoritmo EM garanta que a cada iteração a verossimilhança seja aumentada, em muitos casos, ele não converge mesmo após a realização de um número muito grande de iterações (por exemplo: 150). Isso pode acontecer se o máximo da função de verossimilhança ocorrer perto da fronteira (matrizes de covariância quase singulares ou fora do espaço paramétrico). Um indicador dessa situação para o algoritmo EM sugerido por Jennrich e Schluchter (1986) é uma medida da taxa de convergência dada por:

$$r^{(w)} = \frac{l^{(w)} - l^{(w-1)}}{l^{(w-1)} - l^{(w-2)}} , \quad (3.65)$$

onde $l^{(w)}$ é o valor do logaritmo da função de verossimilhança na w -ésima iteração. Esse valor se aproxima de 1 quando as matrizes de covariância se aproximam de um ponto de singularidade. Por outro lado, os algoritmos de Newton-Raphson e “Scoring” de Fisher podem convergir rapidamente para um valor que esteja fora do espaço paramétrico. Na prática, quando isso ocorre, Fairclough e Helms (1986) sugerem a escolha de um outro modelo para a estrutura da covariância.

Dadas essas considerações, a escolha dos valores iniciais para os processos iterativos é um fator de extrema importância e nesse sentido Laird et al. (1987) sugerem dois métodos para obtenção de valores iniciais.

Para modelos de efeitos aleatórios com a forma geral (2.16), a sugestão é:

$$\alpha^{(0)} = \left(\sum_{i=1}^n \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i' \mathbf{y}_i \right), \quad (3.66)$$

$$\sigma^{2(0)} = \frac{1}{[N - (n-1)q - p]} \left[\sum_{i=1}^n \mathbf{y}_i' \mathbf{y}_i - \alpha^{(0)'} \sum_{i=1}^n \mathbf{X}_i' \mathbf{y}_i - \sum_{i=1}^n \hat{\mathbf{b}}_i' \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i \alpha^{(0)}) \right] \quad (3.67)$$

e

$$\mathbf{D}^{(0)} = \frac{1}{n-1} \sum_{i=1}^n \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i' - \frac{\sigma^{2(0)}}{n} \sum_{i=1}^n (\mathbf{Z}_i' \mathbf{Z}_i)^{-1}, \quad (3.68)$$

onde $\hat{\mathbf{b}}_i = (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i \alpha^{(0)})$.

Quando o modelo é da forma (2.22), outra alternativa é:

$$\sigma^{2(0)} = \frac{1}{(N - nq)} \left(\sum_{i=1}^n \mathbf{y}_i' \mathbf{y}_i - \sum_{i=1}^n \hat{\mathbf{b}}_i' \mathbf{Z}_i' \mathbf{y}_i \right) \quad (3.69)$$

e

$$\mathbf{D}^{(0)} = \frac{1}{(n-1)} \left[\sum_{i=1}^n \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i' - \frac{1}{n} \left(\sum_{i=1}^n \hat{\mathbf{b}}_i \right) \left(\sum_{i=1}^n \hat{\mathbf{b}}_i' \right) \right] - \frac{\sigma^{2(0)}}{n} \sum_{i=1}^n (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \quad (3.70)$$

onde $\hat{\mathbf{b}}_i = (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i' \mathbf{y}_i$.

Notemos que essas duas sugestões de valores iniciais requerem que $\mathbf{Z}_i' \mathbf{Z}_i$ tenha posto completo q . Sempre que $\mathbf{Z}_i' \mathbf{Z}_i$ for singular, como acontece quando $t_i < q$ para alguma unidade experimental, deve-se excluir tais unidades do cálculo de $\sigma^{2(0)}$ e $\mathbf{D}^{(0)}$. Antes de começar as iterações devemos checar se a matriz $\mathbf{D}^{(0)}$ é positiva definida. Isso pode ser

verificado calculando os auto-valores de $\mathbf{D}^{(0)}$ (se algum auto-valor for menor ou igual a zero a matriz $\mathbf{D}^{(0)}$ não é positiva definida).

Quando ajustamos modelos com estrutura de covariância dependente do grupo, como aquele descrito na Seção 2.2, podemos calcular os valores iniciais usando (3.70) para cada grupo.

Uma última consideração é que os algoritmos de Newton-Raphson e “Scoring” de Fisher, quando convergem têm como subproduto estimativas das matrizes de variância assintóticas de $\hat{\alpha}$ e $\hat{\theta}$, enquanto que nos algoritmos EM temos somente a estimativa da variância de $\hat{\alpha}$.

3.3 Soluções Explícitas

Nesta seção vamos descrever brevemente alguns casos particulares do modelo de efeitos aleatórios onde há soluções explícitas para os estimadores de máxima verossimilhança e máxima verossimilhança restrita dos parâmetros envolvidos. Maiores detalhes poderão ser encontrados em Graybill (1976, Cap 12), Laird et al. (1987) e Lange e Laird (1989).

Consideremos o caso de dados balanceados em relação ao tempo, isto é, onde para cada unidade experimental i , $i = 1, \dots, n$, uma variável resposta é observada nas mesmas T ocasiões pré-estabelecidas. Nesse caso, é conveniente representar o modelo de efeitos aleatórios na forma matricial:

$$\mathbf{Y} = \mathbf{Z}\Phi\mathbf{A} + \mathbf{Z}_q\mathbf{B} + \mathbf{E}, \quad (3.71)$$

onde $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n) (T \times n)$ é a matriz contendo as respostas das unidades experimentais, $\mathbf{Z} (T \times r)$ é a matriz de especificação do modelo intra unidades experimentais (essencialmente seus elementos determinam a forma da curva populacional), $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n) (f \times n)$ é a matriz de especificação do modelo entre unidades experimentais (geralmente constituída de 0's e 1's de forma a associar cada unidade experimental à subpopulação a que pertence), $\Phi = (\Phi_1, \dots, \Phi_f) (r \times f)$ é a matriz dos coeficientes polinomiais desconhecidos (cada coluna tem os coeficientes da curva correspondente a uma determinada subpopulação), $\mathbf{Z}_q (T \times q)$ é a matriz de especificação dos efeitos aleatórios, $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_n) (q \times n)$ e $\mathbf{E} = (\mathbf{e}_1, \dots, \mathbf{e}_n) (T \times n)$ são respectivamente a matriz contendo os efeitos aleatórios e a matriz contendo os erros associados a cada unidade experimental. Vamos supor que $0 \leq q \leq r \leq T < n$ e que a matriz $\mathbf{Z}$ possa ser particionada em

$$\mathbf{Z} = [\mathbf{Z}_q : \mathbf{Z}_{\bar{q}}],$$

isto é, as colunas de $\mathbf{Z}_q$ são as primeiras q colunas de $\mathbf{Z}$. Isto implica que o grau da curva populacional é maior ou igual ao da curva de cada unidade experimental. Além disso, vamos supor que $\mathbf{e}_1, \dots, \mathbf{e}_n, \mathbf{b}_1, \dots, \mathbf{b}_n$ são independentes com $\mathbf{e}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_T)$ e

$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$ para $i = 1, \dots, n$ e $\mathbf{D}$ ($q \times q$) é uma matriz de covariância não estruturada. Então podemos escrever:

$$\text{Var}(\mathbf{Y}) = \text{Var}(\text{vec}(\mathbf{Y})) = \mathbf{I}_n \otimes \Sigma,$$

onde $\Sigma = \mathbf{Z}_q \mathbf{D} \mathbf{Z}_q + \sigma^2 \mathbf{I}_T$ e $\text{vec}(\mathbf{Y}) = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_n)'$.

Notemos que esse modelo pode ser escrito na forma (2.16) com $\mathbf{X}_i = \mathbf{a}'_i \otimes \mathbf{Z}$, $\mathbf{Z}_i = \mathbf{Z}_q$ e $\alpha = \text{vec}(\Phi)$.

O logaritmo da função de verossimilhança correspondente é

$$l(\mathbf{Y}; \Phi, \sigma^2, \mathbf{D}) = K - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr}[(\mathbf{Y} - \mathbf{Z}\Phi\mathbf{A})'\Sigma^{-1}(\mathbf{Y} - \mathbf{Z}\Phi\mathbf{A})] \quad (3.72)$$

onde K é uma constante que não depende de Φ , σ^2 e $\mathbf{D}$. E o logaritmo da função de verossimilhança restrita pode ser escrito como:

$$\begin{aligned} l_R(\mathbf{Y}; \sigma^2, \mathbf{D}) &= K - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr}[(\mathbf{Y} - \mathbf{Z}\tilde{\Phi}\mathbf{A})'\Sigma^{-1}(\mathbf{Y} - \mathbf{Z}\tilde{\Phi}\mathbf{A})] \\ &\quad - \frac{f}{2} \ln |\mathbf{Z}'\Sigma^{-1}\mathbf{Z}| - \frac{r}{2} \ln |\mathbf{A}\mathbf{A}'|, \end{aligned} \quad (3.73)$$

onde $\tilde{\Phi} = (\mathbf{Z}'\Sigma^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\Sigma^{-1}\mathbf{Y}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ e K é uma constante que não depende de σ^2 nem de $\mathbf{D}$.

Pode-se mostrar (ver referências em Lange e Laird (1989)) que o estimador de máxima verossimilhança de Φ para o modelo (3.71) é idêntico ao estimador de mínimos quadrados para qualquer q , pois

$$(\mathbf{Z}'\Sigma^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\Sigma^{-1}\mathbf{Y} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}.$$

Consequentemente:

$$\hat{\Phi} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1} \quad (3.74)$$

ou

$$\hat{\alpha} = \left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right)' \left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{y}_i \right). \quad (3.75)$$

Embora o estimador $\hat{\alpha}$ (ou $\hat{\Phi}$) seja invariante com relação aos modelos dessa família, sua matriz de covariância depende da estrutura de covariância de $\mathbf{Y}$. Pode-se mostrar que:

$$\text{Var}(\hat{\alpha}) = \text{Var}(\text{vec}(\hat{\Phi})) = (\mathbf{A}\mathbf{A}')^{-1} \otimes [\sigma^2(\mathbf{Z}'\mathbf{Z})^{-1} + \mathbf{J}_q\mathbf{D}\mathbf{J}_q'], \quad (3.76)$$

onde $\mathbf{J}_q = [\mathbf{I}_q : \mathbf{0}]' (r \times q)$ para $q = 0, 1, \dots, r$.

Os estimadores de máxima verossimilhança e máxima verossimilhança restrita de σ^2 e $\mathbf{D}$ são dados por:

$$\hat{\sigma}^2 = \begin{cases} \frac{1}{n(T-q)} \text{tr}(\mathbf{Q}_q) & \text{para MV} \\ \frac{1}{n(T-q)-f(r-q)} \text{tr}(\mathbf{Q}_q) & \text{para MVR} \end{cases} \quad (3.77)$$

e

$$\hat{\mathbf{D}} = \begin{cases} \frac{1}{n}(\mathbf{Z}'_q\mathbf{Z}_q)^{-1}\mathbf{Z}'_q\mathbf{S}\mathbf{Z}_q(\mathbf{Z}'_q\mathbf{Z}_q)^{-1} - \hat{\sigma}^2(\mathbf{Z}'_q\mathbf{Z}_q)^{-1} & \text{para MV} \\ \frac{1}{n-f}(\mathbf{Z}'_q\mathbf{Z}_q)^{-1}\mathbf{Z}'_q\mathbf{S}\mathbf{Z}_q(\mathbf{Z}'_q\mathbf{Z}_q)^{-1} - \hat{\sigma}^2(\mathbf{Z}'_q\mathbf{Z}_q)^{-1} & \text{para MVR} \end{cases} \quad (3.78)$$

onde $\mathbf{Q}_q = (\mathbf{Y} - \mathbf{Z}\hat{\Phi}\mathbf{A})' [\mathbf{I}_T - \mathbf{Z}_q(\mathbf{Z}'_q\mathbf{Z}_q)^{-1}\mathbf{Z}'_q] (\mathbf{Y} - \mathbf{Z}\hat{\Phi}\mathbf{A})$ e $\mathbf{S} = \mathbf{Y} [\mathbf{I}_n - \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}] \mathbf{Y}'$.

As expressões (3.77) e (3.78) fornecem estimativas de máxima verossimilhança ou máxima verossimilhança restrita somente se $\hat{\mathbf{D}}$ for positiva semi-definida.

Capítulo 4

Aplicações Práticas

O objetivo deste capítulo é ilustrar a aplicação de modelos de efeitos aleatórios em dados longitudinais. Na Seção 4.1 informamos sobre alguns recursos computacionais disponíveis. Na Seção 4.2 apresentamos os resultados obtidos através do ajuste de vários modelos para os exemplos do Capítulo 1.

4.1 Recursos Computacionais

Até o momento, não estão disponíveis nos pacotes estatísticos rotinas para o ajuste de modelos para dados longitudinais não balanceados em relação ao tempo. Somente o BMDP (ver Schluchter (1988b)) anuncia para breve o programa **BMDP5V**. Esse programa ajusta modelos lineares para as médias e oferece várias opções para a estrutura de covariância. Além disso, permite a escolha do método de estimação (MV ou MVR) e do processo iterativo a ser utilizado. O programa BMDP5V tem como saída:

- estimativas de máxima verossimilhança (MV ou MVR) dos parâmetros do modelo e as respectivas estimativas dos erros-padrões assintóticos;
- valor do logaritmo da função de verossimilhança;

- testes de hipóteses sobre os parâmetros associados às médias, usando a estatística de Wald;
- valores observados, valores imputados das respostas incompletas e resíduos.

Contudo, o programa é dirigido para análise de conjuntos de dados incompletos, como aqueles descritos na Seção 2.1.

Como parte deste trabalho, desenvolvemos vários programas para o ajuste de modelos de efeitos aleatórios, usando o módulo CM do pacote científico SOC (EMBRAPA). Esses programas podem ser utilizados em dados não balanceados em relação ao tempo e constam do Apêndice B. Temos cinco programas:

- EMMV.AJU** Estimação através do método de máxima verossimilhança, usando o algoritmo EM.
- SCMV.AJU** Estimação através do método de máxima verossimilhança, usando o algoritmo “Scoring” de Fisher.
- NRMV.AJU** Estimação através do método de máxima verossimilhança, usando o algoritmo de Newton-Raphson.
- EMMVR.AJU** Estimação através do método de máxima verossimilhança restrita, usando o algoritmo EM.
- SCMVR.AJU** Estimação através do método de máxima verossimilhança restrita, usando o algoritmo “Scoring” de Fisher.

4.2 Exemplos de Ajuste de Modelos

Nesta seção apresentamos os resultados obtidos através do ajuste de vários modelos para os exemplos do Capítulo 1. Não temos a intenção de fazer uma análise exaustiva dos dados, mas apenas ilustrar uma possível estratégia de análise e mostrar como os modelos podem ser ajustados. Para efeito de comparação, apresentamos as estimativas dos parâmetros envolvidos, usando o método de máxima verossimilhança e o de máxima verossimilhança restrita e os vários processos iterativos. Os cálculos foram realizados com os programas EMMV.AJU, SCMV.AJU, NRMV.AJU, EMMVR.AJU e SCMVR.AJU. As iterações prosseguiram até que a maior diferença absoluta entre as estimativas de iterações sucessivas fosse menor que 0.001.

4.2.1 EXEMPLO 1

Consideremos os dados da Tabela 1.1 correspondendo à observação da distância entre o centro da pituitária e a fissura pterio-maxilar de 11 meninas e 16 meninos aos 8, 10, 12 e 14 anos. Os dados são balanceados em relação ao tempo. Ajustamos seis modelos diferentes a esses dados, que são descritos a seguir.

Modelo 1

Inicialmente supomos o seguinte modelo para as distâncias (y_i) de cada indivíduo:

$$E(y_i) = (a'_i \otimes I_4)\alpha \quad \text{e} \quad \text{Var}(y_i) = \Sigma \quad \text{para } i = 1, \dots, 27,$$

onde $a'_i = (1, 0)$ se o i -ésimo indivíduo pertence ao grupo de meninas e $a'_i = (0, 1)$ se o i -ésimo indivíduo pertence ao grupo de meninos, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_8)'$ e $\Sigma (4 \times 4)$ é uma matriz de covariância não estruturada igual para todos indivíduos. Nesse caso, $\alpha_1, \alpha_2, \alpha_3$ e α_4 representam respectivamente as respostas médias aos 8, 10, 12 e 14 anos para o grupo de meninas e $\alpha_5, \alpha_6, \alpha_7$ e α_8 para o grupo de meninos.

Utilizamos os programas EMMV.AJU e EMMVR.AJU para estimar os parâmetros

desse modelo. As estimativas obtidas podem ser encontradas nas Tabelas 4.8 e 4.9. O gráfico desses valores médios em função da idade (ver Figura 1.1) sugere que a relação entre a distância e a idade é linear, e possivelmente retas diferentes para cada grupo. Quando utilizamos o método de máxima verossimilhança restrita para estimar a estrutura de covariância, podemos testar a hipótese de existência de efeito linear em cada grupo através da estatística de Wald. Assim, testamos $H_0 : C\alpha = 0$, onde

$$C = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix},$$

resultando em $Q_2 = 2.540$ (gl=2 e p=0.281). Logo, parece haver um efeito linear nas médias de cada grupo ao nível de significância usual ($\alpha < 0.10$). Isso sugere o próximo modelo.

Modelo 2

Supomos que:

$$E(y_i) = X_i\alpha \quad \text{e} \quad \text{Var}(y_i) = \Sigma \quad \text{para } i = 1, \dots, 27,$$

onde $X_i = a_i' \otimes X$,

$$X = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix}$$

e Σ é não estruturada. Nesse modelo, $\alpha = (\alpha_1, \dots, \alpha_4)'$, onde α_1 e α_2 são respectivamente os coeficientes linear e angular da reta associada ao grupo das meninas e α_3 e α_4 os coeficientes linear e angular do grupo dos meninos. Os coeficientes estimados são apresentados na Tabela 4.1.

