

Modelo Linear Misto de Laird-Ware: Predição de Efeitos Aleatórios e Estimação de Parâmetros via Filtro de Kalman

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Paulo Ricardo Bittencourt Guimarães e aprovada pela Comissão Julgadora.

Campinas, 25 de Agosto de 1994.

Clarice Azevedo de Luna Freire
Prof^a. Dr^a. Clarice Azevedo de Luna Freire.

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

Agradecimentos

Gostaria de expressar meus agradecimentos:

- Aos professores do departamento de estatística da UNICAMP pela contribuição à minha formação.
- À professora Clarice, pela orientação, dedicação e amizade, que foram essenciais para a realização deste trabalho.
- Ao professor Jonathan pela amizade, importantes sugestões e ajuda na realização dos programas.
- Ao professor Anselmo que tanto me incentivou.
- À minha querida Lucia pela força e ajuda na correção do Português.
- Aos funcionários da Secretaria de Pós Graduação do IMECC pela colaboração.
- Aos amigos que conquistei durante o mestrado, pelo companheirismo e apoio.
- À UFPR que proporcionou e incentivou o meu licenciamento.
- A CAPES, pelo apoio financeiro.

A meus pais.

Índice

1	Introdução	1
2	Modelo de Laird-Ware: Definições Básicas e Descrição da Metodologia	3
2.1	Introdução	3
2.2	Modelo Geral Linear Misto (Laird-Ware)	7
2.2.1	Representação de alguns planos na forma do modelo de Laird-Ware	9
2.2.2	Construção das matrizes X_i a partir de covariáveis agrupadoras	16
2.2.3	Estruturas de covariâncias	19
2.3	Estimação	25
2.4	Seleção do Modelo	32
2.5	Uma Breve Ilustração	34
3	Os Efeitos Aleatórios do Modelo Laird-Ware: Predição e Adequação do Modelo	37
3.1	Efeitos Aleatórios γ_i	37
3.2	Suposições Distribucionais	42
3.2.1	Análise dos efeitos aleatórios	43
3.2.2	Ajustes finais	48
3.3	Exemplos	50
3.3.1	Potthoff e Roy (1964)	50
3.3.2	Jones (1990)	53
4	O Filtro de Kalman na Estimação e Predição do Modelo Laird-Ware	56
4.1	Introdução	56
4.2	Um Procedimento de Predição Recursiva	58

4.3	Representação em Espaço de Estados e o Filtro de Kalman	63
4.4	Aplicação ao Modelo de Laird-Ware	67
4.4.1	Predição de efeitos aleatórios	68
4.4.2	Estimação de parâmetros	72
5	Aplicações	85
5.1	Introdução	85
5.2	Análise n°1	86
5.3	Análise n° 2	96
5.4	Análise n° 3	99
A	Simetria Composta	107
A.1	Descrição do Modelo	107
A.2	Construção da Estrutura de Covariância	108
A.3	Estimação	109
B	Manual do Programa CAR1	116
B.1	Introdução	116
B.2	O Modelo	117
B.3	Estrutura da Matriz \mathbf{X}_i	118
B.4	Execução do Programa	120
B.5	Exemplo	127

Lista de Figuras

3-1	Gráfico de probabilidade normal ponderado do efeito aleatório	53
3-2	Gráfico de probabilidade normal ponderado do efeito aleatório γ_1	54
3-3	Gráfico de probabilidade normal ponderado do efeito aleatório γ_2	54
3-4	Gráfico de probabilidade normal ponderado do efeito aleatório γ_3	55
5-1	Peso observado dos ratos da dieta 1	87
5-2	Peso observado dos ratos da dieta 2	87
5-3	Peso observado dos ratos da dieta 3	87
5-4	Peso observado dos ratos da dieta 4	88
5-5	Peso observado dos ratos da dieta 5	88
5-6	Retas ajustadas - Modelo 1.1	92
5-7	Modelo final do ganho de peso dos ratos	95
5-8	Gráfico de probabilidade normal ponderado para γ_{i1}	95
5-9	Gráfico de probabilidade normal ponderado para γ_{i2}	96
5-10	Observações das 209 bailarinas	97
5-11	Modelo final ajustado	98
5-12	Gráfico de probabilidade normal ponderado para γ_{i1}	98
5-13	Gráfico de probabilidade normal ponderado para γ_{i2}	99
5-14	Ajuste do modelo 3.2	103
5-15	Modelo final ajustado	104
5-16	Gráfico de probabilidade normal ponderado para γ_{i1}	105
5-17	Gráfico de probabilidade normal ponderado para γ_{i2}	105
5-18	Gráfico de probabilidade normal ponderado para γ	106

Lista de Tabelas

2.1	Estrutura de dados longitudinais	7
3.1	Dados extraídos de Little e Rubin	50
3.2	Efeitos aleatórios estimados para as meninas	51
3.3	Efeitos aleatórios estimados para os meninos	52
5.1	Estimativas, desvios padrão e estatística t dos parâmetros do modelo 1.1	92
5.2	Estimativas, desvios padrão e estatística t dos parâmetros do modelo 1.2	93
5.3	Estimativas, desvios padrão e estatística t dos parâmetros do modelo 1.3	94
5.4	Estimativas, desvios padrão e estatística t dos parâmetros do modelo 1.4	94
5.5	Estimativas, desvios padrão e estatística t dos parâmetros do modelo 2.1	97
5.6	Estimativas, desvios padrão e estatística t dos parâmetros do modelo 3.1	101
5.7	Estimativas, desvios padrão e estatística t dos parâmetros do modelo 3.2	102
5.8	Estimativas, desvios padrão e estatística t dos parâmetros do modelo 3.3	103

Capítulo 1

Introdução

Atualmente, muitos pesquisadores têm realizado experimentos envolvendo dados longitudinais para obter alguma informação sobre o comportamento de determinados fenômenos, principalmente fenômenos das Ciências Biológicas.

O incremento no uso de técnicas estatísticas apropriadas para a análise de dados longitudinais ocorreu graças aos avanços computacionais, que proporcionaram a resolução de problemas freqüentes entre estes experimentos. Estes problemas são causados, principalmente, pela estrutura de dados, que permite um espaçamento desigual entre os tempos de observações; outro problema, no início, de difícil resolução pela carência de instrumentos apropriados, era a modelagem da possível correlação serial entre as observações. E, talvez, o problema mais grave era a possibilidade de presença de dados faltantes, que inviabilizava o uso das técnicas clássicas, que exigiam dados completos.

Aliando os avanços computacionais ao desenvolvimento de um modelo adaptável a dados longitudinais, o modelo de Laird-Ware, o acesso às técnicas foi facilitado. Isto se deve, principalmente, à grande flexibilidade do modelo de Laird-Ware, para tratar os mais variados tipos de problemas que surgem em dados longitudinais. O processo de estimação de parâmetros, no modelo de Laird-Ware, se dá por máxima verossimilhança. Para solucionar o problema de se trabalhar com matrizes de grandes dimensões para estimação dos parâmetros, adaptou-se um processo recursivo conhecido como Filtro de Kalman, que se utiliza sempre de matrizes pequenas, independente do tamanho do conjunto de dados, o que torna o processo recursivo simples e eficiente, sem perder qualidade ao se defrontar

com os problemas de dados longitudinais citados acima. O processo recursivo discutido neste trabalho é o processo desenvolvido por Jones (1993), fazendo uso do Filtro de Kalman e adaptado para realizar a predição da parte aleatória e a estimação da parte fixa do modelo misto de Laird-Ware.

