**IMPUTAÇÃO DE DADOS ATRAVÉS
DE MODELOS DE CURVAS DE
CRESCIMENTO POLINOMIAIS**

Cláudia Tureta Daré

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

Orientadora: Profa. Dra. Lúcia Pereira Barroso

-São Paulo, agosto de 2000-

Aos meus pais,
Carlos e Maria Helena.

IMPUTAÇÃO DE DADOS ATRAVÉS DE MODELOS DE CURVAS DE CRESCIMENTO POLINOMIAIS

Cláudia Tureta Daré

*Este exemplar corresponde à redação final
da dissertação devidamente corrigida e
defendida por Cláudia Tureta Daré e
aprovada pela comissão julgadora.*

São Paulo, 04 de agosto de 2000.

Banco Examinadora:

- Profa. Dra. Lúcia Pereira Barroso (Orientadora) - IME - USP
- Profa. Dr. Rinaldo Artes - IME - USP
- Profa. Dra. Édina Shizue Miazaki - UNB

Agradecimentos

Gostaria de registrar meu agradecimento às pessoas que, de uma forma ou outra, participaram do processo de elaboração deste trabalho. Em especial, agradeço à Profa. Lúcia Pereira Barroso, pela sugestão do tema e pela dedicação e paciência com que me orientou. Aos amigos que estiveram presentes, pela amizade, incentivo, apoio e carinho. Em especial à Ana Cristina Vieira Belem, Marco Antônio Rodrigues, Myléne Valente, Ocileide Custódio, Deusélio Fioresi e Delhi Salinas.

Agradeço especialmente à minha família, pela força, apoio, incentivo e confiança transmitidos durante o desenvolvimento deste trabalho.

Também agradeço ao CNPq pela verba dispensada, sem a qual este projeto não seria possível.

RESUMO

Em estudos longitudinais é freqüente a ocorrência de observações ausentes. Uma abordagem para dados incompletos é a imputação, que consiste em prever as observações ausentes, completando o conjunto de dados para posterior análise.

Neste trabalho, a imputação é feita através do modelo de curvas de crescimento polinomiais sob o modelo misto, com as suposições de normalidade de efeitos e erros aleatórios, independência dos erros e mesma matriz de covariância para efeitos individuais.

O método de imputação adotado é o BLUP (Best Linear Unbiased Predictor), que depende da estimação das matrizes de covariância, para a qual vários métodos são considerados. Além disso, a técnica é aplicada a um conjunto de dados reais e, para outros três conjuntos de dados, os resultados são comparados com os de outros trabalhos.

ABSTRACT

In longitudinal studies is often the occurrence of missing observations. One way to handle incomplete data is the imputation, which consists of predicting the missing observations to complete the data set which will be later analysed.

In this work, the imputation is obtained using the polynomial growth curves models under assumption of normality of effects and random errors, independence of errors and same covariance matrix for individual effects.

The imputation method is the best linear unbiased predictor (BLUP), that depends on estimation of the covariance matrix, for which some methods are considered. The imputation technique is applied to some sets of real data and the results are compared with others works from the literature.

ÍNDICE

1	Introdução	1
2	Modelos de Curvas de Crescimento	3
2.1	Dados longitudinais	3
2.2	Um breve histórico	11
2.3	Modelo de curva de crescimento	16
2.4	Modelos para as médias	23
2.5	Estruturas de covariância	25
3	Estimação dos Parâmetros	33
3.1	Fórmulas explícitas dos estimadores	33
3.1.1	Estimação de máxima verossimilhança	35
3.1.2	Estimação de máxima verossimilhança restrita	39
3.1.3	Estimação para o Modelo III	40
3.2	Algoritmos para maximização da função de verossimilhança	41
3.2.1	Os algoritmos de Newton-Raphson e Scoring de Fisher	43
3.2.2	O Algoritmo EM	47
3.2.3	Valores iniciais e regras de parada	51

4	Imputação	53
4.1	Introdução	53
4.2	Histórico	54
4.3	O BLUP para o modelo misto	55
4.3.1	Imputação no Modelo III	59
5	Aplicações	64
	Conclusões	81
	Apêndice A	83
	Apêndice B	85
	Referências Bibliográficas	99

Capítulo 1

Introdução

Na análise de dados é comum a ocorrência de um problema: a ausência de algumas observações. Isto pode ser decorrente de erros durante a coleta de dados, transcrição, registro ou a ausência propriamente dita.

Este problema pode ser basicamente tratado de três maneiras: descartar os dados referentes às unidades que possuem observações ausentes e analisar apenas a parte completa; analisar o conjunto incompleto quando as técnicas adequadas o permitirem ou fazer a imputação, completando os dados e efetuando a análise como se o conjunto fosse completo.

Neste trabalho o enfoque é para a imputação. É suposto que o mecanismo de não resposta não depende dos dados presentes ou ausentes, ou seja, são MCAR (Missing Completely At Random) segundo Rubin (1976). Considera-se também que os dados, se fossem completos, seriam as observações de uma variável de n unidades experimentais em p condições de avaliação.

É importante ressaltar que o nosso objetivo é a imputação e não a análise dos dados. O método de imputação considerado é o BLUP (Best Linear Unbiased Predictor) e o trabalho é desenvolvido na seguinte seqüência.

O Capítulo 2 apresenta o modelo misto, que engloba os modelos de curvas de cresci-

mento polinomiais. Destaca-se também um outro modelo mais simples, aqui denominado de Modelo III. Várias estruturas que podem ser assumidas para as duas matrizes de covariâncias são também apresentadas. No Capítulo 3 é descrita a estimação dos efeitos fixos e das matrizes de covariâncias do modelo. São consideradas a estimação de máxima verossimilhança e a estimação de máxima verossimilhança restrita sob duas vertentes: formas explícitas e métodos iterativos (destes são abordados os algoritmos EM, Scoring de Fisher e Newton-Raphson). O método de imputação é descrito no Capítulo 4. As aplicações são efetuadas em quatro conjuntos de dados reais, um deles com observações originalmente ausentes e os outros três com simulação de ausência de alguns dados. Estes três últimos conjuntos já foram exaustivamente analisados na literatura, permitindo que os resultados de imputação via BLUP possam ser comparados com os de outros métodos de imputação. As aplicações são apresentadas no Capítulo 5 e são seguidas pelas conclusões e tópicos para futuras pesquisas.

Capítulo 2

Modelos de Curvas de Crescimento

2.1 Dados longitudinais

Estudos longitudinais têm tido ampla aplicação em diversas áreas, principalmente em pesquisas médicas, sociais, biológicas e agronômicas.

Dados longitudinais caracterizam-se pela observação repetida de uma ou mais variáveis respostas na mesma unidade experimental ao longo de diversas condições de avaliação. Nesse contexto, as condições de avaliação ou ocasiões de medidas são comumente denominadas “tempos” e podem ser o tempo propriamente dito, a idade, doses, distâncias ou qualquer outra medida numa escala ordenada; as unidades experimentais ou de investigação também denominadas “indivíduos” representam pessoas, animais, peças, etc. e podem ser classificadas em diferentes grupos (ou tratamentos) segundo um ou mais fatores como raça, sexo, classe comportamental, marca, etc.. As respostas observadas no mesmo indivíduo, em geral são correlacionadas. Estas podem ser categorizadas (presença ou ausência de uma determinada característica) ou contínuas (peso, altura, pressão intra-ocular). É comum, no caso contínuo, supor que as observações seguem uma distribuição normal multivariada.

Em estudos longitudinais, os conjuntos de dados podem ser classificados em:

1. balanceados ou completos em relação ao tempo, quando todos os indivíduos foram observados o mesmo número de vezes nos mesmos instantes de tempo (condições de avaliação);
2. desbalanceados ou incompletos em relação ao tempo, se há indivíduos que não foram observados em um ou mais instantes de tempo, ou se as condições de avaliação variam de indivíduo para indivíduo.

Alguns exemplos de estudos longitudinais são dados a seguir.

Exemplo 2.1 (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. 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. Os dados coletados são apresentados na Tabela 2.1.

Tabela 2.1: *Distância (em mm) entre o centro da pituitária e a fissura pterio-maxilar de 11 meninas e 16 meninos em 4 idades.*

Meninas	Idade (anos)				Meninos	Idade (anos)			
	8	10	12	14		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

Exemplo 2.2 (Elston e Grizzle, 1962). Num estudo realizado na Faculdade de Odontologia da Universidade da Carolina do Norte (E.U.A.), a altura (em mm) do osso ramus (parte ascendente da mandíbula) foi medida em 20 meninos aos 8; 8,5; 9 e 9,5 anos de idade. O objetivo do estudo era estabelecer uma curva de crescimento para ser usada por ortodontistas. Os dados obtidos estão na Tabela 2.2.

Tabela 2.2: *Altura (em mm) do ramus de 20 meninos em 4 idades.*

Indivíduo	Idade (anos)				Indivíduo	Idade (anos)			
	8	8,5	9	9,5		8	8,5	9	9,5
1	47,8	48,8	49,0	49,7	11	51,2	51,4	51,6	51,9
2	46,4	47,3	47,7	48,4	12	48,5	49,2	53,0	55,5
3	46,3	46,8	47,8	48,5	13	52,1	52,8	53,7	55,0
4	45,1	45,3	46,1	47,2	14	48,2	48,9	49,3	49,8
5	47,6	48,5	48,9	49,3	15	49,6	50,4	51,2	51,8
6	52,5	53,2	53,3	53,7	16	50,7	51,7	52,7	53,3
7	51,2	53,0	54,3	54,5	17	47,2	47,7	48,4	49,5
8	49,8	50,0	50,3	52,7	18	53,3	54,6	55,1	55,3
9	48,1	50,8	52,3	54,4	19	46,2	47,5	48,1	48,4
10	45,0	47,0	47,3	48,3	20	46,3	47,6	51,3	51,8

Exemplo 2.3 (Williams e Izenman, 1981). O peso (em cg) de 13 ratos machos foi medido em intervalos sucessivos de 3 dias num período de 21 dias (3, 6, 9, 12, 15, 18 e 21 dias) do nascimento até o desmame. Os dados encontram-se reproduzidos na Tabela 2.3.

Tabela 2.3: *Peso (em cg) de 13 ratos medidos em 7 idades.*

Rato	Idade (dias)						
	3	6	9	12	15	18	21
1	0,190	0,388	0,621	0,823	1,078	1,132	1,191
2	0,218	0,393	0,568	0,729	0,839	0,852	1,004
3	0,211	0,394	0,549	0,700	0,783	0,870	0,925
4	0,209	0,419	0,645	0,850	1,001	1,026	1,069
5	0,193	0,362	0,520	0,530	0,641	0,640	0,751
6	0,201	0,361	0,502	0,530	0,657	0,762	0,888
7	0,202	0,370	0,498	0,650	0,795	0,858	0,910
8	0,190	0,350	0,510	0,666	0,819	0,879	0,929
9	0,219	0,399	0,578	0,699	0,709	0,822	0,953
10	0,225	0,400	0,545	0,690	0,796	0,825	0,836
11	0,224	0,381	0,577	0,756	0,869	0,929	0,999
12	0,187	0,329	0,441	0,525	0,589	0,621	0,796
13	0,278	0,471	0,606	0,770	0,888	1,001	1,105

Exemplo 2.4 (Barroso e Sandoval, 1999). Este conjunto é parte de um estudo mais amplo realizado no INCOR-Instituto do Coração da Universidade de São Paulo e analisado no CEA-Centro de Estatística Aplicada do Departamento de Estatística da Universidade de São Paulo. A frequência cardíaca (em batidas/minuto) foi medida em 10 pacientes em instantes diferentes de tempo durante um exame ergométrico. As medidas foram tomadas nos instantes de inspiração do ar e por isso os instantes são diferentes de paciente para paciente. Os dados estão apresentados na Tabela 2.4.

Tabela 2.4: *Frequência cardíaca (em batidas/minuto) de 10 pacientes em vários instantes.*

Paciente	Instantes (segundos)										
	5	6	7	8	9	10	11	12	13	14	15
1	-	112	-	-	114	-	-	113	-	-	-
2	-	72	-	74	-	-	-	74	-	-	-
3	-	-	-	-	-	76	-	-	-	-	78
4	-	-	103	-	-	89	-	85	-	-	85
5	-	-	-	-	-	-	-	-	95	-	-
6	87	-	-	-	87	-	-	-	-	-	89
7	-	-	-	-	-	-	-	-	-	-	108
8	-	84	-	-	84	-	-	-	-	84	- 85
9	-	-	92	-	-	92	-	-	-	92	-
10	-	-	79	-	-	81	-	-	82	-	-

Exemplo 2.5 (Santos, 1997). Este conjunto de dados é parte de um estudo mais amplo realizado na Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo. A quantidade do hormônio testosterona (em ng/ml) foi medida em 10 éguas em 7 semanas consecutivas. Alguns dados foram perdidos por defeito no “kit” de medição (insuficiência no volume das alíquotas). Os dados estão na Tabela 2.5.

Tabela 2.5: *Quantidade (em ng/ml) de testosterona de 10 éguas em 7 semanas.*

Égua	Instantes de tempo (semanas)						
	1	2	3	4	5	6	7
1	0,067	0,101	0,062	0,045	0,026	0,044	0,015
2	0,051	0,058	0,088	0,072	0,015	0,090	0,111
3	0,034	0,135	0,049	0,206	0,126	0,070	0,015
4	0,171	-	0,052	0,100	0,114	0,074	0,048
5	-	0,105	0,143	0,125	0,015	0,015	0,015
6	0,114	0,089	0,123	0,196	0,214	0,239	0,487
7	0,114	0,772	0,105	0,109	0,064	0,480	0,170
8	0,015	0,015	0,026	0,152	0,015	0,015	-
9	0,087	0,077	0,095	0,077	0,087	0,092	0,162
10	0,040	0,059	0,068	0,108	0,090	0,058	0,089

Os dados das Tabelas 2.1, 2.2 e 2.3 são classificados como balanceados ou completos com relação ao tempo e já foram intensivamente analisados por vários autores, entre os quais podemos citar Lee e Geisser (1975), Rao (1987), Lee (1988), Liski e Nummi (1990) e Soler (1997). Os dados do Exemplo 2.4 são desbalanceados ou incompletos por planejamento, enquanto que o conjunto de dados do Exemplo 2.5 é desbalanceado ou incompleto aleatoriamente, pois embora tenha sido planejado obter as informações para todos os elementos da amostra nos mesmos instantes de tempo, algumas observações foram perdidas por motivos alheios ao planejamento.

Freqüentemente os objetivos de estudos longitudinais são:

1. verificar a existência de interação entre os tratamentos e as condições de avaliação;
2. verificar se existe efeito de tratamento com relação às distribuições das respostas médias;
3. verificar se existe efeito das condições de avaliação;
4. ajustar curvas de respostas médias ou individuais;
5. explicar a variação das respostas ao longo das condições de avaliação;
6. investigar os efeitos de covariadas ou outros fatores no padrão das respostas;
7. estimar e prever os efeitos de tratamentos ou as respostas em tempos futuros, bem como para observações perdidas.

Em resumo, os objetivos de pesquisas longitudinais são caracterizar os padrões de respostas individuais e as mudanças sobre o tempo e investigar os efeitos de covariadas nestes padrões (Ware, 1985).

Um modelo bastante utilizado para a análise de dados longitudinais é o modelo de curva de crescimento. Algumas vantagens desse modelo são:

- comparações mais sensíveis devido ao número reduzido de parâmetros;
- possibilidade de fazer previsões futuras sobre o comportamento da resposta para tempos diferentes daqueles para os quais o estudo foi planejado;
- aquisição de um conhecimento maior sobre o processo de variação da resposta ao longo do tempo;
- interpretação física dos coeficientes das curvas polinomiais;
- interpretação dos resultados consideravelmente simplificada.

Em geral, para atingir os objetivos da análise de dados longitudinais, o ajuste de modelos de curvas de crescimento leva a:

1. testar a hipótese de especificação do modelo, isto é, se o grau da curva assumida é adequada;
2. estimar os coeficientes da curva de crescimento;
3. estimar a matriz de covariância dos coeficientes da curva estimada;
4. testar hipóteses, como por exemplo, a igualdade dos vetores dos coeficientes das curvas de crescimento para vários grupos ou o paralelismo entre as curvas;
5. obter intervalos de confiança simultâneos para funções paramétricas;
6. fazer predições, tanto de valores ausentes (imputação) quanto futuros.

Neste trabalho estamos tratando do caso de dados balanceados por planejamento, mas aleatoriamente incompletos. Para maiores detalhes sobre estudos longitudinais consultar Ware (1985), Singer e Andrade (1986), Andreoni (1989), Dwyer et al. (1992) e Kshirsagar e Smith (1995) entre outros.

2.2 Um breve histórico

Nesta seção, dentro do possível, os artigos são agrupados por conteúdo e apresentados em ordem cronológica. Alguns deles poderiam estar em mais de um grupo e a sequência de tópicos é mais ou menos a seguinte: estimação e testes sobre os efeitos fixos, previsão dos efeitos aleatórios e modelagem da matriz de covariância, análise de dados incompletos, imputação e previsão de observações futuras.

Wishart (1938), interessado em determinar a razão de crescimento do toucinho de porco, foi o primeiro a discutir a análise de dados longitudinais através do modelo de

curva de crescimento. O significado de razão de crescimento e suas mudanças foram comparadas em grupos de tratamentos através da análise de variância univariada.

Box (1950) estudou o caso de selecionar vetores de dados de uma normal multivariada considerando uma estrutura de covariância uniforme.

Rao (1958, 1959) realizou um cuidadoso estudo dos problemas e vantagens decorrentes de se combinar na mesma equação diferentes potências da variável resposta.

Elston e Grizzle (1962) notaram que o método univariado pode ser aplicado ao mesmo problema considerado por Rao (1959), desde que se possa fazer a suposição que as observações em cada unidade são não-correlacionadas e homocedásticas ou que a matriz é uniforme.

Hoel (1964) estudou o efeito de se ignorar a dependência das observações tomadas em diferentes pontos do tempo sobre a validade da inferência estatística.

Em 1964, Potthoff e Roy apresentaram o modelo de curva de crescimento polinomial que pode ser escrito como uma generalização da MANOVA usual. Este modelo é adequado para dados completos, todos os indivíduos de um grupo têm o mesmo valor esperado e curvas polinomiais ajustadas para diferentes grupos têm o mesmo grau, com alguns coeficientes podendo ser zero.

Rao (1959, 1965, 1966) desenvolveu métodos multivariados usando análise de covariância. O modelo de curva de crescimento foi desenvolvido em uma série de artigos por Rao (1959, 1965, 1966, 1967, 1968, 1987) e foi estendido por vários autores, entre eles, Khatri (1966) e Grizzle e Allen (1969), que permitiram uma escolha arbitrária para a matriz de covariância.

Geisser (1970, 1971), Lee e Geisser (1972, 1975), Fearn (1975), Lee (1988, 1991) e Barry (1995) desenvolveram um tratamento Bayesiano do modelo de curva de crescimento generalizado.

Khatri (1973) desenvolveu testes da razão de verossimilhança para matriz de covariância estruturada.

Kleinbaum (1973 a e b) adaptou o modelo de Potthoff e Roy e propôs uma generalização que permite sua aplicação a vetores de observações incompletas por planejamento ou por falta de observações. Seu modelo reduz o modelo de curva de crescimento ao modelo de mínimos quadrados ponderados univariado com matriz de covariância desconhecida e a usual suposição de normalidade.

Ghosh, Grizzle e Sen (1973) aplicaram métodos não-paramétricos no modelo de Potthoff e Roy.

Tubbs, Lewis e Duran (1975) generalizaram o modelo de curva de crescimento de Potthoff e Roy relaxando a suposição de independência entre diferentes indivíduos e consideraram uma matriz de covariância genérica, positiva definida, para o *vec* da matriz de dados.

Lee e Geisser (1975) apresentaram um estudo sobre a previsão parcial ou condicional em modelos de curva de crescimento para várias estruturas de covariância, focalizando o ponto de vista empírico. O desenvolvimento da parte teórica nele utilizado foi dado por Lee e Geisser (1972).

Woolson e Leeper (1980) revisaram a análise de curva de crescimento para dados longitudinais completos e incompletos, baseados no modelo de Potthoff e Roy. Os estudos de dados longitudinais incompletos são apresentados num contexto mais geral do que a modelagem de Kleinbaum.

Crowder e Tredger (1981) consideraram observações em tempos não equidistantes.

Modelos para dados não balanceados descritos por Rao (1965) e Fearn (1975) foram generalizados por Harville (1977), Laird e Ware (1982) e Reinsel (1985).

O modelo de efeitos mistos foi sugerido por Laird e Ware (1982) e também foi considerado por Fairclough e Helms (1986), juntamente com métodos computacionais, para analisar dados que podem ser incompletos.

Reinsel (1984) estendeu o modelo de curva de crescimento multidimensional assumindo curvas individuais e considerando efeitos fixos e aleatórios. Reinsel (1982, 1984,

1985) generalizou o modelo de efeitos aleatórios para incorporar respostas multivariadas, considerando planejamentos não balanceados.

Os estimadores de máxima verossimilhança e máxima verossimilhança restrita para a média e a matriz de covariância são dados por Laird e Ware (1982) e estendidos por Laird, Lange e Stram (1987). Algoritmos mais gerais foram dados por Jennrich e Schluchter (1986), que estenderam o modelo para incorporar formas arbitrárias da estrutura de covariância.

Ware (1985) discutiu o modelo de regressão geral onde a matriz de covariância dentro de indivíduo tem três tipos de estruturas, multivariada geral, efeitos aleatórios e autoregressiva.

Azzalini (1987) considerou o modelo de efeitos aleatórios com covariadas e matriz de covariância estruturada.

Vonesh e Carter (1987) apresentaram um método não iterativo para estimar e comparar parâmetros de locação no modelo de curva de crescimento com coeficientes aleatórios e apresentaram propriedades assintóticas dos parâmetros de locação.

Khatri (1988) investigou o modelo de Potthoff e Roy supondo que os erros não têm distribuição normal, mas sim distribuição elíptica ou esférica.

Laird (1988) fez uma revisão sobre o uso de análises baseadas em verossimilhança para dados longitudinais incompletos.

Lindstrom e Bates (1988) examinaram métodos para melhorar a eficiência computacional do algoritmo EM.

Schluchter (1988) considerou uma larga variedade de tipos de modelo para dados incompletos e desbalanceados, listando vários tipos de estruturas de covariância.

Chi e Reinsel (1989) consideraram modelos de efeitos aleatórios, desenvolveram teste escore para autocorrelação, apresentaram um procedimento explícito para obtenção do estimador de máxima verossimilhança e consideraram previsão de respostas futuras e estimação de Bayes empírica dos efeitos aleatórios.

Lange e Laird (1989) estudaram para uma determinada família de modelos de curva de crescimento, como a mudança no padrão da estrutura de covariância pode afetar a variância estimada dos estimadores da média no caso balanceado.

Stanek III (1990) propôs uma unificação de modelos com medidas repetidas incluindo modelos de curvas de crescimento e ilustrou como transformações simples podem ser usadas para determinar o grau do polinômio em curvas polinomiais.

Um método recursivo para obter estimadores de máxima verossimilhança e máxima verossimilhança restrita para o modelo linear geral para análise de medidas repetidas foi dado por McGilchrist e Cullis (1990) e estendido por Wu e McGilchrist (1995). Este último considerou várias estruturas de covariância.

Von Rosen (1991) apresentou um histórico sobre o modelo de curva de crescimento cobrindo, entre outros, métodos de estimação de parâmetros, versão canônica do modelo, testes, extensões, dados incompletos, enfoque Bayesiano e diferentes estruturas de covariância. O artigo traz uma vasta bibliografia sobre o assunto.

Leeden, Vrijburg e Leeuw (1996) fizeram a revisão de dois procedimentos em análise de dados longitudinais, o primeiro considerando que as medidas são hierárquicas dentro do indivíduo e o segundo sendo a MANOVA com matriz de covariância estruturada.

Referências sobre modelagem da matriz de covariância são Lavange e Helms (1983), Jennrich e Schluchter (1986), Lee (1988), Schluchter (1988), Jones (1993), Keramidas e Lee (1995), Wu e McGilchrist (1995) e BMDP (1992).