Recorrendo a Tabela 4.6, vemos que a estatística observada do teste da razão de verossimilhança generalizada, comparando o Modelo 2 com o Modelo 1 não é significativa ($Q_1 = 2.968$, gl=4 e p=0.563), o que favorece a hipótese de que a relação entre a distância média e a idade seja linear para cada grupo.

Podemos verificar se as retas de cada grupo têm o mesmo coeficiente linear, testando $H_0 : C\alpha = 0$, onde $C = [0 \ 1 \ 0 \ -1]$, através da estatística de Wald. Utilizando os resultados

Tabela 4.1: Estimativas dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 2 do Exemplo 1.

Parâmetro	Programa	
	EMMV	EMMVR
α_1 est	17.425	17.425
	ep	1.128
α_2 est	0.476	0.476
	ep	0.095
α_3 est	15.842	15.842
	ep	0.936
α_4 est	0.827	0.827
	ep	0.079
lver	-110.493	-116.704
Iterações	4	5

do Modelo 2 ajustado por máxima verossimilhança restrita, obtivemos $Q_c = 7.425$ (gl=1 e $p=0.007$), indicando que os coeficientes lineares das retas de cada grupo são diferentes. No caso do método de máxima verossimilhança, podemos ajustar o Modelo 3, que supõe que os coeficientes lineares são iguais, e compará-lo com o Modelo 2 através do teste da razão de verossimilhança generalizada.

Modelo 3 $\rightarrow$ *supõe coef. lineares iguais (α_3)*

Esse modelo é da forma:

$$E(y_i) = X_i \alpha \quad \text{e} \quad \text{Var}(y_i) = \Sigma,$$

onde

$$X_i = \begin{pmatrix} 1 & 0 & 8 \\ 1 & 0 & 10 \\ 1 & 0 & 12 \\ 1 & 0 & 14 \end{pmatrix} \quad \text{para } i = 1, \dots, 11 \quad (\text{meninas}),$$

$$X_i = \begin{pmatrix} 0 & 1 & 8 \\ 0 & 1 & 10 \\ 0 & 1 & 12 \\ 0 & 1 & 14 \end{pmatrix} \quad \text{para } i = 12, \dots, 27 \quad (\text{meninos}),$$

$\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ e Σ é não estruturada. Nesse modelo, α_1 e α_2 são os coeficientes lineares das retas do grupo de meninas e meninos respectivamente, e α_3 é coeficiente angular comum às duas retas. A Tabela 4.2 apresenta as estimativas obtidas dos parâmetros.

Tabela 4.2: Estimativas dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 3 do Exemplo 1.

Parâmetro		Programa	
		EMMV	EMMVR
α_1	est	15.372	15.372
	ep	0.869	0.922
α_2	est	17.418	17.418
	ep	0.845	0.866
α_3	est	0.675	0.675
	ep	0.069	0.070
lver		-113.831	-117.859
Iterações		9	7

Quando comparamos o Modelo 3 com o Modelo 2 através do teste da razão de verossimilhança generalizada, obtivemos $Q_1 = 6.676$ (gl=1 e $p=0.010$), confirmando que esse modelo não é adequado para os dados.

Podemos tentar simplificar a estrutura de covariância Σ através do ajuste de um modelo de efeitos aleatórios.

Modelo 4

Supondo, agora, que a cada indivíduo esteja associada uma reta média individual cujos coeficientes diferem aleatoriamente dos coeficientes da reta média populacional, ajustamos o seguinte modelo:

$$y_i = X_i\alpha + Zb_i + e_i \quad \text{para } i = 1, \dots, 27,$$

onde $\mathbf{X}_i$ e $\boldsymbol{\alpha}$ são iguais aos do Modelo 2, $\mathbf{b}_i = (b_{i1}, b_{i2})' \sim N(\mathbf{0}, \mathbf{D})$, $\mathbf{e}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_4)$,

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} \quad \text{e} \quad \mathbf{Z} = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix}.$$

Nesse modelo,

$$E(\mathbf{y}_i) = \mathbf{X}_i \boldsymbol{\alpha} \quad \text{e} \quad \text{Var}(\mathbf{y}_i) = \boldsymbol{\Sigma} = \mathbf{Z} \mathbf{D} \mathbf{Z}' + \sigma^2 \mathbf{I}_4.$$

As estimativas dos parâmetros podem ser encontradas na Tabela 4.3.

Tabela 4.3: Estimativas dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 4 do Exemplo 1.

Parâmetro		Programa				
		EMMV	SCMV	NRMV ^b	EMMVR	SCMVR
α_1	est	17.373	17.373	17.373	17.373	17.373
(17.373) ^a	ep	1.182	1.182	1.182	1.228	1.228
α_2	est	0.480	0.480	0.480	0.480	0.480
(0.480)	ep	0.100	0.100	0.100	0.104	0.104
α_3	est	16.341	16.341	16.341	16.341	16.341
(16.341)	ep	0.980	0.980	0.980	1.019	1.019
α_4	est	0.784	0.784	0.784	0.784	0.784
(0.784)	ep	0.083	0.083	0.083	0.086	0.086
d_{11}	est	4.589	4.557	4.557	5.804	5.786
(5.415)	ep	—	4.672	4.672	—	5.304
d_{12}	est	-0.201	-0.198	-0.198	-0.291	-0.290
(-0.321)	ep	—	0.379	0.379	—	0.428
d_{22}	est	0.024	0.024	0.024	0.033	0.033
(0.051)	ep	—	0.034	0.034	—	0.038
σ^2	est	1.716	1.716	1.716	1.716	1.716
(1.716)	ep	—	0.330	0.330	—	0.330
lver		-114.657			-120.721	
Iterações		97	2	1	67	4

^a Valores iniciais calculados com as expressões (3.66), (3.69) e (3.70).

^b Valores iniciais da primeira iteração do “Scoring”.

Quando comparamos o Modelo 4 com o Modelo 2 (ver Tabelas 4.6 e 4.7), temos que o Modelo 4 parece ser adequado aos dados. Observando a Tabela 4.3, vemos que das

estimativas dos parâmetros de $\mathbf{D}$, $\hat{d}_{11}$ é a que parece ser mais significativamente diferente de zero, quando comparada ao seu erro-padrão (critério utilizado por Jennrich e Schluchter (1986) para uma possível redução no número de parâmetros do modelo). Esse fato sugere o Modelo 5.

Modelo 5

Consideremos, agora, que:

$$y_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{1}_4 \mathbf{b}_i + \mathbf{e}_i \quad \text{para } i = 1, \dots, 27,$$

onde $\mathbf{X}_i$ e $\boldsymbol{\alpha}$ são iguais ao modelo anterior, $\mathbf{b}_i \sim N(0, \sigma_b^2)$ e $\mathbf{e}_i \sim N(0, \sigma^2 \mathbf{I}_4)$. Assim,

$$E(y_i) = \mathbf{X}_i \boldsymbol{\alpha} \quad \text{e} \quad \text{Var}(y_i) = \boldsymbol{\Sigma} = \sigma_b^2 \mathbf{1}_4 \mathbf{1}_4' + \sigma^2 \mathbf{I}_4.$$

Esse modelo supõe que as variâncias e as covariâncias entre as observações da distância realizadas no mesmo indivíduo são constantes ao longo das idades e as retas individuais diferem somente no coeficiente linear da reta populacional. As estimativas obtidas dos parâmetros estão apresentados na Tabela 4.4. Pelo teste da razão de verossimilhança generalizada esse modelo se ajusta aos dados (ver Tabelas 4.6 e 4.7). Podemos verificar se $\sigma_b^2 = 0$ através do ajuste do próximo modelo.

Modelo 6

Finalmente, podemos verificar se as observações são independentes com variância constante, ajustando o modelo de Regressão linear simples, onde

$$E(y_i) = \mathbf{X}_i \boldsymbol{\alpha} \quad \text{e} \quad \text{Var}(y_i) = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_4 \quad \text{para } i = 1, \dots, 27.$$

Os resultados obtidos estão apresentados na Tabela 4.5. Pelo teste da razão de verossimilhança generalizada, esse modelo não é adequado aos dados.

Tabela 4.4: Estimativas dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 5 do Exemplo 1.

Parâmetro		Programa				
		EMMV	SCMV	NRMV	EMMVR	SCMVR
α_1	est	17.373	17.373	17.373	17.373	17.373
(17.373) ^a	ep	1.162	1.162	1.162	1.184	1.184
α_2	est	0.480	0.480	0.480	0.480	0.480
(0.480)	ep	0.092	0.092	0.092	0.093	0.093
α_3	est	16.341	16.341	16.341	16.341	16.341
(16.341)	ep	0.963	0.963	0.963	0.981	0.981
α_4	est	0.784	0.784	0.784	0.784	0.784
(0.784)	ep	0.077	0.077	0.077	0.078	0.078
σ_b^2	est	3.030	3.030	3.030	3.298	3.298
(3.147)	ep	—	0.955	0.955	—	1.113
σ^2	est	1.875	1.875	1.875	1.922	1.922
(1.947)	ep	—	0.295	0.295	—	0.309
lver		-115.074			-121.309	
Iterações		4	2	3	6	4

^a Valores iniciais calculados com as expressões (3.66), (3.67) e (3.68).

Tabela 4.5: Estimativas dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 6 do Exemplo 1.

Parâmetro		Programa	
		EMMV	EMMVR
α_1	est	17.373	17.373
(17.373)	ep	1.676	1.708
α_2	est	0.480	0.480
(0.480)	ep	0.149	0.152
α_3	est	16.341	16.341
(16.341)	ep	1.390	1.416
α_4	est	0.784	0.784
(0.784)	ep	0.124	0.126
σ^2	est	4.905	5.094
(5.000)	ep	—	—
lver		-139.876	-146.210
Iterações		2	3

Tabela 4.6: Testes de ajuste de modelos para o Exemplo 1 (máxima verossimilhança).

Modelo	Número de parâmetros ($p + k$)	lver ^a	Modelo de comparação	Q_1	gl	p
1	18	-109.009	—	—	—	—
2	14	-110.493	1	2.968	4	0.563
3	13	-113.831	2	6.676	1	0.010
4	8	-114.658	2	8.330	6	0.215
5	6	-115.074	4	0.832	2	0.660
6	5	-139.876	5	49.604	1	$\simeq 0$

^a Valor da parte não constante do logaritmo da função de verossimilhança.

Tabela 4.7: Testes de ajuste de modelos para o Exemplo 1 (máxima verossimilhança restrita).

Modelo	Número de parâmetros ($p + k$)	lver ^a	Modelo de comparação	Q_2	gl	p
1	18	-115.124	—	—	—	—
2	14	-116.704	—	—	—	—
3	13	-117.859	—	—	—	—
4	8	-120.721	2	8.034	6	0.236
5	6	-121.309	4	1.176	2	0.555
6	5	-146.203	5	49.788	1	$\simeq 0$

^a Valor da parte não constante do logaritmo da função de verossimilhança restrita.

Tabela 4.8: Valores médios estimados e estimativas da matriz de covariância/correlação para os modelos do Exemplo 1 (máxima verossimilhança).

Modelo	Idade	Médias				Estimativas das				
		Meninas		Meninos		variâncias e correlações				
		est	ep	est	ep	Variância	Matriz de correlação			
1	8	21.182	0.675	22.875	0.560	5.014	1.000	0.571	0.661	0.522
	10	22.227	0.594	23.813	0.492	3.875		1.000	0.563	0.726
	12	23.091	0.737	25.719	0.611	5.978			1.000	0.728
	14	24.091	0.648	27.469	0.537	4.616				1.000
2	8	21.236	0.589	22.457	0.489	5.119	1.000	0.544	0.653	0.519
	10	22.189	0.546	24.110	0.453	3.923		1.000	0.561	0.719
	12	23.142	0.567	25.764	0.471	5.980			1.000	0.728
	14	24.095	0.646	27.418	0.537	4.618				1.000
3	8	20.770	0.569	22.815	0.480	5.086	1.000	0.562	0.642	0.475
	10	22.119	0.546	24.164	0.453	3.953		1.000	0.542	0.652
	12	23.468	0.557	25.513	0.466	6.061			1.000	0.725
	14	24.818	0.601	26.863	0.517	5.049				1.000
4	8	21.209	0.611	22.616	0.507	4.622	1.000	0.621	0.601	0.573
	10	22.168	0.563	24.184	0.467	4.684		1.000	0.634	0.623
	12	23.127	0.583	25.753	0.483	4.936			1.000	0.659
	14	24.086	0.665	27.322	0.552	5.379				1.000
5	8	21.209	0.628	22.616	0.521	4.905	1.000	0.618	0.618	0.618
	10	22.168	0.571	24.184	0.474	4.905		1.000	0.618	0.618
	12	23.127	0.571	25.753	0.474	4.905			1.000	0.618
	14	24.086	0.628	27.322	0.521	4.905				1.000
6	8	21.209	0.559	22.616	0.463	4.905	1.000	0.000	0.000	0.000
	10	22.168	0.366	24.184	0.303	4.905		1.000	0.000	0.000
	12	23.127	0.366	25.753	0.303	4.905			1.000	0.000
	14	24.086	0.559	27.322	0.463	4.905				1.000

Tabela 4.9: Valores médios estimados e estimativas da matriz de covariância/correlação para os modelos do Exemplo 1 (máxima verossimilhança restrita).

Modelo	Idade	Médias				Estimativas das				
		Meninas		Meninos		variâncias e correlações				
		est	ep	est	ep	Variância	Matriz de correlação			
1	8	21.182	0.702	22.875	0.582	5.415	1.000	0.571	0.661	0.522
	10	22.227	0.617	23.813	0.511	4.185		1.000	0.563	0.726
	12	23.091	0.766	25.719	0.635	6.456			1.000	0.728
	14	24.091	0.673	27.469	0.558	4.986				1.000
2	8	21.236	0.613	22.457	0.527	5.425	1.000	0.568	0.659	0.522
	10	22.189	0.568	24.110	0.491	4.191		1.000	0.581	0.725
	12	23.142	0.589	25.764	0.508	6.263			1.000	0.740
	14	24.095	0.672	27.418	0.574	4.986				1.000
3	8	20.770	0.590	22.815	0.498	5.375	1.000	0.586	0.652	0.490
	10	22.119	0.567	24.164	0.470	4.215		1.000	0.563	0.666
	12	23.468	0.578	25.513	0.484	6.335			1.000	0.737
	14	24.818	0.622	26.863	0.535	5.376				1.000
4	8	21.209	0.635	22.616	0.527	4.950	1.000	0.641	0.612	0.572
	10	22.168	0.585	24.184	0.485	4.963		1.000	0.651	0.633
	12	23.127	0.606	25.753	0.503	5.235			1.000	0.677
	14	24.086	0.691	27.322	0.574	5.768				1.000
5	8	21.209	0.650	22.616	0.539	5.221	1.000	0.632	0.632	0.632
	10	22.168	0.594	24.184	0.493	5.221		1.000	0.632	0.632
	12	23.127	0.594	25.753	0.493	5.221			1.000	0.632
	14	24.086	0.650	27.322	0.539	5.221				1.000
6	8	21.209	0.569	22.616	0.472	5.094	1.000	0.000	0.000	0.000
	10	22.168	0.372	24.184	0.310	5.094		1.000	0.000	0.000
	12	23.127	0.372	25.753	0.310	5.094			1.000	0.000
	14	24.086	0.569	27.322	0.472	5.094				1.000

Como os dados são balanceados em relação ao tempo, existem soluções explícitas para os estimadores de máxima verossimilhança e máxima verossimilhança restrita dos parâmetros associados aos Modelos 1, 4, 5 e 6. Esses modelos são casos particulares do modelo apresentado na Seção 3.3. Existe, também, uma “forma fechada” para os estimadores de máxima verossimilhança do Modelo 2 (ver Khatri (1966), Grizzle e Allen (1969) e Laird et al. (1987)). Contudo, preferimos ajustar os modelos por processos iterativos a fim de estudarmos o comportamento dos algoritmos.

Os algoritmos EM utilizados para estimar os parâmetros dos Modelos 1, 2 e 3 não precisaram de muitas iterações para atingir o critério de convergência estipulado, pois, nesses casos, os algoritmos se reduzem às expressões (3.61) (MV) ou (3.63) (MVR), que têm convergência quadrática para as estimativas procuradas. Por outro lado, os algoritmos EM na formulação para efeitos aleatórios levaram muitas iterações para atingir a convergência (ver Tabela 4.3). Para os algoritmos “Scoring” poucas iterações foram necessárias. O algoritmo de Newton-Raphson divergiu quando utilizamos os valores iniciais indicados na Tabela 4.3. No Modelo 5, todos os algoritmos levaram poucas iterações, inclusive os algoritmos EM.

4.2.2 EXEMPLO 2

Consideremos os dados da Tabela 1.2 referentes à observação do PAC de 18 mexilhões em 4 condições de salinidade da água, onde 25% das observações não foram coletadas. Para esse exemplo ajustamos cinco modelos, cujos resultados são apresentados a seguir.

Modelo 1

O modelo mais geral que pode ser ajustado a esses dados é o modelo de médias de caselas com a matriz de covariância entre unidades experimentais não estruturada. Utilizamos os programas EMMV.AJU e EMMVR.AJU para essa finalidade, especificando como entrada dos programas $\sigma^{2(0)}=0.00001$,

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \dots, \quad \mathbf{X}_9 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \dots, \quad \mathbf{X}_{18} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

e $\mathbf{Z}_i = \mathbf{X}_i$ para $i = 1, \dots, 18$. Nesse modelo, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_4)'$, onde α_j é o valor esperado do PAC na j -ésima variação de salinidade ($j=1$ (-1.7%), 2 (-1.0%), 3 (-0.5%), 4 (0.0%)).

Ambos programas levaram 25 iterações para atingir o critério de convergência estipulado. As estimativas obtidas podem ser encontradas nas Tabelas 4.15 (MV) e 4.17 (MVR).