Apesar de todos esses fatores que contribuíram amplamente para o desenvolvimento das técnicas de análise de dados longitudinais, pouco ainda tem se estudado sobre a verificação das suposições distribucionais do modelo ajustado, fator de vital importância para a aplicação de testes que requerem normalidade. Destacaremos, neste trabalho, além dos fatores citados acima, uma técnica de análise dos efeitos aleatórios do modelo de Laird-Ware, desenvolvida a partir da adaptação de uma técnica clássica, o gráfico de probabilidades normais. Como veremos neste trabalho, o desenvolvimento das técnicas para dados longitudinais foi possível graças à adaptação de técnicas já existentes nas áreas de Modelos Lineares de Regressão e Séries Temporais.

O presente trabalho tem como objetivo principal fornecer um estudo detalhado da utilização do Modelo de Laird-Ware em Dados Longitudinais. É proposto, também, um procedimento de análise de conjuntos de dados longitudinais, que envolve as seguintes etapas: ilustração gráfica das observações, escolha de possíveis modelos (com e sem correlação serial), estimação de parâmetros, redução do modelo, verificação da adequação das suposições distribucionais, e testes dos contrastes.

No capítulo 2, definiremos o modelo de Laird-Ware e desenvolveremos as técnicas associadas a ele. No capítulo 3, discutiremos a predição da parte aleatória do modelo misto de Laird-Ware, e uma técnica de verificação das suposições distribucionais do modelo. No capítulo 4, definiremos o processo recursivo do Filtro de Kalman e a adaptação deste processo, para utilização em dados longitudinais, enfocando o método desenvolvido por Jones (1993). No capítulo 5, aplicaremos os métodos descritos nos capítulos anteriores a alguns conjuntos de dados.

Capítulo 2

Modelo de Laird-Ware: Definições Básicas e Descrição da Metodologia

2.1 Introdução

Nos tempos atuais, principalmente nas ciências biológicas, são realizados inúmeros estudos onde unidades experimentais ou amostrais, classificadas em diferentes sub-populações (grupos), são observadas ao longo do tempo. O tempo é chamado de condição de avaliação, ou seja, cada observação é relacionada a um ponto de tempo. Pode ocorrer que essas unidades sejam observadas em outras condições de avaliação, não necessariamente o tempo, tais como peso, dosagens, dietas, idades, distância. Ilustremos esses casos com os seguintes exemplos:

a) em um estudo para verificar a melhor condição para o abate de suínos, observa-se a quantidade de colesterol no sangue de alguns animais ao longo da sua progressão de peso;

b) em um estudo para verificar a dosagem mais eficaz de uma determinada droga, mede-se a pressão sanguínea de algumas cobaias à medida que se aumenta a dosagem desta droga;

c) para compararmos a eficiência de algumas dietas, observa-se o peso de cada uma das unidades, ao longo de todas as dietas em estudo;

d) considere que estejamos interessados em estudar o comportamento da função renal

em pacientes hipertensos. Então podemos medir, em algumas pessoas selecionadas de uma população de hipertensos, a quantidade de uma determinada substância indicadora da função do rim ao longo das idades;

e) imagine que estejamos interessados em observar o comportamento do batimento cardíaco de um grupo de atletas durante uma prova de maratona. Então medimos, a cada quilômetro percorrido por um atleta, o seu batimento cardíaco, até o final da prova;

f) e, finalmente, para ilustrar o caso onde a condição de avaliação é o tempo, considere um estudo em que estejamos interessados em estudar o comportamento da pressão sanguínea de mulheres grávidas durante as horas do dia. Então, de um grupo de gestantes, em repouso numa clínica, medimos a pressão sanguínea de cada uma, a cada 3 horas em um certo dia.

Estas unidades, a que nos referimos acima, podem ser unidades experimentais ou unidades amostrais (ou observacionais). As primeiras são selecionadas da população em estudo e, então, tratamentos de interesse são alocados aleatoriamente a estas unidades. As unidades amostrais são selecionadas de uma população de unidades disponíveis com alguma característica em comum, como por exemplo, a população de pacientes de uma clínica em tratamento de AIDS, a população de uma lista de clientes de uma loja, etc. Com unidades experimentais realizamos um experimento, enquanto que com unidades amostrais realizamos um estudo observacional. Neste trabalho nos referiremos às unidades amostrais ou experimentais apenas como unidades, às condições de avaliação apenas como tempos, e nos referiremos a um experimento ou estudo observacional apenas como estudo.

Num estudo envolvendo dados longitudinais, unidades possivelmente classificadas em diferentes subpopulações, segundo características inerentes às unidades ou segundo o tratamento que recebem, são observadas ao longo do tempo com os seguintes principais objetivos:

- Encontrar uma 'curva' que explique satisfatoriamente o fenômeno em estudo, ou seja, encontrar uma modelagem razoável do fenômeno;
- Realizar, a partir do modelo ajustado, previsões para unidades populacionais em

algum ponto de tempo, do fenômeno em estudo;

- Verificar se existem diferenças entre os modelos nas diversas subpopulações, quando for o caso.

A análise de dados longitudinais clássica é baseada em planos onde todas as unidades são observadas nos mesmos e em todos os pontos de tempo planejados, ou seja, planos balanceados. Atualmente não precisamos nos prender a estas restrições, pois os métodos estão suficientemente desenvolvidos e as facilidades computacionais nos permitem estruturar um estudo de uma maneira mais conveniente, em termos práticos e econômicos. Os métodos de análise de dados longitudinais têm sido bastante utilizados e conseqüentemente aprimorados para serem adaptados a conjuntos de dados com números diferentes de observações por unidade, e até mesmo diferentes pontos no tempo por unidade, pois isto possibilita uma maior flexibilidade na estruturação do estudo, ou um maior aproveitamento das informações disponíveis.

Podem ocorrer, no decorrer de um estudo longitudinal, perdas de algumas observações de uma determinada unidade, ou seja, por alguma razão, unidades não serem observadas em certos tempos pré-determinados. Estas perdas ocorrem com relativa freqüência, seja por razões aleatórias ou por razão associada à subpopulação. Muitos autores, como Laird (1989) e Jones (1993), ressaltam a importância de se investigar a modalidade de cada perda. No primeiro caso incluem-se, por exemplo, situações em que alguns pacientes deixam de comparecer a sessões de avaliação num ensaio clínico, por alguma razão alheia ao estudo, como falta de tempo ou esquecimento. No segundo caso, usando o mesmo exemplo, pode acontecer que certos pacientes não puderam ser avaliados em determinado momento por terem sofrido alguma 'reação estranha', em decorrência do uso de uma droga utilizada no estudo. Quando a causa da perda das observações tiver sido por motivos alheios ao estudo, pode-se tentar o prosseguimento do estudo, utilizando-se as observações disponíveis de cada unidade. A ocorrência de observações perdidas por razões aleatórias causa problemas na aplicação direta de uma técnica clássica, que exija observações de todas as unidades nos mesmos pontos de tempo, ou seja, dados

balanceados. Frequentemente, neste caso, elimina-se a unidade que apresentou perda de observações. Existem, no entanto, métodos que se utilizam de algum algoritmo de estimação das observações perdidas, como o algoritmo EM (ver Laird & Ware-1982). Quando a causa da perda estiver relacionada com o efeito do tratamento, não podemos, em hipótese alguma, desprezar estas perdas, pois elas podem trazer informações muito importantes para o estudo, além do que, o fato de ignorá-las pode invalidar os resultados.

É natural, em estudos longitudinais, que as observações de uma mesma unidade sejam correlacionadas através do tempo, ou seja, apresentem correlação serial. Neste caso, uma certa observação será influenciada pelas observações anteriores, sendo esta influência cada vez menor à medida que se aumenta o intervalo de tempo entre as observações. Com relação a este fenômeno, Jones (1993) explica: *'Se observações são tomadas em intervalos de tempo muito pequenos, é mais provável que uma observação esteja acima do nível médio de uma unidade, quando as observações anteriores estão acima deste mesmo nível médio'*. Como a correlação serial é parte da estrutura do fenômeno, ela deve ser modelada.