Ware (1985) considerou alguns tipos especiais de estruturas.

Jennrich e Schluchter (1986) discutiram vários tipos de estruturas de covariância usando diferentes algoritmos para obter o estimador de máxima verossimilhança.

Com relação à análise de dados incompletos, imputação e previsão de observações futuras, Hartley e Hocking (1971), Orchard e Woodbury (1972) e Beale e Little (1975) fizeram imputação usando um algoritmo que mais tarde foi formalizado por Dempster, Laird e Rubin (1977) e é conhecido como algoritmo EM.

Laird (1982) introduziu uma classe de estimadores de máxima verossimilhança generalizados que inclui o estimador de máxima verossimilhança restrita e o estimador de máxima verossimilhança ordinário. Fornece equações do algoritmo EM para estimar as componentes de variância no modelo misto. Outras referências sobre dados incompletos são Schwertman e Allen (1979), Schwertman et al. (1980), Schwertman et al. (1985), McGilchrist e Yau (1995), Nummi (1997) e Barroso, Bussab e Knott (1998), cujo interesse era a estimação de parâmetros e/ou a realização de testes de hipóteses. Imputação foi feita por Liski (1984, 1985) e Liski e Nummi (1990) através de métodos Bayesianos e do algoritmo EM. Imputação de dados provenientes de populações finitas através do BLUP (Best Linear Unbiased Predictor) foi sugerida por Barroso (1995) e considerada por Barroso e Salinas (1999).

A previsão de observações futuras foi tratada por Geisser (1974), Reinsel (1984), Rao (1987) e Tian, Shukla e Buncher (1994).

2.3 Modelo de curva de crescimento

Nesta seção apresentamos os modelos descritos na literatura, mais freqüentemente utilizados para a análise de dados longitudinais.

O modelo linear multivariado introduzido por Potthoff e Roy (1964) é adequado para dados completos e é dado por

$$\mathbf{Y}_{p \times n} = \mathbf{X}_{p \times q} \boldsymbol{\xi}_{q \times f} \mathbf{A}_{f \times n} + \mathbf{e}_{p \times n}, \quad (2.1)$$

onde $\mathbf{Y}$ é a matriz de observações, cujas colunas são independentes e possuem distribuição normal p -variada, com média $\mathbf{X}\boldsymbol{\xi}\mathbf{A}$ e matriz de covariância $\boldsymbol{\Sigma}$ arbitrária e positiva definida. As matrizes $\mathbf{X}$ e $\mathbf{A}$ são conhecidas de posto completo q e f , respectivamente; se $\mathbf{A}$ tem posto incompleto utilizam-se matrizes inversas generalizadas ou reparametriza-se o modelo de modo que $\text{posto}(\mathbf{A}) = f$. A matriz $\boldsymbol{\xi}$ é a matriz de parâmetros (desconhecidos) e $\mathbf{e}$

é matriz de erros cujas colunas são independentes. Seja $\mathbf{e}_i$ a i -ésima coluna de $\mathbf{e}$ ($i = 1, 2, \dots, n$); supõe-se que $\mathbf{e}_i \sim \mathbf{N}_p(\mathbf{0}, \Sigma)$. O modelo (2.1), conhecido como o modelo de curva de crescimento polinomial de Potthoff e Roy, é uma generalização do modelo da MANOVA usual

$$\mathbf{Y}_{0_{q \times n}} = \boldsymbol{\xi}_{q \times f} \mathbf{A}_{f \times n} + \mathbf{e}_{0_{q \times n}}, \quad (2.2)$$

onde as colunas de $\mathbf{Y}_0$ são independentes com distribuição normal q -variada têm média $\boldsymbol{\xi} \mathbf{A}$ e matriz de covariância Σ_0 positiva definida.

É possível escrever o modelo (2.1) como (2.2) através da transformação

$$\mathbf{Y}_0 = (\mathbf{X}' \mathbf{G}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{G}^{-1} \mathbf{Y},$$

onde $\mathbf{G}_{p \times p}$ é qualquer matriz simétrica positiva definida tal que $\mathbf{X} \mathbf{G}^{-1} \mathbf{X}'$ tenha posto completo. Nesse caso, Σ_0 é dada por

$$\Sigma_0 = (\mathbf{X}' \mathbf{G}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{G}^{-1} \Sigma \mathbf{G}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{G}^{-1} \mathbf{X})^{-1}.$$

Para fazer essa transformação é necessário escolher $\mathbf{G}$ adequadamente. Quando $p = q$ a análise sobre (2.1) é a mesma do modelo MANOVA usual (2.2), mas se $q < p$ a escolha de $\mathbf{G}$ pode afetar os resultados. Neste caso, Potthoff e Roy (1964) discutem algumas possibilidades, entre elas Σ , que é a escolha ótima pelo critério ENVVM (estimador não-viesado de variância mínima), ou a identidade $\mathbf{I}_p$ ou ainda $\mathbf{S}$ o estimador não-viesado usual quando Σ é desconhecida.

Exemplo 2.6. Suponha que quatro indivíduos são observados em três instantes de tempo (t_1, t_2, t_3) e que estamos ajustando uma reta média aos dados. O modelo de curva de crescimento de Potthoff e Roy é dado por

$$\mathbf{Y}_{p \times n} = \mathbf{X}_{p \times q} \boldsymbol{\xi}_{q \times f} \mathbf{A}_{f \times n} + \mathbf{e}_{p \times n},$$

onde

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & y_{13} & y_{14} \\ y_{21} & y_{22} & y_{23} & y_{24} \\ y_{31} & y_{32} & y_{33} & y_{34} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{bmatrix}, \quad \boldsymbol{\xi} = \begin{bmatrix} \xi_0 \\ \xi_1 \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{e} = \begin{bmatrix} e_{11} & e_{12} & e_{13} & e_{14} \\ e_{21} & e_{22} & e_{23} & e_{24} \\ e_{31} & e_{32} & e_{33} & e_{34} \end{bmatrix}.$$

Se considerarmos que os dois primeiros indivíduos pertencem a um grupo e os outros dois a outro grupo, ou seja $f = 2$, e estamos ajustando retas médias para cada um dos grupos, temos que $\mathbf{Y}$, $\mathbf{X}$ e $\mathbf{e}$ são as mesmas e

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_{01} & \xi_{02} \\ \xi_{11} & \xi_{12} \end{bmatrix} \quad \text{e} \quad \mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Generalizando, suponha que há f grupos (ou tratamentos) distintos, cada qual com n_j indivíduos (unidades experimentais) independentes tal que $\sum_{j=1}^f n_j = n$ e que uma variável Y é medida para cada indivíduo ao longo de p instantes de tempo $t_1, t_2, \dots, t_p$. O modelo linear multivariado generalizado de curva de crescimento pode ser escrito como (2.1), ou seja,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\xi}\mathbf{A} + \mathbf{e}, \quad (2.3)$$

onde $\mathbf{Y}_{p \times n}$ é a matriz das observações, $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_f)$, e $\mathbf{Y}_j$, matriz de dimensão $p \times n_j$ das observações associadas ao j -ésimo grupo, é especificada por

$$\mathbf{Y}_j = \begin{bmatrix} y_{11j} & y_{12j} & \dots & y_{1n_jj} \\ y_{21j} & y_{22j} & \dots & y_{2n_jj} \\ \vdots & \vdots & \ddots & \vdots \\ y_{p1j} & y_{p2j} & \dots & y_{pn_jj} \end{bmatrix},$$

sendo que as observações $y_{1ij}, y_{2ij}, \dots, y_{pij}$, $i = 1, 2, \dots, n_j$ e $j = 1, 2, \dots, f$ podem ser correlacionadas e têm matriz de covariância $\boldsymbol{\Sigma}_{p \times p}$. Assumimos que $\boldsymbol{\Sigma}$ é a mesma para todos

os grupos, ou seja, a matriz de covariância das colunas de $\mathbf{Y}_j$, $j = 1, 2, \dots, f$, é Σ . $\mathbf{A}_{f \times n}$ é a matriz de planejamento, que em geral é dada por $\mathbf{A} = \text{diag}(\mathbf{1}'_{n_j})$, onde $\mathbf{1}'_{n_j}$ é o vetor linha de n_j elementos iguais a 1, $\xi_{q \times f} = (\xi_1, \xi_2, \dots, \xi_f)$, é a matriz de parâmetros (desconhecidos) com $\xi_j = [\xi_{0j}, \xi_{1j}, \dots, \xi_{(q-1)j}]'$ sendo o vetor de coeficientes da curva para o j -ésimo grupo, $(q-1)$ é o grau do polinômio ajustado e $q-1 < p$. $\mathbf{X}_{p \times q}$ é a matriz das potências dos instantes de tempo e $\mathbf{e}_{p \times n}$ é a matriz de erros aleatórios, cujas colunas são supostas independentes e com distribuição $N_p(\mathbf{0}, \Sigma)$, onde Σ é qualquer matriz positiva definida. Seja $\mathbf{y} = \text{vec}(\mathbf{Y})$, onde $\text{vec}(\mathbf{Y})$ indica o empilhamento das colunas de $\mathbf{Y}$. O valor esperado de $\mathbf{Y}$ e a matriz de covariância de $\mathbf{y}$ são

$$E(\mathbf{Y}) = \mathbf{X}\xi\mathbf{A} \quad (2.4)$$

e

$$\mathbf{V} = \text{Var}(\mathbf{y}) = \mathbf{I}_n \otimes \Sigma, \quad (2.5)$$

onde $\mathbf{I}_n$ é a matriz identidade de dimensão n e $\otimes$ denota o produto de Kronecker de duas matrizes.

Nomes como MANOVA Generalizada, GMANOVA, Modelo Linear Multivariado e Modelo de Potthoff e Roy são comumente usados para denominar Modelos de Curvas de Crescimento (von Rosen, 1991).

Em geral $\mathbf{X}_{p \times q}$ é definida como

$$\mathbf{X} = \begin{bmatrix} t_1^0 & t_1 & t_1^2 & \dots & t_1^{q-1} \\ t_2^0 & t_2 & t_2^2 & \dots & t_2^{q-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_p^0 & t_p & t_p^2 & \dots & t_p^{q-1} \end{bmatrix}$$

mas, às vezes, é conveniente escrever $\mathbf{X}_{p \times q}$ como $\mathbf{X}_{0_p \times q}$, onde

$$\mathbf{X}_0 = \begin{bmatrix} P_0(t_1) & P_1(t_1) & \dots & P_{q-1}(t_1) \\ P_0(t_2) & P_1(t_2) & \dots & P_{q-1}(t_2) \\ \vdots & \vdots & \vdots & \vdots \\ P_0(t_p) & P_1(t_p) & \dots & P_{q-1}(t_p) \end{bmatrix},$$

sendo $P_l(t)$ polinômios ortogonais de grau l ($l = 0, 1, 2, \dots, q-1$), definidos sobre os instantes de tempo $t_1, t_2, \dots, t_p$.

Se os instantes são equidistantes, há tabelas de polinômios ortogonais (Pearson e Hartley, 1972), caso contrário é necessário recorrer a alguns programas computacionais como por exemplo o módulo CM do aplicativo NTIA/EMBRAPA.

Obtemos pelo modelo (2.3) que o valor esperado da resposta para os indivíduos submetidos ao j -ésimo tratamento no k -ésimo instante de tempo é o mesmo, ou seja,

$$E(y_{kij}) = \xi_{0j}t_k^0 + \xi_{1j}t_k^1 + \xi_{2j}t_k^2 + \dots + \xi_{(q-1)j}t_k^{q-1}$$

com $j = 1, 2, \dots, f$, $i = 1, 2, \dots, n_j$ e $k = 1, 2, \dots, p$. Notemos que y_{kij} é a observação do i -ésimo indivíduo do j -ésimo tratamento no k -ésimo instante de tempo.

Outra forma de escrever o modelo (2.3) é através de modelos individuais dados por

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\xi} \mathbf{a}_i + \mathbf{e}_i, \quad (2.6)$$

onde $\mathbf{y}_{i_{p \times 1}}$ é o vetor de observações do i -ésimo indivíduo para $i = 1, 2, \dots, n$, $\mathbf{X}_{p \times q}$ e $\boldsymbol{\xi}_{q \times f}$ são definidas como em (2.3), $\mathbf{a}_{i_{f \times 1}}$ é o vetor indicador associado ao i -ésimo indivíduo e $\mathbf{e}_{i_{p \times 1}}$ é o vetor de erros assumidos independentes com distribuição $\mathbf{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$.

Uma forma equivalente e mais flexível para o modelo (2.6) é

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i, \quad (2.7)$$

onde $\mathbf{y}_{i_{p_i \times 1}}$ é o vetor de observações do i -ésimo indivíduo ($i = 1, 2, \dots, n$) e p_i é o número de ocasiões nas quais o indivíduo i foi observado. $\mathbf{X}_{i_{p_i \times q}}$ é a matriz (conhecida) de planejamento, $\boldsymbol{\beta}_{q \times 1}$ é o vetor de parâmetros e $\mathbf{e}_{i_{p_i \times 1}}$ é o vetor de erros aleatórios, supostos

independentes com distribuição $N_{p_i}(\mathbf{0}, \Sigma_i)$, onde Σ_i é qualquer matriz simétrica positiva definida.

O modelo (2.7) incorpora o fato de que nem sempre os dados são balanceados, isto é, nem sempre $\mathbf{X}_i = \mathbf{X}$, $i = 1, 2, \dots, n$, e permite considerar estruturas de covariância diferentes para as unidades experimentais $\Sigma_i = \text{Var}(\mathbf{y}_i)$.

Outro modelo muito utilizado na área é o modelo misto proposto por Laird e Ware (1982) que, para o i -ésimo indivíduo, é dado por

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i, \quad i = 1, 2, \dots, n \quad (2.8)$$

onde $\mathbf{y}_i$, $i = 1, 2, \dots, n$, é o vetor das respostas observadas nos p_i instantes de tempo, $\mathbf{X}_i$ e $\mathbf{Z}_i$ são matrizes conhecidas de ordem $(p_i \times q)$ e $(p_i \times c)$, respectivamente, $\boldsymbol{\alpha}_{q \times 1}$ é o vetor de parâmetros fixos, $\mathbf{b}_{i_{c \times 1}}$ é o vetor de parâmetros aleatórios e $\mathbf{e}_{i_{p_i \times 1}}$ é o vetor de erros aleatórios. Além disso, supõe-se que $\mathbf{b}_i$ e $\mathbf{e}_i$ são independentes e independentes entre si, com $\mathbf{b}_i \sim N_c(\mathbf{0}, \mathbf{G}_i)$, e $\mathbf{e}_i \sim N_{p_i}(\mathbf{0}, \mathbf{R}_i)$, onde $\mathbf{R}_i$ e $\mathbf{G}_i$ são matrizes positivas definidas.

O modelo (2.8) pode ser escrito na forma

$$\mathbf{y} = \mathbf{X} \boldsymbol{\alpha} + \mathbf{Z} \mathbf{b} + \mathbf{e}, \quad (2.9)$$

onde $\mathbf{y}' = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_n)$, $\mathbf{X}' = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_n)$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_i)$, $\boldsymbol{\alpha}$ é o mesmo definido anteriormente, $\mathbf{b}' = (\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n)$ e $\mathbf{e}' = (\mathbf{e}'_1, \mathbf{e}'_2, \dots, \mathbf{e}'_n)$. As dimensões de $\mathbf{y}$, $\mathbf{X}$, $\mathbf{Z}$, $\boldsymbol{\alpha}$, $\mathbf{b}$ e $\mathbf{e}$ são $(\sum_{i=1}^n p_i \times 1)$, $(\sum_{i=1}^n p_i \times q)$, $(\sum_{i=1}^n p_i \times nc)$, $(q \times 1)$, $(nc \times 1)$ e $(\sum_{i=1}^n p_i \times 1)$, respectivamente.

Quando os dados são balanceados com relação ao tempo e as matrizes de planejamento e de covariância são iguais para todos os indivíduos, digamos $\mathbf{X}_i = \tilde{\mathbf{X}}$, $\mathbf{Z}_i = \tilde{\mathbf{Z}}$, $\mathbf{R}_i = \mathbf{R}$ e $\mathbf{G}_i = \mathbf{G}$, para $i = 1, 2, \dots, n$, o modelo é simplificado de modo que $\mathbf{X} = \mathbf{1}_n \otimes \tilde{\mathbf{X}}$ e $\mathbf{Z} = \mathbf{I}_n \otimes \tilde{\mathbf{Z}}$, com dimensões $(np \times q)$ e $(np \times nc)$. As suposições passam a ser

$$E(\mathbf{e}) = \mathbf{0}, \quad \text{Var}(\mathbf{e}) = \tilde{\mathbf{R}} = \mathbf{I}_n \otimes \mathbf{R},$$

$$E(\mathbf{b}) = \mathbf{0}, \quad \text{Var}(\mathbf{b}) = \tilde{\mathbf{G}} = \mathbf{I}_n \otimes \mathbf{G}$$

e

$$\text{Cov}(\mathbf{b}, \mathbf{e}) = \mathbf{0},$$

onde $\tilde{\mathbf{R}}$ e $\tilde{\mathbf{G}}$ são matrizes positivas definidas, de posto completo e dimensões $np \times np$ e $nc \times nc$, respectivamente. A média e a variância de $\mathbf{y}$ são dadas, respectivamente, por

$$\mathbf{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\alpha}$$

e

$$\mathbf{V} = \text{Var}(\mathbf{y}) = \mathbf{Z}\tilde{\mathbf{G}}\mathbf{Z}' + \tilde{\mathbf{R}}.$$

Em geral, é assumido que os erros associados a diferentes instantes para o mesmo indivíduo são independentes e têm a mesma variância, ou seja, $\tilde{\mathbf{R}} = \sigma^2 \mathbf{I}_{np}$.

Os modelos mistos podem ser apresentados na forma de estágios: no primeiro estágio, define-se o modelo linear condicionado aos parâmetros da curva de crescimento individual, isto é, $\boldsymbol{\alpha}$ e $\mathbf{b}$ são parâmetros fixos desconhecidos e $\mathbf{e} \sim \mathbf{N}(\mathbf{0}, \tilde{\mathbf{R}})$; no segundo estágio, os parâmetros das curvas são supostos linearmente dependentes das covariáveis fixadas. Formalmente:

1º estágio

$$\mathbf{y}_i | \mathbf{b}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i,$$

onde $\mathbf{e}_i \sim \mathbf{N}_{p_i}(\mathbf{0}, \mathbf{R}_i)$

2º estágio

$$\mathbf{b}_i \sim \mathbf{N}_c(\mathbf{0}, \mathbf{G}_i).$$

Exemplo 2.7: Suponha que dois indivíduos são observados em três instantes de tempo (t_1, t_2, t_3) e que estamos ajustando uma reta para cada indivíduo. O modelo misto para o i -ésimo indivíduo no k -ésimo instante será dado por

$$y_{ki} = (a + bt_k) + (a_i + b_i t_k) + e_{t_k i}, \quad (2.10)$$

ou seja,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{b} + \mathbf{e}$$

com

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \\ y_{32} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} a \\ b \end{bmatrix},$$

$$\mathbf{Z} = \begin{bmatrix} 1 & t_1 & 0 & 0 \\ 1 & t_2 & 0 & 0 \\ 1 & t_3 & 0 & 0 \\ 0 & 0 & 1 & t_1 \\ 0 & 0 & 1 & t_2 \\ 0 & 0 & 1 & t_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{bmatrix} \quad \text{e} \quad \mathbf{e} = \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{12} \\ e_{22} \\ e_{32} \end{bmatrix}.$$

Com a suposição de erros independentes e mesma matriz de covariância para os efeitos individuais independentes, $\tilde{\mathbf{G}}$ é bloco diagonal e

$$\text{Var}(\mathbf{y}) = \mathbf{Z}\tilde{\mathbf{G}}\mathbf{Z}' + \sigma^2\mathbf{I}_{np}.$$

A partir de agora passaremos a tratar o modelo de curva de crescimento através do modelo misto, para dados balanceados, considerando no Capítulo 4 o caso de dados balanceados por planejamento e incompletos por falta de observações. Nas duas próximas seções ilustramos a modelagem da média e da estrutura de covariâncias.

2.4 Modelos para as médias

O modelo (2.9) permite, por exemplo, expressar a resposta média num tempo t_k em termos de variáveis explicativas medidas no mesmo instante t_k ou em termos do próprio tempo t_k . A especificação da matriz $\mathbf{X}$ é um dos elementos mais importantes na definição do modelo para as médias e suas colunas podem estar associadas:

1. aos fatores que definem a estrutura das subpopulações;
2. ao fator tempo, identificando, por exemplo, a forma da curva a ser ajustada;
3. a possíveis covariáveis, cujo efeito na resposta média desejamos investigar.

Assumimos que todos os indivíduos são observados em p instantes, $t_1, t_2, \dots, t_p$, considerando as respostas como função do tempo. No Capítulo 5 fazemos algumas aplicações e exemplificamos a seguir os modelos que lá serão utilizados.

1. Modelo I: polinômio de grau 1 (retas individuais)

$$\tilde{\mathbf{X}} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_p \end{bmatrix} \quad \text{e} \quad \tilde{\mathbf{Z}} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_p \end{bmatrix}.$$

$$\mathbf{X} = \mathbf{1}_n \otimes \tilde{\mathbf{X}}, \quad \mathbf{Z} = \mathbf{I}_n \otimes \tilde{\mathbf{Z}}, \quad \boldsymbol{\alpha}' = (a, b) \quad \text{e} \quad \mathbf{b}' = (a_1, b_1, a_2, b_2, \dots, a_n, b_n).$$

2. Modelo II: polinômio de grau 2 (parábolas individuais)

$$\tilde{\mathbf{X}} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_p & t_p^2 \end{bmatrix} \quad \tilde{\mathbf{Z}} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_p & t_p^2 \end{bmatrix}$$

$$\mathbf{X} = \mathbf{1}_n \otimes \tilde{\mathbf{X}}, \quad \mathbf{Z} = \mathbf{I}_n \otimes \tilde{\mathbf{Z}}, \quad \boldsymbol{\alpha}' = (a, b, c) \quad \text{e} \quad \mathbf{b}' = (a_1, b_1, c_1, a_2, b_2, c_2, \dots, a_n, b_n, c_n).$$

3. **Modelo III:** consideramos também um modelo no qual a observação $y_{t_k i}$ do i -ésimo indivíduo no k -ésimo instante é uma soma dada por

$$y_{t_k i} = \alpha_{t_k} + b_i + e_{t_k i},$$

onde

α_{t_k} : efeito do tempo (fixo) com p instantes;

b_i : efeito do indivíduo (aleatório) com n observações;

$e_{t_k i}$'s $\sim N(0, \sigma^2)$, independentes;

b_i 's $\sim N(0, \sigma_b^2)$, independentes;

$e_{t_k i}$'s e b_i 's são independentes;

$\mathbf{X} = \mathbf{1}_n \otimes \mathbf{I}_p$ e $\mathbf{Z} = \mathbf{I}_n \otimes \mathbf{1}_p$.

Esse modelo impõe que a estrutura de covariância do vetor $\mathbf{y}$ seja o produto de Kronecker entre a matriz identidade e a matriz simétrica composta, que será definida na próxima seção.

2.5 Estruturas de covariância

É necessário, em estudos longitudinais, considerar que os dados obtidos são correlacionados entre si. Esta correlação pode seguir algum padrão ou não, que é incorporado ao modelo através da escolha da estrutura de covariância. A vantagem em assumir uma matriz de covariância estruturada é que o número de parâmetros é, em geral, menor do que se a matriz fosse não estruturada, permitindo a obtenção de resultados mais eficientes. A estrutura mais adequada depende do entendimento do processo que gerou os dados. Pode-se optar por usar diversas estruturas e escolher a que apresenta um ajuste melhor. Um ponto a ser considerado na escolha é a questão da interpretação.

A seguir apresentamos algumas estruturas para a matriz de covariância individual, extraídas de Lee (1988), Schluchter (1988), Wolfinger (1993), Keramidas e Lee (1995), Wu e McGilchrist (1995) e BMDP (1992).

No modelo misto é possível fazer combinações desses tipos de estruturas para serem assumidas nas matrizes de covariância de efeitos e erros. Consideramos que as matrizes são funções de um vetor de parâmetros θ .