Para verificar se a relação entre o PAC e a variação da salinidade pode ser descrita por uma reta, ajustamos o Modelo 2, mantendo a matriz de covariância não estruturada. Contudo, quando utilizamos o método de máxima verossimilhança restrita, não podemos comparar o Modelo 2 com o Modelo 1 usando o teste da razão de verossimilhança generalizada, pois as estruturas de médias são diferentes. Nesse caso, podemos verificar a existência de efeito linear nas médias usando o Modelo 1, através da estatística de Wald. Assim, testamos $H_0: \mathbf{C}\boldsymbol{\alpha} = \mathbf{0}$, onde $\mathbf{C} = [0.5 \quad -0.5 \quad -0.7 \quad 0.7]$, resultando que $Q_c = 2.494$ (gl=1 e p=0.114), indicando uma tendência linear das médias do PAC ao longo das variações de salinidade.

Modelo 2

Para ajustar uma reta para os valores médios do PAC e uma matriz de covariância não estruturada, utilizamos os programas EMMV.AJU e EMMVR.AJU, especificando $\sigma^{2(0)} = 0.00001$,

$$\mathbf{X}_1 = \begin{pmatrix} 1 & -1.7 \\ 1 & 0.0 \end{pmatrix}, \dots, \mathbf{X}_9 = \begin{pmatrix} 1 & -1.7 \\ 1 & -1.0 \\ 1 & -0.5 \\ 1 & 0.0 \end{pmatrix}, \dots, \mathbf{X}_{18} = \begin{pmatrix} 1 & -0.5 \\ 1 & 0.0 \end{pmatrix},$$

e

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \dots, \mathbf{Z}_9 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \dots, \mathbf{Z}_{18} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

Nesse modelo, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$, onde α_1 e α_2 são, respectivamente, os coeficientes linear e angular da reta populacional. A Tabela 4.10 apresenta as estimativas obtidas para α_1 e α_2 . O teste da razão de verossimilhança generalizada para a comparação do Modelo 1 com o Modelo 2 (ver Tabela 4.14) confirma que o Modelo 2 pode ser adequado.

Tabela 4.10: Estimativas (est) dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 2 do Exemplo 2.

Parâmetro		Programa	
		EMMV	EMMVR
α_1	est	11.950	11.972
(2.966)	ep	0.620	0.639
α_2	est	4.221	4.244
(0.194)	ep	0.522	0.544
lver		-69.763	-71.284
Iterações		29	29

Modelo 3

Supondo, agora, que a cada mexilhão esteja associada uma reta individual com desvios aleatórios em relação à reta populacional, podemos considerar o seguinte modelo de

efeitos aleatórios:

$$y_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{X}_i \mathbf{b}_i + e_i,$$

onde

$$\mathbf{X}_1 = \begin{pmatrix} 1 & -1.7 \\ 1 & 0.0 \end{pmatrix}, \dots, \mathbf{X}_9 = \begin{pmatrix} 1 & -1.7 \\ 1 & -1.0 \\ 1 & -0.5 \\ 1 & 0.0 \end{pmatrix}, \dots, \mathbf{X}_{18} = \begin{pmatrix} 1 & -0.5 \\ 1 & 0.0 \end{pmatrix},$$

$\mathbf{Z}_i = \mathbf{X}_i$ para $i = 1, \dots, 18$. Nesse modelo, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ tem a mesma interpretação que no Modelo 2 e $\mathbf{b}_i = (b_{i1}, b_{i2})'$ contem os desvios aleatórios dos parâmetros das retas individuais em relação aos parâmetros da reta populacional. O teste da razão de verossimilhança generalizada (ver Tabelas 4.14 e 4.16), aplicado em ambos métodos de estimação, indica que o Modelo 3 pode ser adequado aos dados. As estimativas obtidas estão na Tabela 4.11.

Tabela 4.11: Estimativas (est) dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 3 do Exemplo 2.

Parâmetro		Programa				
		EMMV	SCMV	NRMV	EMMVR	SCMVR
α_1	est	11.974	11.974	11.974	11.979	11.979
(11.930) ^a	ep	0.615	0.615	0.616	0.633	0.634
α_2	est	4.079	4.081	4.081	4.090	4.091
(4.030)	ep	0.489	0.491	0.495	0.507	0.509
d_{11}	est	4.097	4.124	4.124	4.520	4.535
(4.611)	ep	—	2.397	2.371	—	2.669
d_{12}	est	1.862	1.893	1.893	2.127	2.145
(1.962)	ep	—	1.694	1.672	—	1.901
d_{22}	est	1.163	1.212	1.212	1.456	1.484
(1.000) ^b	ep	—	1.571	1.598	—	1.767
σ^2	est	3.544	3.522	3.522	3.531	3.519
(3.551)	ep	—	1.056	1.060	—	1.063
lver		-72.313			-73.873	
Iterações		146	3	4	148	5

^a Valores iniciais calculados com as expressões (3.66), (3.69) e (3.70).

^b Valor inicial modificado para tornar $\mathbf{D}^{(0)}$ positiva definida.

Observando a Tabela 4.11, temos que o elemento de $\mathbf{D}$ que parece ser mais significa-

tivamente diferente de zero é d_{11} , quando comparado ao seu erro-padrão. Esse fato sugere o Modelo 4.

Modelo 4

Esse modelo é da forma:

$$y_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{1}_i b_i + e_i \quad \text{para } i = 1, \dots, 18,$$

onde $\mathbf{X}_i$ e $\boldsymbol{\alpha}$ são equivalentes aos do Modelo 3 e $b_i = b_{i1} \sim N(0, \sigma_b^2)$. Nesse modelo, b_i representa o desvio aleatório do coeficiente linear da reta do mexilhão i em relação ao da reta populacional. Os resultados obtidos estão na Tabela 4.12. O Modelo 4 parece se ajustar aos dados pelo teste da razão de verossimilhança generalizada, comparando com o Modelo 3. Finalmente, podemos verificar se $\sigma_b^2 = 0$, isto é, se as observações são independentes, através do Modelo 5.

Tabela 4.12: Estimativas (est) dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 4 do Exemplo 2.

Parâmetro		Programa				
		EMMV	SCMV	NRMV	EMMVR	SCMVR
α_1	est	11.997	11.997	11.997	11.999	11.999
(11.930) ^a	ep	0.531	0.531	0.532	0.544	0.544
α_2	est	4.010	4.010	4.010	4.102	4.102
(4.030)	ep	0.455	0.455	0.456	0.462	0.462
σ_b^2	est	1.860	1.858	1.858	2.020	2.032
(1.970)	ep	—	1.151	1.146	—	1.294
σ^2	est	4.228	4.229	4.229	4.346	4.345
(4.368)	ep	—	0.993	0.991	—	1.048
lver		-73.391			-75.000	
Iterações		10	3	3	10	3

^a Valores iniciais calculados com as expressões (3.66), (3.67) e (3.68).

Modelo 5

Esse modelo é o de Regressão linear simples:

$$y_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{e}_i$$

onde $\mathbf{e}_i \sim N(0, \sigma^2 \mathbf{I}_{t_i})$ para $i = 1, \dots, 18$. As estimativas obtidas estão na Tabela 4.13. Pelo teste da razão de verossimilhança generalizada, esse modelo não se adequa aos dados quando comparado ao Modelo 4.

Tabela 4.13: Estimativas (est) dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 5 do Exemplo 2.

Parâmetro		Programa	
		EMMV	EMMVR
α_1	est	11.930	11.930
(11.930)	ep	0.501	0.511
α_2	est	4.030	4.030
(4.030)	ep	0.533	0.543
σ^2	est	6.099	6.333
(1.000)	ep	—	—
lver		-75.818	-77.519
Iterações		2	4

Tabela 4.14: Testes de ajuste de modelos para o Exemplo 2 (máxima verossimilhança).

Modelo	Número de parâmetros ($p + k$)	lver ^a	Modelo de comparação	Q_1	gl	p
1	14	-68.500	—	—	—	—
2	12	-69.763	1	2.526	2	0.283
3	6	-72.313	2	5.100	6	0.531
4	4	-73.391	3	2.156	2	0.340
5	3	-75.818	4	4.854	1	0.028

^a Valor da parte não constante da função de verossimilhança.

Tabela 4.15: Valores médios estimados e estimativas da matriz de covariância/correlação para os modelos do Exemplo 2 (máxima verossimilhança).

Modelo	Δs_j	Médias		Estimativas das variâncias e correlações				
		est	ep	Variância	Matriz de correlação			
1	-1.7	5.354	0.659	5.076	1.000	0.188	-0.175	0.104
	-1.0	7.380	0.611	5.189		1.000	0.581	0.320
	-0.5	9.306	0.546	4.274			1.000	0.576
	0.0	12.311	0.663	7.908				1.000
2	-1.7	4.783	0.581	5.542	1.000	0.183	-0.164	0.121
	-1.0	7.737	0.407	5.415		1.000	0.645	0.305
	-0.5	9.847	0.455	4.641			1.000	0.519
	0.0	11.958	0.620	8.033				1.000
3	-1.7	5.037	0.562	4.712	1.000	0.219	0.185	0.151
	-1.0	7.894	0.416	5.072		1.000	0.341	0.358
	-0.5	9.934	0.464	6.056			1.000	0.467
	0.0	11.974	0.615	7.646				1.000
4	-1.7	5.028	0.631	6.087	1.000	0.695	0.695	0.695
	-1.0	7.897	0.453	6.087		1.000	0.695	0.695
	-0.5	9.947	0.438	6.087			1.000	0.695
	0.0	11.997	0.531	6.087				1.000
5	-1.7	5.079	0.631	6.099	1.000	0.000	0.000	0.000
	-1.0	7.900	0.373	6.099		1.000	0.000	0.000
	-0.5	9.915	0.352	6.099			1.000	0.000
	0.0	11.930	0.501	6.099				1.000

Tabela 4.16: Testes de ajuste de modelos para o Exemplo 2 (máxima verossimilhança restrita).

Modelo	Número de parâmetros ($p + k$)	lver ^a	Modelo de comparação	Q_2	gl	p
1	14	-70.701	—	—	—	—
2	12	-71.284	—	—	—	—
3	6	-73.873	2	5.178	6	0.521
4	4	-75.000	3	2.254	2	0.324
5	3	-77.519	4	5.038	1	0.025

^a Valor da parte não constante da função de verossimilhança restrita.

Tabela 4.17: Valores médios estimados e estimativas da matriz de covariância/correlação para os modelos do Exemplo 2 (máxima verossimilhança restrita).

Modelo	Δs_j	Médias		Estimativas das variâncias e correlações				
		est	ep	Variância	Matriz de correlação			
1	-1.7	5.366	0.689	5.552	1.000	0.183	-0.165	0.109
	-1.0	7.379	0.635	5.590		1.000	0.575	0.316
	-0.5	9.310	0.568	4.603			1.000	0.571
	0.0	12.311	0.682	8.373				1.000
2	-1.7	4.756	0.606	6.006	1.000	0.207	-0.157	0.097
	-1.0	7.728	0.421	5.598		1.000	0.655	0.318
	-0.5	9.850	0.468	4.877			1.000	0.538
	0.0	11.972	0.639	8.431				1.000
3	-1.7	5.025	0.583	5.051	1.000	0.246	0.191	0.139
	-1.0	7.888	0.429	5.249		1.000	0.359	0.368
	-0.5	9.934	0.478	6.281			1.000	0.467
	0.0	11.979	0.634	8.054				1.000
4	-1.7	5.026	0.645	6.368	1.000	0.682	0.682	0.682
	-1.0	7.897	0.467	6.368		1.000	0.682	0.682
	-0.5	9.948	0.451	6.368			1.000	0.682
	0.0	11.999	0.544	6.368				1.000
5	-1.7	5.079	0.643	6.333	1.000	0.000	0.000	0.000
	-1.0	7.900	0.380	6.333		1.000	0.000	0.000
	-0.5	9.915	0.359	6.333			1.000	0.000
	0.0	11.930	0.511	6.333				1.000

Os algoritmos EM realizaram muitas iterações para estimar a matriz de covariância não estruturada nos Modelos 1 e 2. No Modelo 3, eles foram extremamente lentos para atingir a convergência. Notemos, também, que principalmente as estimativas dos elementos de $\mathbf{D}$ diferem um pouco das estimativas obtidas pelos algoritmos “Scoring” e de Newton-Raphson. Por outro lado, os algoritmos “Scoring” e de Newton-Raphson convergiram em poucas iterações para as estimativas procuradas. No ajuste do Modelo 4 (ver Tabela 4.12), os algoritmos EM também fizeram mais iterações do que os algoritmos “Scoring” e de Newton-Raphson. Devemos ressaltar que, nesse caso, esta diferença pode não ser significativa se levarmos em consideração que cada iteração dos algoritmos EM são bem mais rápidas que as dos outros algoritmos.

4.2.3 EXEMPLO 3

Consideremos os dados da Tabela 1.3, envolvendo a observação da ETM de 13 suínos em função de seus pesos. A Figura 1.3 sugere que o valor médio da ETM possa ser representado por uma reta. Nesse sentido, primeiramente supomos que a cada suíno esteja associada uma reta individual com desvios aleatórios em relação a reta populacional. O modelo de efeitos aleatórios considerado para representar essa relação é:

Modelo 1:
$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{X}_i \mathbf{b}_i + \mathbf{e}_i, \quad i = 1, \dots, 13,$$

onde $\mathbf{y}_i = (y_{i1}, \dots, y_{it_i})'$ corresponde a observação da ETM nos t_i pesos do i -ésimo suíno,

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 92 \\ 1 & 68 \\ 1 & 101 \\ 1 & 86 \end{pmatrix}, \quad \dots, \quad \mathbf{X}_4 = \begin{pmatrix} 1 & 63 \\ 1 & 82 \end{pmatrix}, \quad \dots, \quad \mathbf{X}_{13} = \begin{pmatrix} 1 & 93 \\ 1 & 104 \\ 1 & 86 \\ 1 & 62 \end{pmatrix},$$

com $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$, $\mathbf{b}_i = (b_{i1}, b_{i2})'$ e $\mathbf{e}_i = (e_{i1}, \dots, e_{it_i})'$ para $i = 1, \dots, 13$. Os parâmetros α_1 e α_2 representam os coeficientes linear e angular da reta populacional respectivamente; b_{i1} e b_{i2} representam os desvios dos parâmetros da reta associada ao i -ésimo suíno em relação aos parâmetros da reta populacional. Além disso, supomos que $\mathbf{e}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{t_i})$ e $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$, onde

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}.$$

As estimativas obtidas usando os cinco programas mencionados estão apresentadas na Tabela 4.18. Observando essa tabela, vemos que a estimativa de d_{22} (0.003) é mais significativa, quando comparada ao seu erro-padrão (0.001), do que as estimativas de d_{11} e d_{12} . Esse fato sugere o seguinte modelo:

Modelo 2:
$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i, \quad i = 1, \dots, 13,$$

onde $\mathbf{X}_i$ e $\boldsymbol{\alpha}$ são equivalentes ao modelo anterior,

$$\mathbf{Z}_1 = \begin{pmatrix} 92 \\ 68 \\ 101 \\ 86 \end{pmatrix}, \quad \dots, \quad \mathbf{Z}_4 = \begin{pmatrix} 63 \\ 82 \end{pmatrix}, \quad \dots, \quad \mathbf{Z}_{13} = \begin{pmatrix} 93 \\ 104 \\ 86 \\ 62 \end{pmatrix},$$

e $b_i = b_{i1} \sim N(0, \sigma_*^2)$ é o desvio do coeficiente angular da reta associada ao i -ésimo suíno em relação ao coeficiente angular da reta populacional. Os resultados obtidos com o ajuste desse modelo estão apresentados na Tabela 4.19.

Comparando o Modelo 2 com o Modelo 1 através do teste da razão de verossimilhança generalizada, obtivemos para o método de máxima verossimilhança $Q_1 = 8.740$ (gl=2 e $p=0.0127$) e para o método de máxima verossimilhança restrita $Q_2 = 9.366$ (gl=2 e $p=0.0093$), indicando que o Modelo 2 não é apropriado para os dados.

Tabela 4.18: Estimativas dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 1 do Exemplo 3.

Parâmetro		Programa				
		EMMV	SCMV	NRMV	EMMVR	SCMVR
α_1	est	2.501	2.501	2.501	2.487	2.487
(2.966) ^a	ep	1.174	1.174	1.777	1.226	1.226
α_2	est	0.201	0.201	0.201	0.201	0.201
(0.194)	ep	0.016	0.016	0.016	0.017	0.017
d_{11}	est	12.977	12.975	12.975	14.603	14.602
(16.106)	ep	—	7.121	7.396	—	8.391
d_{12}	est	-0.178	-0.178	-0.178	-0.200	-0.200
(-0.216)	ep	—	0.096	0.099	—	0.113
d_{22}	est	0.003	0.003	0.003	0.003	0.003
(0.003)	ep	—	0.001	0.001	—	0.001
σ^2	est	0.486	0.486	0.486	0.484	0.484
(0.473)	ep	—	0.143	0.147	—	0.142
lver			-33.011		-38.191	
Iterações		17	5	5	16	4

^a Valores iniciais calculados com as expressões (3.66), (3.69) e (3.70).

Nesse exemplo, os algoritmos EM tiveram melhor comportamento do que nos Exemplos 1 e 2, quando ajustamos retas com parâmetros aleatórios. No Modelo 2 todos algoritmos convergiram rapidamente (ver Tabela 4.19), pois os valores iniciais utilizados resultaram em ser bem próximos das soluções procuradas.

Tabela 4.19: Estimativas dos parâmetros (est) e respectivos erros-padrões (ep) para o Modelo 2 do Exemplo 3.

Parâmetro		Programa				
		EMMV	SCMV	NRMV	EMMVR	SCMVR
α_1	est	2.768	2.768	2.768	2.767	2.767
(2.966) ^a	ep	0.833	0.833	0.833	0.844	0.844
α_2	est	0.197	0.197	0.197	0.197	0.197
(0.194)	ep	0.011	0.011	0.011	0.011	0.011
σ_*^2	est	2.6×10^{-4}	2.6×10^{-4}	2.6×10^{-4}	2.9×10^{-4}	2.9×10^{-4}
(2.9×10^{-4})	ep	—	$\simeq 0$	$\simeq 0$	—	$\simeq 0$
σ^2	est	0.965	0.965	0.965	0.991	0.992
(0.991)	ep	—	0.027	0.028	—	0.240
lver		-37.381			-42.874	
Iterações		3	3	3	1	1

^a Valores iniciais calculados com as expressões (3.66), (3.67) e (3.68).