A estrutura básica dos conjuntos de dados longitudinais pode ser representada através da matriz de dados indicada na tabela 2.1. Nesta tabela, k é o número de grupos, m_l se refere ao número de unidades no grupo l , n_{il} é o número de observações da unidade i no grupo l , y_{ilj} é a j -ésima observação da unidade i no grupo l , onde $i = 1, \dots, m_l$, $l = 1, \dots, k$, $j = 1, \dots, n_{il}$. Neste trabalho, por simplicidade, usaremos apenas o índice i para nos referirmos a uma determinada unidade. Apesar da representação retangular da tabela, lembramos que os números de unidades por grupos e os números de observações por unidades podem ser diferentes.

Na seção 2.2 introduziremos a estrutura do modelo geral linear misto de Laird-Ware, que será usado neste trabalho, destacando a possibilidade de se trabalhar com observações perdidas e podendo também fazer uso de várias estruturas de correlação serial. Além disso, descreveremos a representação de diferentes tipos de planos de estudos abrangidos pelo modelo de Laird-Ware. Na seção 2.3 desenvolveremos as etapas de estimação dos

Subpopulação	Unidade	Observações				
		1	2	...	n_{i1}	
1	1	y_{111}	y_{112}	...	$y_{11n_{11}}$	
1	2	y_{211}	y_{212}	...	$y_{21n_{21}}$	
:	:	:	:		:	
1	m_1	y_{m_111}	y_{m_112}	...	$y_{m_11n_{m_11}}$	
2	1	y_{121}	y_{122}	...	$y_{12n_{12}}$	
2	2	y_{221}	y_{222}	...	$y_{22n_{22}}$	
:	:	:	:		:	
2	m_2	y_{m_221}	y_{m_222}	...	$y_{m_22n_{m_22}}$	
:	:	:	:		:	
k	1	y_{1k1}	y_{1k2}	...	$y_{1kn_{1k}}$	
k	2	y_{2k1}	y_{2k2}	...	$y_{2kn_{2k}}$	
:	:	:	:		:	
k	m_k	y_{m_kk1}	y_{m_kk2}	...	$y_{m_kkn_{m_kk}}$	

Tabela 2.1: Estrutura de dados longitudinais

parâmetros do modelo de Laird-Ware, derivando os estimadores de máxima verossimilhança. Na seção 2.4 descreveremos duas formas de selecionar o modelo mais adequado para um certo conjunto de dados: o teste da razão de verossimilhança e a medida de seleção AIC ('Akaike's Information Criterion'). Veremos também, nesta seção, como realizar testes de hipóteses para redução do modelo e comparação dos tratamentos. E finalmente, na seção 2.5, daremos uma breve ilustração de um estudo observacional com dados longitudinais, totalmente não-balanceado.

2.2 Modelo Geral Linear Misto (Laird-Ware)

Agora apresentaremos um modelo proposto por Laird e Ware (1982) a partir do trabalho de Harville (1977), que engloba uma variedade de modelos com efeitos fixos e aleatórios, e ainda, com variedade de estruturas de correlação dos resíduos.

O modelo de Laird-Ware, definido para uma unidade genérica, indexada por i , é dado por:

$$\tilde{y}_i = \mathbf{X}_i \tilde{\beta} + \mathbf{Z}_i \tilde{\gamma}_i + \tilde{\epsilon}_i, \quad (2.1)$$

onde

* \tilde{y}_i é um vetor coluna, $n_i \times 1$, das n_i observações tomadas da unidade i ao longo do tempo,

$$\tilde{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix};$$

* $\tilde{\beta}$ é um vetor $b \times 1$ de parâmetros fixos desconhecidos, onde a dimensão b é fixa para qualquer unidade,

$$\tilde{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_b \end{bmatrix};$$

* \mathbf{X}_i , $n_i \times b$, é uma matriz de desenho, que faz a seleção de elementos do vetor $\tilde{\beta}$;

* $\tilde{\gamma}_i$ é um vetor $g \times 1$ de efeitos aleatórios, que representam alguma característica, particular à unidade i , onde a dimensão g é fixa para qualquer unidade. Estes efeitos aleatórios são assumidos independentes entre unidades diferentes e com distribuição normal, ou seja, $\tilde{\gamma}_i \sim iN(0, \sigma^2 \mathbf{B})$;

* \mathbf{Z}_i , $n_i \times g$, é uma matriz de desenho, que faz a seleção dos elementos do vetor $\tilde{\gamma}_i$;

* $\tilde{\epsilon}_i$ é um vetor aleatório, $n_i \times 1$, representando a falta de ajuste, assumido independente entre diferentes unidades, normalmente distribuído, ou seja, $\tilde{\epsilon}_i \sim N(0, \sigma^2 \mathbf{W}_i)$, e

ainda, independente do vetor de efeitos aleatórios, γ_i ,

$$\tilde{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_i} \end{bmatrix} .$$

Observa-se que no modelo de Laird-Ware cada unidade tem o seu próprio modelo, possibilitando que diferentes unidades possam ter diferentes números de observações, assim como diferentes tempos de observação. No entanto, o vetor de parâmetros $\tilde{\beta}$ está presente nos modelos de todas as unidades, e os vetores de efeitos aleatórios $\tilde{\gamma}_i$ tem distribuição de probabilidade comum. Estes são pontos chaves do modelo de Laird-Ware, que promovem a conexão dos modelos individuais. Convém chamar a atenção para o fato de que precisa existir uma certa coerência entre as matrizes \mathbf{X}_i e \mathbf{Z}_i , como será exemplificado na próxima seção.

2.2.1 Representação de alguns planos na forma do modelo de Laird-Ware

Nesta seção iremos apresentar exemplos na forma do modelo de Laird-Ware (2.1), alguns destes representando os planos mais frequentes na literatura. O enfoque será dado às definições das matrizes \mathbf{X}_i e \mathbf{Z}_i , à coerência entre as duas e à interpretação dos modelos. Nada será comentado sobre a estrutura de covariâncias entre observações, assunto este a ser detalhado em seção seguinte.

1) Modelo Saturado ou Curva de Crescimento

Ilustremos este caso a partir de um exemplo dado por Potthoff e Roy (1964), onde se deseja modelar o comportamento de alguma variável, y , para meninos e para meninas, nas idades de 8, 10, 12 e 14 anos. Poderíamos modelar as observações de um menino da seguinte forma:

$$\tilde{y}_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \vdots \\ \beta_8 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \gamma_i + \epsilon_i \quad ,$$

e as observações de uma menina,

$$\tilde{y}_i = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \vdots \\ \beta_8 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \gamma_i + \epsilon_i \quad .$$

Assim, cada grupo tem uma curva de crescimento diferente e arbitrária. A curva de crescimento esperada, nas idades 8, 10, 12 e 14, para os meninos, é dada pelos parâmetros β_1 a β_4 enquanto que, para as meninas, é dada pelos parâmetros β_5 a β_8 . O vetor de efeitos aleatórios γ_i proporciona um incremento vertical às respectivas curvas populacionais, para cada unidade. Outra possível parametrização para este mesmo conjunto de dados é a seguinte:

$$\tilde{y}_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \vdots \\ \beta_8 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \gamma_i + \epsilon_i$$

para o grupo de meninos e

$$\tilde{y}_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \vdots \\ \beta_8 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \gamma_i + \epsilon_i$$

para o grupo de meninas. Agora, β_5 a β_8 são as diferenças entre as duas curvas esperadas, nas idades 8, 10, 12 e 14.