Enumeramos a seguir diversas estruturas de covariância consideradas na literatura. Nem todas serão consideradas neste trabalho, mas são listadas para que o leitor tenha uma visão de vários tipos de estrutura.

1. Diagonal

$$\begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix},$$

onde $\sigma^2 > 0$ e $\theta = \sigma^2$. Esta é a estrutura mais simples para a matriz de covariância, pois considera que as observações na mesma unidade experimental são independentes e têm variância constante. A grande maioria dos artigos, de uma forma ou de outra, considera esse tipo de estrutura.

2. Componentes de Variância ou “Banded” diagonal principal

$$\begin{bmatrix} \sigma_1 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_p \end{bmatrix},$$

onde $\sigma_k > 0, k = 1, 2, \dots, p$ e $\theta = (\sigma_1, \sigma_2, \dots, \sigma_p)'$. Esta estrutura considera que as observações no mesmo indivíduo são independentes, mas podem ter variâncias diferentes.

3. “Banded” Toeplitz ou média móvel de primeira ordem, MA(1)

$$\sigma^2 \begin{bmatrix} 1 & \rho & 0 & 0 & \dots & 0 \\ \rho & 1 & \rho & 0 & \dots & 0 \\ 0 & \rho & 1 & \rho & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix},$$

com $\sigma^2 > 0$ e $|\rho| < 1$, $\boldsymbol{\theta} = (\sigma^2, \rho)'$. Aqui há suposição de que as observações adjacentes são igualmente correlacionadas e que as variâncias são constantes. Além disso, observações espaçadas por mais de um período de tempo são não correlacionadas. Azzalini (1987) e Schluchter (1988) usaram esta estrutura. Duas extensões são dadas a seguir.

4. Simétrica composta com grupos heterogêneos

$$\begin{bmatrix} \sigma_1 & \sigma_2 & 0 & 0 & \dots & 0 \\ \sigma_2 & \sigma_1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{p-1} & \sigma_p \\ 0 & 0 & 0 & \dots & \sigma_p & \sigma_{p-1} \end{bmatrix},$$

com $\sigma_k > 0$ para $k = 1, 2, \dots, p$ e vetor de parâmetros $\boldsymbol{\theta} = (\sigma_1, \sigma_2, \dots, \sigma_p)$.

5. “Banded”

$$\begin{bmatrix} \sigma_1 & \sigma_{p+1} & 0 & 0 & \dots & 0 \\ \sigma_{p+1} & \sigma_2 & \sigma_{p+2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{p-1} & \sigma_{p+(p-1)} \\ 0 & 0 & 0 & \dots & \sigma_{p+(p-1)} & \sigma_p \end{bmatrix},$$

onde $\sigma_k > 0$ para $k = 1, 2, \dots, p$ com vetor de parâmetros $\boldsymbol{\theta} = (\sigma^2, \sigma_1, \sigma_2, \dots, \sigma_{p+(p-1)})'$.

Esta estrutura é uma extensão das anteriores, pois permite que as variâncias sejam

diferentes e que, embora cada observação seja correlacionada somente com as observações adjacentes, esta correlação pode estar mudando com o tempo.

6. Simétrica Composta

$$\begin{bmatrix} \sigma_1 + \sigma_2 & \sigma_1 & \sigma_1 & \dots & \sigma_1 \\ \sigma_1 & \sigma_1 + \sigma_2 & \sigma_1 & \dots & \sigma_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1 & \sigma_1 & \sigma_1 & \dots & \sigma_1 + \sigma_2 \end{bmatrix},$$

com $\sigma_1 + \sigma_2 > 0, -1/(p-1) < \sigma_1/(\sigma_1 + \sigma_2) < 1$, para que a matriz seja positiva definida, e $\boldsymbol{\theta} = (\sigma_1, \sigma_2)'$. Lee (1988), Schluchter (1988), Stanek III (1990) e Wu e McGilchrist (1995) usaram a estrutura simétrica composta (ou equivalentemente, a uniforme).

7. Uniforme

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix},$$

com $\sigma^2 > 0, -1/(p-1) < \rho < 1$, para garantir que a matriz seja positiva definida, e $\boldsymbol{\theta} = (\sigma^2, \rho)'$. Se fizermos $\rho = 1/(\sigma^2 + \sigma_1)$ a estrutura uniforme é equivalente a simétrica composta. Esta estrutura supõe que as variáveis têm a variância constante e que são igualmente correlacionadas.

8. Auto-regressiva de primeira ordem, AR(1), ou correlação serial

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{p-1} \\ \rho & 1 & \rho & \dots & \rho^{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \rho^{p-3} & \dots & 1 \end{bmatrix},$$

com $\sigma^2 > 0$, $|\rho| < 1$ e $\boldsymbol{\theta} = (\sigma^2, \rho)'$. A estrutura AR(1) permite modelar uma situação comum na prática, onde a correlação entre as observações realizadas na mesma unidade experimental descrece à medida que a distância entre elas aumenta. Esse decrescimento segue um padrão. A estrutura de correlação serial é muito utilizada. Algumas referências são Potthoff e Roy (1964), Lee e Geisser (1975), Lee (1988), Schluchter (1988), Chi e Reinsel (1989), Keramidis e Lee (1990, 1995), Wolfinger (1993), Wu e McGilchrist (1995) e BMDP (1992).

9. Toeplitz ou auto-regressiva geral

$$\begin{bmatrix} \sigma_p & \sigma_1 & \sigma_2 & \dots & \sigma_{p-1} \\ \sigma_1 & \sigma_p & \sigma_1 & \dots & \sigma_{p-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{p-1} & \sigma_{p-2} & \sigma_{p-3} & \dots & \sigma_p \end{bmatrix},$$

onde $\sigma_p > 0$. Adotar este tipo de estrutura implica que as variâncias das observações são constantes assim como as covariâncias (correlações) entre duas medidas quaisquer separadas pelo mesmo intervalo de tempo. Esta estrutura pode ser vista como uma estrutura de média móvel de ordem igual ao tamanho da matriz (Wolfinger, 1993). Segundo Lee (1988), esta estrutura parece ser a mais natural para dados no modelo de curva de crescimento, uma vez que tais dados são freqüentemente compostos por muitas séries curtas. A estrutura auto-regressiva geral também foi considerada por Rao (1967) e Krishnaiah (1975).

10. Não-estruturada

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \dots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \sigma_{3p} & \dots & \sigma_{pp} \end{bmatrix},$$

onde $\sigma_{kk} > 0$, para $k = 1, 2, \dots, p$ e com vetor de parâmetros $\boldsymbol{\theta} = (\sigma_{11}, \sigma_{12}, \dots, \sigma_{pp})'$. Esta estrutura é adequada quando não há bases empíricas ou teóricas para adotar

uma estrutura especial (padronizada) e os dados são completos ou balanceados. Quando a matriz não estruturada é adotada podem aparecer problemas de estimação e computacionais, pois o número de parâmetros a serem estimados, $p(p + 1)/2$, pode ser muito grande e tende a crescer com p , afetando a eficiência dos resultados bem como impossibilitando a estimação de alguns parâmetros, quando os dados são altamente incompletos. Essa estrutura é considerada por grande número de autores.

11. Variância Comum

$$\begin{bmatrix} \sigma_1 & \sigma_1 & \sigma_1 & \dots & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \dots & \sigma_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_1 & \sigma_1 & \sigma_1 & \dots & \sigma_1 \end{bmatrix},$$

onde $\sigma_1 > 0$ e $\boldsymbol{\theta} = \sigma_1$. Esta estrutura é bem restrita e pode não representar adequadamente os dados, pois impõe que a correlação entre duas observações, próximas ou não, seja igual a 1.

12. Fator Analítico

$$\Delta\Delta' + \Psi,$$

onde $\Delta_{p \times k}$ é uma matriz de pesos conhecida e $\Psi_{p \times p}$ é uma matriz diagonal. Essa é a estrutura imposta pela Análise Fatorial, onde Δ é a matriz de cargas fatoriais e Ψ é a matriz das variâncias específicas.

13. Linear

$$\sum_{i=1}^a \theta_i \mathbf{G}_i,$$

onde as $\mathbf{G}_i$'s são matrizes simétricas conhecidas e θ_i 's são os parâmetros da matriz de covariância. Algumas estruturas de covariância mais comuns, entre elas a uniforme, simétrica composta, AR(1) e não-estruturada, têm a forma da estrutura linear. Schluchter (1988) sugeriu o uso deste tipo de estrutura.

14. Espacial

$$\sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \dots & \rho^{d_{1p}} \\ \rho^{d_{12}} & 1 & \rho^{d_{23}} & \dots & \rho^{d_{2p}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{d_{1p}} & \rho^{d_{2p}} & \rho^{d_{3p}} & \dots & 1 \end{bmatrix},$$

onde $\sigma^2 > 0$, $-1/(p-1) < \rho < 1$ e d_{ij} é a distância euclidiana entre a i -ésima e a j -ésima coordenadas espaciais ou $d_{ij} = |t_i - t_j|$, sendo t_i e t_j os tempos associados com o i -ésimo e j -ésimo elementos da matriz de correlação. É uma diferente generalização da estrutura AR(1) que inclui séries de tempos não-equidistantes e dados de espaços de duas ou mais dimensões.

As estruturas AR(1) e MA(1) são casos particulares da estrutura auto-regressiva geral. Além disso, a estrutura MA(1) pode ser considerada um caso particular das estruturas “Banded” e “Banded” Toeplitz. A diagonal é caso particular da “Banded” diagonal principal, a “Banded” Toeplitz da simétrica composta com grupos heterogêneos e da “Banded” e a uniforme da AR(1), que por sua vez é caso particular da Toeplitz.

As estruturas mais usuais são diagonal, simétrica composta (uniforme), AR(1) e não-estruturada.

Wolfinger (1993) apresentou as seguintes estruturas de covariância para os modelos mistos: diagonal, componentes de variância, simétrica composta, não-estruturada, AR(1), “Banded”, Toeplitz, “Banded” Toeplitz, simétrica composta com grupos heterogêneos, AR(1) + variância comum, AR(1) + diagonal e espacial.

As estruturas apresentadas podem ser utilizadas para a modelagem das matrizes $\mathbf{R}$ e $\mathbf{G}$ nos modelos de efeitos mistos (2.9) e não são necessariamente do mesmo tipo, ou seja, as estruturas para as matrizes de covariância podem ser impostas separadamente ou como uma combinação destas.

As quatro combinações mais comuns, também consideradas por Azzalini (1987), Chi e Reinsel (1989) e Keramidas e Lee (1995) são: variância comum e não-estruturada para

a matriz dos efeitos aleatórios $\mathbf{G}$ combinadas com AR(1) e diagonal para a matriz dos erros $\mathbf{R}$.

Diante de tantas possibilidades para a estrutura da matriz de covariância temos que selecionar aquela que é a mais apropriada ao modelo. Esta escolha pode ser feita com base num conhecimento a priori sobre os dados ou ser baseada em procedimentos como erro quadrático médio mínimo, erro absoluto médio mínimo, testes da razão de verossimilhança e critérios de informação. Esta escolha pode ser abordada também nas fases de análise exploratória e diagnóstico dos dados.

Lange e Laird (1989) estudaram o efeito que a suposição da estrutura de covariância tem sobre a variância dos estimadores dos parâmetros.

Wolfinger (1993) utilizou técnicas estatísticas para comparar as várias estruturas adotadas para a matriz de covariância. A seleção do modelo misto geral é feito em três estágios usando máxima verossimilhança (MV) e máxima verossimilhança restrita (MVR). Kermidas e Lee (1995) sugeriram determinadas estruturas a priori, usando resultados de simulação para cada uma e escolhendo a que leva ao menor erro quadrático médio e ao menor erro absoluto médio.

Dawson, Gennings e Carter (1997) mostraram como determinar uma forma inicial para a estrutura da matriz de covariância através de duas técnicas gráficas. Pan, Fangan e von Rosen (1997) utilizaram o estudo de influência local de uma forma de perturbação na matriz de covariância para avaliar a adequação dessa estrutura ao modelo de curva de crescimento.

A discussão mais aprofundada de técnicas de escolha da estrutura de covariância não faz parte do escopo deste trabalho.

Capítulo 3

Estimação dos Parâmetros

Neste capítulo tratamos da estimação dos parâmetros do modelo misto (2.9), ou seja, da estimação dos efeitos fixos α e do vetor θ de parâmetros da matriz de covariância.

Apresentamos os métodos de estimação de máxima verossimilhança (MV) e de máxima verossimilhança restrita (MVR). Para dados balanceados e sob certas condições, existem fórmulas explícitas para os estimadores, que são apresentadas na Seção 3.1. Em outras situações, processos iterativos são comumente utilizados para a obtenção de estimativas. Na Seção 3.2 descrevemos os três algoritmos mais populares, que são: EM, Newton-Raphson e Scoring de Fisher.

3.1 Fórmulas explícitas dos estimadores

Existem certas situações em modelos de curvas de crescimento sob as quais é possível obter fórmulas explícitas para os estimadores de máxima verossimilhança e de máxima verossimilhança restrita (para maiores detalhes ver Graybill (1976), Laird, Lange e Stram (1987), Lange e Laird (1989), Andreoni (1989) e Nummi (1997)).

Assim, consideramos que os dados sejam balanceados no tempo, isto é, que para todos os indivíduos pesquisados, os dados sejam observados nos mesmos instantes de tempo.

Então os estimadores explícitos são baseados em uma parte completa das observações para a qual n_c denota o número de indivíduos e p_c o número de instantes. Além disso, consideramos que o grau do polinômio ajustado é menor ou igual a $p_c - 1$ e que $p_c < n_c$.

A maioria dos autores considera que os erros dentro de indivíduos são independentes e que a matriz de covariância dos efeitos aleatórios é não estruturada. Entre eles estão Vonesh e Carter (1987) e Andreoni (1989). McGilchrist e Yau (1995) consideram que a matriz de covariância dos erros é conhecida a menos de uma constante e Nummi (1997) estende os resultados da estimação de verossimilhança para o caso em que a variância dos erros tem uma estrutura parcimoniosa.

Como já foi descrito no Capítulo 2, o modelo misto de efeitos fixos e aleatórios é dado por

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{b} + \mathbf{e},$$

onde $\mathbf{y}' = (y'_1, y'_2, \dots, y'_{n_c})$, $\mathbf{X}' = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_{n_c})$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_i)$, $\mathbf{b}' = (b'_1, b'_2, \dots, b'_{n_c})$, $\mathbf{e}' = (e'_1, e'_2, \dots, e'_{n_c})$ e $\boldsymbol{\alpha}$ é o mesmo para todos os indivíduos, com as suposições

$$\begin{aligned} E(\mathbf{e}) &= \mathbf{0}, & \text{Var}(\mathbf{e}) &= \tilde{\mathbf{R}}, \\ E(\mathbf{b}) &= \mathbf{0} & \text{e} & \text{Var}(\mathbf{b}) = \tilde{\mathbf{G}}. \end{aligned}$$

A matriz de covariância de $\mathbf{y}$ é $\mathbf{V} = \mathbf{Z}\tilde{\mathbf{G}}\mathbf{Z}' + \tilde{\mathbf{R}}$.

Estamos considerando primeiro que os indivíduos são independentes, que as curvas individuais ajustadas têm o mesmo grau e que as matrizes de covariância são as mesmas, isto é,

$$\begin{aligned} \mathbf{X}_i &= \tilde{\mathbf{X}}, & \mathbf{Z}_i &= \tilde{\mathbf{Z}}, \\ \tilde{\mathbf{R}} &= \mathbf{I}_{n_c} \otimes \mathbf{R}, & \tilde{\mathbf{G}} &= \mathbf{I}_{n_c} \otimes \mathbf{G}, \\ \mathbf{V} &= \mathbf{I}_{n_c} \otimes \boldsymbol{\Sigma} & \text{e} & \boldsymbol{\Sigma} = \tilde{\mathbf{Z}}\mathbf{G}\tilde{\mathbf{Z}}' + \mathbf{R}, \end{aligned}$$

onde $\otimes$ indica o produto de Kronecker.

Esse modelo pode ser escrito na forma matricial

$$\mathbf{Y} = \mathbf{1}_{n_c}\boldsymbol{\alpha}'\tilde{\mathbf{X}}' + \mathbf{B}\tilde{\mathbf{Z}}' + \mathbf{E}, \quad (3.1)$$

onde $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n_c})'$ com $\mathbf{y} = \text{vec}(\mathbf{Y}')$, $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{n_c})'$ com $\mathbf{b} = \text{vec}(\mathbf{B}')$ e $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n_c})'$ com $\mathbf{e} = \text{vec}(\mathbf{E}')$.

Esse modelo é um caso particular do modelo

$$\mathbf{Y} = \mathbf{A}\Phi \mathbf{T}' + \Delta \mathbf{T}'_c + \mathbf{E}, \quad (3.2)$$

onde as dimensões de $\mathbf{Y}$, $\mathbf{A}$, Φ , $\mathbf{T}$, Δ , $\mathbf{T}_c$ e $\mathbf{E}$ são $(n_c \times p_c)$, $(n_c \times f)$, $(f \times q)$, $(p_c \times q)$, $(n_c \times c)$, $(p_c \times c)$ e $(n_c \times p_c)$, respectivamente.

Se $\mathbf{T}_c$ é a matriz formada pelas primeiras c colunas de $\mathbf{T}$, o grau das curvas individuais é menor ou igual ao grau da curva populacional.

Assim, os estimadores serão obtidos para o modelo mais geral (3.2) e depois identificados para o modelo de interesse (3.1). O desenvolvimento dos estimadores de máxima verossimilhança e máxima verossimilhança restrita é apresentado em detalhes em Andreoni (1989). Aqui apresentamos uma derivação diferente do estimador de máxima verossimilhança, dada em Nummi (1997) e detalhada na Seção 3.1.1.

3.1.1 Estimação de máxima verossimilhança

O método de máxima verossimilhança é baseado na maximização da função de verossimilhança. Os parâmetros são estimados simultaneamente e são dados pelo ponto de máximo da função densidade conjunta. Assim, além das suposições feitas sobre os momentos dos efeitos e erros aleatórios, é necessário também fazer suposições sobre a forma das distribuições dessas quantidades aleatórias. Nesse contexto, estamos supondo que essas distribuições são normais.

Maximizar a função de verossimilhança é equivalente a maximizar o logaritmo dessa função.

Suponha que $\mathbf{R} = \sigma^2 \mathbf{I}_{p_c}$, ou seja, os erros são independentes mesmo que correspondam a um mesmo indivíduo e têm variância constante. Além disso, suponha que $\mathbf{G}$ é não estruturada. Nesse caso, o logaritmo da função de verossimilhança, multiplicado por 2, é

dado por

$$2l(\mathbf{Y}; \Phi, \sigma^2, \mathbf{G}) = K - n_c \log|\Sigma| - \text{tr}[(\mathbf{Y} - \mathbf{A}\Phi \mathbf{T}')'(\mathbf{Y} - \mathbf{A}\Phi \mathbf{T}')\Sigma^{-1}], \quad (3.3)$$

onde K é uma constante que não depende dos parâmetros. A função (3.3) depende de σ^2 e $\mathbf{G}$ através de $\Sigma = \mathbf{T}_c \mathbf{G} \mathbf{T}_c' + \sigma^2 \mathbf{I}_{p_c}$.

Não há solução não-iterativa direta para $\mathbf{G}$ e uma reparametrização é necessária. Usando a igualdade matricial $(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB})^{-1}\mathbf{BVA}^{-1}$ (Rao, 1973) e o fato que $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$ para dimensões adequadas (Mardia, Kent e Bibby, 1979) e depois de alguma álgebra (ver resultados A1 e A2 do Apêndice A), vale que

$$2l(\mathbf{Y}; \Phi, \sigma^2, \Psi) = K - \text{tr}\{\mathbf{P}(\Phi)[\sigma^{-2}\mathbf{I}_{p_c} - \sigma^{-2}\mathbf{T}_c(\mathbf{T}_c'\mathbf{T}_c)^{-1}\mathbf{T}_c' + \mathbf{T}_c(\mathbf{T}_c'\mathbf{T}_c)^{-1}\Psi^{-1}(\mathbf{T}_c'\mathbf{T}_c)^{-1}\mathbf{T}_c']\} \\ - n_c \log|\Psi| - n_c \log|\mathbf{T}_c'\mathbf{T}_c| - n_c(p_c - c)\log(\sigma^2), \quad (3.4)$$

onde K é uma constante que não depende dos parâmetros, $\mathbf{P}(\Phi) = (\mathbf{Y} - \mathbf{A}\Phi \mathbf{T}')'(\mathbf{Y} - \mathbf{A}\Phi \mathbf{T}')$ e $\Psi = \sigma^2(\mathbf{T}_c'\mathbf{T}_c)^{-1} + \mathbf{G}$.

O estimador de mínimos quadrados ordinários e o estimador de máxima verossimilhança de Φ são idênticos (Rao (1967) e Reinsel (1982, 1985)). Este é dado por

$$\hat{\Phi} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}, \quad (3.5)$$

cuja variância é dada por

$$\text{Var}(\text{vec}(\hat{\Phi})) = (\mathbf{A}'\mathbf{A})^{-1} \otimes [\mathbf{J}_c \mathbf{G} \mathbf{J}_c' + \sigma^2(\mathbf{T}'\mathbf{T})^{-1}], \quad (3.6)$$

onde $\mathbf{J}_c = (\mathbf{I}_c : \mathbf{0})'$, de dimensão $(q \times c)$.

Maximizar (3.4) é equivalente a maximizar a função de verossimilhança perfilada $2l(\mathbf{Y}; \hat{\Phi}, \sigma^2, \Psi)$, que depende de Ψ através da função

$$f(\Psi) = -n_c \log|\Psi| - \text{tr}[\Psi^{-1}(\mathbf{T}_c'\mathbf{T}_c)^{-1}\mathbf{T}_c'\mathbf{P}(\hat{\Phi})\mathbf{T}_c(\mathbf{T}_c'\mathbf{T}_c)^{-1}],$$

que é maximizada (ver resultado A3 no Apêndice A e Seber (1984), A.7.1) em

$$\hat{\Psi} = \frac{1}{n_c} (\mathbf{T}'_c \mathbf{T}_c)^{-1} \mathbf{T}'_c \mathbf{P}(\hat{\Phi}) \mathbf{T}_c (\mathbf{T}'_c \mathbf{T}_c)^{-1}, \quad (3.7)$$

onde

$$\mathbf{P}(\hat{\Phi}) = (\mathbf{Y} - \mathbf{A} \hat{\Phi} \mathbf{T}')' (\mathbf{Y} - \mathbf{A} \hat{\Phi} \mathbf{T}').$$

O estimador $\hat{\Psi}$ pode ser simplificado para

$$\hat{\Psi} = \frac{1}{n_c} (\mathbf{T}'_c \mathbf{T}_c)^{-1} \mathbf{T}'_c \mathbf{S} \mathbf{T}_c (\mathbf{T}'_c \mathbf{T}_c)^{-1}, \quad (3.8)$$

onde $\mathbf{S} = \mathbf{Y}' [\mathbf{I}_{n_c} - \mathbf{A}(\mathbf{A}' \mathbf{A})^{-1} \mathbf{A}'] \mathbf{Y}$.

Resta agora maximizar $2l(\mathbf{Y}; \hat{\Phi}, \sigma^2, \hat{\Psi})$ com respeito a σ^2 , o que resulta em

$$\hat{\sigma}^2 = \frac{\text{tr}[\mathbf{P}(\hat{\Phi})(\mathbf{I}_{p_c} - \mathbf{T}_c(\mathbf{T}'_c \mathbf{T}_c)^{-1} \mathbf{T}'_c)]}{n_c(p_c - c)}. \quad (3.9)$$

Agora, dados os estimadores de Ψ e σ^2 , é possível estimar $\mathbf{G}$, ou seja,

$$\hat{\mathbf{G}} = \hat{\Psi} - \hat{\sigma}^2 (\mathbf{T}'_c \mathbf{T}_c)^{-1}. \quad (3.10)$$

Um problema neste caso é que a matriz $\mathbf{G}$ pode não ser positiva definida. Chi e Reinsel (1989) sugerem a reparametrização de $\mathbf{G}$ pela decomposição de Cholesky, o que resultará em uma matriz não-negativa.