4.2.4 Comentários Finais

Como era esperado, a utilização do método de máxima verossimilhança restrita leva a maiores erros-padrões das estimativas dos parâmetros envolvidos do que no método de máxima verossimilhança; já que é levada em consideração a perda de graus de liberdade devida a estimação dos parâmetros associados às médias.

Quanto aos processos iterativos propostos, os algoritmos “Scoring” e de Newton-Raphson se mostraram mais eficientes para obter as estimativas dos parâmetros dos modelos. Possivelmente, os algoritmos EM apresentados devem ter um melhor comportamento quando utilizados com os métodos de aceleração da convergência (ver Laird et al. (1987)).

Apesar de os modelos de efeitos aleatórios serem bastante flexíveis para análise de dados longitudinais não balanceados em relação ao tempo, sua utilização deve ser encarada mais sob o ponto de vista exploratório. Maiores estudos sobre a metodologia apresentada são necessários quando temos amostras pequenas, bem como outras medidas de adequabilidade dos modelos. Ainda mais, a escolha de um modelo deve ser acompanhada de uma análise de resíduos ($y_i - X_i\hat{\alpha}$ e $\hat{b}_i$, para $i = 1, \dots, n$, em modelos de efeitos aleatórios) e da verificação da suposição de normalidade. Algumas técnicas de diagnóstico são encontradas em Dempster e Ryan (1985).

Apêndice A

Derivadas

Neste apêndice apresentamos os cálculos das derivadas parciais de primeira e segunda ordem do logaritmo da função de verossimilhança (Seção A.2) e da função de verossimilhança restrita (Seção A.3). Esses cálculos envolvem alguns resultados sobre álgebra e diferenciação matricial que constam da Seção A.1.

A.1 Alguns Resultados sobre Matrizes

Aqui são apresentados alguns resultados sobre álgebra matricial e diferenciação matricial utilizados neste trabalho. Esses resultados podem ser encontrados em Searle (1982), por exemplo.

1. **Traço:** Se $\mathbf{A}$ é uma matriz quadrada $m \times m$, então $\text{tr } \mathbf{A} = \sum_{i=1}^m a_{ii}$ e $\text{tr } \mathbf{AB} = \text{tr } \mathbf{BA}$ para $\mathbf{A} m \times k$ e $\mathbf{B} k \times m$.
2. **Diferenciação Matricial:** Se y é um escalar, $\mathbf{x}$ um vetor $m \times 1$, $\mathbf{z}$ um vetor $n \times 1$ e $\mathbf{X}$ uma matriz $m \times n$, então $\partial y / \partial \mathbf{x}$ e $[\partial y / \partial \mathbf{x}]'$ são vetores $m \times 1$ e $\partial \mathbf{x} / \partial \mathbf{z}'$, $\partial \mathbf{X} / \partial y$, $\partial^2 y / \partial \mathbf{z}' \partial \mathbf{x}$ e $\partial y / \partial \mathbf{X}$ são matrizes $m \times n$. Assim, para $\mathbf{x}$, $\mathbf{y}$ e $\mathbf{z}$ vetores, $\mathbf{A}$ e $\mathbf{B}$ matrizes de constantes conhecidas e $\mathbf{X}$, $\mathbf{X}_1$ e $\mathbf{X}_2$ matrizes dependendo de um escalar y , todos compatíveis para as operações que serão consideradas, temos:

$$2.1 \quad \partial \mathbf{X}_1 \mathbf{X}_2 / \partial y = (\partial \mathbf{X}_1 / \partial y) \mathbf{X}_2 + \mathbf{X}_1 (\partial \mathbf{X}_2 / \partial y)$$

$$2.2 \quad \partial \mathbf{A} \mathbf{X} \mathbf{B} / \partial y = \mathbf{A}' (\partial \mathbf{X} / \partial y) \mathbf{B}$$

$$2.3 \quad \partial \mathbf{x}' \mathbf{A} / \partial \mathbf{x} = \mathbf{A}$$

$$2.4 \quad \partial \mathbf{A} \mathbf{x} / \partial \mathbf{x} = \mathbf{A}'$$

$$2.5 \quad \partial \mathbf{x}' \mathbf{A} \mathbf{x} / \partial \mathbf{x} = 2 \mathbf{A} \mathbf{x} \text{ para } \mathbf{A} \text{ simétrica}$$

$$2.6 \quad \partial \ln |\mathbf{X}| / \partial y = \text{tr}[\mathbf{X}^{-1} (\partial \mathbf{X} / \partial y)]$$

$$2.7 \quad \partial \mathbf{X}^{-1} / \partial y = -\mathbf{X}^{-1} (\partial \mathbf{X} / \partial y) \mathbf{X}^{-1}$$

$$2.8 \quad \partial \mathbf{y}' / \partial \mathbf{x} = (\partial \mathbf{z}' / \partial \mathbf{x}) (\partial \mathbf{y}' / \partial \mathbf{z})$$

$$2.9 \quad \partial \text{tr} \mathbf{X} \mathbf{A} / \partial \mathbf{X} = \mathbf{A} + \mathbf{A}' - \text{diag}(\mathbf{A}) \text{ para } \mathbf{X} \text{ simétrica, onde } \text{diag}(\mathbf{A}) \text{ denota uma matriz diagonal com os elementos da diagonal iguais aos elementos da diagonal de } \mathbf{A}.$$

$$2.10 \quad \partial \ln |\mathbf{X}| / \partial \mathbf{X} = 2 \mathbf{X}^{-1} - \text{diag}(\mathbf{X}^{-1}) \text{ para } \mathbf{X} \text{ simétrica.}$$

A.2 Derivadas do Logaritmo da Função de Verossimilhança

Vimos na Seção 3.1 que o logaritmo da função de verossimilhança é dado por:

$$l(\mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = K - \frac{1}{2} \left[\sum_{i=1}^n \ln |\boldsymbol{\Sigma}_i| + \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \right], \quad (\text{A.1})$$

onde K é uma constante que não depende de $\boldsymbol{\alpha}$ nem de $\boldsymbol{\theta}$.

Derivando-se $l(\mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta})$ em relação a $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)$, temos:

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\alpha}} &= -\frac{1}{2} \sum_{i=1}^n \frac{\partial [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})]}{\partial \boldsymbol{\alpha}} \\ &= -\frac{1}{2} \sum_{i=1}^n \frac{\partial (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})'}{\partial \boldsymbol{\alpha}} \frac{\partial [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})]}{\partial (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})} \\ &= \frac{1}{2} \sum_{i=1}^n 2 \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \\ &= \sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \end{aligned} \quad (\text{A.2})$$

Usando a notação $\dot{\Sigma}_{i(g)} = \partial \Sigma_i / \partial \theta_g$ temos para $g = 1, \dots, k$:

$$\begin{aligned}
 \frac{\partial l}{\partial \theta_g} &= -\frac{1}{2} \left[\sum_{i=1}^n \frac{\partial \ln |\Sigma_i|}{\partial \theta_g} + \sum_{i=1}^n (y_i - X_i \alpha)' \frac{\partial \Sigma_i^{-1}}{\partial \theta_g} (y_i - X_i \alpha) \right] \\
 &= \frac{1}{2} \left[-\sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} + \sum_{i=1}^n (y_i - X_i \alpha)' \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} (y_i - X_i \alpha) \right] \\
 &= \frac{1}{2} \left[-\sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} + \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} (y_i - X_i \alpha) (y_i - X_i \alpha)' \right] \\
 &= \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} [(y_i - X_i \alpha) (y_i - X_i \alpha)' - \Sigma_i] \\
 &= \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} [(y_i - X_i \alpha) (y_i - X_i \alpha)' - \Sigma_i] \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \quad (A.3)
 \end{aligned}$$

A matriz de segundas derivadas parciais em relação a α é:

$$\begin{aligned}
 \frac{\partial^2 l}{\partial \alpha' \partial \alpha} &= \sum_{i=1}^n \frac{\partial [X_i' \Sigma_i^{-1} (y_i - X_i \alpha)]}{\partial \alpha'} \\
 &= -\sum_{i=1}^n \frac{\partial [X_i' \Sigma_i^{-1} X_i \alpha]}{\partial \alpha'} \\
 &= -\sum_{i=1}^n X_i' \Sigma_i^{-1} X_i \quad (A.4)
 \end{aligned}$$

Para obtermos $\partial^2 l / \partial \alpha_j \partial \theta_g$ observemos que de (A.2) podemos escrever:

$$\frac{\partial l}{\partial \alpha_j} = \sum_{i=1}^n x'_{ij} \Sigma_i^{-1} (y_i - X_i \alpha) \quad \text{para } j = 1, \dots, p,$$

onde x_{ij} é a j -ésima coluna de X_i . Assim, para $j = 1, \dots, p$ e $g = 1, \dots, k$:

$$\begin{aligned}
 \frac{\partial^2 l}{\partial \alpha_j \partial \theta_g} &= \frac{\partial^2 l}{\partial \theta_g \partial \alpha_j} \\
 &= \sum_{i=1}^n \frac{[x'_{ij} \Sigma_i^{-1} (y_i - X_i \alpha)]}{\partial \theta_g} \\
 &= \sum_{i=1}^n x'_{ij} \frac{\partial \Sigma_i^{-1}}{\partial \theta_g} (y_i - X_i \alpha) \\
 &= -\sum_{i=1}^n x'_{ij} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} (y_i - X_i \alpha) \quad (A.5)
 \end{aligned}$$

Usando a notação $\ddot{\Sigma}_{i(g,h)} = \partial^2 \Sigma_i / \partial \theta_g \partial \theta_h$, temos para $g, h = 1, \dots, k$ que:

$$\frac{\partial^2 l}{\partial \theta_g \partial \theta_h} = \frac{1}{2} \frac{\partial}{\partial \theta_g} \left\{ \sum_{i=1}^n \text{tr} \Sigma_i^{-1} [(y_i - X_i \alpha) (y_i - X_i \alpha)' - \Sigma_i] \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \right\}$$

$$\begin{aligned}
&= \frac{1}{2} \frac{\partial}{\partial \theta_g} \left[- \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} + \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \right] \\
&= \frac{1}{2} \left[- \sum_{i=1}^n \text{tr} \frac{\partial(\Sigma_i^{-1} \dot{\Sigma}_{i(h)})}{\partial \theta_g} + \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \frac{\partial(\Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1})}{\partial \theta_g} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \right]
\end{aligned}$$

Como

$$\begin{aligned}
\frac{\partial(\Sigma_i^{-1} \dot{\Sigma}_{i(h)})}{\partial \theta_g} &= \frac{\partial \Sigma_i^{-1}}{\partial \theta_g} \dot{\Sigma}_{i(h)} + \Sigma_i^{-1} \frac{\partial \dot{\Sigma}_{i(h)}}{\partial \theta_g} \\
&= -\Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} + \Sigma_i^{-1} \ddot{\Sigma}_{i(g,h)}
\end{aligned}$$

e

$$\begin{aligned}
\frac{\partial(\Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1})}{\partial \theta_g} &= \frac{\partial(\Sigma_i^{-1} \dot{\Sigma}_{i(h)})}{\partial \theta_g} \Sigma_i^{-1} + \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \frac{\partial \Sigma_i^{-1}}{\partial \theta_g} \\
&= \Sigma_i^{-1} \ddot{\Sigma}_{i(g,h)} \Sigma_i^{-1} - \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} - \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1},
\end{aligned}$$

temos que

$$\begin{aligned}
\frac{\partial^2 l}{\partial \theta_g \partial \theta_h} &= -\frac{1}{2} \sum_{i=1}^n \text{tr} \left(\Sigma_i^{-1} \ddot{\Sigma}_{i(g,h)} - \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \right) \\
&\quad + \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \Sigma_i^{-1} \ddot{\Sigma}_{i(g,h)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \\
&\quad - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \\
&\quad - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}).
\end{aligned}$$

Usando o fato que $(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})$ é escalar e que Σ_i^{-1} , $\dot{\Sigma}_{i(g)}$ e $\dot{\Sigma}_{i(h)}$ são matrizes simétricas, temos que

$$(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) = (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}).$$

Logo,

$$\begin{aligned}
\frac{\partial^2 l}{\partial \theta_g \partial \theta_h} &= -\frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \ddot{\Sigma}_{i(g,h)} + \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \\
&\quad + \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \ddot{\Sigma}_{i(g,h)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})'
\end{aligned}$$

$$\begin{aligned}
 & - \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \\
 = & \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \ddot{\Sigma}_{i(g,h)} \Sigma_i^{-1} [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \Sigma_i] \\
 & + \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \\
 & - \frac{2}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \\
 = & \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \Sigma_i] \Sigma_i^{-1} \ddot{\Sigma}_{i(g,h)} \\
 & - \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} [2(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \Sigma_i] \\
 = & - \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \dot{\Sigma}_{i(g)} \Sigma_i^{-1} [2(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \Sigma_i] \Sigma_i^{-1} \dot{\Sigma}_{i(h)} \\
 & + \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \Sigma_i] \Sigma_i^{-1} \ddot{\Sigma}_{i(g,h)} \quad (\text{A.6})
 \end{aligned}$$

Quando $\Sigma_i = \sum_{g=1}^k \theta_g \mathbf{G}_{ig}$, onde $\mathbf{G}_{ig}$'s são matrizes de constantes conhecidas, temos:

$$\dot{\Sigma}_{i(g)} = \frac{\partial \Sigma_i}{\partial \theta_g} = \mathbf{G}_{ig} \quad \text{e} \quad \ddot{\Sigma}_{i(g,h)} = \frac{\partial^2 \Sigma_i}{\partial \theta_g \partial \theta_h} = 0.$$

Assim, as expressões (A.3), (A.5) e (A.6) podem ser simplificadas, resultando em:

$$\frac{\partial l}{\partial \theta_g} = \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \Sigma_i] \Sigma_i^{-1} \mathbf{G}_{ig} \quad \text{para } g = 1, \dots, k, \quad (\text{A.7})$$

$$\frac{\partial^2 l}{\partial \alpha_j \partial \theta_g} = - \sum_{i=1}^n \mathbf{x}_{ij}' \Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \quad \text{para } j = 1, \dots, p \quad \text{e} \quad g = 1, \dots, k \quad (\text{A.8})$$

e

$$\frac{\partial^2 l}{\partial \theta_g \partial \theta_h} = - \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} [2(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \Sigma_i] \Sigma_i^{-1} \mathbf{G}_{ih} \quad \text{para } g, h = 1, \dots, k. \quad (\text{A.9})$$

A.3 Derivadas do Logaritmo da Função de Verossimilhança Restrita

Na Seção 3.1 vimos que o logaritmo da função de verossimilhança restrita é dado por:

$$l_R(\mathbf{y}; \boldsymbol{\theta}) = K - \frac{1}{2} \left[\sum_{i=1}^n \ln |\Sigma_i| + \ln \left| \sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} \mathbf{X}_i \right| + \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}) \right], \quad (\text{A.10})$$

onde $\tilde{\boldsymbol{\alpha}} = (\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} \mathbf{y}_i)$ e K é uma constante que não depende de $\boldsymbol{\theta}$.