Nas modelagens acima, se a menina ou menino não fosse observado numa das quatro idades, as matrizes \mathbf{X}_i e \mathbf{Z}_i perderiam as linhas correspondentes a esta observação.

2) Modelo Polinomial

Uma outra alternativa é definir a curva esperada como um polinômio. Para o exemplo do item anterior, poderíamos definir uma reta para os meninos,

$$\tilde{y}_i = \begin{bmatrix} 1 & 8 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 1 & 12 & 0 & 0 \\ 1 & 14 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \gamma_i + \tilde{\epsilon}_i \quad ,$$

e outra reta para as meninas,

$$\tilde{y}_i = \begin{bmatrix} 1 & 8 & 1 & 8 \\ 1 & 10 & 1 & 10 \\ 1 & 12 & 1 & 12 \\ 1 & 14 & 1 & 14 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \gamma_i + \tilde{\epsilon}_i \quad ,$$

onde β_{11} e β_{12} são, respectivamente, o intercepto e a inclinação para a reta esperada do grupo dos meninos e $(\beta_{11} + \beta_{21})$ e $(\beta_{12} + \beta_{22})$ são, respectivamente, o intercepto e a inclinação para a reta esperada do grupo das meninas. Para cada unidade haverá um incremento vertical γ_i às retas de seu grupo, ou seja, a reta esperada para uma unidade em particular é a reta esperada de seu grupo, deslocada de uma quantidade γ_i .

Podemos também considerar além de um incremento vertical (incremento no inter-

cepto), um incremento na inclinação da reta populacional. Desta forma teremos

$$\tilde{y}_i = \begin{bmatrix} 1 & 8 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 1 & 12 & 0 & 0 \\ 1 & 14 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{bmatrix} + \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix} \tilde{\gamma}_i + \tilde{\epsilon}_i$$

para a reta esperada dos meninos, e

$$\tilde{y}_i = \begin{bmatrix} 1 & 8 & 1 & 8 \\ 1 & 10 & 1 & 10 \\ 1 & 12 & 1 & 12 \\ 1 & 14 & 1 & 14 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{bmatrix} + \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix} \tilde{\gamma}_i + \tilde{\epsilon}_i$$

para a reta esperada das meninas.

A formulação geral do modelo polinomial para uma certa unidade do primeiro grupo é:

$$\tilde{y}_i = \begin{bmatrix} 1 & t_{i1} & \dots & t_{i1}^{p-1} & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & t_{in_i} & \dots & t_{in_i}^{p-1} & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{1p} \\ \beta_{21} \\ \vdots \\ \beta_{2p} \\ \vdots \\ \beta_{k1} \\ \vdots \\ \beta_{kp} \end{bmatrix} + \begin{bmatrix} 1 & t_{i1} & \dots & t_{i1}^{g-1} \\ \vdots & \vdots & & \vdots \\ 1 & t_{in_i} & \dots & t_{in_i}^{g-1} \end{bmatrix} \begin{bmatrix} \gamma_{i1} \\ \gamma_{i2} \\ \vdots \\ \gamma_{ig} \end{bmatrix} + \tilde{\epsilon}_i, \quad (2.2)$$

onde p é a ordem do polinômio da parte fixa do modelo, ou seja, o polinômio é composto de p termos, sendo completo e de grau $p - 1$. Para uma unidade, do j -ésimo grupo, ($j = 2, \dots, k$), a formulação geral do modelo polinomial será a mesma, havendo apenas uma modificação na matriz \mathbf{X}_i :

$$\mathbf{X}_i = \begin{bmatrix} \underbrace{1 \quad t_{i1} \quad \dots \quad t_{i1}^{p-1}}_{1^\circ \text{ bloco}} \quad 0 \quad \dots \quad 0 & \underbrace{1 \quad t_{i1} \quad \dots \quad t_{i1}^{p-1}}_{j - \text{ésimo bloco}} \quad 0 \quad \dots \quad 0 \\ \vdots \quad \vdots \quad \quad \quad \vdots \quad \vdots \quad \quad \quad \vdots \quad \vdots \quad \quad \quad \vdots \quad \vdots \quad \quad \quad \vdots \\ 1 \quad t_{in_i} \quad \dots \quad t_{in_i}^{p-1} \quad 0 \quad \dots \quad 0 & 1 \quad t_{in_i} \quad \dots \quad t_{in_i}^{p-1} \quad 0 \quad \dots \quad 0 \end{bmatrix} .$$

Daí se determina a dimensão do vetor β : haverá p parâmetros para cada um dos k grupos, onde os parâmetros dos grupos $2, 3, \dots, k$ serão diferenças em relação aos parâmetros do grupo 1. A quantidade g é o número de efeitos aleatórios. Note que g , onde $g \leq p$, é também a ordem do polinômio que define a combinação dos efeitos aleatórios γ_i .

3) Misto

No plano misto, diferentes unidades são observadas em diferentes subconjuntos de pontos de tempo para os quais o estudo foi globalmente dimensionado. Considere, por exemplo, um ensaio clínico onde o objetivo é avaliar o efeito de uma droga até 6 semanas depois da ministração desta droga. Um conjunto de pacientes é observado nas semanas 1, 2 e 3, outro conjunto nas semanas 3, 4 e 5 e outro nas semanas 4, 5 e 6. Um possível modelo a ser considerado para este estudo terá a seguinte forma:

$$\tilde{y}_2 = \begin{bmatrix} 1 & 1 & \dots & 1^{p-1} \\ 1 & 2 & & 2^{p-1} \\ 1 & 3 & \dots & 3^{p-1} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \gamma_{i1} \\ \gamma_{i2} \end{bmatrix} + \tilde{\epsilon}_i$$

para o paciente 2 observado nas semanas 1,2 e 3,

$$\tilde{y}_5 = \begin{bmatrix} 1 & 3 & \dots & 3^{p-1} \\ 1 & 4 & & 4^{p-1} \\ 1 & 5 & \dots & 5^{p-1} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \gamma_{i1} \\ \gamma_{i2} \end{bmatrix} + \epsilon_i$$

para o paciente 5 observado nas semanas 3, 4 e 5, e

$$\tilde{y}_9 = \begin{bmatrix} 1 & 4 & \dots & 4^{p-1} \\ 1 & 5 & & 5^{p-1} \\ 1 & 6 & \dots & 6^{p-1} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} \gamma_{i1} \\ \gamma_{i2} \end{bmatrix} + \epsilon_i$$

para o paciente 9 observado nas semanas 4, 5 e 6 após a ministração da droga. Desta forma, uma curva polinomial explicará o padrão de comportamento dos pacientes até 6 semanas após a ministração da droga. Note, ainda, que no modelo de cada unidade foi definido um vetor de efeitos aleatórios bivariado. A parcela $\mathbf{Z}_i \tilde{\gamma}_i$ determina um incremento aos dois primeiros parâmetros do vetor $\tilde{\beta}$ (β_1 e β_2). Este plano misto é o modelo polinomial, definido anteriormente, com a particularidade de cada unidade ser observada em subconjuntos dos pontos de tempo.

4) Medidas Repetidas

Neste plano, originalmente, deseja-se aplicar k tratamentos a todas as unidades, cuidando para que a ordem de aplicação desses tratamentos seja aleatorizada para cada unidade. Variando a ordem de aplicação de unidade para unidade, o efeito da ordenação dos tratamentos é minimizado. É importante também que exista, entre as aplicações, um intervalo de tempo suficiente para evitar que um tratamento influencie o resultado em um tratamento seguinte. Assim, podemos modelar as respostas de uma unidade como:

$$\tilde{y}_i = \mathbf{I}_k \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \gamma_i + \tilde{\epsilon}_i \quad ,$$

onde \mathbf{I}_k é matriz identidade de ordem k . Note agora que a matriz \mathbf{X}_i (igual a \mathbf{I}_k), pré-multiplicada ao vetor $\tilde{\beta}$, fornecerá um vetor com as respostas esperadas, referentes aos k tratamentos. O vetor que tem a função de ‘personalizar’ estas médias para a unidade i , em forma de incremento, é o vetor $\mathbf{1} \gamma_i$.