Generalização da matriz de covariância dos erros

Nesta parte fazemos a extensão dos resultados apresentados anteriormente para o caso em que a matriz de covariância dos erros tem uma estrutura mais geral que a identidade, mas ainda tem uma forma simples que depende de um parâmetro ρ , isto é, $\mathbf{R} = \sigma^2 \mathbf{R}(\rho)$ e $\Sigma = \mathbf{T}_c \mathbf{G} \mathbf{T}'_c + \mathbf{R}$. As estruturas Banded Toeplitz, simétrica composta, uniforme e correlação serial estão nessa classe.

Para obter o estimador de máxima verossimilhança, primeiro consideramos ρ fixado e encontramos $\hat{\Phi}(\rho)$, $\hat{\Psi}(\rho)$ e $\hat{\sigma}^2(\rho)$. Seja $\mathbf{R}^{-1/2}$ a matriz tal que $\mathbf{R}^{-1/2} \mathbf{R}^{-1/2} = \mathbf{R}^{-1}$ e

transforme os dados de maneira que a matriz $\mathbf{Y}$ seja pós-multiplicada por $\mathbf{R}^{-1/2}$, isto é, $E(\mathbf{Y}\mathbf{R}^{-1/2}) = \mathbf{A}\Phi\mathbf{T}'_*$, onde $\mathbf{T}'_* = \mathbf{T}'\mathbf{R}^{-1/2}$.

Assim,

$$\Sigma_* = \mathbf{R}^{-1/2}\mathbf{T}_c\mathbf{G}\mathbf{T}'_c\mathbf{R}^{-1/2} + \sigma^2\mathbf{R}^{-1/2}\mathbf{R}\mathbf{R}^{-1/2} = \mathbf{T}_{c*}\mathbf{G}\mathbf{T}'_{c*} + \sigma^2\mathbf{I}_{p_c},$$

onde $\mathbf{T}'_{c*} = \mathbf{T}'_c\mathbf{R}^{-1/2}$.

Agora basta obter os estimadores de $\sigma^2(\rho)$, $\Psi(\rho)$ e $\Phi(\rho)$ para a variável transformada, dados por

$$\hat{\sigma}^2(\rho) = \frac{\text{tr}\{\mathbf{P}(\hat{\Phi}(\rho))[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{T}_c(\mathbf{T}'_c\mathbf{R}^{-1}\mathbf{T}_c)^{-1}\mathbf{T}'_c\mathbf{R}^{-1}]\}}{n_c(p_c - c)}, \quad (3.11)$$

$$\hat{\Psi}(\rho) = \frac{1}{n_c}(\mathbf{T}'_c\mathbf{R}^{-1}\mathbf{T}_c)^{-1}\mathbf{T}'_c\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{T}_c(\mathbf{T}'_c\mathbf{R}^{-1}\mathbf{T}_c)^{-1} \quad (3.12)$$

e

$$\hat{\Phi}(\rho) = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{R}^{-1}\mathbf{T}(\mathbf{T}'\mathbf{R}^{-1}\mathbf{T})^{-1}, \quad (3.13)$$

com as restrições de que $\hat{\sigma}^2(\rho) > 0$ e $\hat{\Psi}(\rho) \geq \hat{\sigma}^2(\rho)(\mathbf{T}'_c\mathbf{R}^{-1}\mathbf{T}_c)^{-1}$.

O próximo passo é maximizar o logaritmo da função de verossimilhança perfilada $l(\hat{\Phi}(\rho), \hat{\Psi}(\rho), \hat{\sigma}^2(\rho), \rho)$, que depende somente de ρ . Se essa função for complicada, é possível utilizar um dos algoritmos para obtenção de pontos de máximo.

Estimadores no modelo simplificado

Nesta subseção fazemos a identificação dos estimadores sob o modelo de interesse (3.1). Como caso particular desse modelo inclui-se a modelagem individual de polinômios de mesmo grau. Então

$$\hat{\alpha}' = (\mathbf{1}'_{n_c}\mathbf{1}_{n_c})^{-1}\mathbf{1}'_{n_c}\mathbf{Y}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \quad (3.14)$$

ou

$$\hat{\alpha} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}},$$

onde $\bar{\mathbf{y}}$ é o vetor de dimensão $(p_c \times 1)$ das médias amostrais,

$$\hat{\sigma}^2 = \frac{1}{n_c(p_c - c)} \text{tr}[\mathbf{P}(\hat{\boldsymbol{\alpha}})(\mathbf{I}_{p_c} - \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}')] \quad (3.15)$$

e

$$\hat{\mathbf{G}} = \frac{1}{n_c}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\mathbf{S}\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1} - \hat{\sigma}^2(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}, \quad (3.16)$$

onde $\mathbf{P}(\hat{\boldsymbol{\alpha}}) = (\mathbf{Y} - \mathbf{1}_{n_c}\hat{\boldsymbol{\alpha}}'\tilde{\mathbf{X}})'(\mathbf{Y} - \mathbf{1}_{n_c}\hat{\boldsymbol{\alpha}}'\tilde{\mathbf{X}})$ e $\mathbf{S} = \mathbf{Y}'[\mathbf{I}_{n_c} - \frac{1}{n_c}\mathbf{1}_{n_c}\mathbf{1}_{n_c}']\mathbf{Y}$.

3.1.2 Estimação de máxima verossimilhança restrita

Os estimadores de máxima verossimilhança dos parâmetros das matrizes de covariância em geral subestimam os parâmetros porque não levam em conta a perda de graus de liberdade resultante por se estimar os efeitos fixos. Para corrigir isso, Patterson e Thompson (1971) sugeriram o estimador de máxima verossimilhança restrita (ver também Harville (1974, 1977)).

A função de verossimilhança restrita não depende da matriz de efeitos fixos $\boldsymbol{\alpha}$. A idéia é maximizar a função, obtendo os estimadores de máxima verossimilhança restrita e depois obter o estimador de mínimos quadrados generalizados de $\boldsymbol{\alpha}$.

Seguindo os mesmos passos da seção anterior, temos que 2 vezes o logaritmo da função de verossimilhança restrita é dada por

$$2l_R(\mathbf{Y}; \sigma^2, \mathbf{G}) = K - n_c \log|\boldsymbol{\Sigma}| - \text{tr}[(\mathbf{Y} - \mathbf{A}\hat{\boldsymbol{\Phi}}\mathbf{T}')'(\mathbf{Y} - \mathbf{A}\hat{\boldsymbol{\Phi}}\mathbf{T}')\boldsymbol{\Sigma}^{-1}] - f \log|\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T}| - q \log|\mathbf{A}'\mathbf{A}|, \quad (3.17)$$

onde K é uma constante que não depende dos parâmetros, $\hat{\boldsymbol{\Phi}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}$, e (3.17) depende de σ^2 e $\mathbf{G}$ através de $\boldsymbol{\Sigma} = \mathbf{T}_c\mathbf{G}\mathbf{T}_c' + \sigma^2\mathbf{I}_{p_c}$.

Analogamente aos resultados obtidos na seção anterior, também apresentados por Andreoni (1989), tem-se que os estimadores de máxima verossimilhança restrita são

$$\hat{\boldsymbol{\Psi}}_R = \frac{1}{n_c - f}(\mathbf{T}_c'\mathbf{T}_c)^{-1}\mathbf{T}_c'\mathbf{S}\mathbf{T}_c(\mathbf{T}_c'\mathbf{T}_c)^{-1}, \quad (3.18)$$

$$\hat{\sigma}_R^2 = \frac{\text{tr}[\mathbf{P}(\hat{\Phi})(\mathbf{I}_{p_c} - \mathbf{T}_c(\mathbf{T}_c'\mathbf{T}_c)^{-1}\mathbf{T}_c')]}{n_c(p_c - c) - f(q - c)} \quad (3.19)$$

e

$$\hat{\mathbf{G}}_R = \hat{\Psi}_R - \hat{\sigma}_R^2(\mathbf{T}_c'\mathbf{T}_c)^{-1}, \quad (3.20)$$

onde $\mathbf{P}(\hat{\Phi}) = (\mathbf{Y} - \mathbf{A}\hat{\Phi}\mathbf{T}')'(\mathbf{Y} - \mathbf{A}\hat{\Phi}\mathbf{T}')$ e $\mathbf{S} = \mathbf{Y}'[\mathbf{I}_{n_c} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{Y}$.

Note que os resultados são parecidos com os obtidos pelo método de máxima verossimilhança; a diferença está nos denominadores das expressões. No modelo simplificado, basta fazer essa mudança.

3.1.3 Estimação para o Modelo III

O Modelo III, definido na Seção 2.4, é um caso particular do modelo misto segundo o qual uma observação é dada pela soma de três parcelas: um efeito fixo representando o tempo, um efeito aleatório representando o indivíduo e um erro aleatório. Esse modelo impõe que a estrutura de covariância dos dados seja bloco diagonal, onde as matrizes dos blocos da diagonal têm a forma da matriz simétrica composta dada por

$$\begin{bmatrix} \sigma^2 + \sigma_b^2 & \sigma^2 & \sigma^2 & \dots & \sigma^2 \\ \sigma^2 & \sigma^2 + \sigma_b^2 & \sigma^2 & \dots & \sigma^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma^2 & \sigma^2 & \sigma^2 & \dots & \sigma^2 + \sigma_b^2 \end{bmatrix}.$$

Os estimadores explícitos das componentes dessa matriz, obtidos pelo método da ANOVA, são baseados em uma parte completa das observações e são dados por

$$\hat{\sigma}^2 = \frac{A - B}{(n_c - 1)(p_c - 1)} \quad (3.21)$$

e

$$\hat{\sigma}_b^2 = \begin{cases} \frac{p_c B - A}{p_c(n_c - 1)(p_c - 1)}, & \text{se esse valor é } \geq 0 \\ 0, & \text{caso contrário} \end{cases}, \quad (3.22)$$

onde n_c e p_c são, respectivamente, o número de unidades experimentais e o número de instantes do conjunto completo considerado para o cálculo das estimativas. Além disso,

$$A = \sum_{i=1}^{n_c} \sum_{k=1}^{p_c} (y_{t_k i} - \bar{y}_{t_k.})^2 \quad \text{e} \quad B = \sum_{i=1}^{n_c} \sum_{k=1}^{p_c} (\bar{y}_{.i} - \bar{y}_{..})^2,$$

onde

$\bar{y}_{t_k.}$ é a média das observações das n_c unidades experimentais no instante t_k ;

$\bar{y}_{.i}$ é a média das observações do i -ésimo indivíduo nos p_c instantes de tempo;

$\bar{y}_{..}$ é a média de todas as observações, ou seja, das n_c unidades experimentais nos p_c instantes de tempo.

O estimador $\hat{\sigma}_b^2$ é truncado no zero para garantir que a estimativa não seja negativa. Quando as distribuições dos efeitos e erros aleatórios são supostas normais, esses estimadores coincidem com o estimador de máxima verossimilhança restrita e com o estimador MINQUE (Barroso, 1995).

As quantidades A e B podem ser escritas em forma matricial, dadas por

$$A = \mathbf{y}_c' (\mathbf{I}_{n_c p_c} - \mathbf{X}_c (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c') \mathbf{y}_c$$

e

$$B = \frac{1}{p_c} \mathbf{y}_c' (\mathbf{I}_{n_c p_c} - \mathbf{X}_c (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c') \mathbf{Z}_c \mathbf{Z}_c' \mathbf{y}_c,$$

onde $\mathbf{y}_c$ é a parte completa do vetor $\mathbf{y}$, $\mathbf{X}_c = \mathbf{1}_{n_c} \otimes \mathbf{I}_{p_c}$ e $\mathbf{Z}_c = \mathbf{I}_{n_c} \otimes \mathbf{1}_{p_c}$.

3.2 Algoritmos para maximização da função de verossimilhança

Nem sempre é possível obter fórmulas explícitas para os estimadores de máxima verossimilhança e máxima verossimilhança restrita. É vasta a literatura sobre a utilização

de algoritmos iterativos utilizados com essa finalidade. Os três algoritmos mais populares são Newton-Raphson, Scoring de Fisher e o algoritmo EM (esperança e maximização).

Dentre as aplicações a dados longitudinais destacamos Laird (1982), que introduziu uma classe de estimadores de máxima verossimilhança generalizados que inclui o estimador de máxima verossimilhança restrita, e forneceu as equações do algoritmo EM para estimar as componentes da variância no modelo misto.

Laird e Ware (1982) discutiram em detalhes o uso do EM para obter os estimadores considerando os erros homocedásticos e independentes. Esse trabalho teve continuidade em Laird, Lange e Stram (1987), que apresentaram dois enfoques para o uso do EM.

Jennrich e Schluchter (1986) focalizaram o modelo de curva de crescimento com dados não balanceados ou incompletos e descreveram os algoritmos de Newton-Raphson e Scoring de Fisher para computar o estimador de máxima verossimilhança e o EM híbrido com o Scoring para os dois tipos de estimadores.

Chi e Reinsel (1989) consideraram um modelo mais geral do que Laird e Ware (1982) pois permitiram que os erros dentro dos indivíduos fossem correlacionados. Eles usaram o algoritmo Scoring de Fisher.

O algoritmo EM, apresentado por Dempster, Laird e Rubin (1977), tem a propriedade de que a função de verossimilhança não diminui com o passar das iterações. Entretanto ele é mais lento do que os outros dois algoritmos. Lindstrom e Bates (1988) consideraram métodos para tornar o EM computacionalmente mais eficiente e implementaram o algoritmo de Newton-Raphson que, segundo eles, é preferível ao EM.

Uma vantagem do EM é que ele pode ser usado quando o mecanismo de não resposta é ignorável ou não (Rubin, 1987).

Outros algoritmos também são apresentados na literatura. Harville (1977) fez uma revisão sobre a estimação de máxima verossimilhança e máxima verossimilhança restrita das componentes da variância apresentando algoritmos gerais para sua computação e relacionando com outros métodos como a ANOVA, a teoria MINQUE e estimação Bayesiana.

Outro método computacional recursivo para obter os estimadores para o modelo linear geral foi dado por McGilchrist e Cullis (1990) e estendido por Wu e McGilchrist (1995).

Para implementação dos algoritmos, consideramos o modelo (2.9).

3.2.1 Os algoritmos de Newton-Raphson e Scoring de Fisher

Os algoritmos de Newton-Raphson e Scoring de Fisher são processos iterativos baseados na expansão de Taylor da função de verossimilhança até o termo de ordem 2. Mais detalhes podem ser encontrados em Bard (1974) e Andreoni (1989). Assim, é necessário obter as primeiras e segundas derivadas da função de verossimilhança, ou de seu logaritmo, que para o modelo

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i, \quad i = 1, 2, \dots, n \quad (3.23)$$

com

$$\begin{aligned} E(\mathbf{e}_i) &= \mathbf{0}, & \text{Var}(\mathbf{e}_i) &= \sigma^2 \mathbf{I}_{p_i}, \\ E(\mathbf{b}_i) &= \mathbf{0}, & \text{Var}(\mathbf{b}_i) &= \mathbf{G} \quad (\text{não estruturada}), \\ \Sigma_i &= \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{p_i}, \end{aligned}$$

é dado por

$$l(\mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = K - \frac{1}{2} \sum_{i=1}^n \log |\Sigma_i| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}), \quad (3.24)$$

onde K é uma constante que não depende dos parâmetros.

As primeiras e segundas derivadas são dadas por

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

e

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

Se a matriz $\mathbf{G}$ pode ser escrita como uma combinação linear de matrizes conhecidas tais que $\mathbf{G} = \sum_{g=1}^a \theta_g \mathbf{G}_g$ e conseqüentemente $\Sigma_i = \sum_{g=1}^a \theta_g \mathbf{G}_{ig}$ com $\mathbf{G}_{ig} = \mathbf{Z}_i \mathbf{G}_g \mathbf{Z}_i'$ para

$g = 1, 2, \dots, a - 1$ e $\mathbf{G}_{ig} = \mathbf{I}_{pi}$ para $g = a$ e $\theta_a = \sigma^2$, os elementos dos vetores e matrizes acima são

$$\begin{aligned} \mathbf{s}_\alpha &= \sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}), \\ (q \times 1) \end{aligned} \quad (3.25)$$

$$\begin{aligned} \mathbf{s}_\theta : [\mathbf{s}_\theta]_g &= \left[\frac{1}{2} \sum_{i=1}^n \text{tr}[\Sigma_i^{-1} ((\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \Sigma_i) \Sigma_i^{-1} \mathbf{G}_{ig}] \right], \\ (a \times 1) \end{aligned} \quad (3.26)$$

$$\begin{aligned} \mathbf{H}_{\alpha\alpha} &= - \sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} \mathbf{X}_i, \\ (q \times q) \end{aligned} \quad (3.27)$$

$$\begin{aligned} \mathbf{H}_{\alpha\theta} &= \mathbf{H}'_{\theta\alpha} : [\mathbf{H}_{\alpha\theta}]_{jg} = \left[- \sum_{i=1}^n \mathbf{x}_{ij}' \Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) \right] \\ (q \times a) \end{aligned} \quad (3.28)$$

e

$$\begin{aligned} \mathbf{H}_{\theta\theta} : [\mathbf{H}_{\theta\theta}]_{gh} &= \left[- \frac{1}{2} \sum_{i=1}^n \text{tr}[\Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} (2(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})' - \Sigma_i) \Sigma_i^{-1} \mathbf{G}_{ih}] \right], \\ (a \times a) \end{aligned} \quad (3.29)$$

para $g, h = 1, 2, \dots, a$, e onde $\mathbf{x}_{ij}$ é a j -ésima coluna da matriz $\mathbf{X}_i$.

O algoritmo de Newton-Raphson calcula o valor dos parâmetros na $(w + 1)$ -ésima iteração ($\boldsymbol{\alpha}^{(w+1)}$ e $\boldsymbol{\theta}^{(w+1)}$) com base nos valores obtidos na w -ésima iteração ($\boldsymbol{\alpha}^{(w)}$ e $\boldsymbol{\theta}^{(w)}$) usando

$$\begin{bmatrix} \boldsymbol{\alpha}^{(w+1)} \\ \boldsymbol{\theta}^{(w+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}^{(w)} \\ \boldsymbol{\theta}^{(w)} \end{bmatrix} - \begin{bmatrix} \mathbf{H}_{\alpha\alpha} & \mathbf{H}_{\alpha\theta} \\ \mathbf{H}_{\theta\alpha} & \mathbf{H}_{\theta\theta} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{s}_\alpha \\ \mathbf{s}_\theta \end{bmatrix}, \quad (3.30)$$

onde o vetor escore $\mathbf{s}$ e a matriz Hessiana $\mathbf{H}$ são calculados nos valores correntes $\boldsymbol{\alpha}^{(w)}$ e $\boldsymbol{\theta}^{(w)}$.

Outro algoritmo de maximização bastante utilizado é o algoritmo Scoring de Fisher, que é uma modificação do algoritmo de Newton-Raphson. A diferença na expressão (3.30) é que a matriz Hessiana $\mathbf{H}$, ou seja, a matriz de informação empírica é substituída por sua esperança, ou seja, a matriz de informação de Fisher. Então, o algoritmo é dado por

$$\begin{bmatrix} \boldsymbol{\alpha}^{(w+1)} \\ \boldsymbol{\theta}^{(w+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}^{(w)} \\ \boldsymbol{\theta}^{(w)} \end{bmatrix} + \begin{bmatrix} \mathbf{I}_{\alpha\alpha} & \mathbf{I}_{\alpha\theta} \\ \mathbf{I}_{\theta\alpha} & \mathbf{I}_{\theta\theta} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{s}_\alpha \\ \mathbf{s}_\theta \end{bmatrix}, \quad (3.31)$$

onde

$$\mathbf{I}_{\alpha\alpha} = -\mathbf{E}[\mathbf{H}_{\alpha\alpha}] = \sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i, \quad (3.32)$$

$$\mathbf{I}_{\alpha\theta} = \mathbf{I}'_{\theta\alpha} = -\mathbf{E}[\mathbf{H}_{\alpha\theta}] = \mathbf{0} \quad (3.33)$$

e

$$\mathbf{I}_{\theta\theta} = -\mathbf{E}[\mathbf{H}_{\theta\theta}] : [\mathbf{I}_{\theta\theta}]_{gh} = \left[\frac{1}{2} \sum_{i=1}^n \text{tr}[\boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} \mathbf{G}_{ih}] \right]. \quad (3.34)$$

Como $\mathbf{I}_{\alpha\theta} = \mathbf{0}$, a expressão (3.31) pode ser simplificada e particionada em

$$\boldsymbol{\alpha}^{(w+1)} = \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{y}_i \right) \quad (3.35)$$

e

$$\begin{bmatrix} \boldsymbol{\theta}^{(w+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}^{(w)} \end{bmatrix} + \begin{bmatrix} \mathbf{I}_{\theta\theta} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{s}_\theta \end{bmatrix}. \quad (3.36)$$

Como estamos supondo que $\boldsymbol{\Sigma}_i$ tem a estrutura linear, a expressão (3.36) pode ser escrita como

$$\left[\boldsymbol{\theta}^{(w+1)} \right]_g = \left[\sum_{i=1}^n \text{tr}[\boldsymbol{\Sigma}_i^{-1(w)} \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{G}_{ih}] \right]^{-1} \left[\sum_{i=1}^n \text{tr}[\boldsymbol{\Sigma}_i^{-1(w)} \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1(w)} \mathbf{r}_i^{(w+1)} \mathbf{r}_i^{(w+1)'}] \right], \quad (3.37)$$

onde $\mathbf{r}_i^{(w+1)} = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w+1)}$.

Aos leitores interessados no desenvolvimento dessas derivadas e esperanças, sugerimos a leitura de Andreoni (1989), onde todos os detalhes são fornecidos.

Esses dois algoritmos podem também ser utilizados para a obtenção dos estimadores de máxima verossimilhança restrita. Para isso, basta seguir o processo análogo, somente aplicado à função de máxima verossimilhança restrita. Os resultados são os seguintes:

$$l_R(\mathbf{y}; \boldsymbol{\theta}) = K - \frac{1}{2} \sum_{i=1}^n \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \log \left| \sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\alpha}}), \quad (3.38)$$

onde K é uma constante que não depende dos parâmetros e

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

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

$$\mathbf{H}_R = [\partial^2 l_R / \partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}],$$

onde

$$\mathbf{s}_R : [\mathbf{s}_R]_g = \left[\frac{1}{2} \sum_{i=1}^n \text{tr}[\Sigma_i^{-1} \mathbf{G}_{ig} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\alpha})(\mathbf{y}_i - \mathbf{X}_i \tilde{\alpha})' - \mathbf{P}_i \mathbf{G}_{ig}] \right] \quad (3.39)$$

e

$$\mathbf{H}_R : [\mathbf{H}_R]_{gh} = \left[- \sum_{i=1}^n \text{tr}[\Sigma_i^{-1} \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih} \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \tilde{\alpha})(\mathbf{y}_i - \mathbf{X}_i \tilde{\alpha})'] + \frac{1}{2} \sum_{i=1}^n \text{tr}[\mathbf{P}_i \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih}] \right], \quad (3.40)$$

para $g, h = 1, 2, \dots, a$ e $\mathbf{P}_i = \Sigma_i^{-1} - \Sigma_i^{-1} \mathbf{X}_i (\sum_{i=1}^n \mathbf{X}_i' \Sigma_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i' \Sigma_i^{-1}$.

A matriz de informação de Fisher correspondente é

$$\mathbf{I}_R = -\mathbf{E}[\mathbf{H}_R] : [\mathbf{I}_R]_{gh} = \left[\frac{1}{2} \sum_{i=1}^n \text{tr}[\mathbf{P}_i \mathbf{G}_{ig} \mathbf{P}_i \mathbf{G}_{ih}] \right]. \quad (3.41)$$

As expressões para os algoritmos de Newton-Raphson e Scoring de Fisher para a estimação de máxima verossimilhança restrita são, respectivamente,

$$\begin{bmatrix} \boldsymbol{\theta}^{(w+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}^{(w)} \end{bmatrix} - \begin{bmatrix} \mathbf{H}_R \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{s}_R \end{bmatrix} \quad (3.42)$$

e

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

onde, como no caso anterior, as matrizes $\mathbf{H}_R$ e $\mathbf{I}_R$ são calculadas no valor corrente $\boldsymbol{\theta}^{(w)}$.