As derivadas de $l_R(\mathbf{y}; \boldsymbol{\theta})$ em relação a θ_g ($g = 1, \dots, k$) são mais facilmente calculadas se escrevermos $l_R(\mathbf{y}; \boldsymbol{\theta})$ na forma utilizada por Harville (1974, 1977):

$$l_R(\mathbf{y}; \boldsymbol{\theta}) = K - \frac{1}{2} \left[\ln |\mathbf{V}| + \ln |\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}| + (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}}) \right], \quad (\text{A.11})$$

com $\mathbf{y} = (\mathbf{y}_1', \dots, \mathbf{y}_n')'$, $\mathbf{V} = \text{diag}(\Sigma_1, \dots, \Sigma_n)$, $\mathbf{X} = (\mathbf{X}_1', \dots, \mathbf{X}_n')'$ e

$$\tilde{\boldsymbol{\alpha}} = (\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} \mathbf{y}_i) = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}.$$

Também podemos escrever:

$$\begin{aligned} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}}) &= \mathbf{V}^{-1}[\mathbf{y} - \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}] \\ &= \mathbf{V}^{-1}[\mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}] \mathbf{y} \\ &= \mathbf{P} \mathbf{y} \end{aligned} \quad (\text{A.12})$$

com $\mathbf{P} = \mathbf{V}^{-1}[\mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}]$ e

$$\begin{aligned} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}}) &= (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}})' \mathbf{V}^{-1} \mathbf{V} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}}) \\ &= \mathbf{y}' \mathbf{P} \mathbf{V} \mathbf{P} \mathbf{y} \\ &= \mathbf{y}' \mathbf{V}^{-1} [\mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}] [\mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}] \mathbf{y} \\ &= \mathbf{y}' \mathbf{V}^{-1} [\mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}] \mathbf{y} \\ &= \mathbf{y}' \mathbf{P} \mathbf{y} \end{aligned} \quad (\text{A.13})$$

Para calcular as derivadas de $l_R(\mathbf{y}; \boldsymbol{\theta})$ em relação a $\theta_1, \dots, \theta_k$ utilizaremos a seguinte

notação:

$$\dot{\mathbf{V}}_{(g)} = \frac{\partial \mathbf{V}}{\partial \theta_g} \quad \text{e} \quad \ddot{\mathbf{V}}_{(g,h)} = \frac{\partial^2 \mathbf{V}}{\partial \theta_g \partial \theta_h}$$

e o seguinte resultado

$$\begin{aligned} \frac{\partial \mathbf{P}}{\partial \theta_g} &= \frac{\partial \mathbf{V}^{-1} [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]}{\partial \theta_g} \\ &= \frac{\partial \mathbf{V}^{-1}}{\partial \theta_g} [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}] + \mathbf{V}^{-1} \frac{\partial}{\partial \theta_g} [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}] \\ &= -\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}] \\ &\quad - \mathbf{V}^{-1} \left[\mathbf{X} \frac{\partial (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}}{\partial \theta_g} \mathbf{X}'\mathbf{V}^{-1} + \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}' \frac{\partial \mathbf{V}^{-1}}{\partial \theta_g} \right] \\ &= -\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{P} \\ &\quad + \mathbf{V}^{-1} \left[\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}' \frac{\partial \mathbf{V}^{-1}}{\partial \theta_g} \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \right. \\ &\quad \left. + \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} \right] \\ &= -\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{P} \\ &\quad - \mathbf{V}^{-1} \left[\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \right. \\ &\quad \left. - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} \right] \\ &= -\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{P} \\ &\quad + \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}] \\ &= -\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{P} + \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{P} \\ &= -\mathbf{V}^{-1} [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}] \dot{\mathbf{V}}_{(g)} \mathbf{P} \\ &= -\mathbf{P} \dot{\mathbf{V}}_{(g)} \mathbf{P} \end{aligned} \tag{A.14}$$

Assim, temos:

$$\begin{aligned} \frac{\partial l_R}{\partial \theta_g} &= -\frac{1}{2} \frac{\partial}{\partial \theta_g} [\ln |\mathbf{V}| + \ln |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}\mathbf{y}] \\ &= -\frac{1}{2} \left[\frac{\partial \ln |\mathbf{V}|}{\partial \theta_g} + \frac{\partial \ln |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|}{\partial \theta_g} + \frac{\partial (\mathbf{y}'\mathbf{P}\mathbf{y})}{\partial \theta_g} \right] \\ &= -\frac{1}{2} [\text{tr } \mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} - \text{tr } (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} \mathbf{X} - \mathbf{y}'\mathbf{P} \dot{\mathbf{V}}_{(g)} \mathbf{P}\mathbf{y}] \\ &= -\frac{1}{2} [\text{tr } \mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} - \text{tr } \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} - \mathbf{y}'\mathbf{P} \dot{\mathbf{V}}_{(g)} \mathbf{P}\mathbf{y}] \end{aligned}$$

$$\begin{aligned}
 &= -\frac{1}{2} \left[\text{tr} \mathbf{V}^{-1} [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}] \dot{\mathbf{V}}_{(g)} - \mathbf{y}'\mathbf{P}\dot{\mathbf{V}}_{(g)}\mathbf{P}\mathbf{y} \right] \\
 &= -\frac{1}{2} \text{tr} \mathbf{P}\dot{\mathbf{V}}_{(g)} + \frac{1}{2} \mathbf{y}'\mathbf{P}\dot{\mathbf{V}}_{(g)}\mathbf{P}\mathbf{y} \quad \text{para } g = 1, \dots, k
 \end{aligned}$$

ou ainda, usando (A.12)

$$\begin{aligned}
 \frac{\partial l_R}{\partial \theta_g} &= -\frac{1}{2} \text{tr} \mathbf{P}\dot{\mathbf{V}}_{(g)} + \frac{1}{2} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}})' \mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}}) \\
 &= -\frac{1}{2} \text{tr} \mathbf{P}\dot{\mathbf{V}}_{(g)} + \frac{1}{2} \text{tr} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}})' \mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}}) \\
 &= -\frac{1}{2} \text{tr} \mathbf{P}\dot{\mathbf{V}}_{(g)} + \frac{1}{2} \text{tr} \mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}}) (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}})' \\
 &= \frac{1}{2} \text{tr} \left[\mathbf{V}^{-1} \dot{\mathbf{V}}_{(g)} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}}) (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}})' - \mathbf{P}\dot{\mathbf{V}}_{(g)} \right] \quad \text{para } g = 1, \dots, k. \quad (\text{A.15})
 \end{aligned}$$

A segunda derivada parcial de $l_R(\mathbf{y}; \boldsymbol{\theta})$ é:

$$\begin{aligned}
 \frac{\partial^2 l_R}{\partial \theta_g \partial \theta_h} &= -\frac{1}{2} \frac{\partial \text{tr} \mathbf{P}\dot{\mathbf{V}}_{(h)}}{\partial \theta_g} + \frac{1}{2} \mathbf{y}' \frac{\partial (\mathbf{P}\dot{\mathbf{V}}_{(h)}\mathbf{P})}{\partial \theta_g} \mathbf{y} \\
 &= -\frac{1}{2} \text{tr} \left[\frac{\partial \mathbf{P}\dot{\mathbf{V}}_{(h)}}{\partial \theta_g} \right] + \frac{1}{2} \mathbf{y}' \left[\frac{\partial \mathbf{P}\dot{\mathbf{V}}_{(h)}}{\partial \theta_g} \mathbf{P} + \mathbf{P}\dot{\mathbf{V}}_{(h)} \frac{\partial \mathbf{P}}{\partial \theta_g} \right] \mathbf{y} \\
 &= -\frac{1}{2} \text{tr} \left[\frac{\partial \mathbf{P}}{\partial \theta_g} \dot{\mathbf{V}}_{(h)} + \mathbf{P} \frac{\partial \dot{\mathbf{V}}_{(h)}}{\partial \theta_g} \right] \\
 &\quad + \frac{1}{2} \mathbf{y}' \left[\frac{\partial \mathbf{P}}{\partial \theta_g} \dot{\mathbf{V}}_{(h)} \mathbf{P} + \mathbf{P} \frac{\partial \dot{\mathbf{V}}_{(h)}}{\partial \theta_g} \mathbf{P} + \mathbf{P}\dot{\mathbf{V}}_{(h)} \frac{\partial \mathbf{P}}{\partial \theta_g} \right] \mathbf{y}
 \end{aligned}$$

Usando (A.14), obtemos

$$\begin{aligned}
 \frac{\partial^2 l_R}{\partial \theta_g \partial \theta_h} &= -\frac{1}{2} \text{tr} [-\mathbf{P}\dot{\mathbf{V}}_{(g)}\mathbf{P}\dot{\mathbf{V}}_{(h)} + \mathbf{P}\ddot{\mathbf{V}}_{(g,h)}] \\
 &\quad + \frac{1}{2} \mathbf{y}' [-\mathbf{P}\dot{\mathbf{V}}_{(g)}\mathbf{P}\dot{\mathbf{V}}_{(h)}\mathbf{P} + \mathbf{P}\ddot{\mathbf{V}}_{(g,h)}\mathbf{P} - \mathbf{P}\dot{\mathbf{V}}_{(h)}\mathbf{P}\dot{\mathbf{V}}_{(g)}\mathbf{P}] \mathbf{y}
 \end{aligned}$$

Como $\mathbf{y}'\mathbf{P}\dot{\mathbf{V}}_{(g)}\mathbf{P}\dot{\mathbf{V}}_{(h)}\mathbf{P}\mathbf{y}$ é escalar e $\mathbf{P}$, $\dot{\mathbf{V}}_{(g)}$, $\dot{\mathbf{V}}_{(h)}$ são simétricas, temos:

$$\mathbf{y}'\mathbf{P}\dot{\mathbf{V}}_{(g)}\mathbf{P}\dot{\mathbf{V}}_{(h)}\mathbf{P}\mathbf{y} = \mathbf{y}'\mathbf{P}\dot{\mathbf{V}}_{(h)}\mathbf{P}\dot{\mathbf{V}}_{(g)}\mathbf{P}\mathbf{y}.$$

Logo,

$$\frac{\partial^2 l_R}{\partial \theta_g \partial \theta_h} = -\frac{1}{2} \text{tr} [\mathbf{P}\ddot{\mathbf{V}}_{(g,h)} - \mathbf{P}\dot{\mathbf{V}}_{(g)}\mathbf{P}\dot{\mathbf{V}}_{(h)}]$$

$$\begin{aligned}
& + \frac{1}{2} \mathbf{y}' \mathbf{P} [\ddot{\mathbf{V}}_{(g,h)} - 2 \dot{\mathbf{V}}_{(g)} \mathbf{P} \dot{\mathbf{V}}_{(h)}] \mathbf{P} \mathbf{y} \\
& = -\frac{1}{2} \text{tr} [\mathbf{P} \ddot{\mathbf{V}}_{(g,h)} - \mathbf{P} \dot{\mathbf{V}}_{(g)} \mathbf{P} \dot{\mathbf{V}}_{(h)}] \\
& \quad + \frac{1}{2} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}})' \mathbf{V}^{-1} [\ddot{\mathbf{V}}_{(g,h)} - 2 \dot{\mathbf{V}}_{(g)} \mathbf{P} \dot{\mathbf{V}}_{(h)}] \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}}) \\
& = \frac{1}{2} \text{tr} \mathbf{V}^{-1} [\ddot{\mathbf{V}}_{(g,h)} - 2 \dot{\mathbf{V}}_{(g)} \mathbf{P} \dot{\mathbf{V}}_{(h)}] \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}}) (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\alpha}})' \\
& \quad - \frac{1}{2} \text{tr} [\mathbf{P} \ddot{\mathbf{V}}_{(g,h)} - \mathbf{P} \dot{\mathbf{V}}_{(g)} \mathbf{P} \dot{\mathbf{V}}_{(h)}] \quad \text{para } g, h = 1, \dots, k. \quad (\text{A.16})
\end{aligned}$$

Para reescrevermos (A.15) e (A.16) na notação do modelo (2.1), devemos observar que: $\mathbf{V} = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_n) \Rightarrow \mathbf{V}^{-1} = \text{diag}(\boldsymbol{\Sigma}_1^{-1}, \dots, \boldsymbol{\Sigma}_n^{-1}) \Rightarrow \dot{\mathbf{V}}_{(g)} = \text{diag}(\dot{\boldsymbol{\Sigma}}_{1(g)}, \dots, \dot{\boldsymbol{\Sigma}}_{n(g)})$ e $\ddot{\mathbf{V}}_{(g,h)} = \text{diag}(\ddot{\boldsymbol{\Sigma}}_{1(g,h)}, \dots, \ddot{\boldsymbol{\Sigma}}_{n(g,h)})$, onde $\dot{\boldsymbol{\Sigma}}_{i(g)} = \partial \boldsymbol{\Sigma}_i / \partial \theta_g$ e $\ddot{\boldsymbol{\Sigma}}_{i(g,h)} = \partial^2 \boldsymbol{\Sigma}_i / \partial \theta_g \partial \theta_h$, e ainda que $\mathbf{P} = \mathbf{V}^{-1} [\mathbf{I} - \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}]$ tem a forma:

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{P}_{12} & \dots & \mathbf{P}_{1n} \\ \mathbf{P}_{21} & \mathbf{P}_2 & \dots & \mathbf{P}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{n1} & \mathbf{P}_{n2} & \dots & \mathbf{P}_n \end{pmatrix},$$

onde

$$\mathbf{P}_i = \boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1}$$

e

$$\mathbf{P}_{ij} = -\boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}_j' \boldsymbol{\Sigma}_j^{-1} \quad \text{para } i \neq j \text{ e } i, j = 1, \dots, n.$$

Então, (A.15) será dada por:

$$\frac{\partial l_R}{\partial \theta_g} = \frac{1}{2} \sum_{i=1}^n \text{tr} [\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(g)} \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}) (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' - \mathbf{P}_i \dot{\boldsymbol{\Sigma}}_{i(g)}] \quad \text{para } g = 1, \dots, k, \quad (\text{A.17})$$

e (A.16) por:

$$\begin{aligned}
\frac{\partial^2 l_R}{\partial \theta_g \partial \theta_h} & = \frac{1}{2} \sum_{i=1}^n \text{tr} \boldsymbol{\Sigma}_i^{-1} [\ddot{\boldsymbol{\Sigma}}_{i(g,h)} - 2 \dot{\boldsymbol{\Sigma}}_{i(g)} \mathbf{P}_i \dot{\boldsymbol{\Sigma}}_{i(h)}] \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}) (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' \\
& \quad - \frac{1}{2} \sum_{i=1}^n \text{tr} [\mathbf{P}_i \ddot{\boldsymbol{\Sigma}}_{i(g,h)} - \mathbf{P}_i \dot{\boldsymbol{\Sigma}}_{i(g)} \mathbf{P}_i \dot{\boldsymbol{\Sigma}}_{i(h)}] \quad \text{para } g, h = 1, \dots, k. \quad (\text{A.18})
\end{aligned}$$

Quando $\boldsymbol{\Sigma}_i = \sum_{g=1}^k \theta_g \mathbf{G}_{ig}$, onde $\mathbf{G}_{ig}$'s são matrizes de constantes conhecidas, temos:

$$\dot{\boldsymbol{\Sigma}}_{i(g)} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \theta_g} = \mathbf{G}_{ig} \quad \text{e} \quad \ddot{\boldsymbol{\Sigma}}_{i(g,h)} = \frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial \theta_g \partial \theta_h} = \mathbf{0}.$$

Assim, as equações (A.17) e (A.18) podem ser simplificadas, resultando em:

$$\frac{\partial l_R}{\partial \theta_g} = \frac{1}{2} \sum_{i=1}^n \text{tr} [\Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})(\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' - \mathbf{P}_i \mathbf{G}_{ig}] \text{ para } g = 1, \dots, k \quad (\text{A.19})$$

e

$$\begin{aligned} \frac{\partial^2 l_R}{\partial \theta_g \partial \theta_h} = & - \sum_{i=1}^n \text{tr} \Sigma_i^{-1} \mathbf{G}_{ig} \mathbf{P}_i \dot{\Sigma}_{i(h)} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})(\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' \\ & + \frac{1}{2} \sum_{i=1}^n \text{tr} \mathbf{P}_i \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \text{ para } g, h = 1, \dots, k. \end{aligned} \quad (\text{A.20})$$

Apêndice B

Programas

Foram desenvolvidos programas com o módulo CM do pacote científico SOC (EMBRAPA) para o ajuste de modelos de efeitos aleatórios descritos na Seção 2.2 com estrutura de covariância da forma:

$$\Sigma_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{t_i} \quad (\text{B.1})$$

Mais especificamente, temos cinco programas:

EMMV.AJU	Estimação por máxima verossimilhança usando o algoritmo EM.
SCMV.AJU	Estimação por máxima verossimilhança usando o algoritmo “Scoring” de Fisher.
NRMV.AJU	Estimação por máxima verossimilhança usando o algoritmo de Newton-Raphson.
EMMVR.AJU	Estimação por máxima verossimilhança restrita usando o algoritmo EM.
SCMVR.AJU	Estimação por máxima verossimilhança restrita usando o algoritmo “Scoring” de Fisher.

Os programas SCMV.AJU, NRMV.AJU e SCMVR.AJU são dirigidos especificamente para o ajuste de modelos de efeitos aleatórios com matriz de covariância da forma (B.1), onde $\mathbf{D}$ é não estruturada. Os programas EMMV.AJU e EMMVR.AJU podem ajustar modelos semelhantes, onde:

- (i) $\mathbf{D} = \text{diag}(\mathbf{D}_1, \dots, \mathbf{D}_f)$, onde f é o número de subpopulações e $\mathbf{D}_1, \dots, \mathbf{D}_f$ são matrizes de covariância não estruturadas (ver (2.21)).
- (ii) $\Sigma_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' = \mathbf{E}_i \mathbf{D} \mathbf{E}_i'$, onde $\mathbf{D}$ é não estruturada (Basta especificar $\sigma^{2(0)} = 0.00001$).
- (iii) $\Sigma_i = \sigma^2 \mathbf{I}_{t_i}$ (modelos de Regressão — Basta especificar $\mathbf{Z}_i = \mathbf{0}_{t_i}$, $i = 1, \dots, n$).

O critério de parada das iterações é que a maior diferença entre os valores das estimativas de iterações sucessivas seja menor que a precisão desejada.

B.1 Organização do Arquivo de Dados

A matriz de dados deve conter colunas correspondentes às seguintes variáveis:

Código	Indica o número da unidade experimental. A primeira unidade experimental deve ter código igual 1 e a n -ésima código igual a n .
X_{ij}	Indica os valores da j -ésima coluna das matrizes X_i 's, $j = 1, \dots, p$.
Z_{ij}	Indica os valores da j -ésima coluna das matrizes Z_i 's, $j = 1, \dots, q$.
y_i	Indica os valores observados da variável resposta.

Cada linha corresponde aos dados de uma unidade experimental em uma determinada ocasião de observação.

Exemplo: O conjunto de dados para o ajuste dos modelos no Exemplo 4.3 tem a forma:

1	1	92	20.00
1	1	68	14.33
1	1	101	23.00
1	1	86	18.00
2	1	64	15.33
2	1	89	21.67
2	1	98	23.00
2	1	78	18.33
.	.	.	.
.	.	.	.
.	.	.	.
13	1	93	20.33
13	1	104	23.67
13	1	86	19.00
13	1	62	15.00

Notemos que a primeira coluna do conjunto de dados é a coluna de códigos, a segunda e a terceira definem as matrizes X_i 's e Z_i 's e a quarta coluna contém a variável resposta.

Essa matriz de dados deve ser transformada num arquivo do tipo SOC através do procedimento GENESE.