5) Corte Transversal (‘Cross-Sectional’)

Em um plano de corte transversal se planeja observar cada unidade numa única ocasião. É o planejamento mais simples e tem o especial atrativo de não requerer esforços no sentido de acompanhar as unidades ao longo do tempo, o que o torna mais econômico. Este plano, pela sua simplicidade, poderia ser analisado através de métodos amplamente divulgados na literatura, como análise de regressão ou análise de variância. Podemos representá-lo da seguinte forma:

$$y_i = \begin{bmatrix} 1 & t_i & \dots & t_i^{p-1} & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{1p} \\ \beta_{21} \\ \vdots \\ \beta_{2p} \\ \vdots \\ \beta_{k1} \\ \vdots \\ \beta_{kp} \end{bmatrix} + \begin{bmatrix} 1 & t_i & \dots & t_i^{g-1} \end{bmatrix} \begin{bmatrix} \gamma_{i1} \\ \gamma_{i2} \\ \vdots \\ \gamma_{ig} \end{bmatrix} + \epsilon_i \quad ,$$

onde $g \leq p$. Esta representação corresponde a um estudo onde as unidades são divididas

em k grupos. A unidade i , descrita acima, faz parte do primeiro grupo. Note que a matriz \mathbf{X}_i terá apenas uma linha e definirá a combinação de parâmetros fixos que irá explicar a única observação da unidade i , y_i , tomada no tempo t . Para melhor explicar esta observação y_i , acrescentamos uma combinação de efeitos aleatórios $\mathbf{Z}_i \tilde{\gamma}_i$ que irá conter informação inerente à unidade i .

Em geral, para um indivíduo do j -ésimo grupo o modelo será idêntico, exceto que a matriz (linha) \mathbf{X}_i terá o j -ésimo bloco de p elementos não nulos ($j = 2, \dots, k$):

$$\mathbf{X}_i = \left[\begin{array}{cccccccccccc} \underbrace{1^\circ \text{bloco}} & & & & & & \underbrace{j - \text{ésimo bloco}} & & & & & \\ 1 & t_i & \dots & t_i^{p-1} & 0 & \dots & 0 & 1 & t_i & \dots & t_i^{p-1} & 0 & \dots & 0 \end{array} \right] .$$

$\beta_{j1}, \dots, \beta_{jp}$ são os incrementos aos coeficientes do polinômio do grupo 1, para a definição dos coeficientes do polinômio do grupo j , para $j = 2, 3, \dots, k$.

2.2.2 Construção das matrizes \mathbf{X}_i a partir de covariáveis agrupadoras

Consideremos o modelo de Laird-Ware polinomial, na presença de vários grupos, quando os elementos de $\tilde{\beta}$ são interpretados como a diferença em relação a um grupo padrão. Em muitos casos, a matriz \mathbf{X}_i tem várias colunas de vetores nulos, ou colunas de potências dos tempos de observações. O programa computacional CAR1, desenvolvido por Jones e Boadi-Boateng (1992), especialmente para análise de dados longitudinais, foi utilizado neste trabalho. Este programa utiliza variáveis denominadas ‘covariáveis agrupadoras’, cuja única função é a construção das matrizes \mathbf{X}_i . Apesar de os detalhes sobre o programa se encontrarem em forma de manual, no apêndice, nesta seção é destacada a definição de covariáveis agrupadoras .

As covariáveis agrupadoras são associadas a variáveis indicadoras dicotômicas (assumem valor 0 ou 1), e têm a função de definir o grupo a que pertence uma certa unidade, indicando a presença ou não de determinado tratamento ou característica. Através das covariáveis agrupadoras e de uma matriz que chamaremos de matriz básica, será construída a matriz \mathbf{X}_i . Esta matriz básica será denotada por \mathbf{X}_y :

$$\mathbf{X}_y = \begin{bmatrix} 1 & t_{y1} & t_{y1}^2 & \cdots & t_{y1}^{p-1} \\ \vdots & t_{y2} & t_{y2}^2 & \cdots & t_{y2}^{p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_{yn_y} & t_{yn_y}^2 & \cdots & t_{yn_y}^{p-1} \end{bmatrix} \quad (2.3)$$

Esta matriz caracteriza um polinômio de ordem p , ou seja, completo de grau $p - 1$. Ao final do processo, a matriz \mathbf{X}_i terá as colunas de \mathbf{X}_y aumentada de mais colunas conforme explicado a seguir.

Se tivermos k grupos, $c = k - 1$ será o número de covariáveis agrupadoras. A covariável j é associada a um determinado valor p_j , onde $p_j \leq p$. Assim como p define a ordem do polinômio da matriz básica e também do primeiro grupo, p_1, p_2, \dots, p_c definem a ordem do polinômio para explicar os demais grupos. Desta forma, podemos definir polinômios de ordens diferentes para cada tratamento. Além disso, p_j é o número de colunas da matriz básica que serão multiplicadas pela covariável j para construir a matriz \mathbf{X}_i .

A unidade i tem associada a si, um valor para cada uma das c covariáveis. Para a primeira covariável multiplica-se o valor que esta assume (0 ou 1) pelas p_1 primeiras colunas de \mathbf{X}_y . Para a segunda covariável multiplica-se o valor que esta assume pelas p_2 primeiras colunas de \mathbf{X}_y , e assim por diante até a c -ésima covariável. Portanto, a matriz \mathbf{X}_i será formada pela matriz \mathbf{X}_y aumentada pelas $p_1 + p_2 + \dots + p_c$ colunas resultantes das multiplicações descritas acima.

Consideremos um exemplo, onde um modelo define uma reta para cada tratamento, em um estudo envolvendo 3 tratamentos. Teremos, portanto, com 3 tratamentos, 2 variáveis agrupadoras, sempre fazendo referência aos dois últimos tratamentos. Para as unidades relacionadas ao tratamento 1, os valores das duas variáveis agrupadoras serão $v_1 = 0$ e $v_2 = 0$, indicando a ausência do tratamento 2 e do tratamento 3. Para as unidades relacionadas ao tratamento 2, teremos $v_1 = 1$ e $v_2 = 0$, indicando a presença do tratamento 2 e ausência do tratamento 3. Para as unidades relacionadas ao tratamento 3, $v_1 = 0$ e $v_2 = 1$, indicando ausência do tratamento 2 e presença do tratamento 3.