Como essas expressões não dependem de α , somente o vetor $\boldsymbol{\theta}$ é obtido de forma iterativa. Após a convergência do algoritmo, o vetor α pode ser estimado por

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

com Σ_i é substituído pela última estimativa obtida no processo iterativo.

Esses algoritmos fornecem também a estimativa da variância de $\hat{\alpha}$ e $\hat{\theta}$. Para dados não balanceados, a estimativa da variância assintótica é baseada na informação empírica. As estimativas com base na informação de Fisher só são válidas se os dados incompletos são MCAR (Missing Completely At Random) e nesse caso as baseadas na informação empírica são preferíveis (Rubin (1987) e Laird (1988)).

3.2.2 O Algoritmo EM

Outro processo iterativo bastante popular é o algoritmo EM, que consiste da realização alternada de dois passos até a convergência: o passo E (esperança) e o passo M (maximização).

O algoritmo é utilizado com dados incompletos. Os dados completos são particionados em dados presentes e dados ausentes. Quando a distribuição dos dados completos é da família exponencial regular, o passo E consiste em calcular a esperança condicional das estatísticas suficientes para os parâmetros com dados completos, condicionada aos dados presentes e às estimativas atuais dos parâmetros. Nesse passo, os dados são “completados”.

O passo M consiste na obtenção dos estimadores de máxima verossimilhança (restrita) com base nos dados “completados” no passo E.

Há uma variação do algoritmo EM, conhecida como algoritmo ER. Neste último, em lugar de se obter estimadores de máxima verossimilhança obtém-se estimadores robustos. Essa modificação não será considerada neste trabalho, mas o leitor interessado pode encontrar detalhes em Silva (1989).

Segundo Laird, Lange e Stram (1987) há dois enfoques para o uso do algoritmo EM em dados longitudinais. O primeiro considera os dados observados como sendo os dados presentes e os dados completos como os dados presentes mais os ausentes, ou seja, todos os que deveriam ter sido observados. O segundo enfoque, ligado ao modelo misto, considera

os dados observados como acima, mas os dados completos incluem também os efeitos aleatórios e os erros. Segundo os autores esse é o enfoque mais geral e é o que aqui será considerado.

Para o modelo (2.9), com a suposição adicional de normalidade para efeitos e erros aleatórios, a distribuição conjunta dos dados completos é dada por

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \\ \mathbf{e}_i \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{X}_i \boldsymbol{\alpha} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2 \mathbf{I}_{p_i} + \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' & \mathbf{Z}_i \mathbf{G} & \sigma^2 \mathbf{I}_{p_i} \\ \mathbf{G} \mathbf{Z}_i' & \mathbf{G} & \mathbf{0} \\ \sigma^2 \mathbf{I}_{p_i} & \mathbf{0} & \sigma^2 \mathbf{I}_{p_i} \end{pmatrix} \right], \quad (3.44)$$

para $i = 1, 2, \dots, n$.

O logaritmo da função de verossimilhança correspondente é

$$l(\mathbf{y}, \mathbf{b}, \mathbf{e}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = K - \frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i - \frac{n}{2} \log|\mathbf{G}| - \frac{1}{2} \text{tr}[\mathbf{G}^{-1} \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i'], \quad (3.45)$$

onde K é uma constante que não depende dos parâmetros e $N = \sum_{i=1}^n p_i$. As estatísticas suficientes para σ^2 e $\mathbf{G}$ são $t_1 = \sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i$ e $\mathbf{T}_2 = \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i'$, respectivamente, e os estimadores de máxima verossimilhança a serem considerados no passo M são

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \mathbf{e}_i' \mathbf{e}_i}{N}, \quad (3.46)$$

$$\hat{\mathbf{G}} = \frac{\sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i'}{n} \quad (3.47)$$

e

$$\hat{\boldsymbol{\alpha}} = \left(\sum_{i=1}^n \mathbf{X}_i' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{y}_i \right). \quad (3.48)$$

No passo E, usando o Resultado A4 do Apêndice A, temos que as esperanças condicionais das estatísticas suficientes são dadas por

$$E(t_1 | \mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \sum_{i=1}^n E(\mathbf{e}_i' | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) E(\mathbf{e}_i | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) + \sum_{i=1}^n \text{tr}[\text{Var}(\mathbf{e}_i | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta})], \quad (3.49)$$

$$E(\mathbf{T}_2 | \mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \sum_{i=1}^n E(\mathbf{b}_i | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) E(\mathbf{b}_i' | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) + \sum_{i=1}^n \text{Var}(\mathbf{b}_i | \mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}). \quad (3.50)$$

Para obter as parcelas das expressões (3.46) e (3.47) basta aplicar o Resultado A5 do Apêndice A e que resulta em:

1. $E(\mathbf{b}_i|\mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \mathbf{G}\mathbf{Z}_i'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha}),$
2. $\text{Var}(\mathbf{b}_i|\mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \mathbf{G} - \mathbf{G}\mathbf{Z}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i\mathbf{G},$
3. $E(\mathbf{e}_i|\mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \sigma^2\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha}),$
4. $\text{Var}(\mathbf{e}_i|\mathbf{y}_i; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \sigma^2\mathbf{I}_{p_i} - (\sigma^2)^2\boldsymbol{\Sigma}_i^{-1}.$

Combinando os dois passos do algoritmo, as expressões recursivas dos estimadores dos parâmetros são dadas abaixo

$$\sigma^{2(w+1)} = \frac{1}{N} \sum_{i=1}^n \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha}^{(w)} - \mathbf{Z}_i\mathbf{b}_i^{(w)} \right)' \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha}^{(w)} - \mathbf{Z}_i\mathbf{b}_i^{(w)} \right) + \frac{\sigma^{2(w)}}{N} \sum_{i=1}^n \text{tr}[\mathbf{I}_{p_i} - \sigma^{2(w)}\boldsymbol{\Sigma}_i^{-1(w)}], \quad (3.51)$$

$$\mathbf{G}^{(w+1)} = \mathbf{G}^{(w)} + \frac{1}{n} \sum_{i=1}^n \mathbf{b}_i^{(w)} \mathbf{b}_i^{(w)'} - \frac{\mathbf{G}^{(w)}}{n} \left(\sum_{i=1}^n \mathbf{Z}_i'\boldsymbol{\Sigma}_i^{-1(w)}\mathbf{Z}_i \right) \mathbf{G}^{(w)} \quad (3.52)$$

e

$$\boldsymbol{\alpha}^{(w+1)} = \left(\sum_{i=1}^n \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1(w)}\mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1(w)}\mathbf{y}_i \right), \quad (3.53)$$

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

Para a obtenção do estimador de máxima verossimilhança restrita é considerada uma versão Bayesiana do modelo misto, onde o vetor $\boldsymbol{\alpha}$ tem distribuição normal com vetor de média $\mathbf{0}$ e matriz de covariância $\boldsymbol{\Gamma}$ e é independente dos vetores de efeitos aleatórios $\mathbf{b}_i$'s e dos erros $\mathbf{e}_i$'s.

Harville (1974) mostrou que a função de verossimilhança limite quando $\boldsymbol{\Gamma}^{-1} \rightarrow \mathbf{0}$ é equivalente à função de verossimilhança restrita. Assim, a obtenção do estimador de máxima verossimilhança restrita é análoga à do estimador de máxima verossimilhança, tomando a distribuição dada por (3.54) abaixo e fazendo $\boldsymbol{\Gamma}^{-1}$ tender a zero.

$$\begin{pmatrix} \mathbf{y} \\ \boldsymbol{\alpha} \\ \mathbf{b} \\ \mathbf{e} \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2\mathbf{I}_N + \mathbf{Z}\mathbf{H}\mathbf{Z}' + \mathbf{X}\boldsymbol{\Gamma}\mathbf{X}' & \mathbf{X}\boldsymbol{\Gamma} & \mathbf{Z}\mathbf{H} & \sigma^2\mathbf{I}_N \\ \boldsymbol{\Gamma}\mathbf{X}' & \boldsymbol{\Gamma} & \mathbf{0} & \mathbf{0} \\ \mathbf{H}\mathbf{Z}' & \mathbf{0} & \mathbf{H} & \mathbf{0} \\ \sigma^2\mathbf{I}_N & \mathbf{0} & \mathbf{0} & \sigma^2\mathbf{I}_N \end{pmatrix} \right], \quad (3.54)$$

onde $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_n)'$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$, $\mathbf{H} = \mathbf{I}_n \otimes \mathbf{G}$ e $\Gamma^{-1} \rightarrow \mathbf{0}$.

O logaritmo da função de verossimilhança restrita limite é dada por

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

onde K é uma constante que não depende dos parâmetros.

As estatísticas suficientes são as mesmas do caso de máxima verossimilhança e as suas esperanças condicionais do passo E são:

$$E(t_1|\mathbf{y}; \boldsymbol{\theta}) = \sum_{i=1}^n E(\mathbf{e}'_i|\mathbf{y}; \boldsymbol{\theta}) E(\mathbf{e}_i|\mathbf{y}; \boldsymbol{\theta}) + \sum_{i=1}^n \text{tr}[\text{Var}(\mathbf{e}_i|\mathbf{y}; \boldsymbol{\theta})], \quad (3.56)$$

e

$$E(\mathbf{T}_2|\mathbf{y}; \boldsymbol{\theta}) = \sum_{i=1}^n E(\mathbf{b}_i|\mathbf{y}; \boldsymbol{\theta}) E(\mathbf{b}'_i|\mathbf{y}; \boldsymbol{\theta}) + \sum_{i=1}^n \text{Var}(\mathbf{b}_i|\mathbf{y}; \boldsymbol{\theta}), \quad (3.57)$$

com

1. $E(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\theta}) = (\sum_{i=1}^n \mathbf{X}'_i \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^n \mathbf{X}'_i \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i) = \hat{\boldsymbol{\alpha}},$
2. $\text{Var}(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\theta}) = (\sum_{i=1}^n \mathbf{X}'_i \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i)^{-1},$
3. $E(\mathbf{b}_i|\mathbf{y}; \boldsymbol{\theta}) = \mathbf{G} \mathbf{Z}'_i \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\alpha}}),$
4. $\text{Var}(\mathbf{b}_i|\mathbf{y}; \boldsymbol{\theta}) = \mathbf{G} - \mathbf{G} \mathbf{Z}'_i \mathbf{P}_i \mathbf{Z}_i \mathbf{G},$
5. $E(\mathbf{e}_i|\mathbf{y}; \boldsymbol{\theta}) = \sigma^2 \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\alpha}}),$
6. $\text{Var}(\mathbf{e}_i|\mathbf{y}; \boldsymbol{\theta}) = \sigma^2 \mathbf{I}_{p_i} - (\sigma^2)^2 \mathbf{P}_i,$

para $i = 1, 2, \dots, n$ e $\mathbf{P}_i = \boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i (\sum_{i=1}^n \mathbf{X}'_i \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}'_i \boldsymbol{\Sigma}_i^{-1}$ e onde já foi incorporado que Γ^{-1} tende a zero.

Como no caso anterior, combinando-se os dois passos do algoritmo, tem-se que as expressões recursivas dos estimadores dos parâmetros são

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

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

onde

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

$$\boldsymbol{\Sigma}_i^{-1(w)} = [\sigma^{2(w)} \mathbf{I}_{p_i} + \mathbf{Z}_i \mathbf{G}^{(w)} \mathbf{Z}_i']^{-1},$$

$$\mathbf{b}_i^{(w)} = \mathbf{G}^{(w)} \mathbf{Z}_i' \boldsymbol{\Sigma}_i^{-1(w)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(w)})$$

e

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

Todos os detalhes das derivações dos resultados apresentados nesta seção podem ser encontrados em Andreoni (1989).

3.2.3 Valores iniciais e regras de parada

Quando utilizamos processos iterativos precisamos fornecer valores iniciais para as estimativas dos parâmetros a fim de tornar possível o início dos cálculos no processo. A escolha dos valores iniciais é um aspecto importante da utilização desses processos. Esses valores podem ser escolhidos arbitrariamente ou podem ser baseados nos dados. Laird, Lange e Stram (1987), Andreoni (1989), Liski e Nummi (1990) e Nummi (1995) apresentam algumas sugestões para o cálculo desses valores.

Para o modelo de efeitos aleatórios Laird, Lange e Stram (1987) sugerem

$$\boldsymbol{\alpha}^{(0)} = \left(\sum_{i=1}^n \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i' \mathbf{y}_i \right), \quad (3.60)$$

$$\sigma^{2(0)} = \frac{1}{[N - (n-1)r - q]} \left[\sum_{i=1}^n \mathbf{y}_i' \mathbf{y}_i - \boldsymbol{\alpha}^{(0)'} \sum_{i=1}^n \mathbf{X}_i' \mathbf{y}_i - \sum_{i=1}^n \hat{\mathbf{b}}_i' \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(0)}) \right] \quad (3.61)$$

e

$$\mathbf{G}^{(0)} = \frac{1}{(n-1)} \sum_{i=1}^n \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i' - \frac{\sigma^{2(0)}}{n} \sum_{i=1}^n (\mathbf{Z}_i' \mathbf{Z}_i)^{-1}, \quad (3.62)$$

onde $N = \sum_{i=1}^n p_i$ e $\hat{\mathbf{b}}_i = (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}^{(0)})$.

É exigido que $\sigma^{2(0)}$ seja não negativa e que a matriz $\mathbf{G}^{(0)}$ seja positiva semi-definida. Além disso, é necessário que a matriz tenha posto completo e por isso os valores iniciais são baseados nos dados completos.

É também necessário adotar uma regra de parada para o processo. As mais comuns são as seguintes:

1. a diferença absoluta entre as estimativas dos parâmetros em dois passos consecutivos é menor do que um valor fixado a priori;
2. a diferença absoluta entre os valores da função de verossimilhança (restrita) em dois passos consecutivos é menor do que um valor fixado a priori;
3. a diferença relativa entre as estimativas dos parâmetros em dois passos consecutivos é menor do que um valor fixado a priori (a relação é tomada com base nos valores no penúltimo passo);
4. o número de iterações é fixado e
5. combinação das regras anteriores.

O BMDP5V utiliza como padrão a combinação de 2) e 4). Nas aplicações deste trabalho foi utilizada uma combinação de 1) e 4).

Capítulo 4

Imputação

4.1 Introdução

Neste capítulo tratamos da imputação de dados desbalanceados ou incompletos aleatoriamente, como é o caso do Exemplo 2.5, e sempre que nos referirmos a dados incompletos, estamos considerando este caso. A idéia é que o planejamento foi feito para que as observações fossem tomadas de n indivíduos em p instantes de tempo, sendo estes os mesmos para todos os indivíduos, e que algumas informações foram perdidas.

A ocorrência de dados desse tipo é comum ou porque são detectadas discrepâncias e falhas (erros de registro, transcrição, digitação ou respostas forjadas) e as observações são descartadas, ou pela falta de informação propriamente dita.

Basicamente há três formas de se tratar dados incompletos: analisar os dados desbalanceados, descartar os dados referentes a indivíduos que possuem alguma informação ausente ou utilizar algum método para fazer a imputação de dados.

Método de imputação é todo procedimento de previsão de valores individuais, isto é, qualquer método que atribui um valor específico a cada observação perdida, utilizando uma relação na qual estão presentes as informações dos dados amostrados.

O nosso interesse é realizar a imputação de dados através de modelos de curvas de

crescimento polinomiais, traduzidos pelo modelo misto (2.9). A imputação será feita pelo BLUP (Best Linear Unbiased Predictor), desenvolvido em Barroso (1995) e Barroso, Bussab e Knott (1998).

Alguns autores classificam a imputação ou previsão como previsão condicional ou estendida. A primeira considera as observações (todas ou parte delas) para prever os valores ausentes, ou seja, a previsão da parte ausente é feita através da parte observada. Este tipo pode ser encontrado, por exemplo em Lee e Geisser (1972, 1975), Rao (1987), Lee (1988) e Kanda (1992). A segunda emprega todas as observações para prever valores futuros, ou seja, a previsão é produzida além dos limites de tempo observado pela amostra (Rao (1987), Lee (1988) e Kanda (1992)).

4.2 Histórico

Lee e Geisser (1975) apresentaram alguns métodos de imputação sob o enfoque Bayesiano para o modelo de curva de crescimento de Potthoff e Roy. Dos métodos considerados alguns impunham estruturas de covariância específicas e outros não.

Rao (1987) realizou a previsão para o modelo de curva de crescimento de Potthoff e Roy através de vários métodos, alguns utilizando somente as observações anteriores ao tempo em que a imputação estava sendo feita e outros usando todos os dados presentes.

Lee (1988) realizou a previsão e estimação para o modelo de curva de crescimento de Potthoff e Roy sob as estruturas de covariância uniforme e correlação serial.

Liski e Nummi (1990) apresentaram dois métodos de imputação para o modelo de curva de crescimento de Potthoff e Roy: o algoritmo EM e o EM passo único (primeiro passo do algoritmo). Estes permitem que mais de uma observação seja imputada simultaneamente, além de utilizar todas as observações presentes. Os previsores também podem ser estendidos para casos nos quais o interesse não é apenas a previsão da última observação, mas o conjunto de várias unidades.

Kanda (1992) apresentou um método de predição para o modelo de curva de crescimento de Potthoff e Roy com matriz de covariância estruturada. Considerou as estruturas uniforme e correlação serial para a predição condicional e estendida.

Tian, Shukla e Buncher (1994) consideraram quatro dos métodos propostos por Rao (1987). Além disso propuseram o uso de funções “spline” cúbicas para fazer imputação.

Barroso (1995) e Barroso, Bussab e Knott (1998) introduziram o método de imputação via BLUP para o modelo misto usual para populações finitas. Consideraram os casos univariado e multivariado.

Salinas (1998) e Barroso e Salinas (1999) consideraram vários métodos de imputação para o modelo misto usual, com o objetivo de aplicar a técnica do bootstrap a dados incompletos.

4.3 O BLUP para o modelo misto

O método adotado neste trabalho é o de imputação feita pelo BLUP para o modelo misto, quando há a ocorrência de dados ausentes. O BLUP é linear pois é uma função linear dos dados observados, é não-viesado, isto é, a esperança do preditor é igual à esperança da quantidade prevista e é ótimo porque tem o menor erro quadrático médio na classe dos preditores lineares não-viesados. Aqui cabe uma ressalva quanto à distinção que alguns autores fazem em relação aos termos preditor e estimador; o primeiro é usado para efeitos aleatórios e o segundo para efeitos fixos.

Nesta seção não fazemos distinção entre esses termos e apresentamos o BLUP para os efeitos do modelo e para as observações ausentes, juntamente com as expressões dos correspondentes erros quadráticos médios.

O modelo misto é dado por (2.9), ou seja,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{b} + \mathbf{e}, \quad (4.1)$$

com as mesmas suposições e restrições apresentadas no Capítulo 2, com

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \tilde{\mathbf{R}} + \mathbf{Z}\tilde{\mathbf{G}}\mathbf{Z}'. \quad (4.2)$$

No caso de dados incompletos, suponha que m das np observações são ausentes e particione o vetor $\mathbf{y}$ nas partes observada $\mathbf{y}_o$ e ausente $\mathbf{y}_a$. O modelo pode ser escrito como

$$\begin{bmatrix} \mathbf{y}_o \\ \mathbf{y}_a \end{bmatrix} = \begin{bmatrix} \mathbf{X}_o \\ \mathbf{X}_a \end{bmatrix} \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{Z}_o \\ \mathbf{Z}_a \end{bmatrix} \mathbf{b} + \begin{bmatrix} \mathbf{e}_o \\ \mathbf{e}_a \end{bmatrix}.$$

O BLUP de $\mathbf{y}_a$ é dado por

$$\hat{\mathbf{y}}_a = (\mathbf{I}_m - \mathbf{E}'\mathbf{Q}\mathbf{E}\mathbf{E}'\tilde{\mathbf{R}}^{-1}\mathbf{E})^{-1}\mathbf{E}'\mathbf{Q}\mathbf{F}\mathbf{F}'\tilde{\mathbf{R}}^{-1}\mathbf{F}\mathbf{F}'\mathbf{y}, \quad (4.3)$$

onde

$$\mathbf{Q} = \tilde{\mathbf{R}}[\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \tilde{\mathbf{R}}^{-1} - \mathbf{v}^{-1}]\tilde{\mathbf{R}};$$

$\mathbf{E}$: matriz de dimensão $(np \times m)$ indicadora dos dados ausentes;

$\mathbf{F}$: matriz de dimensão $(np \times (np - m))$ indicadora dos dados presentes.

Cada coluna de $\mathbf{F}$ indica um dado presente e é formada por 0's e um 1, localizado na mesma linha de $\mathbf{y}$ onde está o dado que está sendo indicado. A matriz $\mathbf{E}$ é análoga, somente indica dados ausentes.

A construção das matrizes indicadoras, $\mathbf{F}$ e $\mathbf{E}$, é ilustrada no exemplo a seguir.

Exemplo 4.1: Considere o Exemplo 2.7, onde $np = 6$, e suponha que duas observações são ausentes, a primeira e a quinta. Então,

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \\ y_{32} \end{bmatrix}, \quad \mathbf{y}_o = \begin{bmatrix} y_{21} \\ y_{31} \\ y_{12} \\ y_{32} \end{bmatrix}, \quad \mathbf{y}_a = \begin{bmatrix} y_{11} \\ y_{22} \end{bmatrix},$$

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{E} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

O erro quadrático médio de $\hat{\mathbf{y}}_a$ é dado por

$$\begin{aligned} \text{EQM}(\hat{\mathbf{y}}_a) = & [(\mathbf{I}_m - \mathbf{E}'\mathbf{Q}\mathbf{E}\tilde{\mathbf{R}}^{-1}\mathbf{E})^{-1}\mathbf{E}'\mathbf{Q}\mathbf{F}\mathbf{F}'\tilde{\mathbf{R}}^{-1}\mathbf{F}\mathbf{F}' - \mathbf{E}']\mathbf{V}[\mathbf{F}\mathbf{F}'\tilde{\mathbf{R}}^{-1}\mathbf{F}\mathbf{F}'\mathbf{Q}\mathbf{E} \\ & (\mathbf{I}_m - \mathbf{E}'\tilde{\mathbf{R}}^{-1}\mathbf{E}\mathbf{E}'\mathbf{Q}\mathbf{E})^{-1} - \mathbf{E}]. \end{aligned} \quad (4.4)$$

Se for de interesse obter o BLUP dos efeitos fixos e aleatórios, basta completar os dados, isto é, construir o vetor

$$\hat{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_o \\ \hat{\mathbf{y}}_a \end{bmatrix}$$

e o BLUP tem a mesma expressão que o correspondente para dados completos, só que aplicada ao vetor $\hat{\mathbf{y}}$ no lugar de $\mathbf{y}$. Especificamente o BLUP dos efeitos é dado por

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\hat{\mathbf{y}} \quad (4.5)$$

e

$$\hat{\mathbf{b}} = (\mathbf{Z}'\tilde{\mathbf{R}}^{-1}\mathbf{Z} + \tilde{\mathbf{G}}^{-1})^{-1}(\mathbf{Z}'\tilde{\mathbf{R}}^{-1} - \mathbf{Z}'\tilde{\mathbf{R}}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\hat{\mathbf{y}}. \quad (4.6)$$

Seus correspondentes erros quadráticos médios são

$$\text{EQM}(\hat{\boldsymbol{\alpha}}) = \mathbf{B}\mathbf{V}\mathbf{B}' \quad (4.7)$$

e

$$\text{EQM}(\hat{\mathbf{b}}) = \mathbf{C}\mathbf{V}\mathbf{C}' - \mathbf{C}\mathbf{Z}\tilde{\mathbf{G}} - \tilde{\mathbf{G}}\mathbf{Z}'\mathbf{C}' + \tilde{\mathbf{G}}, \quad (4.8)$$

onde

$$\mathbf{B} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\tilde{\mathbf{R}}[\mathbf{I}_{np} + \mathbf{E}\mathbf{E}'\tilde{\mathbf{R}}^{-1}\mathbf{E}(\mathbf{I}_m - \mathbf{E}'\mathbf{Q}\mathbf{E}\mathbf{E}'\tilde{\mathbf{R}}^{-1}\mathbf{E})^{-1}\mathbf{E}'\mathbf{Q}]\mathbf{F}\mathbf{F}'\tilde{\mathbf{R}}^{-1}\mathbf{F}\mathbf{F}',$$

$$\begin{aligned} \mathbf{C} &= \mathbf{A}\mathbf{Z}'[\mathbf{I}_{np} - \tilde{\mathbf{R}}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\tilde{\mathbf{R}}] \\ &\quad [\mathbf{I}_{np} + \mathbf{E}\mathbf{E}'\tilde{\mathbf{R}}^{-1}\mathbf{E}(\mathbf{I}_m - \mathbf{E}'\mathbf{Q}\mathbf{E}\mathbf{E}'\tilde{\mathbf{R}}^{-1}\mathbf{E})^{-1}\mathbf{E}'\mathbf{Q}]\mathbf{F}\mathbf{F}'\tilde{\mathbf{R}}^{-1}\mathbf{F}\mathbf{F}' \end{aligned}$$

e

$$\mathbf{A} = (\mathbf{Z}'\tilde{\mathbf{R}}^{-1}\mathbf{Z} + \tilde{\mathbf{G}}^{-1})^{-1} = \tilde{\mathbf{G}}[\mathbf{I}_c - \mathbf{Z}'(\mathbf{Z}\tilde{\mathbf{G}}\mathbf{Z}' + \tilde{\mathbf{R}})^{-1}\mathbf{Z}\tilde{\mathbf{G}}].$$

Se considerarmos que os erros do modelo são independentes com variância constante, ou seja, $\tilde{\mathbf{R}} = \sigma^2\mathbf{I}_{np}$, o BLUP de $\hat{\mathbf{y}}_a$ e seu correspondente erro quadrático médio passam a ser dados por

$$\hat{\mathbf{y}}_a = (\mathbf{I}_m - \mathbf{E}'\mathbf{Q}\mathbf{E})^{-1}\mathbf{E}'\mathbf{Q}\mathbf{F}\mathbf{F}'\mathbf{y}, \quad (4.9)$$

onde $\mathbf{E}$ e $\mathbf{F}$ são as mesmas dadas anteriormente e

$$\mathbf{Q} = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{I}_{np} - \mathbf{V}^{-1}.$$

e

$$\text{EQM}(\hat{\mathbf{y}}_a) = [(\mathbf{I}_m - \mathbf{E}'\mathbf{Q}\mathbf{E})^{-1}\mathbf{E}'\mathbf{Q}\mathbf{F}\mathbf{F}' - \mathbf{E}']\mathbf{V}[\mathbf{F}\mathbf{F}'\mathbf{Q}\mathbf{E}(\mathbf{I}_m - \mathbf{E}'\mathbf{Q}\mathbf{E})^{-1} - \mathbf{E}]. \quad (4.10)$$

Os previsores e erros quadráticos médios apresentados nesta seção dependem dos parâmetros das matrizes de covariância, que nem sempre são conhecidos. Na prática,

o método tradicional sugere que eles sejam substituídos por estimativas obtidas de estimadores consistentes, o que torna o erro quadrático médio maior (Kackar e Harville, 1984). A sugestão para obter tais estimativas é utilizar algum dos métodos propostos no Capítulo 3.