B.2 Dados de Entrada

Para a execução dos programas são necessários os seguintes dados:

DADOS	Matriz contendo o arquivo de dados (arquivo SOC).
n	Número de unidades experimentais.
cod	Número da <i>coluna</i> da matriz DADOS que contem os códigos das unidades experimentais.
cix	Número da <i>coluna</i> da matriz DADOS onde começa a especificação das matrizes X_i 's.
cfx	Número da <i>coluna</i> da matriz DADOS onde termina a especificação das matrizes X_i 's.
ciz	Número da <i>coluna</i> da matriz DADOS onde começa a especificação das matrizes Z_i 's.
cfz	Número da <i>coluna</i> da matriz DADOS onde termina a especificação das matrizes Z_i 's.
cy	Número da <i>coluna</i> da matriz DADOS que contem as observações da variável resposta (y_i 's).
alfa_0	Vetor ($p \times 1$) contendo o valor inicial para $\hat{\alpha}$ (1).
sigma2_0	Valor inicial para $\hat{\sigma}^2$ (2).
D_0	Matriz ($q \times q$) contendo o valor inicial da matriz $\hat{D}$ (3).
max_it	Número máximo de iterações desejadas.
precisao	Precisão desejada para as estimativas (Ex: 0.001).
imp_it=0	Para imprimir somente os resultados da iteração 0 e da iteração final.
imp_it=1	Para imprimir os resultados de todas as iterações.
imp_bi=0	Para não imprimir i , r_i , $\hat{b}_i$ e $\widehat{\text{Var}}(\hat{b}_i - b_i)$.
imp_bi=1	Para imprimir i , r_i , $\hat{b}_i$ e $\widehat{\text{Var}}(\hat{b}_i - b_i)$.

Observações:

- (1) Não especificar se for usar EMMVR.AJU ou SCMVR.AJU.
- (2) Deve ser um valor positivo.
- (3) A matriz D_0 deve ser positiva definida.

B.3 Dados de Saída

As saídas que os programas fornecem são:

<code>max_it</code>	Número máximo de iterações permitidas.
<code>precisao</code>	Precisão especificada.

Para as iterações solicitadas (ver `imp_it` na Seção B.2):

<code>iteracao</code>	Número da iteração.
<code>alfa</code>	Estimativa de α (programas: EMMV.AJU, SCMV.AJU, EMMVR.AJU e SCMVR.AJU).
<code>sigma2</code>	Estimativa de σ^2 (programas: EMMV.AJU e EMMVR.AJU).
<code>D</code>	Estimativa de D (programas: EMMV.AJU e EMMVR.AJU).
<code>theta</code>	Vetor cujas primeiras $q(q+1)/2$ linhas contêm as estimativas dos elementos da parte triangular superior de D e o último elemento é a estimativa de σ^2 (programas: SCMV.AJU e SCMVR.AJU).
<code>phi</code>	Vetor contendo nas primeiras p linhas a estimativa de α , nas seguintes $q(q+1)/2$ linhas as estimativas dos elementos da parte triangular superior de D e na última linha a estimativa de σ^2 (programa: NRMV.AJU).
<code>lver</code>	Valor da parte não constante do logaritmo da função de verossimilhança.

Para a última iteração realizada, temos:

<code>var_alfa</code>	Matriz $(p \times p)$ contendo a estimativa da matriz de covariância de $\hat{\alpha}$ (programas: EMMV.AJU, SCMV.AJU, EMMVR.AJU e SCMVR.AJU).
<code>var_thet</code>	Matriz $[q(q+1)/2+1 \times q(q+1)/2+1]$ contendo a estimativa da matriz de covariância dos elementos de $\hat{\mathbf{D}}$ e $\hat{\sigma}^2$ (programas: SCMV.AJU e SCMVR.AJU).
<code>var_phi</code>	Matriz contendo as estimativas da matriz de covariância de $\hat{\alpha}$, dos elementos de $\hat{\mathbf{D}}$ e $\hat{\sigma}^2$ (programa: NRMV.AJU).

Opcionalmente, temos para a última iteração realizada (se `imp_it = 1`):

<code>ind</code>	Número da unidade experimental.
<code>ri</code>	Vetor de resíduos.
<code>bvb</code>	Matriz cuja primeira coluna contem as “estimativas” dos efeitos aleatórios $(\hat{\mathbf{b}}_i, i = 1, \dots, n)$, e as outras colunas contêm as correspondentes estimativas de $\widehat{\text{Var}}(\hat{\mathbf{b}}_i - \mathbf{b}_i)$.

B.4 Exemplos

Exemplo 1: Os resultados do Modelo 2 (EMMVR) do Exemplo 2 foram obtidos através dos seguintes comandos:

```

/*****
/*                               ENTRADA DOS DADOS                               */
*****/

/*  Leitura do arquivo de dados  */
DADOS=leia "c:iso24";

/*  Parametros de controle dos dados  */

n=18;                /* numero de unidades experimentais */
cod=2;               /* coluna de codigos                */
cix=3;               /* coluna inicial de X              */
cfx=4;               /* coluna final de X                */
ciz=7;               /* coluna inicial de Z              */
cfz=10;              /* coluna final de Z                */
cy=6;                /* coluna de y                      */

/*  Estimativas iniciais dos parametros  */
/*  alfa_0 = { 11.9304 , 4.0302 }; */ /* nao colocar alfa_0 se for usar */
/*                               /* EMMVR.AJU ou SCMVR.AJU          */
sigma2_0 = 0.00001;
D_0 = { 6 0 0 0, 0 6 0 0, 0 0 6 0, 0 0 0 6};

/*  Parametros de controle das iteracoes  */
max_it=25;           /* numero maximo de iteracoes permitidas */
precisao=0.001;      /* precisao desejada                      */

/*  Parametros de controle da impressao  */
imp_it=1;            /* se 1 ==> imprime todas iteracoes      */
/*                               /* se 0 ==> imprime iteracoes 0 e a final */
imp_bi=1;            /* se 1 ==> imprime i,ri,biest e v(biest-bi) */
/*                               /* se 0 ==> nao imprime i,ri,biest e v(biest-bi) */

/* Impressao dos titulos */
anote "ISO24 - MODELO 2 - RETA - COV NAO ESTRUTURADA - MVR";

```

```

/*****
/*      Escolha do metodo de estimacao e do processo iterativo      */
/*
/*
/* Opcoes:
/*
/* EMMV.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/*      usando o algoritmo EM
/* SCMV.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/*      usando o algoritmo Scoring de Fisher
/* NRMV.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/*      usando o algoritmo de Newton-Raphson
/* EMMVR.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/*      restrita usando o algoritmo EM
/* SCMVR.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/*      restrita usando o algoritmo Scoring de Fisher
*****/
exec "EMMVR.AJU";

fim;

```

Exemplo 2: Os resultados do Modelo 1 (SCMV) do Exemplo 3 foram obtidos através dos seguintes comandos:

```

/*****
/*
/*      ENTRADA DOS DADOS
/*
*****/

/*  Leitura do arquivo de dados  */
DADOS=leia "c:suino";

/*  Parametros de controle dos dados  */

n=13;          /* numero de unidades experimentais */
cod=1;         /* coluna de codigos
cix=2;         /* coluna inicial de X
cfx=3;         /* coluna final de X
ciz=2;         /* coluna inicial de Z
cfz=3;         /* coluna final de Z
cy=4;         /* coluna de y

```

```

/* Estimativas iniciais dos parametros */
alfa_0 = { 2.96584 , 0.194122 }; /* nao colocar alfa_0 se for usar */
                                   /* EMMVR.AJU ou SCMVR.AJU          */
sigma2_0 = 0.473413;
D_0 = { 16.105833 (-0.216484) , (-0.216484) 0.003183 };

/* Parametros de controle das iteracoes */
max_it=25;           /* numero maximo de iteracoes permitidas */
precisao=0.001;      /* precisao desejada          */

/* Parametros de controle da impressao */
imp_it=1;           /* se 1 ==> imprime todas iteracoes          */
                   /* se 0 ==> imprime iteracoes 0 e a final    */
imp_bi=1;           /* se 1 ==> imprime i,ri,biest e v(biest-bi)  */
                   /* se 0 ==> nao imprime i,ri,biest e v(biest-bi) */

/* Impressao dos titulos */
anote "SUINOS - MODELO 1 - RETAS ALEATORIAS - MV";

/*****
/* Escolha do metodo de estimacao e do processo iterativo */
/*
/*
/* Opcoes:
/*
/* EMMV.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/* usando o algoritmo EM */
/* SCMV.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/* usando o algoritmo Scoring de Fisher */
/* NRMV.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/* usando o algoritmo de Newton-Raphson */
/* EMMVR.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/* restrita usando o algoritmo EM */
/* SCMVR.AJU Estimacao atraves do metodo de maxima verossimilhanca */
/* restrita usando o algoritmo Scoring de Fisher */
*****/
exec "SCMV.AJU";

fim;

```


B.5 Listagens dos Programas

```

/*****
/*          PROGRAMA EMMV.AJU          */
/*****
/*  Programa em CM do SOC para estimacao dos parametros de      */
/*  modelos de efeitos aleatorios atraves do metodo de maxima    */
/*  verossimilhanca usando o algoritmo EM                        */
/*****

/*  Realocacao dos dados */
mat = {DADOS[,cod] DADOS[,cix:cfx] DADOS[,cy] DADOS[,ciz:cfz]};
DADOS=;
cod = 1;
cfx = cfx - cix + 2;
cix = 2;
cy  = cfx + 1;
cfz = (cfz-ciz);
ciz = cy + 1;
cfz = cfz + ciz;

/*  Calcula dimensoes de alfa, D e mat */
p = nlin(alfa_0);      /* p = dimensao de alfa      */
q = nlin(D_0);         /* q = dimensao de D      */
no = nlin(mat);        /* no = numero total de observacoes */

/*  Impressao dos parametros de controle das iteracoes */
anote " ";
anote " ";
anote "Metodo de Maxima Verossimilhanca";
anote "Algoritmo EM";
anote " ";
anote " ";
anote "Parametros de controle das iteracoes:";
imprime max_it 8:0 (" ");
imprime precisao 8:6 (" ");

/*  Inicio do processo iterativo */
alfa = alfa_0;
sigma2 = sigma2_0;
D = D_0;
invd = inv(D_0);
p_par = 1000;
iteracao = 0;

```

```

enquanto ((iteracao<max_it)&&(p_par>=precisao)) {
  alfa_0 = alfa;
  sigma2_0 = sigma2;
  D_0 = D;
  srzb = 0;
  striw = 0;
  sbb = D - D;
  szwz = D - D;
  sxwx = ident(p) - ident(p);
  sxwy = {(1:p)}' - {(1:p)}';
/* calcula a verossimilhanca da iteracao anterior */
/* e os proximos D, sigma2 e alfa */
  lver = 0;
  i = 1;
  enquanto (i<=n) {
    ind = cata(mat,cod,i,"=");
    xi = ind[,cix:cfx];
    zi = ind[,ciz:cfz];
    yi = ind[,cy];
    wi = (ident(nlin(zi)) - zi*inv(sigma2_0*invd + zi'*zi)*zi')/sigma2_0;
    ri = yi - xi*alfa_0;
    bi = D_0*zi'*wi*ri;
    lver = lver + log(det(wi)) - ri'*wi*ri;
    srzb = srzb + (ri-zi*bi)'*(ri-zi*bi);
    striw = striw + traco(ident(nlin(zi)) - sigma2_0*wi);
    sbb = sbb + bi*bi';
    szwz = szwz + zi'*wi*zi;
    i = i + 1;
  }
  lver = (0.5)*lver;
  se ((imp_it==1)||((iteracao==0)) {
    anote"=====";
    imprime iteracao 3:0 (" ");
    imprime alfa 10:5 (" ");
    imprime sigma2 10:5 (" ");
    imprime D 10:5;
    anote "Valor da parte nao constante do log da funcao";
    anote "de verossimilhanca:";
    imprime lver 10:5 (" ");
  }
/* Calcula novos sigma2, D e alfa */
sigma2 = (srzb + sigma2_0*striw)/no;
D = (sbb - D_0*szwz*D_0)/n + D_0;
invd = inv(D);

```

```

i = 1;
enquanto (i<=n) {
  ind = cata(mat,cod,i,"=");
  xi = ind[,cix:cfx];
  zi = ind[,ciz:cfz];
  yi = ind[,cy];
  wi = (ident(nlin(zi)) - zi*inv(sigma2*invd + zi'*zi)*zi')/sigma2;
  sxwx = sxwx + xi'*wi*xi;
  sxwy = sxwy + xi'*wi*yi;
  i = i + 1;
}
var_alfa = inv(sxwx);
alfa = var_alfa*sxwy;
p_par = max({(max(abs(D-D_0)))
              (abs(sigma2-sigma2_0))
              (max(abs(alfa-alfa_0))))});
iteracao = iteracao + 1;
}
/* calcula log da verossimilhanca da ultima iteracao */
lver = 0;
i = 1;
enquanto (i<=n) {
  ind = cata(mat,cod,i,"=");
  xi = ind[,cix:cfx];
  zi = ind[,ciz:cfz];
  yi = ind[,cy];
  wi = (ident(nlin(zi)) - zi*inv(sigma2*invd + zi'*zi)*zi')/sigma2;
  ri = yi - xi*alfa;
  lver = lver + log(det(wi)) - ri'*wi*ri;
  se (imp_bi==1) {
    bi = D*zi'*wi*ri;
    pi = wi - wi*xi*var_alfa*xi'* wi;
    vbi = D - D*zi'*pi*zi*D;
    se (i==1) { r = ri;
                bvb = {bi vbi}; }
    cc { r = {r,ri};
        bvb = {bvb , bi vbi}; }}
  i = i + 1;
}
lver = (0.5)*lver;
anote"=====";
imprime iteracao 3:0 (" ");
imprime alfa 10:5 (" ");
imprime sigma2 10:5 (" ");

```

```

imprime D 10:5;
anote "Valor da parte nao constante do log da funcao";
anote "de verossimilhanca:";
imprime lver 10:5 (" ");
imprime var_alfa 10:5;
se (imp_bi==1) {
  anote "Impressao dos residuos ri, estimativas dos efeitos";
  anote "aleatorios biest e variancia dos efeitos";
  anote "aleatorios v(biest-bi)";
  imprime { mat[,cod] r } 10:5 ("ind,ri");
  imprime bvb 12:5 ("biest,v(biest-bi)");
}
/*                                     *** FIM ***

```

```

/*****
/*          PROGRAMA SCMV.AJU          */
*****/

/*  Programa em CM do SOC para estimacao dos parametros de  */
/*  modelos de efeitos aleatorios atraves do metodo de maxima */
/*  verossimilhanca usando o algoritmo de Scoring de Fisher  */
*****/

/*  Realocacao dos dados */
mat = {DADOS[,cod] DADOS[,cix:cfx] DADOS[,cy] DADOS[,ciz:cfz]};
DADOS=;
cod = 1;
cfx = cfx - cix + 2;
cix = 2;
cy  = cfx + 1;
cfz = (cfz-ciz);
ciz = cy + 1;
cfz = cfz + ciz;

/*  Calcula dimensoes de alfa, D e theta */
p = cfx - cix + 1;      /* p = dimensao de alfa */
q = nlin(D_0);          /* q = dimensao de D   */
k = q*(q+1)/2 + 1;      /* k = dimensao de theta */

/*  Calcula a derivada de D em relacao a seus parametros e coloca */
/*  em A e preenche theta_0 com D_0 e sigma2_0                      */
theta_0 = { (1:k) }';
i = 1;
j = 1;
g = 1;
enquanto (g<=(k-1)) { Ag = D_0 - D_0;
                      Ag[i,j] = 1;
                      Ag[j,i] = 1;
                      theta_0[g,] = D_0[i,j];
                      se (g==1) A = Ag;
                        cc  A = {A,Ag};
                      se (j==q) { i = i+1;
                                  j = i; }
                        cc  j = j+1;
                      g = g+1; }

theta_0[k,] = sigma2_0;
sigma2_0=;

/*  Impressao dos parametros de controle das iteracoes */

```

```

anote " ";
anote " ";
anote "Metodo de Maxima Verossimilhanca";
anote "Algoritmo Scoring de Fisher";
anote " ";
anote " ";
anote "Parametros de controle das iteracoes";
imprime max_it 8:0 (" ");
imprime precisao 8:6 (" ");

/* Inicio do processo iterativo */
alfa = alfa_0;
theta = theta_0;
D = D_0;
D_0=;
invd = inv(D);
p_alfa = 1000;
p_theta = 1000;
iteracao = 0;
enquanto ((iteracao<max_it)&&((p_alfa>=precisao)|| (p_theta>=precisao)))
{
    alfa_0 = alfa;
    theta_0 = theta;
    lver = 0;
    IS = ident(k)-ident(k); /* IS = esperanca da matriz hessiano */
    s = {(1:k)}' - {(1:k)}'; /* s = vetor escore */
    i = 1;
    enquanto (i<=n)
    {
        ind = cata(mat,cod,i,"");
        xi = ind[,cix:cfx];
        zi = ind[,ciz:cfz];
        yi = ind[,cy];
        wi =(ident(nlin(zi)) - zi*inv(theta[k,]*invd + zi'*zi)*zi')/theta[k,];
        ri = yi - xi*alfa;
        lver = lver + log(det(wi)) - ri'*wi*ri;
        g = 1;
        enquanto (g<=k)
        {
            se (g==k) Gig = ident(nlin(wi));
            cc { Ag = A[((g-1)*q+1):(g*q),];
                Gig = zi*Ag*zi'; }
            /* Calcula s=vetor escore (k x 1) */
            s[g,] = s[g,] + traco(wi*ri*ri'*wi*Gig - wi*Gig);
        }
    }
}

```

```

    h = g;
    enquanto (h<=k)
    {
        se (h==k) Gih = ident(nlin(wi));
        cc      { Ah = A[((h-1)*q+1):(h*q),];
                  Gih = zi*Ah*zi'; }
        IS[g,h] = IS[g,h] + traco(wi*Gig*wi*Gih);
        IS[h,g]=IS[g,h];
        h=h+1;
    }
    g=g+1;
}
i=i+1;
}
lver = (0.5)*lver;
if ((imp_it==1)|| (iteracao==0))
{
    anote"=====";
    imprime iteracao 3:0 (" ");
    imprime alfa 10:5 (" ");
    imprime theta 10:5 (" ");
    anote "Valor da parte nao constante do log da funcao";
    anote "de verossimilhanca:";
    imprime lver 10:5 (" ");
}
IS = inv(IS);
/* Calcula novo theta */
theta = theta_0 + IS*s;
/* Monta nova D */
D = ident(q) - ident(q);
g = 1;
enquanto (g<=(k-1))
{
    Ag = A[((g-1)*q+1):(g*q),];
    D = D + theta[g,]*Ag;
    g = g + 1;
}
/* Calcula novo alfa */
sxwx = ident(p) - ident(p);
sxwy = {(1:p)}' - {(1:p)}';
invd = inv(D);
i = 1;
enquanto (i<=n)
{