Como foi definido anteriormente, $p = 2$, e, ainda, $p_1 = 2$ e $p_2 = 2$. Portanto, a matriz básica será:

$$\mathbf{X}_y = \begin{bmatrix} 1 & t_{y1} \\ 1 & t_{y2} \\ \vdots & \vdots \\ 1 & t_{yn_y} \end{bmatrix}$$

Esta escolha das variáveis agrupadoras, conjuntamente com a matriz básica, implica na definição de uma reta para cada tratamento. Assim, teremos para a unidade 3 relacionada ao tratamento 1, com 4 observações no 1º, 3º, 7º, e 23º dias do estudo, matriz de desenho, ($v_1 = 0$ e $v_2 = 0$),

$$\mathbf{X}_3 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 & 0 \\ 1 & 7 & 0 & 0 & 0 & 0 \\ 1 & 23 & 0 & 0 & 0 & 0 \end{bmatrix} ;$$

para a unidade 8, relacionada ao tratamento 2, observada apenas no 8º dia do estudo, ($v_1 = 1$ e $v_2 = 0$),

$$\mathbf{X}_8 = \begin{bmatrix} 1 & 8 & 1 & 8 & 0 & 0 \end{bmatrix} ,$$

e finalmente para a unidade 17 relacionada ao tratamento 3, com 6 observações no 8º, 9º, 10º, 13º, 23º, e 25º dias do estudo, ($v_1 = 0$ e $v_2 = 1$),

$$\mathbf{X}_{17} = \begin{bmatrix} 1 & 8 & 0 & 0 & 1 & 8 \\ 1 & 9 & 0 & 0 & 1 & 9 \\ 1 & 10 & 0 & 0 & 1 & 10 \\ 1 & 13 & 0 & 0 & 1 & 13 \\ 1 & 23 & 0 & 0 & 1 & 23 \\ 1 & 25 & 0 & 0 & 1 & 25 \end{bmatrix} .$$

Desta forma, o vetor de parâmetros $\tilde{\beta}$ será o seguinte:

$$\tilde{\beta} = \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \\ \beta_{31} \\ \beta_{32} \end{bmatrix}, \quad (2.4)$$

onde, β_{11} e β_{12} são o intercepto e o coeficiente angular (inclinação), respectivamente, para a reta do tratamento 1, $(\beta_{11} + \beta_{21})$ e $(\beta_{12} + \beta_{22})$ são o intercepto e a inclinação para a reta do tratamento 2 e $(\beta_{11} + \beta_{31})$ e $(\beta_{12} + \beta_{32})$ são o intercepto e a inclinação para a reta do tratamento 3.

2.2.3 Estruturas de covariâncias

O modelo misto de Laird-Ware expressa o vetor de observações de uma unidade, \tilde{y}_i , como a soma de três parcelas, sendo duas destas aleatórias e independentes:

$$\tilde{\mathbf{Z}}_i \tilde{\gamma}_i + \tilde{\epsilon}_i, \quad ,$$

onde

$$\tilde{\gamma}_i \sim N(0, \sigma^2 \mathbf{B}) \quad e \quad \tilde{\epsilon}_i \sim N(0, \sigma^2 \mathbf{W}_i) \quad ,$$

e, portanto,

$$Var(\tilde{y}_i) = \sigma^2 (\tilde{\mathbf{Z}}_i \mathbf{B} \tilde{\mathbf{Z}}_i' + \mathbf{W}_i) = \sigma^2 \mathbf{V}_i \quad . \quad (2.5)$$

Esta estrutura de covariâncias evidencia duas fontes de variabilidade, \mathbf{B} e \mathbf{W}_i . Enunciarremos aqui algumas das estruturas mais utilizadas na modelagem da variância de dados longitudinais. As estruturas abaixo pertencem à i -ésima unidade de um grupo qualquer.

Denotemos por q o número de parâmetros na matriz V_i .

1) Independente ($q = 0$)

É a forma mais simples para a matriz V_i . Assumimos que as observações de uma mesma unidade são independentes entre si, com variância constante. Na estrutura de covariâncias temos apenas um parâmetro, a variância do resíduo, σ^2 . Não assumimos a existência de efeitos aleatórios γ_i . Portanto, as partes da estrutura de covariância serão definidas da seguinte forma:

$$\begin{aligned} W_i &= I_i \\ B &= 0 \\ \Rightarrow V_i &= I_i \end{aligned}$$

onde I_i é uma matriz identidade $n_i \times n_i$. Note que este modelo não é mais o modelo misto, pois não estamos considerando a existência de efeitos aleatórios γ_i . Este é o modelo de Gauss-Markov, com suposição de normalidade, para uma unidade,

$$y_i = X_i \beta + \epsilon_i \quad , \quad \epsilon_i \sim N(0, \sigma^2 I) \quad , \quad (2.6)$$

quando poderíamos ainda expressar o conjunto de todas as observações segundo o modelo de Gauss-Markov, como veremos na seção seguinte.

2) Simetria Composta ($q = 1$)

Esta estrutura está destacada no apêndice 1. Ela apresenta covariância constante entre quaisquer observações de uma mesma unidade, devido a erros independentes e um efeito aleatório, unidimensional ($g = 1$). Esta estrutura surge, por exemplo, em modelos de curvas de crescimento onde as curvas esperadas, para indivíduos específicos, apresentam um desvio vertical aleatório em relação à curva esperada populacional.

$$\begin{aligned} W_i &= I_i \quad , \\ B &= b \quad , \\ Z_i &= \mathbf{1} \quad , \end{aligned}$$

$$\Rightarrow \mathbf{V}_i = \underset{\sim}{\mathbf{1}} \underset{\sim}{b} \underset{\sim}{\mathbf{1}}' + \mathbf{I}_i = \begin{bmatrix} b & \dots & \dots & b \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ b & \dots & \dots & b \end{bmatrix} + \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} = \begin{bmatrix} b+1 & b & \dots & b \\ b & b+1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \dots & b & b+1 \end{bmatrix}$$

onde b é um escalar, $\sigma^2 b$ é a variância do efeito aleatório γ_i , e $\underset{\sim}{\mathbf{1}}$ é um vetor, $n_i \times 1$, com todos os n_i elementos iguais a um. Os parâmetros σ^2 e b são os únicos da estrutura de covariâncias.

3) Efeitos aleatórios com erros independentes ($q = \frac{1}{2}g(g+1)$)

Nesta estrutura consideramos os erros independentes, sendo a covariância entre as observações de uma unidade definida totalmente pelos efeitos aleatórios entre unidades.

$$\mathbf{W}_i = \mathbf{I}_i \quad ,$$

$$\Rightarrow \mathbf{V}_i = \mathbf{Z}_i \mathbf{B} \mathbf{Z}_i' + \mathbf{I}_i \quad .$$

Os parâmetros na estrutura de covariância, além de σ^2 , são os elementos da matriz \mathbf{B} , de dimensão $g \times g$.

4) Auto-regressiva de 1ª ordem

Aqui não consideramos a existência de efeitos aleatórios (ou seja, este não é um modelo misto). Nesta estrutura de covariâncias o vetor dos resíduos para cada unidade é visto como uma série temporal curta, seguindo um processo estacionário auto-regressivo de 1ª ordem (a covariância entre duas observações decresce à medida que aumenta o intervalo de tempo entre elas). Devemos considerar duas situações especiais:

a) Tempos igualmente espaçados

$$\mathbf{W}_i = \begin{bmatrix} 1 & \rho & \dots & \rho^{n_i-1} \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho^{n_i-1} & \dots & \rho & 1 \end{bmatrix} ,$$

$$\mathbf{B} = \mathbf{0} \quad ,$$

$$\Rightarrow \mathbf{V}_i = \mathbf{W}_i \quad .$$

Aqui, os parâmetros da estrutura de covariâncias são σ^2 e ρ , onde ρ é um coeficiente auto-regressivo, assumindo algum valor entre 0 e 1, e ρ^j define a correlação entre duas observações distantes j unidades de tempo . Logo, de acordo com a matriz \mathbf{W}_i acima, $q = 1$.

Uma situação mais geral seria a definição de estrutura em faixas, onde não houvesse, necessariamente, potências de um valor ρ :

$$\mathbf{W}_i = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{n_i-1} \\ \rho_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_1 \\ \rho_{n_i-1} & \dots & \rho_1 & 1 \end{bmatrix} .$$

Neste caso, o número de parâmetros na matriz \mathbf{W}_i seria igual a $n_i - 1$, e assim q é igual ao número total de tempos (equiespaçados) de observações, menos 1. Esta estrutura matricial, onde cada diagonal da matriz tem elementos iguais, é conhecida como Toeplitz.

b) Tempos desigualmente espaçados

Quando observações são desigualmente espaçadas é necessário considerar um modelo que leve em consideração as diferenças que ocorrem entre os pontos de tempo, ou seja, um modelo de tempo contínuo. Um processo auto-regressivo de 1ª ordem de tempo contínuo

tem função de correlação

$$\rho(\delta t) = \exp\{-\alpha_0 |\delta t|\} \quad , \quad (2.7)$$

onde δt é o intervalo de tempo entre duas observações e α_0 é uma constante positiva.