4.3.1 Imputação no Modelo III

O Modelo III apresentado na Seção 2.4 é um caso particular do modelo misto de efeitos fixos e aleatórios. Esse modelo impõe que a estrutura de covariância seja bloco diagonal, com os blocos seguindo o padrão da matriz simétrica composta.

Neste trabalho, por simplicidade, quando usamos este modelo, fazemos a estimação das componentes da variância através do método ANOVA truncado, apresentado na Seção 3.1.3, embora outros métodos pudessem ser utilizados. Na verdade, como já foi dito, esses estimadores coincidem com o estimador de máxima verossimilhança restrita sob normalidade de efeitos e erros aleatórios.

Para este modelo, a expressão do previsor dada por (4.3) ganha formas interpretáveis que são apresentadas a seguir.

Uma observação ausente no último instante

Nas aplicações que serão apresentadas no Capítulo 5, consideramos o caso em que somente uma observação é ausente e isto ocorre no último instante de coleta dos dados. Nesse caso, sem perda de generalidade, considere que a observação ausente é a do n -ésimo indivíduo e que $y_{t_p n}$ denota essa observação. O BLUP é dado por

$$\hat{y}_{t_p n} = \bar{y}_{t_p \cdot} + \left(\frac{K(p-1)}{K(p-1) + 1} \right) (\bar{y}_{\cdot n} - \bar{y}_c), \quad (4.11)$$

onde

$$K = \frac{\sigma_b^2}{\sigma^2};$$

$$\bar{y}_c = \frac{1}{(n-1)(p-1)} \sum_{i=1}^{n-1} \sum_{k=1}^{p-1} y_{t_k i};$$

$$\bar{y}_{t_p.} = \frac{1}{(n-1)} \sum_{i=1}^{n-1} y_{t_p i};$$

$$\bar{y}_{.n} = \frac{1}{(p-1)} \sum_{k=1}^{p-1} y_{t_k n},$$

ou seja, K é a razão entre as componentes da variância, $\bar{y}_c$ é a média da parte completa dos dados, isto é, a média das observações dos $n-1$ indivíduos completos nos $p-1$ primeiros instantes, $\bar{y}_{t_p.}$ é a média das observações presentes no último instante e $\bar{y}_{.n}$ é a média das observações presentes do n -ésimo indivíduo.

O previsor dado por (4.11) é a média das observações no último instante mais uma correção que depende da média do indivíduo com o dado ausente e a média dos demais indivíduos. Se não há efeito de indivíduo, essa correção desaparece.

O erro quadrático médio, para esse caso, ganha a forma

$$\text{EQM}(\hat{y}_{t_p n}) = \frac{n}{(n-1)} \left(\frac{Kp+1}{K(p-1)+1} \right) \sigma^2. \quad (4.12)$$

Várias observações ausentes em vários instantes

O caso mais geral que pode ocorrer é aquele em que m observações são perdidas em qualquer instante de tempo, de qualquer indivíduo.

Nesse caso o BLUP das observações ausentes é dado por

$$\hat{\mathbf{y}}_a = (-\mathbf{E}'\mathbf{P}\mathbf{E})^{-1}\mathbf{u}, \quad (4.13)$$

onde $\mathbf{u}$ é um vetor de dimensão $(m \times 1)$, em que cada observação ausente corresponde a uma de suas linhas. Por exemplo, se a observação do j -ésimo indivíduo no l -ésimo

instante é perdida, o elemento correspondente no vetor $\mathbf{u}$ é dado por

$$u_{lj} = (1 - a) \sum_{i=1}^n y_{t_{li}} + a \left((n-1) \sum_{k=1}^p y_{t_{kj}} - \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{\substack{k=1 \\ k \neq l}}^p y_{t_{ki}} \right), \quad (4.14)$$

onde $a = \frac{K}{pK+1}$.

A expressão (4.14) tem interpretação análoga à da expressão (4.11).

A matriz $-\mathbf{E}'\mathbf{P}\mathbf{E}$ tem dimensão $(m \times m)$ e é composta por blocos que correspondem a cada indivíduo com dado ausente. Cada bloco tem dimensão igual ao número de dados ausentes do indivíduo a ele associado. Os blocos têm as seguintes formas:

1. os blocos da diagonal são matrizes do tipo

$$(n-1) \begin{bmatrix} (1-a) & -a & \dots & -a \\ -a & (1-a) & \dots & -a \\ \vdots & \vdots & \ddots & \vdots \\ -a & -a & \dots & (1-a) \end{bmatrix},$$

ou seja, $(n-1)[\mathbf{I} - a\mathbf{J}]$, onde $\mathbf{I}$ é a matriz identidade e $\mathbf{J}$ é a matriz quadrada de 1's, ambas de dimensão igual ao número de dados ausentes do indivíduo a que corresponde.

2. os blocos fora da diagonal são matrizes que correspondem ao cruzamento de dois indivíduos com dado ausente. Suas dimensões são dadas pelos números de dados ausentes desses indivíduos. Os elementos da matriz são

- $-(1-a)$ se o cruzamento da linha e coluna da matriz $-\mathbf{E}'\mathbf{P}\mathbf{E}$ corresponde a tempos iguais;
- a se o cruzamento da linha e coluna da matriz $-\mathbf{E}'\mathbf{P}\mathbf{E}$ corresponde a tempos diferentes.

Para ilustrar, considere o exemplo a seguir.

Exemplo 4.2. Considere que foi planejado que quatro indivíduos fossem observados em três instantes, mas que algumas observações foram perdidas aleatoriamente, conforme esquema abaixo.

indivíduo	tempo		
	t_1	t_2	t_3
1	y_{11}	y_{21}	y_{31}
2	y_{12}	?	?
3	y_{13}	y_{23}	?
4	?	y_{24}	?

onde ? indica a falta da observação. Neste caso, $n = 4$, $p = 3$ e $m = 5$.

A matriz $-\mathbf{E}'\mathbf{P}\mathbf{E}$ é dada por

$$\begin{bmatrix} (n-1)(1-a) & -(n-1)a & a & a & a \\ -(n-1)a & (n-1)(1-a) & -(1-a) & a & -(1-a) \\ a & -(1-a) & (n-1)(1-a) & a & -(1-a) \\ a & a & a & (n-1)(1-a) & -(n-1)a \\ a & -(1-a) & -(1-a) & -(n-1)a & (n-1)(1-a) \end{bmatrix}$$

e o vetor $\mathbf{u}$ por

$$\mathbf{u} = \begin{bmatrix} u_{22} \\ u_{32} \\ u_{33} \\ u_{14} \\ u_{34} \end{bmatrix} = \begin{bmatrix} (1-a)(y_{21} + y_{23} + y_{24}) + a[(n-1)y_{12} - (y_{11} + y_{31} + y_{13})] \\ (1-a)y_{31} + a[(n-1)y_{12} - (y_{11} + y_{21} + y_{13} + y_{23} + y_{24})] \\ (1-a)y_{31} + a[(n-1)(y_{13} + y_{23}) - (y_{11} + y_{21} + y_{12} + y_{24})] \\ (1-a)(y_{11} + y_{12} + y_{13}) + a[(n-1)y_{24} - (y_{21} + y_{31} + y_{23})] \\ (1-a)y_{31} + a[(n-1)y_{24} - (y_{11} + y_{21} + y_{12} + y_{13} + y_{23})] \end{bmatrix}.$$

Para fazer a imputação, o primeiro passo é estimar as componentes da variância σ^2 e σ_b^2 , e conseqüentemente o valor de a , e depois aplicar a fórmula (4.13).

Como o Modelo III é um caso particular do modelo mais geral (2.9), também é possível fazer a imputação pela fórmula (4.9). A vantagem é que usando (4.13) temos que inverter uma matriz de dimensão $(m \times m)$ e usando (4.9) uma de dimensão $(np \times np)$. Lembre que $m < np$, pois m é o número de dados ausentes e np o número de dados completos. No Exemplo 4.2 passaríamos de uma matriz de dimensão (5×5) para uma de dimensão (12×12) .

Capítulo 5

Aplicações

Neste capítulo apresentamos os resultados de aplicações de imputação via BLUP. Essas aplicações são de dois tipos: primeiro consideramos três conjuntos de dados já bastante analisados na literatura (Exemplos 2.1, 2.2 e 2.3) e comparamos nossos resultados com os de outros autores; segundo, consideramos, a título de ilustração do método de imputação, um conjunto de dados reais (Exemplo 2.5), apresentando os valores imputados sob vários modelos. As aplicações foram feitas no módulo CM do aplicativo NTIA/EMBRAPA. Os programas encontram-se no Apêndice B.

Aplicação 1

No primeiro caso os conjuntos são completos. A idéia é simular que algumas das observações são ausentes, fazer a imputação e comparar os valores imputados com os valores reais, que na verdade são conhecidos. O exercício consiste em omitir a última observação do primeiro indivíduo, fazer a imputação com base em algum método, retornar essa observação aos dados, omitir a última observação do segundo indivíduo e repetir o procedimento até o n -ésimo indivíduo. O exercício é repetido para diferentes métodos de imputação e a escolha do melhor é feita com base em alguma medida de variabilidade.

Esse procedimento é conhecido como LOO (leave-one-out) ou CV (cross validation). As medidas de variabilidade mais comumente utilizadas são Erro Quadrático Médio (EQM), ou equivalentemente a raiz quadrada deste (REQM), e o Erro Absoluto Médio (EAM). O EQM foi usado por Lee e Geisser (1975), Rao (1987), Lee (1988), Barroso (1995) e Tian, Shukla e Buncher (1994); o REQM por Liski e Nummi (1990) e Kanda (1992) e o EAM por Lee e Geisser (1975) e Lee (1988).

Sistematizando, neste trabalho, a escolha entre os vários métodos de imputação é feita da seguinte maneira:

1. retirar a p -ésima observação do primeiro indivíduo;
2. prever y_{t_p1} através do método de imputação;
3. retornar a observação aos dados e repetir o procedimento para os outros indivíduos;
4. calcular

$$\widehat{\text{EQM}}(\hat{y}_a) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_{t_pi} - y_{t_pi})^2,$$

onde y_{t_pi} é o valor observado e $\hat{y}_{t_pi}$ é valor imputado e

5. escolher o método de imputação que apresenta o menor $\widehat{\text{EQM}}(\hat{y}_a)$.

Além disso, consideramos que a imputação no último instante pode ser feita com base nos dados presente nesse instante mais somente os dados no penúltimo instante, ou nos dois instantes anteriores ao último ou, assim por diante até considerar todos os $p - 1$ instantes anteriores.

Escolhemos prever a observação no último instante pois esse foi o exercício efetuado por outros autores e assim, poderíamos comparar os resultados encontrados. Os resultados de imputação via BLUP são aqui comparados com os resultados obtidos por Lee e Geisser (1975), Rao (1987), Lee (1988), Liski e Nummi (1990) e Kanda (1992).

Listamos abaixo os métodos usados por esses autores, destacando que as descrições detalhadas dos mesmos podem ser encontradas nas referências.

1. Lee e Geisser (1975) utilizaram os seguintes métodos de imputação:
 - (a) média aproximada, moda, quase mínimos quadrados e ponto p_1 para o modelo de covariância não-estruturada;
 - (b) média aproximada, moda, quase mínimos quadrados e ponto p_1 para o modelo com a estrutura simples de Rao. (Rao, 1967);
 - (c) previsor marginal, mínimos quadrados simples, ad hoc;
 - (d) média aproximada para o modelo de correlação serial;
 - (e) quase mínimos quadrados e ponto p_1 para os modelos de covariância com as estruturas de correlação serial e fator de Rao (Rao, 1967).
2. Rao (1987) considerou os seguintes métodos:
 - (a) previsão linear: regressão direta e regressão inversa;
 - (b) modelos de curvas de crescimento: regressão individual, regressão sobre os coeficientes polinomiais, calibração e previsão de Bayes empírica;
 - (c) componentes principais e fator analítico.
3. Lee (1988) considerou principalmente dois tipos de estrutura para a matriz de covariância geral: uniforme e correlação serial. Em alguns casos considerou também a matriz não estruturada. Para estimar a correlação usou máxima verossimilhança (MLE) e mínima variância preditiva (MPVE). Impôs a mesma estrutura de covariância para grupos diversos ou permitiu que elas fossem diferentes. Os previsores utilizados foram os seguintes:
 - (a) média aproximada;
 - (b) previsor ad hoc;
 - (c) previsor combinado;

- (d) previsor estendido.
4. Kanda (1992) considerou o modelo de curva de crescimento e utilizou o previsor média aproximada para as estruturas de covariância geral uniforme e serial. Baseou suas comparações em estimativas do Erro Quadrático Médio teórico expandido até ordem n^{-1} .
5. Liski e Nummi (1990) empregaram os seguintes métodos de imputação:
- (a) algoritmo EM;
- (b) EM - passo único (o primeiro passo do algoritmo EM).

As Tabelas 5.1, 5.2, 5.3 e 5.4 mostram as estimativas do Erro Quadrático Médio correspondentes aos Exemplos 2.1 e 2.2. As Tabelas 5.5. e 5.6 correspondem ao Exemplo 2.3 e mostram essas estimativas multiplicadas por n , pois os valores do EQM são muito baixos.

Nos artigos de Lee e Geisser (1975), Lee (1988) e Kanda (1992), as previsões são feitas somente com todos os instantes, o que é indicado nas tabelas por t_1 a t_p . No artigo de Liski e Nummi (1990) são considerados todos os instantes e também parte deles, exceto para os dados do Exemplo 2.3 (Tabelas 5.5 e 5.6), em que as técnicas não são aplicadas para t_1 a t_7 e t_2 a t_7 . Rao (1987) aplica seus métodos ao conjunto todo de dados ou a parte deles, quando possível. Essa mesma regra é seguida para a aplicação do BLUP.

Os comentários baseados nas tabelas, ou seja, em dados empíricos são os seguintes:

Exemplo 2.1. Dados de Potthoff e Roy

- os melhores resultados foram obtidos por Kanda, não fazendo diferença se as matrizes de covariância são as mesmas para os grupos de meninas e meninos; Lee, entretanto, obteve melhores resultados quando considerou matrizes diferentes;

- os resultados obtidos para os modelos de grau 2 foram em geral piores do que os dos modelos de grau 1;
- o melhor de todos os resultados (1,571) foi obtido por Kanda, considerando matrizes de covariância diferentes para os dois grupos e impondo a estrutura uniforme para ambas;
- em termos gerais o BLUP apresentou bons resultados;
- o melhor dentre os resultados da imputação via BLUP (2,251) foi obtido através do Modelo III quando os instantes 2 e 3 foram usados para prever o dado do instante 4. Isso é bastante positivo pois o Modelo III, além de ser simples, de fácil aplicação, é também interpretável;
- houve problemas quando o algoritmo de Newton-Raphson (máxima verossimilhança ou máxima verossimilhança restrita) foi utilizado: ou o valor do EQM foi muito alto ou houve a necessidade de inverter uma matriz singular. Nesse caso, a literatura sugere o uso da decomposição QR (Strang, 1988) que leva à utilização da pseudo inversa ou o uso da decomposição de Cholesky (Lindstrom e Bates, 1988); pode-se tentar também mudar o valor inicial dos parâmetros. As soluções adotadas não produziram resultados satisfatórios.

Exemplo 2.2. Dados de Elston e Grizzle

- em termos gerais, os melhores resultados foram obtidos por Lee, quando este considerou a estrutura de correlação serial;
- os piores resultados são aqueles obtidos pelo BLUP no modelo de grau 2, quando um dos métodos iterativos e somente parte dos dados são usados. Também para o modelo de grau 1, os resultados do BLUP via algum dos algoritmos, não foram satisfatórios;

- os resultados via BLUP para o modelo de grau 2 com soluções explícitas estão próximos dos resultados médios de outros autores;
- dentre os modelos de grau 2, os resultados via BLUP com soluções explícitas estão entre os melhores;
- o melhor resultado (0,498) foi obtido por Rao através do procedimento de Bayes empírico;
- o melhor resultado obtido via BLUP (0,521) refere-se ao Modelo III quando somente o penúltimo instante foi usado para fazer a previsão. Este resultado é bem próximo daquele obtido por Rao;

Exemplo 2.3. Dados de Williams e Izenman

- todos os autores apresentaram algum método com bom resultado;
- o pior resultado foi encontrado através da aplicação do BLUP com o algoritmo de Newton-Raphson e o método de máxima verossimilhança;
- ao contrário dos exemplos anteriores, aplicação do algoritmo de Newton-Raphson com o método de máxima verossimilhança restrita produziu alguns resultados satisfatórios;
- embora menores, também houve problemas com o emprego do algoritmo de Newton-Raphson;
- o melhor resultado (0,025) foi encontrado por Lee com o método da média aproximada com correlação serial, para o modelo de grau 2;
- o melhor resultado do BLUP (0,028) é bem próximo do resultado de Lee e de novo foi obtido no Modelo III com a utilização de um único instante.

Aplicação 2

O segundo tipo de aplicação foi feito sobre os dados do Exemplo 2.5. A imputação foi feita via BLUP, variando-se o modelo para as médias (Modelos I, II e III) e a forma de estimação dos parâmetros das matrizes de covariância. Os valores imputados sob cada uma dessas variações são dados na Tabela 5.7. A notação y_{24} indica a observação do indivíduo 4 no instante 2; as outras são análogas.

Pode-se verificar que é ampla a variedade de valores. O Modelo III resulta na imputação de valores altos para o animal 4 no instante 2. Isso ocorre porque um dos animais apresenta um valor bastante alto no instante 2, elevando a média nesse ponto (ver Tabela 2.5).

O valor imputado, sob o Modelo III não é afetado pelo fato das variâncias serem estimadas com os dados de indivíduos completos ou os dados de instantes completos; o mesmo não ocorre com os resultados de soluções explícitas.

Outro fato interessante é que não faz diferença se as estimativas dos parâmetros das matrizes de covariância são obtidas por máxima verossimilhança ou máxima verossimilhança restrita. Além disso, os valores obtidos pelos algoritmos são próximos.

Para escolher qual método de imputação é o “melhor” dentre os estudados, foi utilizado o critério “leave-one-out”. A Tabela 5.8 apresenta as estimativas dos erros quadráticos médios para cada método usado. Para o Modelo I, os valores obtidos sob máxima verossimilhança ou máxima verossimilhança restrita são bem próximos. O mesmo não ocorre com o Modelo II. No Modelo I, quando o método de estimação de soluções explícitas é utilizado, os resultados são piores do que os obtidos através dos métodos iterativos. De todos eles, o melhor método de imputação é o BLUP aplicado ao Modelo III.