```

```

ind = cata(mat,cod,i,"=");
xi = ind[,cix:cfx];
zi = ind[,ciz:cfz];
yi = ind[,cy];
wi = (ident(nlin(zi)) -
      zi*inv(theta[k,]*invd + zi'*zi)*zi')/theta[k,];
sxwx = sxwx + xi'*wi*xi;
sxwy = sxwy + xi'*wi*yi;
i = i + 1;
}
var_alfa = inv(sxwx);
alfa = var_alfa*sxwy;
/* p_alfa = sqrt(csquad(alfa-alfa_0))/sqrt(csquad(alfa_0));*/
/* p_theta = sqrt(csquad(theta-theta_0))/sqrt(csquad(theta_0));*/
p_alfa = max(abs(alfa-alfa_0));
p_theta = max(abs(theta-theta_0));
iteracao = iteracao + 1;
}

/*calcula para a ultima iteracao o log da verossimilhanca e var_theta*/
IS = ident(k)-ident(k);
lver = 0;
i = 1;
enquanto (i<=n)
{
ind = cata(mat,cod,i,"=");
xi = ind[,cix:cfx];
zi = ind[,ciz:cfz];
yi = ind[,cy];
wi = (ident(nlin(zi)) -zi*inv(theta[k,]*invd + zi'*zi)*zi')/theta[k,];
ri = yi - xi*alfa;
lver = lver + log(det(wi)) - ri'*wi*ri;
g = 1;
enquanto (g<=k)
{
se (g==k) Gig = ident(nlin(wi));
cc { Ag = A[((g-1)*q+1):(g*q),];
    Gig = zi*Ag*zi'; }
h = g;
enquanto (h<=k)
{
se (h==k) Gih = ident(nlin(wi));
cc { Ah = A[((h-1)*q+1):(h*q),];
    Gih = zi*Ah*zi'; }

```



```

IS[g,h] = IS[g,h] + traco(wi*Gig*wi*Gih);
IS[h,g]=IS[g,h];
h=h+1;
}
g=g+1;
}
se (imp_bi==1) { bi = D*zi'*wi*ri;
                  pi = wi - wi*xi*var_alfa*xi'* wi;
                  vbi = D - D*zi'*pi*zi*D;
                  se (i==1) { r = ri;
                              bvb = {bi vbi}; }
                              cc      { r = {r,ri};
                                          bvb = {bvb , bi vbi}; }}

i = i + 1;
}
lver = (0.5)*lver;
var_theta = 2*inv(IS);
anote "=====
imprime iteracao 3:0 (" ");
imprime alfa 10:5 (" ");
imprime var_alfa 10:5 (" ");
imprime theta 13:7 (" ");
imprime var_theta;
anote "Valor da parte nao constante do log da funcao";
anote "de verossimilhanca:";
imprime lver 10:5 (" ");
if (imp_bi==1)
{
  anote "Impressao dos residuos ri, estimativas dos efeitos";
  anote "aleatorios biest e variancia dos efeitos";
  anote "aleatorios v(biest-bi)";
  imprime { mat[,cod] r } 10:5 ("ind,ri");
  imprime bvb 12:5 ("biest,v(biest-bi)");
}
/*
*** FIM ***
*/

```

```

/*****
/*          PROGRAMA NRMV.AJU          */
/*****
/*  Programa em CM do SOC para estimacao dos parametros de      */
/*  modelos de efeitos aleatorios atraves do metodo de maxima    */
/*  verossimilhanca usando o algoritmo de Newton-Raphson        */
/*****

/*  Realocacao dos dados */
mat = {DADOS[,cod] DADOS[,cix:cfx] DADOS[,cy] DADOS[,ciz:cfz]};
DADOS=;
cod = 1;
cfx = cfx - cix + 2;
cix = 2;
cy  = cfx + 1;
cfz = (cfz-ciz);
ciz = cy + 1;
cfz = cfz + ciz;

/*  Calcula dimensoes de alfa, D e theta */
p = cfx - cix + 1;      /* p = dimensao de alfa */
q = nlin(D_0);          /* q = dimensao de D      */
k = q*(q+1)/2 + 1;      /* k = dimensao de theta */

/*  Calcula a derivada de D em relacao a seus parametros e coloca */
/*  em A e preenche phi_0 com alfa_0, D_0 e sigma2_0                */
phi_0 = { alfa_0 , {(1:k)}' };
i = 1;
j = 1;
g = 1;
enquanto (g<=(k-1)) { Ag = D_0 - D_0;
                      Ag[i,j] = 1;
                      Ag[j,i] = 1;
                      phi_0[(p+g),] = D_0[i,j];
                      se (g==1) A = Ag;
                      cc      A = {A,Ag};
                      se (j==q) { i = i+1;
                                  j = i; }
                      cc      j = j+1;
                      g = g+1; }
phi_0[(p+k),] = sigma2_0;
sigma2_0=;

/*  Impressao dos parametros de controle das iteracoes */

```

```

anote " ";
anote " ";
anote "Metodo de Maxima Verossimilhanca";
anote "Algoritmo de Newton-Raphson";
anote " ";
anote " ";
anote "Parametros de controle das iteracoes";
imprime max_it 8:0 (" ");
imprime precisao 8:6 (" ");

/* Inicio do processo iterativo */
phi = phi_0;
D = D_0;
D_0=;
p_phi = 1000;
iteracao = 0;
enquanto ((iteracao<max_it)&&(p_phi>=precisao))
{
  phi_0 = phi;
  invd = inv(D);
  lver = 0;
  s_alfa = {(1:p)}' - {(1:p)}';
  s_theta = {(1:k)}' - {(1:k)}';
  HS_alfa = ident(p) - ident(p);
  HS_theta = ident(k) - ident(k);
  HS_at = s_alfa*s_theta';
  i = 1;
  enquanto (i<=n)
  {
    ind = cata(mat,cod,i,"=");
    xi = ind[,cix:cfx];
    zi = ind[,ciz:cfz];
    yi = ind[,cy];
    wi =(ident(nlin(zi)) -
          zi*inv(phi[(p+k),]*invd + zi'*zi)*zi')/phi[(p+k),];
    ri = yi - xi*phi[1:p,];
    lver = lver + log(det(wi)) - ri'*wi*ri;
    s_alfa = s_alfa + xi'*wi*ri;
    HS_alfa = HS_alfa - xi'* wi*xi;
    g = 1;
    enquanto (g<=k)
    {
      se (g==k) Gig = ident(nlin(wi));
      cc { Ag = A[((g-1)*q+1):(g*q),];

```

```

        Gig = zi*Ag*zi'; }
s_theta[g,] = s_theta[g,] + traco(wi*ri*ri'*wi*Gig - wi*Gig);
j = 1;
enquanto (j<=p)
{
    HS_at[j,g] = HS_at[j,g] - xi[,j]'*wi*Gig*wi*ri;
    j = j + 1;
}
h = g;
enquanto (h<=k)
{
    se (h==k) Gih = ident(nlin(wi));
    cc { Ah = A[((h-1)*q+1):(h*q),];
        Gih = zi*Ah*zi'; }
    HS_theta[g,h] = HS_theta[g,h] - 2*ri'*wi*Gih*wi*Gig*wi*ri
        + traco(wi*Gig*wi*Gih);
    HS_theta[h,g]=HS_theta[g,h];
    h=h+1;
}
g=g+1;
}
i=i+1;
}
lver = (0.5)*lver;
se ((imp_it==1)|| (iteracao==0))
{
    anote "===== ";
    imprime iteracao 3:0 (" ");
    imprime phi 10:5 (" ");
    anote "Valor da parte nao constante do log da funcao";
    anote "de verossimilhanca:";
    imprime lver 10:5 (" ");
}
HS = {HS_alfa HS_at , (HS_at') (0.5*HS_theta)};
s = { s_alfa , (0.5*s_theta) };
HS = inv(HS);
/* Calcula novo phi */
phi = phi - HS*s;
/* Monta nova D */
D = ident(q) - ident(q);
g = 1;
enquanto (g<=(k-1))
{
    Ag = A[((g-1)*q+1):(g*q),];

```

```

    D = D + phi[(p+g),]*Ag;
    g = g + 1;
}
/* p_phi = sqrt(csquad(phi-phi_0))/sqrt(csquad(phi_0));*/
p_phi = max(abs(phi-phi_0));
iteracao = iteracao + 1;
}
/* calcula para a ultima iteracao o log da verossimilhanca */
/* e a matriz de covariancia de phi */
H_alf = - HS[1:p,1:p];
HS_alfa = HS_alfa - HS_alfa;
HS_theta = HS_theta - HS_theta;
HS_at = HS_at - HS_at;
invd = inv(D);
lver = 0;
i = 1;
enquanto (i<=n)
{
    ind = cata(mat,cod,i,"=");
    xi = ind[,cix:cfx];
    zi = ind[,ciz:cfz];
    yi = ind[,cy];
    wi = (ident(nlin(zi)) -
           zi*inv(phi[(p+k),]*invd + zi'*zi)*zi')/phi[(p+k),];
    ri = yi - xi*phi[1:p,];
    lver = lver + log(det(wi)) - ri'*wi*ri;
    HS_alfa = HS_alfa - xi'*wi*xi;
    g = 1;
    enquanto (g<=k)
    {
        se (g==k) Gig = ident(nlin(wi));
        cc { Ag = A[((g-1)*q+1):(g*q),];
            Gig = zi*Ag*zi'; }
        j = 1;
        enquanto (j<=p)
        {
            HS_at[j,g] = HS_at[j,g] - xi[,j]'*wi*Gig*wi*ri;
            j = j + 1;
        }
        h = g;
        enquanto (h<=k)
        {
            se (h==k) Gih = ident(nlin(wi));
            cc { Ah = A[((h-1)*q+1):(h*q),];

```

```

        Gih = zi*Ah*zi'; }
    HS_theta[g,h] = HS_theta[g,h] - 2*ri'*wi*Gih*wi*Gig*wi*ri
        + traco(wi*Gig*wi*Gih);
    HS_theta[h,g]=HS_theta[g,h];
    h=h+1;
}
g=g+1;
}
se (imp_bi==1) { bi = D*zi'*wi*ri;
    pi = wi - wi*xi*H_alf*xi'*wi;
    vbi = D - D*zi'*pi*zi*D;
    se (i==1) { r = ri;
        bvb = {bi vbi}; }
        cc      { r = {r,ri};
            bvb = {bvb , bi vbi}; }}

    i=i+1;
}
lver = (0.5)*lver;
HS = {HS_alfa HS_at , (HS_at') (0.5*HS_theta)};
var_phi = - inv(HS);
anote "===== ";
imprime iteracao 3:0 (" ");
imprime phi 10:5 (" ");
imprime var_phi;
anote "Valor da parte nao constante do log da funcao";
anote "de verossimilhanca:";
imprime lver 10:5 (" ");
se (imp_bi==1)
{
    anote "Impressao dos residuos ri, estimativas dos efeitos";
    anote "aleatorios biest e variancia dos efeitos";
    anote "aleatorios v(biest-bi)";
    imprime { mat[,cod] r } 10:5 ("ind,ri");
    imprime bvb 12:5 ("biest,v(biest-bi)");
}
/*                                     *** FIM ***                               */

```

```

/*****
/*          PROGRAMA EMMVR.AJU          */
/*****
/*  Programa em CM do SOC para estimacao dos parametros de      */
/*  modelos de efeitos aleatorios atraves do metodo de maxima    */
/*  restrita usando o algoritmo EM                                */
/*****

/*  Realocacao dos dados */
mat = {DADOS[,cod] DADOS[,cix:cfx] DADOS[,cy] DADOS[,ciz:cfz]};
DADOS=;
cod = 1;
cfx = cfx - cix + 2;
cix = 2;
cy  = cfx + 1;
cfz = (cfz-ciz);
ciz = cy + 1;
cfz = cfz + ciz;

/*  Calcula dimensoes de alfa, D e mat */
p = cfx - cix + 1;          /* p = dimensao de alfa */
q = nlin(D_0);              /* q = dimensao de D    */
no = nlin(mat);             /* no = numero total de observacoes */

/*  Calcula alfa_0 */
sxwx = ident(p) - ident(p);
sxwy = {(1:p)}' - {(1:p)}';
invd = inv(D_0);
i = 1;
enquanto (i<=n)
{
  ind = cata(mat,cod,i,"=");
  xi = ind[,cix:cfx];
  zi = ind[,ciz:cfz];
  yi = ind[,cy];
/*wi=inv(zi*D_0*zi'+sigma2_0*ident(nlin(zi)))**/
  wi = (ident(nlin(zi)) - zi*inv(sigma2_0*invd + zi'*zi)*zi')/sigma2_0;
  sxwx = sxwx + xi'*wi*xi;
  sxwy = sxwy + xi'*wi*yi;
  i = i + 1;
}
var_alfa = inv(sxwx);
alfa_0 = var_alfa*sxwy;

```

```

/* Impressao dos parametros de controle das iteracoes */
anote " ";
anote " ";
anote "Metodo de Maxima Verossimilhanca Restrita";
anote "Algoritmo EM";
anote " ";
anote " ";
anote "Parametros de controle das iteracoes";
imprime max_it 8:0 (" ");
imprime precisao 8:6 (" ");

/* Inicio do processo iterativo */
alfa = alfa_0;
sigma2 = sigma2_0;
D = D_0;
p_par = 1000;
iteracao = 0;
enquanto ((iteracao<max_it)&&(p_par>=precisao))
{
    alfa_0 = alfa;
    sigma2_0 = sigma2;
    D_0 = D;
    srzb = 0;
    strip = 0;
    sbb = D - D;
    szpz = D - D;
/* calcula a verossimilhanca da iteracao anterior */
/* e os proximos D, sigma2 e alfa */
lver=0;
i=1;
enquanto (i<=n)
{
    ind = cata(mat,cod,i,"=");
    xi = ind[,cix:cfx];
    zi = ind[,ciz:cfz];
    yi = ind[,cy];
/* wi=inv(zi*D_0*zi'+sigma2_0*ident(nlin(zi)))* */
    wi = (ident(nlin(zi)) - zi*inv(sigma2_0*invd + zi'*zi)*zi')/sigma2_0;
    pi = wi - wi*xi*var_alfa*xi'*wi;
    ri = yi - xi*alfa_0;
    bi = D_0*zi'*wi*ri;
    lver = lver + log(det(wi)) - ri'*wi*ri;
    srzb = srzb + (ri-zi*bi)'*(ri-zi*bi);
    strip = strip + traco(ident(nlin(zi)) - sigma2_0*pi);
}

```



```

    sbb = sbb + bi*bi';
    szpz = szpz + zi'*pi*zi;
    i=i + 1;
}
lver = (0.5)*(lver + log(det(var_alfa)));
se ((imp_it==1)|| (iteracao==0))
{
    anote"=====";
    imprime iteracao 3:0 (" ");
    imprime alfa 10:5 (" ");
    imprime sigma2 10:5 (" ");
    imprime D 10:5;
    anote "Valor da parte nao constante do log da";
    anote "funcao de verossimilhanca restrita:";
    imprime lver 10:5 (" ");
}
/* Calcula novos D, sigma2 e alfa */
sigma2 = (srzb + sigma2_0*strip)/no;
D = (sbb - D_0*szpz*D_0)/n + D_0;
sxwx = ident(p) - ident(p);
sxwy = {(1:p)}' - {(1:p)}';
invd = inv(D);
i = 1;
enquanto (i<=n)
{
    ind = cata(mat,cod,i,"=");
    xi = ind[,cix:cfx];
    zi = ind[,ciz:cfz];
    yi = ind[,cy];
/* wi=inv(zi*D*zi'+sigma2*ident(nlin(zi)));*/
    wi = (ident(nlin(zi)) - zi*inv(sigma2*invd + zi'*zi)*zi')/sigma2;
    sxwx = sxwx + xi'*wi*xi;
    sxwy = sxwy + xi'*wi*yi;
    i = i + 1;
}
var_alfa = inv(sxwx);
alfa = var_alfa*sxwy;
p_par = max( {(max(abs(D-D_0))
              (abs(sigma2-sigma2_0))
              (max(abs(alfa-alfa_0))))});
iteracao = iteracao + 1;
}
/* calcula log da verossimilhanca da ultima iteracao */
lver = 0;

```

```

i = 1;
enquanto (i<=n)
{
  ind = cata(mat,cod,i,"");
  xi = ind[,cix:cfx];
  zi = ind[,ciz:cfz];
  yi = ind[,cy];
/*wi=inv(zi*D*zi'+sigma2*ident(nlin(zi)))**/
  wi = (ident(nlin(zi)) - zi*inv(sigma2*invd + zi'*zi)*zi')/sigma2;
  ri = yi - xi*alfa;
  lver = lver + log(det(wi)) - ri'*wi*ri;
  se (imp_bi==1)
  {
    bi = D*zi'*wi*ri;
    pi = wi - wi*xi*var_alfa*xi'* wi;
    vbi = D - D*zi'*pi*zi*D;
    se (i==1) { r = ri;
               bvb = {bi vbi}; }
    cc { r = {r,ri};
        bvb = {bvb , bi vbi}; }}
  i = i + 1;
}
lver = (0.5)*(lver + log(det(var_alfa)));
anote"=====";
imprime iteracao 3:0 (" ");
imprime alfa 10:5 (" ");
imprime sigma2 10:5 (" ");
imprime D 10:5;
anote "Valor da parte nao constante do log da";
anote "funcao de verossimilhanca restrita:";
imprime lver 10:5 (" ");
imprime var_alfa 10:5;
se (imp_bi==1)
{
  anote "Impressao dos residuos ri, estimativas dos efeitos";
  anote "aleatorios biest e variancia dos efeitos";
  anote "aleatorios v(biest-bi)";
  imprime { mat[,cod] r } 10:5 ("ind,ri");
  imprime bvb 12:5 ("biest,v(biest-b)");
}
/*                                     *** FIM ***                               */