Nesse caso, a matriz \mathbf{W}_i será dada por

$$\mathbf{W}_i = \begin{bmatrix} 1 & \exp\{-\alpha_0(t_2 - t_1)\} & \cdots & \exp\{-\alpha_0(t_{n_i} - t_1)\} \\ \exp\{-\alpha_0(t_2 - t_1)\} & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ \exp\{-\alpha_0(t_{n_i} - t_1)\} & \cdots & \cdots & 1 \end{bmatrix} \quad ,$$

$$\mathbf{B} = 0 \quad ,$$

$$\Rightarrow \mathbf{V}_i = \mathbf{W}_i \quad ,$$

sendo o número de parâmetros em $\mathbf{V}_i : q = 1$.

Na situação (a), com tempos igualmente espaçados, se por acaso a i -ésima unidade possuir observações perdidas, deveremos reduzir a matriz \mathbf{W}_i , eliminando a linha e a coluna correspondentes ao tempo em que esta observação deveria ter sido colhida. Na situação (b), a matriz \mathbf{W}_i já é apropriada para a correspondência às observações existentes, sejam quais forem os pontos de tempo utilizados.

Existe a possibilidade de incluirmos na matriz \mathbf{W}_i mais um componente de variação, se assumirmos a existência de erros observacionais. Neste caso, será acrescida a cada elemento da diagonal desta matriz a quantidade σ_0^2 , ou melhor, em vez do elemento 1 na diagonal, aparecerá $1 + \sigma_0^2$, e assim teremos mais um parâmetro no modelo. Desta forma estamos assumindo erros observacionais com variância $\sigma^2 \times \sigma_0^2$. A justificativa para a inclusão de erros observacionais é que a estrutura AR(1) não é exata, ou seja, existe uma pequena flutuação em torno de uma representação AR(1).

5) Efeitos aleatórios e resíduos auto-regressivos de 1ª ordem ($q = \frac{1}{2}g(g + 1) + 1$)

$$\begin{aligned} & \mathbf{W}_i \text{ como em (4) } , \\ & \mathbf{B} \text{ sem nenhuma estrutura } , \end{aligned}$$

$$\Rightarrow \mathbf{V}_i = \mathbf{Z}_i \mathbf{B} \mathbf{Z}'_i + \mathbf{W}_i \quad .$$

Na matriz \mathbf{V}_i teremos os $\frac{1}{2}g(g + 1)$ parâmetros da matriz \mathbf{B} e os parâmetros da matriz \mathbf{W}_i .

6) Desestruturada ($q = \frac{1}{2}n(n + 1)$)

Esta estrutura permite a representação de alguns modelos multivariados clássicos, possivelmente, com dados incompletos. O modelo multivariado para medidas repetidas e o modelo multivariado de curvas de crescimento podem ser vistos como casos especiais do modelo de Laird-Ware com \mathbf{V}_i desestruturada. Nestes casos, a princípio, espera-se observar todas as unidades nos mesmos n pontos de tempo. Isto explica a dimensão $n \times n$ da matriz \mathbf{W}_i , ilustrada abaixo, que define uma correlação para cada par de observações tomadas da unidade i ,

$$\mathbf{W}_i = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1n} \\ \theta_{12} & \theta_{22} & \cdots & \theta_{2n} \\ \vdots & \vdots & & \vdots \\ \theta_{1n} & \theta_{2n} & \cdots & \theta_{nn} \end{bmatrix} ,$$

$$\mathbf{B} = 0 \quad ,$$

$$\mathbf{V}_i = \mathbf{W}_i \quad .$$

Se a unidade i sofrer perda de observações, a matriz \mathbf{W}_i será reduzida eliminando as linhas e as colunas correspondentes aos tempos em que estas observações deveriam ter sido colhidas. Quando n é grande e os dados são altamente não balanceados ou incompletos,

esta estrutura será ineficiente devido ao grande número de parâmetros, $q = \frac{1}{2}n(n + 1)$, além de σ^2 .

2.3 Estimação

O enfoque desta seção é a estimação dos parâmetros do modelo de Laird-Ware. Inicialmente, os estimadores de máxima verossimilhança para σ^2 e $\underline{\beta}$ são definidos. Um procedimento (Jones, 1993) para encontrar o máximo da função de verossimilhança é descrito, já que para os demais parâmetros não há uma forma fechada para os seus estimadores de máxima verossimilhança. O estimador de máxima verossimilhança de $\underline{\beta}$ é equivalente ao estimador de mínimos quadrados de $\underline{\beta}$, numa reexpressão do modelo de Laird-Ware como um modelo de Aitken, supondo a estrutura de covariâncias conhecida. Alguns resultados sobre a estimabilidade de combinações lineares do vetor $\underline{\beta}$ serão lembrados, dadas as propriedades dos estimadores de mínimos quadrados.

Conforme foi exposto, o modelo de Laird-Ware define, para o vetor de observações de certa unidade, \underline{y}_i , a decomposição:

$$\underline{y}_i = \mathbf{X}_i \underline{\beta} + \mathbf{Z}_i \underline{\gamma}_i + \underline{\epsilon}_i \quad ,$$

onde

- $\underline{\gamma}_i$ iid $\sim N(0, \sigma^2 \mathbf{B})$;
- $\underline{\epsilon}_i$ iid $\sim N(0, \sigma^2 \mathbf{W}_i)$;
- $\underline{\epsilon}_i$ independente de $\underline{\gamma}_i$.

Desta forma, o vetor de esperanças e a matriz de covariâncias do vetor de observações da unidade i são dados, respectivamente, por:

$$\mathbf{X}_i \underline{\beta} \quad e \quad \sigma^2 \mathbf{V}_i = (\mathbf{Z}_i \mathbf{B} \mathbf{Z}_i' + \mathbf{W}_i) \sigma^2 \quad ,$$

e com isso, o modelo de probabilidade do vetor de observações \underline{y}_i será:

$$\underline{y}_i \sim N(\underline{X}_i \underline{\beta}, \sigma^2 \underline{V}_i) \quad , \quad (2.8)$$

cuja função de densidade é

$$f(\underline{y}_i) = \frac{1}{(2\pi)^{\frac{n_i}{2}} (|\sigma^2 \underline{V}_i|)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\underline{y}_i - \underline{X}_i \underline{\beta})'(\sigma^2 \underline{V}_i)^{-1}(\underline{y}_i - \underline{X}_i \underline{\beta})\right\} \quad . \quad (2.9)$$

A função de verossimilhança é o produto dessas funções sobre todas as unidades, considerando agora, função dos parâmetros:

$$L(\underline{\beta}, \sigma^2, \underline{V}_i) = \prod_i \frac{1}{(2\pi)^{\frac{n_i}{2}} (|\sigma^2 \underline{V}_i|)^{\frac{1}{2}}} \exp\left\{(\underline{y}_i - \underline{X}_i \underline{\beta})'(\sigma^2 \underline{V}_i)^{-1}(\underline{y}_i - \underline{X}_i \underline{\beta})\right\} \quad (2.10)$$

e, realizando a transformação $-2 \ln L$, teremos

$$\ell(\underline{\beta}, \sigma^2, \underline{V}_i) = -2 \ln L = \sum_i \left[n_i \ln(2\pi) + \ln(|\sigma^2 \underline{V}_i|) + (\underline{y}_i - \underline{X}_i \underline{\beta})'(\sigma^2 \underline{V}_i)^{-1}(\underline{y}_i - \underline{X}_i \underline{\beta}) \right] \quad . \quad (2.11)$$