Tabela 5.1: *EQM para o modelo de grau 1 (dados de Potthoff e Roy)*

Método		Instantes de tempo		
		t_1 a t_4	t_2 a t_4	t_3 a t_4
Lee e Geisser	Média aproximada/não-estruturada	3,734	–	–
	Moda/não-estruturada	3,462	–	–
	Quase mínimos quadrados/não-estruturada	3,057	–	–
	Ponto p_1 /não-estruturada	5,536	–	–
	Média aproximada/Rao simples	3,375	–	–
	Moda/Rao simples	2,877	–	–
	Quase mínimos quadrados/Rao simples	2,970	–	–
	Ponto p_1 /Rao simples	5,509	–	–
	Previsor marginal	5,417	–	–
	Mínimos quadrados simples	4,000	–	–
	Ad hoc	2,250	–	–
	Média aproximada/serial	2,680	–	–
Lee	Média aproximada/serial (iguais)	2,364	–	–
	Ad hoc/serial (iguais)	2,607	–	–
	Média aproximada/uniforme (iguais)	2,235	–	–
	Previsor combinado/uniforme (iguais)	2,090	–	–
	Média aproximada/serial (diferentes)	1,999	–	–
	Ad hoc/serial (diferentes)	2,099	–	–
	Previsor combinado/serial (diferentes)	1,985	–	–
	Média aproximada/uniforme (diferentes)	2,179	–	–
	Previsor combinado/uniforme (diferentes)	1,984	–	–
Liski e Nummi	EM	3,280	3,176	3,108
	EM - passo único	3,736	3,126	3,108

Tabela 5.1.: Continuação

Método		Instantes de tempo		
		t_1 a t_4	t_2 a t_4	t_3 a t_4
Rao	Regressão direta	4,430	3,588	3,665
	Regressão inversa	6,211	5,227	4,929
	Regressão individual	3,998	12,426	+
	Regressão s/coeficientes polinomiais	3,288	3,585	+
	Calibração	3,680	8,358	+
	Bayes empírico	2,322	+	+
Kanda	<i>Matrizes de covariância iguais</i>			
	Média aproximada/serial (condicional)	1,800	–	–
	Média aproximada/uniforme (condicional)	1,585	–	–
	Média aproximada/serial (estendida)	1,904	–	–
	Média aproximada/uniforme (estendida)	1,656	–	–
	<i>Matrizes de covariância diferentes</i>			
	Média aproximada/serial (condicional)	1,731	–	–
	Média aproximada/uniforme (condicional)	1,571	–	–
	Média aproximada/serial (estendida)	1,839	–	–
	Média aproximada/uniforme (estendida)	1,642	–	–
BLUP	Solução explícita/MV	3,209	3,471	+
	Solução explícita/MVR	3,182	3,518	+
	Newton-Raphson/MV	*	*	*
	Newton-Raphson/MVR	384,174	377,984	*
	Scoring de Fisher/MV	2,994	2,505	4,269
	Scoring de Fisher/MVR	3,072	2,493	3,753
	EM/MV	2,578	2,631	3,664
	EM/MVR	2,589	3,011	3,664
	Modelo III	2,738	2,251	3,177

– os autores não consideraram a imputação aplicada a esta parte dos dados;

+ não é possível aplicar a técnica a apenas esta parte dos dados

* matriz singular

Tabela 5.2: EQM para o modelo de grau 2 (dados de Potthoff e Roy)

Método		Instantes de tempo	
		t_1 a t_4	t_2 a t_4
Rao	Regressão individual	47,398	+
	Regressão s/ coef. polinomiais	4,430	+
	Calibração	9,483	+
Liski e Nummi	EM	3,587	3,265
	EM - passo único	3,756	3,265
BLUP	Solução explícita/MV	2,964	+
	Solução explícita/MVR	2,987	+
	Newton-Raphson/MV	*	*
	Newton-Raphson/MVR	*	77,632
	Scoring de Fisher/MV	2,682	3,614
	Scoring de Fisher/MVR	48,189	26,710
	EM/MV	2,785	3,612
	EM/MVR	2,785	3,329

+ não é possível aplicar a técnica a apenas esta parte dos dados

* matriz singular

Tabela 5.3: *EQM para o modelo de grau 1 (dados de Elston e Grizzle)*

Método		Instantes de tempo		
		t_1 a t_4	t_2 a t_4	t_3 a t_4
Rao	Regressão direta	0,769	0,577	0,566
	Regressão inversa	0,808	0,608	0,618
	Regressão individual	0,584	0,812	+
	Regressão s/coeficientes polinomiais	0,716	0,577	+
	Calibração	0,638	0,751	+
	Bayes empírico	0,498	+	+
	Componentes principais ($k = 1$)	1,541	+	+
	Componentes principais ($k = 2$)	0,643	+	+
	Componentes principais ($k = 3$)	0,769	+	+
	Fator analítico ($k = 1$)	1,480	+	+
	Fator analítico ($k = 2$)	0,781	+	+
Lee e Geisser	Média aproximada/não-estruturada	0,858	–	–
	Moda/não-estruturada	0,842	–	–
	Quase mínimos quadrados/não-estruturada	1,591	–	–
	Ponto p_1 /não-estruturada	0,799	–	–
	Média aproximada/Rao simples	0,697	–	–
	Moda/Rao simples	0,726	–	–
	Quase mínimos quadrados/Rao simples	1,769	–	–
	Ponto p_1 /Rao simples	0,731	–	–
	Quase mínimos quadrados/serial geral	1,523	–	–
	Ponto p_1 /serial geral	0,736	–	–
	Média aproximada/serial	0,510	–	–
	Quase mínimos quadrados/fator de Rao	1,672	–	–
	Ponto p_1 /fator de Rao	0,890	–	–
	Previsor marginal	7,860	–	–
	Mínimos quadrados simples	0,640	–	–
	Ad hoc	0,652	–	–

Tabela 5.3: Continuação

Método		Instantes de tempo		
		t_1 a t_4	t_2 a t_4	t_3 a t_4
Lee	Média aproximada/serial (MLE)	0,518	–	–
	Ad hoc/serial (MLE)	0,567	–	–
	Estendida/serial (MLE)	0,518	–	–
	Média aproximada/serial (MPVE)	0,503	–	–
	Ad hoc/serial (MPVE)	0,615	–	–
	Estendida/serial (MPVE)	0,484	–	–
	Média aproximada/uniforme (MLE)	1,469	–	–
	Previsor combinado/uniforme (MLE)	0,830	–	–
	Estendida/uniforme (MLE)	1,340	–	–
	Média aproximada/uniforme (MPVE)	1,469	–	–
	Previsor combinado/uniforme (MPVE)	0,753	–	–
	Estendida/uniforme (MPVE)	1,340	–	–
Kanda	Estendida/não estruturada (MLE)	0,641	–	–
	Estendida/não estruturada (MPVE)	0,641	–	–
	Média aproximada/serial (condicional)	0,783	–	–
	Média aproximada/serial (estendida)	0,828	–	–
Liski e Nummi	Média aproximada/uniforme (condicional)	0,954	–	–
	Média aproximada/uniforme (estendida)	0,826	–	–
Liski e Nummi	EM	0,835	0,669	0,563
	EM - passo único	0,867	0,677	0,563
BLUP	Solução explícita/MV	0,567	0,674	+
	Solução explícita/MVR	0,560	0,525	+
	Newton-Raphson/MV	247,265	1,211	*
	Newton-Raphson/MVR	797,137	391,363	2,821
	Scoring de Fisher/MV	0,577	0,740	0,548
	Scoring de Fisher/MVR	0,744	48,230	0,636
	EM/MV	0,576	0,731	0,548
	EM/MVR	0,574	0,712	0,548
	Modelo III	1,474	1,073	0,521

– os autores não consideraram a imputação aplicada a esta parte dos dados;

+ não é possível aplicar a técnica a apenas esta parte dos dados

* matriz singular

Tabela 5.4: *EQM para o modelo de grau 2 (dados de Elston e Grizzle)*

Método		Instantes de tempo	
		t_1 a t_4	t_2 a t_4
Rao	Regressão individual	2,989	+
	Regressão s/coeficientes polinomiais	0,769	+
	Calibração	2,172	+
Liski e Nummi	EM	0,839	0,701
	EM - passo único	0,903	0,701
BLUP	Solução explícita/MV	0,804	+
	Solução explícita/MVR	0,801	+
	Newton-Raphson/MV	251,757	8681,120
	Newton-Raphson/MVR	1309,089	571,391
	Scoring de Fisher/MV	1,208	3,015
	Scoring de Fisher/MVR	2,662	2,277
	EM/MV	0,788	0,668
	EM/MVR	2,277	0,787

+ não é possível aplicar a técnica a apenas esta parte dos dados

Tabela 5.5: *EQM para o modelo de grau 1 (dados de Williams e Izenman)*

Método		Instantes de tempo					
		t_1 a t_7	t_2 a t_7	t_3 a t_7	t_4 a t_7	t_5 a t_7	t_6 a t_7
Rao	Regressão direta	0,095	0,079	0,047	0,037	0,031	0,027
	Regressão inversa	0,103	0,081	0,048	0,040	0,034	0,028
	Regressão individual	0,206	0,158	0,111	0,066	0,055	+
	Regressão s/ coef. pol.	0,049	0,043	0,039	0,036	0,031	+
	Calibração	0,035	0,035	0,034	0,034	0,033	+
	Bayes empírico	0,194	0,143	0,097	0,054	+	+
	Comp.principais ($k = 1$)	0,038	–	–	–	–	–
	Comp.principais ($k = 2$)	0,048	–	–	–	–	–
	Comp.principais ($k = 3$)	0,043	–	–	–	–	–
	Fator analítico ($k = 1$)	0,038	–	–	–	–	–
	Fator analítico ($k = 2$)	0,061	–	–	–	–	–
	Fator analítico ($k = 3$)	0,071	–	–	–	–	–
Liski e	EM	–	–	0,044	0,035	0,030	0,028
Nummi	EM - passo único	–	–	0,047	0,035	0,030	0,028
BLUP	Solução expl./MV	0,116	0,079	0,057	0,042	0,034	+
	Solução expl./MVR	0,117	0,080	0,057	0,042	0,034	+
	Newton-Raphson/MV	*	*	*	*	13,012	*
	Newton-Raphson/MVR	0,148	0,108	0,075	0,735	0,052	*
	Scoring de Fisher/MV	0,102	0,080	0,071	0,081	0,144	0,306
	Scoring de Fisher/MVR	0,103	0,080	0,069	0,076	0,138	0,302
	EM/MV	0,090	0,052	0,035	0,078	0,142	0,302
	EM/MVR	0,178	0,119	0,083	0,078	0,144	0,302
	Modelo III	0,056	0,043	0,034	0,033	0,036	0,028

- os autores não consideraram a imputação aplicada a esta parte dos dados;
- + não é possível aplicar a técnica a apenas esta parte dos dados
- * matriz singular

Tabela 5.6: *EQM para o modelo de grau 2 (dados de Williams e Izenman)*

Método		Instantes de tempo				
		t_1 a t_7	t_2 a t_7	t_3 a t_7	t_4 a t_7	t_5 a t_7
Rao	Regressão individual	0,104	0,095	0,096	0,229	+
	Regressão s/coef. pol.	0,060	0,040	0,039	0,037	+
	Calibração	0,037	0,040	0,052	0,094	+
	Bayes empírico	0,087	0,075	0,069	+	+
Lee	Média aproximada/serial	0,025	–	–	–	–
	Ad hoc/serial	0,043	–	–	–	–
	Média aproximada/uniforme	0,120	–	–	–	–
	Previsor combinado/uniforme	0,073	–	–	–	–
	Estendida/serial	0,044	–	–	–	–
	Estendida/uniforme	0,066	–	–	–	–
	Estendida/não-estruturada	0,103	–	–	–	–
Liski e Nummi	EM	–	–	0,044	0,037	0,030
	EM - passo único	–	–	0,047	0,037	0,030
BLUP	Solução expl./MV	0,066	0,050	0,040	0,040	+
	Solução expl./MVR	0,065	0,050	0,040	0,040	+
	Newton-Raphson/MV	*	*	*	*	*
	Newton-Raphson/MVR	0,073	0,059	0,064	0,193	0,287
	Scoring Fisher/MV	0,055	0,048	0,062	0,046	0,032
	Scoring Fisher/MVR	0,057	0,048	0,042	0,163	0,031
	EM/MV	0,056	0,046	0,042	0,047	0,053
	EM/MVR	0,058	0,048	0,043	0,051	0,033

– os autores não consideraram a imputação aplicada a esta parte dos dados;

+ não é possível aplicar a técnica a apenas esta parte dos dados

* matriz singular

Tabela 5.7: *Valores imputados (dados de Santos)*

Modelo	MV			MVR		
	y_{24}	y_{15}	y_{78}	y_{24}	y_{15}	y_{78}
Modelo I (grau 1)						
Soluções explícitas						
indivíduos completos	0,087	0,072	0,057	0,092	0,080	0,056
tempos completos	0,121	0,127	0,034	0,120	0,125	0,022
Scoring de Fisher	0,093	0,082	0,056	0,096	0,088	0,056
Newton Raphson	0,088	0,074	0,049	0,089	0,077	0,051
EM	0,092	0,082	0,055	0,092	0,082	0,053
Modelo II (grau 2)						
Soluções explícitas						
indivíduos completos	0,064	0,078	0,071	0,071	0,078	0,067
tempos completos	0,131	0,099	0,106	0,129	0,101	0,125
Scoring de Fisher	0,067	0,068	0,049	0,072	0,069	0,046
Newton Raphson	0,053	0,067	0,055	0,065	0,076	0,056
EM	0,087	0,052	0,022	0,068	0,069	0,048
Modelo III						
indivíduos completos	–	–	–	0,153	0,047	0,077
tempos completos	–	–	–	0,154	0,050	0,082

Tabela 5.8: *EQM para os Modelos I, II e III (dados de Santos)*

Modelo	MV	MVR
Modelo I (grau 1)		
Soluções explícitas	0,057	0,055
Scoring de Fisher	0,036	0,037
Newton Raphson	0,037	0,037
EM	0,036	0,035
Modelo II (grau 2)		
Soluções explícitas	0,061	0,037
Scoring de Fisher	0,073	0,073
Newton Raphson	0,037	0,073
EM	0,035	0,088
Modelo III	–	0,032

Conclusões

A principal contribuição deste trabalho é a implementação do método de imputação via BLUP (Best Linear Unbiased Predictor) em modelos de curvas de crescimento polinomiais, no módulo CM do aplicativo NTIA/EMBRAPA.

Nesse método há a possibilidade de variação do modelo para as médias e do método de estimação dos parâmetros das matrizes de covariância.

Os resultados da imputação via BLUP foram comparados com os resultados obtidos por outros autores. As principais conclusões são:

- não há um método que seja o melhor; nos exemplos pode-se ver que, dependendo do conjunto de dados, o melhor resultado é ora obtido por um método, ora por outro;
- os resultados encontrados via BLUP sob o Modelo III sempre apresentam-se bons. Isso, como já foi dito, é bastante satisfatório pois esse modelo é simples, interpretável e de fácil aplicação (o programa encontra-se no Apêndice B);
- para se efetuar a imputação de dados, vários métodos devem ser tentados; a escolha do melhor pode ser feita através do critério leave-one-out.

A continuidade deste trabalho pode consistir em desenvolver um algoritmo para obtenção dos estimadores de máxima verossimilhança e máxima verossimilhança restrita, quando a matriz de covariância dos erros tem uma estrutura parcimoniosa. O caminho teórico já foi dado neste trabalho, resta considerar certas estruturas.

Outro ponto que pode ser considerado para futura investigação é o desenvolvimento das técnicas e aplicação ao caso multivariado.

Pode-se também considerar a possibilidade do desenvolvimento de programas para outros aplicativos, como por exemplo o S-Plus ou Ox.

Por fim, outros modelos para a média, além dos modelos polinomiais, podem ser considerados.

Apêndice A

Resultado A1

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB})^{-1}\mathbf{BVA}^{-1}$$

Temos que

$$\begin{aligned}\Sigma^{-1} &= (\mathbf{T}_c \mathbf{G} \mathbf{T}_c' + \sigma^2 \mathbf{I}_{p_c})^{-1} \\ &= \sigma^{-2} \mathbf{I}_{p_c} - \sigma^{-2} \mathbf{T}_c \mathbf{G} (\mathbf{G} + \sigma^{-2} \mathbf{G} \mathbf{T}_c' \mathbf{T}_c \mathbf{G})^{-1} \mathbf{G} \mathbf{T}_c' \sigma^{-2} \\ &= \sigma^{-2} \mathbf{I}_{p_c} - \sigma^{-2} \mathbf{T}_c (\mathbf{I}_c + \sigma^{-2} \mathbf{G} \mathbf{T}_c' \mathbf{T}_c)^{-1} \mathbf{G} \mathbf{T}_c' \sigma^{-2} \\ &= \sigma^{-2} \mathbf{I}_{p_c} - \sigma^{-2} \mathbf{T}_c [(\mathbf{T}_c' \mathbf{T}_c)^{-1} + \sigma^{-2} \mathbf{G}]^{-1} \mathbf{G} \mathbf{T}_c' \sigma^{-2} \\ &= \sigma^{-2} \mathbf{I}_{p_c} - \sigma^{-2} \mathbf{T}_c (\mathbf{T}_c' \mathbf{T}_c)^{-1} [\sigma^2 (\mathbf{T}_c' \mathbf{T}_c)^{-1} + \mathbf{G}]^{-1} \mathbf{G} \mathbf{T}_c' \\ &= \sigma^{-2} \mathbf{I}_{p_c} - \sigma^{-2} \mathbf{T}_c (\mathbf{T}_c' \mathbf{T}_c)^{-1} \Psi^{-1} [\Psi - \sigma^2 (\mathbf{T}_c' \mathbf{T}_c)^{-1}] \mathbf{T}_c' \\ &= \sigma^{-2} \mathbf{I}_{p_c} - \sigma^{-2} \mathbf{T}_c (\mathbf{T}_c' \mathbf{T}_c)^{-1} \mathbf{T}_c' + \mathbf{T}_c (\mathbf{T}_c' \mathbf{T}_c)^{-1} \Psi^{-1} (\mathbf{T}_c' \mathbf{T}_c)^{-1} \mathbf{T}_c',\end{aligned}$$

onde $\Psi = \sigma^2 (\mathbf{T}_c' \mathbf{T}_c)^{-1} + \mathbf{G}$

Resultado A2

Usando o fato que $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$ para dimensões adequadas, temos

$$\begin{aligned}|\Sigma| &= |\mathbf{T}_c \mathbf{G} \mathbf{T}_c' + \sigma^2 \mathbf{I}_{p_c}| \\ &= |\sigma^2 (\sigma^{-2} \mathbf{T}_c \mathbf{G} \mathbf{T}_c' + \mathbf{I}_{p_c})| \\ &= (\sigma^2)^{p_c} |\sigma^{-2} \mathbf{T}_c \mathbf{G} \mathbf{T}_c' + \mathbf{I}_{p_c}|\end{aligned}$$

$$\begin{aligned}
&= (\sigma^2)^{p_c} |\sigma^{-2} \mathbf{T}_c' \mathbf{T}_c \mathbf{G} + \mathbf{I}_{p_c}| \\
&= (\sigma^2)^{p_c} |\sigma^{-2} (\mathbf{T}_c' \mathbf{T}_c) [\mathbf{G} + \sigma^2 (\mathbf{T}_c' \mathbf{T}_c)^{-1}]| \\
&= (\sigma^2)^{p_c} |\sigma^{-2} (\mathbf{T}_c' \mathbf{T}_c) \boldsymbol{\Psi}| \\
&= (\sigma^2)^{p_c - c} |\mathbf{T}_c' \mathbf{T}_c| |\boldsymbol{\Psi}|.
\end{aligned}$$

Resultado A3

Considere a função

$$f(\boldsymbol{\Sigma}) = \log|\boldsymbol{\Sigma}| + \text{tr}[\boldsymbol{\Sigma}^{-1} \mathbf{A}]$$

se $\mathbf{A} > 0$ então, sujeito a $\boldsymbol{\Sigma} > 0$, $f(\boldsymbol{\Sigma})$ é minimizada unicamente em $\boldsymbol{\Sigma} = \mathbf{A}$. (Ver prova em Seber (1984))

Resultado A4

(Ver Searle, 1971)

Seja $\mathbf{u}$ um vetor aleatório e $\mathbf{u}' \mathbf{Q} \mathbf{u}$ uma forma quadrática. Então

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

Resultado A5

(Ver Mardia, Kent e Bibby, 1979)

Se

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim N \left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right],$$

então

$$\mathbf{x}_2 | \mathbf{x}_1 \sim N \left[\boldsymbol{\mu}_2 + \mathbf{V}_{21} \mathbf{V}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1), \mathbf{V}_{22} - \mathbf{V}_{21} \mathbf{V}_{11}^{-1} \mathbf{V}_{12} \right].$$

Apêndice B

Exemplo: o exemplo de programa abaixo é a aplicação aos dados do Exemplo 2.5. Ele está dividido em três partes:

1. entrada dos dados e definição do modelo;
2. estimação das matrizes de covariância;
3. cálculo dos valores imputados.

Na primeira parte, os dados ausentes são substituídos por zeros. As matrizes **E**, **F**, **X** e **Z** são definidas. Na segunda parte, um dos seguintes programas é usado:

- com todos os dados disponíveis: EMMV.txt, SCMV.txt, NRMV.txt, EMMVR.txt, SCMVR.txt, NRMVR.txt;
- com uma parte completa dos dados: EXPLIMV.txt, EXPLIMVR.txt, MOD3.txt.

Os primeiros, com exceção de NRMVR.txt foram desenvolvidos e são apresentados em Andreoni (1989). O NRMVR.txt e os outros, que podem ser usados descartando-se os indivíduos com dados ausentes (eliminando linhas) ou instantes com dados ausentes (eliminando colunas), estão adiante. Além disso, é apresentado também o programa que calcula os valores iniciais para os algoritmos denominado CHUTE.txt.

PARTE 1

/* Exemplo: Testosterona - Modelo III */

assuma saida= "c : saida.txt";

```
my = {1 0.067 0.101 0.062 0.045 0.026 0.044 0.015,
      2 0.051 0.058 0.088 0.072 0.015 0.090 0.111,
      3 0.034 0.135 0.049 0.206 0.126 0.070 0.015,
      4 0.171 0.000 0.052 0.100 0.114 0.074 0.048,
      5 0.000 0.105 0.143 0.125 0.015 0.015 0.015,
      6 0.114 0.089 0.123 0.196 0.214 0.239 0.487,
      7 0.114 0.772 0.105 0.109 0.064 0.480 0.170,
      8 0.015 0.015 0.026 0.152 0.015 0.015 0.000,
      9 0.087 0.077 0.095 0.077 0.087 0.092 0.162,
      10 0.040 0.059 0.068 0.108 0.090 0.058 0.089};
```

p=ncoln(y)-1; /* número de instantes */

n=nlin(my); /* número de indivíduos */

m=3; /* número de observações ausentes*/

REF={1:(n*p), ident(n*p)};

E1=cata(REF', 1, (23 29 56), "==");

F1=cata(REF', 1, (1:22 24:28 30:55 57:71), "==");

Ee=E1[,2:(n*p+1)];

E=Ee ';

Ff=F1[,2:(n*p+1)];

F=Ff ';

```

Xi=ident(p);
Zi=criamat(p,1,1);
n1=criamat(n,1,1);
X=n1@Xi;
Z=ident(n)@Zi;

```

Se o modelo a ser ajustado é o de retas individuais e as observações são tomadas em tempos eqüidistantes, basta substituir os comandos que definem as matrizes X_i e Z_i por

```

Xi= {1 (-3), 1 (-2), 1 (-1), 1 0, 1 1, 1 2, 1 3};
Zi = Xi;

```

Se o modelo a ser ajustado é o de parábolas individuais e as observações são tomadas em tempos eqüidistantes, basta substituir os comandos que definem as matrizes X_i e Z_i por

```

Xi= {1 (-3) 9, 1 (-2) 4, 1 (-1) 1, 1 0 0, 1 1 1, 1 2 4, 1 3 9};
Zi = Xi;

```

PARTE 2

MOD3.txt

```
/* estimar as componentes da variância, excluindo linhas */

mya=cata(my,1,{1:3 6:7 9:10},"==");
myc=mya[ ,2:(p+1)];
nc=7;
pc=7;

yc=vec(myc);
nc1=criamat(nc,1,1);
pc1=criamat(pc,1,1);
xc=nc1@ident(pc);
zc=ident(nc)@pc1;
A=yc'*(ident(nc*pc)-xc*inv(xc'*xc)*xc')*yc;
B=(yc'*(ident(nc*pc)-xc*inv(xc'*xc)*xc')*zc*zc'*yc)/pc;

imprime A;
imprime B;

sigma2=(A-B)/((nc-1)*(pc-1));
sigmab2=(pc*B-A)/(pc*(nc-1)*(pc-1));
se (sigmab2 < 0) sigmab2=0;
cc sigmab2=sigmab2;
G=sigmab2;

imprime sigma2;
imprime G;
```

Para estimar as componentes da variância cortando colunas, basta substituir as primeiras 4 linhas do programa por

```
myc=my[,4:7];  
nc=10; pc=4;
```

EXPLIMV.txt

```
/* estimar as componentes da variância, excluindo linhas */  
  
mya=cata(my,1,{1:3 6:7 9:10},"==");  
myc=mya[,2:(p+1)];  
nc=7; pc=7;  
  
Tc=Xi;  
T=Tc;  
c=ncol(Tc);  
A1=criamat(nc, 1, 1);  
  
Mphi=inv(A1'*A1)*A1'*myc*T*inv(T'*T);  
MP=(myc-A1*Mphi*T')*(myc-A1*Mphi*T');  
S=myc*(ident(nc)-A1*inv(A1'*A1)*A1')*myc;  
Mpsi=(inv(Tc'*Tc)*Tc'*S*Tc*inv(Tc'*Tc))/nc;  
  
sigma2=traco(MP*(ident(pc)-Tc*inv(Tc'*Tc)*Tc'))/(nc*(pc-c));  
G=Mpsi-sigma2*inv(Tc'*Tc);  
  
imprime sigma2;  
imprime G;
```

Para estimar as componentes da variância cortando colunas, basta substituir as primeiras 5 linhas do programa por `myc=my[,4:7];`

```
nc=10;
```

```
pc=4;
```

```
Tc = { 1 0, 1 1, 1 2, 1 3 };      ou      Tc = { 1 0 0, 1 1 1, 1 2 4, 1 3 9 };
```