```

```

/*****
/*          PROGRAMA SCMRV.AJU          */
/*****
/*  Programa em CM do SOC para estimacao dos parametros de      */
/*  modelos de efeitos aleatorios atraves do metodo de maxima    */
/*  restrita usando o algoritmo Scoring de Fisher                */
/*****

/*  Realocacao dos dados */
mat = {DADOS[,cod] DADOS[,cix:cfx] DADOS[,cy] DADOS[,ciz:cfz]};
DADOS=;
cod = 1;
cfx = cfx - cix + 2;
cix = 2;
cy  = cfx + 1;
cfz = (cfz-ciz);
ciz = cy + 1;
cfz = cfz + ciz;

/*  Calcula dimensoes de alfa, D e theta */
p = cfx - cix + 1;          /* p = dimensao de alfa */
q = nlin(D_0);              /* q = dimensao de D    */
k = q*(q+1)/2 + 1;          /* k = dimensao de theta */

/* Calcula a derivada de D em relacao a seus parametros e coloca */
/* em A e preenche theta_0 com D_0 e sigma2_0                        */
theta_0 = { (1:k) }';
i = 1;
j = 1;
g = 1;
enquanto (g<=(k-1)) { Ag = D_0 - D_0;
                      Ag[i,j] = 1;
                      Ag[j,i] = 1;
                      theta_0[g,] = D_0[i,j];
                      se (g==1) A = Ag;
                          cc    A = {A,Ag};
                      se (j==q) { i = i+1;
                                  j = i; }
                          cc    j = j+1;
                      g = g+1; }
theta_0[k,] = sigma2_0;
sigma2_0=;

/* Calcula alfa_0 */

```

```

sxwx = ident(p) - ident(p);
sxwy = {(1:p)}' - {(1:p)}';
invd = inv(D_0);
i = 1;
enquanto (i<=n)
{
    ind = cata(mat,cod,i,"=");
    xi = ind[,cix:cfx];
    zi = ind[,ciz:cfz];
    yi = ind[,cy];
    wi = (ident(nlin(zi)) -
           zi*inv(theta_0[k,]*invd + zi'*zi)*zi')/theta_0[k,];
    sxwx = sxwx + xi'*wi*xi;
    sxwy = sxwy + xi'*wi*yi;
    i = i + 1;
}
var_alfa = inv(sxwx);
alfa_0 = var_alfa*sxwy;

/* Impressao dos parametros de controle das iteracoes */
anote " ";
anote " ";
anote "Metodo de Maxima Verossimilhanca Restrita";
anote "Algoritmo Scoring de Fisher";
anote " ";
anote " ";
anote "Parametros de controle das iteracoes";
imprime max_it 8:0 (" ");
imprime precisao 8:6 (" ");

/* Inicio do processo iterativo */
alfa = alfa_0;
theta = theta_0;
D = D_0;
D_0=;
p_alfa = 1000;
p_theta = 1000;
iteracao = 0;
enquanto ((iteracao<max_it)&&((p_alfa>=precisao)|| (p_theta>=precisao)))
{
    alfa_0 = alfa;
    theta_0 = theta;
    lver = 0;
    IS = ident(k)-ident(k); /* IS = esperanca da matriz hessiano */

```

```

s = {(1:k)}' - {(1:k)}'; /* s = vetor escore */
i = 1;
enquanto (i<=n)
{
    ind = cata(mat,cod,i,"");
    xi = ind[,cix:cfx];
    zi = ind[,ciz:cfz];
    yi = ind[,cy];
    wi =(ident(nlin(zi)) - zi*inv(theta[k,]*invd + zi'*zi)*zi')/theta[k,];
    pi = wi - wi*xi*var_alfa*xi'*wi;
    ri = yi - xi*alfa;
    lver = lver + log(det(wi)) - ri'*wi*ri;
    g = 1;
    enquanto (g<=k)
    {
        se (g==k) Gig = ident(nlin(wi));
        cc { Ag = A[((g-1)*q+1):(g*q),];
            Gig = zi*Ag*zi'; }
        /* Calcula s=vetor escore (k x 1) */
        s[g,] = s[g,] + traco(wi*ri*ri'*wi*Gig - pi*Gig);
        h = g;
        enquanto (h<=k)
        {
            se (h==k) Gih = ident(nlin(wi));
            cc { Ah = A[((h-1)*q+1):(h*q),];
                Gih = zi*Ah*zi'; }
            IS[g,h] = IS[g,h] + traco(pi*Gig*pi*Gih);
            IS[h,g]=IS[g,h];
            h=h+1;
        }
        g=g+1;
    }
    i=i+1;
}
lver = (0.5)*(lver + log(det(var_alfa)));
if ((imp_it==1)||(iteracao==0))
{
    anote "=====";
    imprime iteracao 3:0 (" ");
    imprime alfa 10:5 (" ");
    imprime theta 10:5 (" ");
    anote "Valor da parte nao constante do log da";
    anote "funcao de verossimilhanca restrita:";
    imprime lver 10:5 (" ");
}

```

```

    }
    IS = inv(IS);
    /* Calcula novo theta */
    theta = theta_0 + IS*s;
    /* Monta nova D */
    D = ident(q) - ident(q);
    g = 1;
    enquanto (g<=(k-1))
    {
        Ag = A[((g-1)*q+1):(g*q),];
        D = D + theta[g,]*Ag;
        g = g + 1;
    }
    /* Calcula novo alfa */
    sxwx = ident(p) - ident(p);
    sxwy = {(1:p)}' - {(1:p)}';
    invd = inv(D);
    i = 1;
    enquanto (i<=n)
    {
        ind = cata(mat,cod,i,"");
        xi = ind[,cix:cfx];
        zi = ind[,ciz:cfz];
        yi = ind[,cy];
        wi = (ident(nlin(zi)) -zi*inv(theta[k,]*invd + zi'*zi)*zi')/theta[k,];
        sxwx = sxwx + xi'*wi*xi;
        sxwy = sxwy + xi'*wi*yi;
        i = i + 1;
    }
    var_alfa = inv(sxwx);
    alfa = var_alfa*sxwy;
    /* p_alfa = sqrt(csquad(alfa-alfa_0))/sqrt(csquad(alfa_0));*/
    /* p_theta = sqrt(csquad(theta-theta_0))/sqrt(csquad(theta_0));*/
    p_alfa = max(abs(alfa-alfa_0));
    p_theta = max(abs(theta-theta_0));
    iteracao = iteracao + 1;
}

/*calcula para a ultima iteracao o log da verossimilhanca e var_theta*/
IS = ident(k)-ident(k); /* IS = esperanca da matriz hessiano */
lver = 0;
i = 1;
enquanto (i<=n)
{

```

```

ind = cata(mat,cod,i,"");
xi = ind[,cix:cfx];
zi = ind[,ciz:cfz];
yi = ind[,cy];
wi = (ident(nlin(zi)) -zi*inv(theta[k,]*invd + zi'*zi)*zi')/theta[k,];
pi = wi - wi*xi*var_alfa*xi'*wi;
ri = yi - xi*alfa;
lver = lver + log(det(wi)) - ri'*wi*ri;
g = 1;
enquanto (g<=k)
{
  se (g==k) Gig = ident(nlin(wi));
  cc { Ag = A[((g-1)*q+1):(g*q),];
      Gig = zi*Ag*zi'; }
  h = g;
  enquanto (h<=k)
  {
    se (h==k) Gih = ident(nlin(wi));
    cc { Ah = A[((h-1)*q+1):(h*q),];
        Gih = zi*Ah*zi'; }
    IS[g,h] = IS[g,h] + traco(pi*Gig*pi*Gih);
    IS[h,g]=IS[g,h];
    h=h+1;
  }
  g=g+1;
}
se (imp_bi==1) { bi = D*zi'*wi*ri;
  pi = wi - wi*xi*var_alfa*xi'* wi;
  vbi = D - D*zi'*pi*zi*D;
  se (i==1) { r = ri;
    bvb = {bi vbi}; }
  cc { r = {r,ri};
      bvb = {bvb , bi vbi}; }}

i = i + 1;
}
lver = (0.5)*(lver + log(det(var_alfa)));
var_theta = 2*inv(IS);
anote "===== ";
imprime iteracao 3:0 (" ");
imprime alfa 10:5 (" ");
imprime var_alfa 10:5 (" ");
imprime theta 13:7 (" ");
imprime var_theta;
anote "Valor da parte nao constante do log da";

```

```

anote "funcao de verossimilhanca restrita:";
imprime lver 10:5 (" ");
if (imp_bi==1)
{
  anote "Impressao dos residuos ri, estimativas dos efeitos";
  anote "aleatorios biest e variancia dos efeitos";
  anote "aleatorios v(biest-bi)";
  imprime { mat[,cod] r } 10:5 ("ind,ri");
  imprime bvb 12:5 ("biest,v(biest-bi)");
}
/*                                     *** FIM ***

```


Referências Bibliográficas

- [1] Albert, A. (1976). When is a sum of squares an analysis of variance? *Annals of Statistics*, **4**: 775–8.
- [2] Andrade, D. F. e Helms, R. W. (1984a). *ML estimation and LR tests for the multivariate normal distribution with patterned mean and covariance matrix. Complete and incomplete data cases*. University of North Carolina at Chapel Hill, Institute of Statistics Mimeo Series, No 1455. 156p. Tese (Doutorado).
- [3] Andrade, D. F. e Helms, R. W. (1984b). Maximum likelihood estimates in the multivariate normal with patterned mean and covariance via the EM algorithm. *Communications in Statistics, Series A — Theory and Methods*, **13**: 2239–51.
- [4] Bard, Y. (1974). *Nonlinear Parameter Estimation*. New York, Academic Press. 341p.
- [5] Bartlett, M. S. (1937). Some examples of statistical methods of research in agriculture and applied biology. *Journal of the Royal Statistical Society, Supplement*, **4**: 137–70.
- [6] Beale, E. M. L. e Little, R. J. A. (1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society, Series B*, **37**: 129–46.
- [7] Berk, K. (1987). Computing for incomplete repeated measures. *Biometrics*, **43**: 385–98.

- [8] Bock, R. D. (1975). *Multivariate Statistical Methods in Behavioral Research*. New York, McGraw-Hill. 623p.
- [9] Box, G. E. P. e Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco, Holden-Day. 575p.
- [10] Cook, N. R. e Ware, J. H. (1983). Design and analysis methods for longitudinal research. *Annual Review of Public Health*, **4**:1-24.
- [11] Dempster, A. P.; Laird, N. M. e Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**:1-38.
- [12] Dempster, A. P. e Ryan, L. M. (1985). Weighted normal plots. *Journal of the American Statistical Association*, **80**:845-50.
- [13] Dempster, A. P.; Rubin, D. B. e Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, **76**:341-53.
- [14] Elston, R. C. e Grizzle, J. E. (1962). Estimation of time-response curves and their bands. *Biometrics*, **18**:148-59.
- [15] Fairclough, D. L. e Helms, R. W. (1986). A mixed linear model with linear covariance structure: A sensitivity analysis of maximum likelihood estimators. *Journal of Statistical Computation and Simulation*, **25**:205-36.
- [16] Graybill, F. A. (1976). *Theory and Application of the Linear Model*. North Scituate, Duxbury Press. 704p.
- [17] Graybill, F. A. e Hultquist, R. A. (1961). Theorems concerning Eisenhart's model II. *Annals of Mathematical Statistics*, **32**:261-9.
- [18] Greenhouse, S. W. e Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, **24**:95-112.

- [19] Grizzle, J. E. e Allen, D. M. (1969). Analysis of growth and dose-response curves. *Biometrics*, **25**:357-81.
- [20] Hartley, H. O. e Hocking, R. R. (1971). The analysis of incomplete data. *Biometrics*, **27**:783-808.
- [21] Hartley, H. O. e Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, **54**:93-108.
- [22] Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**:383-5.
- [23] Harville, D. A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics*, **4**:384-95.
- [24] Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**:320-40.
- [25] Hemmerle, W. J. e Hartley, H. O. (1973). Computing maximum likelihood estimates for the mixed AOV model using the W transform. *Technometrics*, **15**:819-31.
- [26] Henderson, C. R.; Kempthorne, O.; Searle, S. R. e Von Krosigk, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **15**:192-218.
- [27] Jennrich, R. I. e Sampson, P. F. (1976). Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, **18**:11-7.
- [28] Jennrich, R. I. e Schluchter, M. D. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, **42**:805-20.
- [29] Khatri, C. G. (1966). A note on the MANOVA model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics*, **18**:75-86.

- [30] Kleinbaum, D. G. (1973a). Testing linear hypotheses in generalized multivariate linear models. *Communications in Statistics*, **1**:433–57.
- [31] Kleinbaum, D. G. (1973b). A generalization of the growth curve model which allows missing data. *Journal of Multivariate Analysis*, **3**:117–24.
- [32] Laird, N. M. (1982). Computation of variance components using the EM algorithm. *Journal of Statistical Computation and Simulation*, **14**:295–303.
- [33] Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, **7**:305–15.
- [34] Laird, N. M.; Lange, N. e Stram, D. (1987). Maximum likelihood computations with repeated measures: Application of the EM algorithm. *Journal of the American Statistical Association*, **82**:97–105.
- [35] Laird, N. M. e Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, **38**:963–74.
- [36] Lange, N. e Laird, N. M. (1989). The effect of covariance structure on variance estimation in balanced growth curve models with random parameters. *Journal of the American Statistical Association*, no prelo.
- [37] Lavange, L. M. e Helms, R. W. (1983). *The analysis of incomplete longitudinal data with modeled covariance structures*. University of North Carolina at Chapel Hill, Institute of Statistics Mimeo Series, No 1449. 134p. Tese (Doutorado).
- [38] Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York, John-Wiley. 584p.
- [39] Lindley, D. V. e Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B*, **34**:1–41.

- [40] Lindstrom, M. J. e 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**:1014-22.
- [41] Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, **83**:1198-202.
- [42] Louis, T. A. (1988). General methods for analysing repeated measures. *Statistics in Medicine*, **7**:329-45.
- [43] Louis, T. A. e Spiro, A. III (1986). Fitting and assessing first-order auto-regressive models with covariates. Research Report, Department of Biostatistics, Harvard School of Public Health.
- [44] Mardia, K. V.; Kent, J. T. e Bibby, J. M. (1979). *Multivariate Analysis*. London, Academic Press. 521p.
- [45] Mood, A. M.; Graybill, F. A. e Boes, D. C. (1974). *Introduction to the Theory of Statistics*. 2.ed. New York, McGraw-Hill. 564p.
- [46] Morrison, D. F. (1976). *Multivariate Statistical Methods*. 2.ed. New York, McGraw-Hill. 415p.
- [47] Orchard, T. e Woodbury, M. A. (1972). A missing information principle: Theory and applications. *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, **1**:697-715.
- [48] Patterson, H. D. e Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika*, **58**:545-54.
- [49] Potthoff, R. F. e Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**:313-26.

- [50] Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its applications to the analysis of growth curves. *Biometrika*, **52**: 447-58.
- [51] Rao, C. R. (1975). Simultaneous estimation of parameters in different linear models and its applications to biometric problems. *Biometrics*, **31**: 545-54.
- [52] Rochon, J. e Helms, R. W. (1989). Maximum likelihood estimation for incomplete repeated-measurements under an ARMA covariance structure. *Biometrics*, **45**: 207-18.
- [53] Rubin, D. B. (1976a). Noniterative least squares estimates, standard errors and F-tests for analysis of variance with missing data. *Journal of the Royal Statistical Society, Series B*, **38**: 270-4.
- [54] Rubin, D. B. (1976b). Inference and missing data. *Biometrika*, **63**: 581-92.
- [55] Schluchter, M. D. (1988a). Analysis of incomplete multivariate data using linear models with structured covariance matrices. *Statistics in Medicine*, **7**: 317-24.
- [56] Schluchter, M. D. (1988b). *Unbalanced Repeated Measures Models with Structured Covariance Matrices* -- BMDP5V. Technical Report # 86, BMDP.
- [57] Schwertman, N. C. (1978). A note on the Greenhouse-Geisser correction for incomplete data split-plot analysis. *Journal of the American Statistical Association*, **73**: 393-6.
- [58] Searle, S. R. (1970). Large sample variance of maximum likelihood estimators of variance components using unbalancead data. *Biometrics*: 505-24.
- [59] Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. New York, John Wiley. 438p.
- [60] Singer, J. M. (1977). *Análise de curvas de crescimento*. São Paulo, IME-USP. 113p. Dissertação (Mestrado).

- [61] Singer, J. M. e Andrade, D. F. (1986). *Análise de Dados Longitudinais*. Campinas, VII Simpósio Nacional de Probabilidade e Estatística.
- [62] Singer, J. M. e Mauro, E. S. R. (1987). *Estudo de respostas eletrofisiológicas do nervo conectivo cérebro-visceral de Perna-Perna exposto a diferentes condições de "stress" osmótico*. São Paulo, IME-USP. 64p. SEA RAE # 8708.
- [63] SOC Software científico elaborado pelo Núcleo Tecnológico de Informática para Agropecuária da EMBRAPA (NTIA/ EMBRAPA) – Caixa Postal 5010 – CEP 13081 – Campinas, SP.
- [64] Szatrowski, T. H. (1983). Missing data in the one-population multivariate normal patterned mean and covariance matrix testing and estimation problem. *Annals of Statistics*, **3**:947-58.
- [65] Timm, N. H. (1980). Multivariate analysis of variance of repeated measures. *Handbook of Statistics, Analysis of Variance*. Vol. 1, P. R. Krishnaiah (ed.), 41-87. New York, North Holland.
- [66] Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, **39**:95-101.
- [67] Wilson, P. D. (1988). Autoregressive growth curves and Kalman filtering. *Statistics in Medicine*, **7**:373-86.
- [68] Winer, B. J. (1971). *Statistical Principles In Experimental Design*. 2.ed. New York, McGraw-Hill. 907p.