EXPLIMVR.txt

```
/* estimar as componentes da variância, excluindo linhas */
```

```
mya=cata(my,1,{1:3 6:7 9:10},"==");
```

```
myc=mya[,2:(p+1)];
```

```
nc=7;
```

```
pc=7;
```

```
Tc=Xi;
```

```
T=Tc;
```

```
c=ncol(Tc);
```

```
q=ncol(T);
```

```
A1=criamat(nc, 1, 1);
```

```
nf=ncol(A1);
```

```
Mphi=inv(A1'*A1)*A1'*myc*T*inv(T'*T);
```

```
MP=(myc-A1*Mphi*T')*(myc-A1*Mphi*T');
```

```
S=myc'*(ident(nc)-A1*inv(A1'*A1)*A1')*myc;
```

```
Mpsi=(inv(Tc'*Tc)*Tc'*S*Tc*inv(Tc'*Tc))/(nc-nf);
```

```
sigma2=traco(MP*(ident(pc)-Tc*inv(Tc'*Tc)*Tc'))/(nc*(pc-c)-nf*(q-c));
```

```
G=Mpsi-sigma2*inv(Tc'*Tc);
```

```
imprime sigma2;
```

```
imprime G;
```

Para estimar as componentes da variância cortando colunas, basta substituir as primeiras 5 linhas do programa por `myc=my[ ,4:7];`

```
nc=10;
```

```
pc=4;
```

```
Tc = { 1 0, 1 1, 1 2, 1 3 };      ou      Tc = { 1 0 0, 1 1 1, 1 2 4, 1 3 9 };
```

CHUTE.txt

```
Nt=N*p-1;
```

```
cx=ncol(Xi);
```

```
cz=ncol(Zi);
```

```
i=1;
```

```
enquanto (i<=n)
```

```
{ codi=criamat(p,1,i);
```

```
se (i==1) cod=codi;
```

```
cc cod={cod,codi};
```

```
i=i+1;}
```

```
Za=n1@Zi;
```

```
my1=my[,2:(p+1)];
```

```
y=vec(my);
```

```
id=REF'[1];
```

```
dadosa={id cod X Za y};
```

```
dados=cata(dadosa, 1,(1:22 24:28 30:55 57:71), "==" );
```

```

sxx=criamat(cx, cx,0);
sxy=criamat(cx, 1,0);
i=1;
enquanto (i!=n)
{ ind=cata(dados,2,i,"==");
Xia=ind[ ,3:(cx+2)];
yia=ind[ , (cx+cz+3)];
sxx=sxx + Xia'*Xia;
sxy=sxy + Xia'*yia;
i=i+1;}
alfa=inv(sxx)*sxy;
imprime alfa;

syy=0;
sbz=0;
sxx=criamat(cz, cz,0);
sxy=criamat(cz, 1,0);
i=1;
czia=cx+3;
czib=cx+cx+2;
enquanto (i!=n)
Xia=ind[ ,3:(cx+2)];
Zia=ind[ ,czia:czib];
yia=ind[ , (cx+cz+3)];

bi=inv(Zia'*Zia)*Zia'*(yia-Xia*alfa);
syy=syy + yia'*yia;

```



```

sbz=sbz+bi'* Zia'*(yia-Xia*alfa);
szz=szz+inv(Zia'*Zia);
sbb=sbb+bi*bi';
i=i+1;}
sigma2=(syy-alfa'*sxy-sbz)/(Nt-(n-1)*cz-cx);
imprime sigma2;

G=sbb/(n-1) - sigma2*szz/n;
imprime G;

```

NRMVR.txt

```

k=cz*(cz+1)/2+1; /* calculo das derivadas de G */ theta={{1:k}}';

i=1;
j=1;
l=1;
enquanto (l<=(k-1))
{Al = G-G;
Al[i,j]=1;
Al[j,i]=1;
theta[l, ] =G[i,j];
se (l==1) A=Al;
cc A={A,Al};
se (j==cz) {i=i+1;
j=i;}
cc j=j+1;

```

```

l=l+1;}
theta[k, ] =sigma2;

/* calculo de alfa */
sxwx=ident(cx)-ident(cx);
sxwy={ (1:cx) }'-{ (1:cx) }';
invg=inv(G);
i=1;
cza=cx+3;
czib=cx+cz+2;
enquanto (i<=n)
{ind =cata(dados, 2, i, "==");
Xia=ind[ ,3:(cx+2)];
Zia=ind[ , cza:czib];
yia=ind[ , (cx+cz+3)];
si=(ident(nlin(Zia))-Zia*inv(theta[k, ]*invg+Zia'*Zia)*Zia')/theta[k, ];
sxwx=sxwx+Xia'*si*Xia;
sxwy=sxwy+Xia'*si*yia;
i=i+1;}
var_alfa=inv(sxwx);
alfa=var_alfa*sxwy;

/* processo iterativo */
p_theta=1000;
iteracao=0;
enquanto ((iteracao<50)&&(p_theta>=0.001))
{theta_0=theta;

```

```

HS=ident(k)-ident(k);
s={ (1:k) }'-{ (1:k) }';
i=1;
enquanto (i<=n)
{ind =cata(dados, 2, i, "==");
Xia=ind[ ,3:(cx+2)];
Zia=ind[ , czia:czib];
yia=ind[ , (cx+cz+3)];
si=(ident(nlin(Zia))-Zia*inv(theta[k, ]*invg+Zia'*Zia)*Zia')/theta[k, ];
ri=yia-Xia*alfa;
pi=si-si*Xia*var_alfa*Xia'*si;

l=1;
enquanto (l<=k)
{ se (l==k) Gil=ident(nlin(si));
cc {Al = A[((l-1)*cz+1):(l*cz),];
Gil=Zia*Al*Zia';}

/* calculo do vetor escore */
s[l, ] = s[l, ] + traco(si*ri*ri'*si*Gil-pi*Gil);
h=l;
enquanto (h<=k)
{se (h==k) Gih=ident(nlin(si));
cc {Ah=A[((h-1)*cz+1):(h*cz),];
Gih=Zia*Ah*Zia';}
HS[l,h] = HS[l,h]-2*traco(si*Gil*pi*Gih*si*ri*ri')+traco(pi*Gil*pi*Gih);
HS[h,l]=HS[l,h];
h=h+1;}

```

```

l=l+1;}
i=i+1;}
HSI=inv(HS);

/* calculo dos novos theta e sigma2 */
theta =theta_0 + HSI*s;
sigma2=theta[k, ];

/* calculo da nova G */
G=ident(cz) - ident(cz);
l=1;
enquanto (l<=(k-1))
{Al=A[((l-1)*cz+1):(l*cz), ];
G = G + theta[l, ]*Al;
l=l+1;}
p_theta=max(abs(theta-theta_0));
iteracao=iteracao+1;}

/* calculo do novo alfa */
sxwx=ident(cx)-ident(cx);
sxwy={(1:cx)}'-{(1:cx)}';
invG=inv(G);
i=1;
enquanto (i<=n)
{ind =cata(dados, 2, i, "==" );
Xia=ind[ ,3:(cx+2)];
Zia=ind[ , czia:czib];
yia=ind[ , (cx+cz+3)];

```

```

si=(ident(nlin(Zia))-Zia*inv(theta[k, ]*invg+Zia'*Zia)*Zia')/theta[k, ];
sxwx=sxwx+Xia'*si*Xia;
sxwy=sxwy+Xia'*si*yia;
i=i+1;}
var_alfa=inv(sxwx);
alfa=var_alfa*sxwy;

imprime iteracao;
imprime p_theta;
imprime alfa;
    se (sigma2 <= 0) sigma2=0.01;
    cc sigma2=sigma2;
imprime sigma2;
    dG=abs(min(diag(G)));
    g2=dG+1;
    q=cz;
    g2=g2*ident(q);
    aa,ag=auto(G);
    r=1; enquanto (r<=q)
    se (aa[r,]<=0) G=G+g2;
    cc G[r,r]=G[r,r];
    r=r+1;
    cz=q;
imprime G;

```

PARTE 3

```
/* fazer a imputação */  
  
my1=my[ ,2:(p+1)];  
y=vec(my1);  
Rtil=sigma2*ident(n*p);  
Gtil=ident(n)@G;  
V=Z*Gtil*Z'+ Rtil;  
Vi=inv(V);  
W=X' *Vi*X;  
Wi=inv(W);  
Ri=inv(Rtil);  
  
Q=Rtil*(Vi*X*Wi*X'*Vi + Ri - Vi)*Rtil;  
Mb=E'*Q*E*E'*Ri*E;  
Ma=E'*Q*F*F'*Ri*F*F'*y;  
ya=inv(ident(m)-Mb)*Ma;  
  
imprime ya;
```

Referências Bibliográficas

- [1] Andreoni, S.(1989). *Modelos de efeitos aleatórios para análise de dados longitudinais não balanceados em relação ao tempo*. Dissertação de Mestrado, IME-USP.
- [2] Azzalini, A.(1987). Growth curves analysis for patterned covariance matrices. *News Perspectives in Theoretical and Applied Statistics*. New York, John Wiley, 63-73.
- [3] Bard, Y.(1974). *Nonlinear parameter estimation*. Academic Press, New York, 341 p.
- [4] Barroso, L.P.(1995). *Imputação de dados em painéis para populações finitas*. Tese de Doutorado, IME-USP.
- [5] Barroso, L.P.; Bussab, W.O. and Knott, M.(1998). Best linear unbiased predictor in the mixed model with incomplete data. *Communications in Statistics - Theory and Methods*, **27**(1), 121-129.
- [6] Barroso, L.P. and Salinas, D.T.P. (1999). Bootstrap for imputed data in panels: a numerical evaluation. *InterStat*, December, 1.
- [7] Barroso, L.P. e Sandoval, M.C.(1999). *Estudo das variáveis do teste de esforço cardiopulmonar na diferenciação de classes de pacientes*. RAE-CEA-98PO6.
- [8] Barry, D.(1995). A bayesian model for growth curve analysis. *Biometrics*, **51**, 639-655.

- [9] Beale, E.M. and Little, R.J.A.(1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society B*, **37**, 129-146.
- [10] BMDP (1992). *Statistical Software Manual*. University of California Press, Berkeley, vol **2**.
- [11] Box, G.E.P.(1950). Problems in the analysis of growth and wear curves. *Biometrics*, **6**, 362-389.
- [12] Chi, E.M and Reinsel, G.C.(1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association*, **84**, 452-459.
- [13] Crowder, M.J. and Tredger, J.A.(1981). The use of exponentially damped polynomial for biological recovery data. *Applied Statistics*, **30**(2), 147-152.
- [14] Dawson, K.S, Gennings,C. and Carter, W.H.(1997). Two graphical techniques useful in detecting correlation structure in repeated measures data. *The American Statistician*, **51**(3), 275-283.
- [15] Dempster, A.P.; Laird, N.M. and Rubin,D.B.(1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, **39**, 1-38.
- [16] Dwyer, J.H.; Feinleib, M.; Lippert, P. and Hoffmeister, H.(1992). *Statistical models for longitudinal studies of health*. Oxford University Press, New York, 385 p.
- [17] Elston, R.C. and Grizzle, J.E.(1962). Estimation of time response and their confidence bands. *Biometrics*, **18**, 148-159.
- [18] Fairclough, D.L. and Helms, R.W.(1986). A mixed linear model with linear covariance: a sensitivity analysis of maximum likelihood estimators. *Journal of Statistical Computation and Simulation*, **25**, 205-236.

- [19] Fearn, T.(1975). A Bayesian approach to growth curves. *Biometrika*, **62**, 89-100.
- [20] Geisser, S.(1970). Bayesian analysis of growth curves. *Sankya A*, **32**, 53-64.
- [21] Geisser, S.(1971). The inferential use of predictive distribuitons. *Foundations of Statistical Inference*, Holt, Rinehart and Winstson, 456-469.
- [22] Geisser, S.(1974). Predictive approach to random effect model. *Biometrika*, **61**(1), 101-107.
- [23] Ghosh, M.N.; Grizzle, J.E. and Sen, P.K.(1973). Nonparametric methods in longitudinal studies. *Journal of the American Satistical Association*, **68**, 341, 29-36.
- [24] Graybill, F.A.(1976). *Theory and application of linear model*. Duxbury Press, North Scituate, 704 p.
- [25] Grizzle, J.E. and Allen, O.M.(1969). Analysis of growth and dose response curves. *Biometrics*, **25**, 357-381.
- [26] Hartley, H.O. and Hocking, R.R.(1971). The analysis of incomplete data (with discussion). *Biometrics*, **27**, 783-823.
- [27] Harville, D.A.(1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383-385.
- [28] Harville, D.A.(1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**(358), 320-336.
- [29] Hoel, P.G.(1964). Methods for comparing growth type curves. *Biometrics*, **20**, 859-872.
- [30] Jennrich, R.I. and Schluchter, M.D. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, **42**, 805-820.

- [31] Jones, R .H.(1993). *Longitudinal data with serial correlation: a state-space approach*. Chapman and Hall, London.
- [32] Kackar, R.N. and Harville, R.(1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, **79**, 853-862.
- [33] Kanda, T.(1992). MSE's of prediction in growth curve model with covariance structure. *Annals of the Institute of Statistical Mathematics*, **44** (3), 519-528.
- [34] Keramidas, E.M. and Lee, J.C.(1990). Forecasting technological substitutions with concurrent short times series. *Journal of the American Statistical Association*, **85**, 625-632.
- [35] Keramidas, E.M. and Lee, J.C.(1995). Selection of a covariance structure for growth curves. *Biometrical Journal*, **37**, 783-797.
- [36] Khatri, C.G.(1966). A note on a Manova model applied to problems in growth curves. *Annals of the Institute of Statistical Mathematics*, **18**, 75-86.
- [37] Khatri, C.G.(1973). Testing some covariance structures under a growth curve model. *Journal of Multivariate Analysis*, **3**, 102-116.
- [38] Khatri, C.G.(1988). Robustness study for linear growth model. *Journal of Multivariate Analysis*, **24**, 66-87.
- [39] Kleinbaum, D.G.(1973a). A generalization of the growth curve model. *Journal of Multivariate Analysis*, **3**, 117-124.
- [40] Kleinbaum, D.G.(1973b). Testing linear hypotheses in generalized multivariate linear models. *Communications in Statistics*, **1**, 433-457.

- [41] Krishnaiah, P.R.(1975). Simultaneous tests for multiple comparisons of growth curve when errors are autocorrelated. In *Proceeding of the 40th Session of the International Statistical Institute*, 62-66.
- [42] Kshirsagar, A.M. and Smith, W.B.(1995). *Growth curves*. Marcel Dekker, New York, 359 p.
- [43] Laird, N.M.(1982). Computation of components using the EM algorithm. *Journal of the American Statistical Association*, **14**, 295-303.
- [44] Laird, N.M.(1988). Missing data in longitudinal studies. *Statistics in Medicine*. **7**, 305-315.
- [45] Laird, N.M. and Ware, J.H.(1982). Random effects models for longitudinal data. *Biometrics*, **38**, 963-974.
- [46] Laird, N.M.; Lange, N. and Stram, D.O.(1987). Maximum likelihood computations with repeated measures: application of the EM algorithm. *Journal of the American Statistical Association*, **82**, 97-105.
- [47] Lange, N. and 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*, **84**, 241-247.
- [48] Lavange, L.M. and Helms, R.W.(1983). The analysis of incomplete longitudinal data with time series covariance structures. *Joint Statistical Meetings*, Toronto, Canada.
- [49] Lee, J.C.(1988). Prediction and estimation of growth curves with special covariance structures. *Journal of the American Statistical Association*, **83**, 432-440.
- [50] Lee, J.C.(1991). Tests and models selection of growth curve model. *Biometrics*, **47**, 147-159.

- [51] Lee, J.C. and Geisser, S.(1972). Growth curve prediction. *Sankya A*, **34**, 393-412.
- [52] Lee, J.C. and Geisser, S.(1975). Applications of growth curve prediction. *Sankya A*, **37**, 239-256.
- [53] Leeden, v.d.R.; Vrijburg, K. and Leeuw, J. (1996). A review of two different approaches for the analysis of growth data using longitudinal linear models: comparing hierarchical linear regression (ML3, HLM) and repeated measures designs with structured covariance matrices(BMDP5V). *Computational Statistics and Data Analysis*, **21**(5) , 583-605.
- [54] Lindstrom, M.J. and Bates, D.M.(1988). Newton-Raphson and EM algorithms for linear mixed effects models for repeated measures data. *Journal of the American Statistical Association*, **83**, 1014-1022.
- [55] Liski, E.P.(1984). Missing data problems in growth curves. In *Compstat 1984* Viena: Physica-Verlag, 479-484.
- [56] Liski, E.P. (1985). Estimation from incomplete data in growth curve models. *Communications in Statistics Simulation and Computation*, **14**, 775-792.
- [57] Liski, E.P. and Nummi, T.(1990). Prediction in growth curve models using the EM algorithm. *Computational Statistics and Data Analysis*, **10**, 99-108.
- [58] Mardia, K.V.; Kent, J.T. and Bibby, J.M.(1979). *Multivariate Analysis*. Academic Press, New York, 518 p.
- [59] McGilchrist, C.A. and Cullis, B.R.(1990). REML estimation for repeated measures analysis. *Journal of Statistical Computational and Simulation*, **38**, 151-164.
- [60] McGilchrist, C.A. and Yau, K.K.W.(1995). The derivation of BLUP, ML, REML estimation methods for generalized linear mixed models. *Communications in Statistics Theory and Methods*, **24**(12), 2963-2980.

- [61] NTIA Versão 4.2.1 (1995). *Software científico elaborado pelo Núcleo Tecnológico de Informática para Agropecuária da EMBRAPA(NTIA/EMBRAPA)*. Campinas, SP.
- [62] Nummi, T. (1997). Estimation in a random effects growth curve model. *Journal of Applied Statistics*, **24**(2), 157-168.
- [63] Orchard, T. and Woodbury, M.A.(1972). A missing information principle: theory and applications. *Proceedings of the 6th Berkeley Symposium on the Mathematical Statistics and Probability*, **1**, 697-715.
- [64] Pan, J.X.; Fang, K.T. and von Rosen, D.(1997). Local influence assessment in the growth curve model with unstructured covariance. *Journal of Statistical Planning and Inference*, **62**, 263-278.
- [65] Patterson, H.D. and Thompson, R.(1971). Recovery of inter-block information when blocks sizes are unequal. *Biometrika*, **58**, 545-554.
- [66] Pearson, E.S. and Hartley, H.O.(1972). *Biometrika Tables for Statisticians*. Cambridge University, Cambridge, vols 1 and 2.
- [67] Potthoff, R.F. and Roy, S.N.(1964). A generalized multivariate analysis of variance model useful specially for growth curve problems. *Biometrika*, **51**, 313-326.
- [68] Rao, C.R.(1958). Some statistical methods for comparison of growth curves. *Biometrics*, **14**, 1-17.
- [69] Rao, C.R.(1959). Some problems involving linear hypothesis in multivariate analysis. *Biometrika*, **46**, 49-58.
- [70] Rao, C.R.(1965). The theory of least square when parameters are stochastic and its application to the analysis of growth curve. *Biometrika*, **52**, 447-458.

- [71] Rao, C.R.(1966). Covariance adjustment and related problems in multivariate analysis. In *Multivariate Analysis II*. P.R.Krishnaiah. Academic Press, New York, 87-103.
- [72] Rao, C.R.(1967). Least square theory using estimated dispersion matrix and applications to measurement of signals. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 355-372.
- [73] Rao, C.R.(1968). A note on a previous lemma in the tehory of least squares and some further results. *Sankya A*, **30**, 259-266.
- [74] Rao, C.R.(1973). *Linear statistical inference and its applications*. Wiley, New York, 625 p.
- [75] Rao, C.R.(1987). Prediction of future observations in growth curve models. *Statistical Science*, **2**, 434-471.
- [76] Reinsel, G.(1982). Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure. *Journal of the American Statistical Association*, **77**, 190-195.
- [77] Reinsel, G.(1984). Estimation and prediction in a multivariate random effects generalized linear model. *Journal of the American Statistical Association*, **79**, 406-414.
- [78] Reinsel, G.(1985). Mean squared error properties of empirical Bayes estimators in multivariate random effects general linear models. *Journal of the American Statistical Association*, **80**, 642-650.
- [79] Rubin, D.B.(1976). Inference and missing data. *Biometrika*, **63**, 581-592.
- [80] Rubin, D.B.(1987). *Multiple Imputation for nonresponse in surveys*. Wiley, New York, 258 p.

- [81] Salinas, D.T.P.(1998). *Bootstrap não-paramétrico aplicado a dados incompletos*. Dissertação de Mestrado, IME-USP.
- [82] Santos, I.C.C.dos (1997). *Interação materno fetal e o parto na égua: perfis de esteróides plasmáticos na fêmea a termo e no neonato*. Dissertação de Mestrado, Departamento de Reprodução Animal da Faculdade de Medicina Veterinária e Zootecnia da USP.
- [83] Schluchter, M.D.(1988). Analysis of incomplete data using linear models with structured covariance matrices. *Statistics in Medicine*, **7**, 317-324.
- [84] Schwertman, N.C. and Allen, D.M.(1979). Smoothing and indefinite variance covariance matrix. *Journal of Statistical Computation and Simulation*, **9**, 183-194.
- [85] Schwertman, N.C.; Husely, J.R. and Allen, D.M.(1980). Computation of the estimated parameters and Wald statistics for the generalized growth curve model. *Communications in Statistics Simulation and Computation*, **9**, 675-693.
- [86] Schwertman, N.C.; Flynn, W.; Stein, S. and Schenk, K.L.(1985). A Monte Carlo study of alternative procedures for testing the hypotheses of parallelism for complete and incomplete growth curve data. *Journal of Statistical Computation and Simulation*, **21**, 1-37.
- [87] Searle, S.R.(1971). *Linear models*. John Wiley and Sons, New York, 532 p.
- [88] Seber, G.A.F.(1984). *Multivariate observation*. John Wiley and Sons, New York.
- [89] Silva, P.L.N.(1989). *Crítica e imputação de dados quantitativos utilizando o SAS*. Dissertação de Mestrado, IMPA-RJ.
- [90] Singer, J.M. e Andrade, D.F.(1986). *Análise de dados longitudinais*. VII Simpósio Nacional de Probabilidade e Estatística, Campinas-SP.

- [91] Soler, J.M.P.(1997). *Contribuições ao estudo do modelo de Potthoff e Roy para curvas de crescimento*. Tese de Doutorado, IME-USP.
- [92] Stanek III, E.J.(1990). A two-step method for understanding and fitting growth curve models. *Statistics in Medicine*, **9**, 1491-1513.
- [93] Strang, G.(1988). *Linear algebra and its applications*. Harcourt Brace Jovanovich, Forth Worth, 505 p.
- [94] Tian, J.J, Shukla, R. and Buncher, C.R.(1994). On prediction of future observation in growth curve model. *Statistics in Medicine*, **13**, 2205-2217.
- [95] Tubbs, J.D.; Lewis, T.O. and Duran, B.S.(1975). A note on the analysis of the MANOVA model and its applications to growth curves. *Communications in Statistics*, **4**(7), 643-653.
- [96] von Rosen, D.(1991). The growth curve model: a review. *Communications in Statistics Theory and Methods*, **20**(9) ,2791-2822.
- [97] Vonesh, E.F. and Carter, R.L.(1987). Efficient inference for random coefficient growth curve models with unbalanced data. *Biometrics*, **43**, 617-628.
- [98] Ware, J.M.(1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, **39**, 95-101.
- [99] Wishart, J.(1938). Growth rate determinations in nutrition studies with the bacon pig and their analysis. *Biometrika*, **30**, 16-28.
- [100] Williams, J.S. and Izenman, A.J. (1981). *A class of linear spectral models and analysis of for the study of longitudinal data*. Technical Report, Dep. Statistics, Colorado State University.

- [101] Wolfinger, R.(1993). Covariance structure selection in general mixed models. *Communications in Statistics Simulation and Computation*, **22**(4), 1079-1106.
- [102] Woolson, R.F. and Leeper, J.D.(1980). Growth curve analysis of complete and incomplete longitudinal data. *Communications in Statistics Theory and Methods* A,9(14), 1491-1515.
- [103] Wu, Y. and McGilchrist, C.A.(1995). Recursive computation for repeated measures models with patterned variance matrices. *Journal of Statistical Computation and Simulation*, **51**, 333-341.