Mas, sabemos que

$$\ln(|\sigma^2 \underline{V}_i|) = \ln(\sigma^{2n_i} |\underline{V}_i|) = n_i \ln(\sigma^2) + \ln(|\underline{V}_i|) \quad (2.12)$$

e fazendo $n = \sum_i n_i$, o número total de observações em todas as unidades, teremos

$$\ell(\underline{\beta}, \sigma^2, \underline{V}_i) = n \ln(\sigma^2) + n \ln(2\pi) + \sum_i \ln(|\underline{V}_i|) + \frac{1}{\sigma^2} \sum_i (\underline{y}_i - \underline{X}_i \underline{\beta})' \underline{V}_i^{-1} (\underline{y}_i - \underline{X}_i \underline{\beta}) \quad . \quad (2.13)$$

Derivando a expressão acima para encontrarmos o estimador da variância residual,

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{1}{(\sigma^2)^2} \sum_i (\underline{y}_i - \underline{X}_i \underline{\beta})' \underline{V}_i^{-1} (\underline{y}_i - \underline{X}_i \underline{\beta}) \quad ,$$

e, igualando a zero,

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (\underline{y}_i - \underline{X}_i \underline{\beta})' \underline{V}_i^{-1} (\underline{y}_i - \underline{X}_i \underline{\beta}) \quad . \quad (2.14)$$

Substituindo σ^2 , em (2.13), pela expressão $\hat{\sigma}^2$, em (2.14), obtemos a função ℓ descrita apenas pelos parâmetros de $\tilde{\beta}$ e de \mathbf{V}_i ,

$$\ell(\tilde{\beta}, \mathbf{V}_i) = n \ln(\hat{\sigma}^2) + n \ln(2\pi) + \sum_i \ln(|\mathbf{V}_i|) + n \quad . \quad (2.15)$$

Os parâmetros desconhecidos em (2.15) são o vetor $\tilde{\beta}$, os elementos da matriz \mathbf{B} e alguns parâmetros que são usados na definição das matrizes \mathbf{W}_i . O valor de $\tilde{\beta}$ que minimiza (2.15) será definido através dos próximos resultados.

Resultado 2.1:

Sejam \tilde{a} um vetor de constantes, \tilde{x} um vetor aleatório e \mathbf{A} uma matriz simétrica, de dimensões compatíveis às operações de produtos exibidas, então

$$\frac{\partial(\tilde{a}'\tilde{x})}{\partial \tilde{x}} = \tilde{a} \quad (2.16)$$

e

$$\frac{\partial(\tilde{x}' \mathbf{A} \tilde{x})}{\partial \tilde{x}} = 2\mathbf{A} \tilde{x} \quad (2.17)$$

□

Resultado 2.2:

Para dados valores dos parâmetros em \mathbf{B} e \mathbf{W}_i , o estimador de $\tilde{\beta}$ que minimiza $\ell(\tilde{\beta}, \mathbf{V}_i)$, em (2.15) (e conseqüentemente maximiza a função de verossimilhança) é a solução do sistema em $\tilde{\beta}$

$$\left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right) \tilde{\beta} = \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \tilde{y}_i \right) \quad ,$$

$$\hat{\tilde{\beta}} = \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \tilde{y}_i \right) \quad , \quad (2.18)$$

assumindo a existência da inversa.

Prova:

A expressão (2.15) é função do vetor $\tilde{\beta}$, apenas na parcela:

$$n \ln(\hat{\sigma}^2) \quad ,$$

onde

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (y_i - \mathbf{X}_i \tilde{\beta})' \mathbf{V}_i^{-1} (y_i - \mathbf{X}_i \tilde{\beta}) \quad .$$

O valor de $\tilde{\beta}$ que minimiza $\hat{\sigma}^2$ é dado pela solução do sistema

$$\frac{\partial \hat{\sigma}^2}{\partial \tilde{\beta}} = 0 \quad . \quad (2.19)$$

Podemos reexpressar $\hat{\sigma}^2$, dada acima, por uma soma:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_i \{ y_i' \mathbf{V}_i^{-1} y_i - (\mathbf{X}_i \tilde{\beta})' \mathbf{V}_i^{-1} y_i - y_i' \mathbf{V}_i^{-1} \mathbf{X}_i \tilde{\beta} + (\mathbf{X}_i \tilde{\beta})' \mathbf{V}_i^{-1} \mathbf{X}_i \tilde{\beta} \} = \\ &= \frac{1}{n} \sum_i \{ y_i' \mathbf{V}_i^{-1} y_i - 2 y_i' \mathbf{V}_i^{-1} \mathbf{X}_i \tilde{\beta} + \tilde{\beta}' \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \tilde{\beta} \} = \\ &= \frac{1}{n} \{ \sum_i y_i' \mathbf{V}_i^{-1} y_i - 2 [\sum_i y_i' \mathbf{V}_i^{-1} \mathbf{X}_i] \tilde{\beta} + \tilde{\beta}' [\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i] \tilde{\beta} \} \quad . \end{aligned}$$

Usando o resultado 1,

$$\frac{\partial \hat{\sigma}^2}{\partial \tilde{\beta}} = \frac{1}{n} \{ -2 \sum_i y_i' \mathbf{V}_i^{-1} \mathbf{X}_i + 2 [\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i] \tilde{\beta} \} \quad .$$

Portanto o sistema (2.19) é:

$$- \sum_i y_i' \mathbf{V}_i^{-1} \mathbf{X}_i + \sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \tilde{\beta} = 0 \quad ,$$

cuja solução é

$$\hat{\tilde{\beta}} = (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} y_i) \quad ,$$

desde que exista a inversa.

□

Note que não é necessário assumir que as matrizes $\mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i$ sejam não singulares, mas, apenas a soma $\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i$.

A variância deste estimador pode ser facilmente derivada:

$$\begin{aligned}
 \text{Var}(\hat{\beta}) &= (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} [\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \text{Var}(\tilde{y}_i) (\mathbf{X}_i' \mathbf{V}_i^{-1})'] (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \\
 &= (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} [\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} (\sigma^2 \mathbf{V}_i) \mathbf{V}_i^{-1} \mathbf{X}_i] (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \\
 &= \sigma^2 (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \\
 &= \sigma^2 (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} .
 \end{aligned} \tag{2.20}$$

Até este ponto, estão definidos estimadores de β e σ^2 , cujas expressões envolvem o conhecimento das matrizes \mathbf{V}_i . A quantidade de parâmetros empregados na definição das matrizes \mathbf{V}_i é pequena para várias estruturas comumente utilizadas, conforme visto na seção anterior. Uma alternativa para a estimação dos parâmetros do modelo de Laird-Ware seria a escolha de vários valores para os parâmetros da matriz \mathbf{V}_i , com imediato cálculo da função de verossimilhança, numa procura do seu máximo. Portanto, um procedimento de estimação dos parâmetros pode ser resumido da seguinte forma:

Escolhemos arbitrariamente valores iniciais para os parâmetros das matrizes \mathbf{B} e \mathbf{W}_i . Com esses valores construímos a matriz \mathbf{V}_i

$$\mathbf{V}_i = \mathbf{Z}_i \mathbf{B} \mathbf{Z}_i' + \mathbf{W}_i \quad ,$$

e calculamos

$$\hat{\beta} = (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} (\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \tilde{y}_i) \quad ,$$

calculamos

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (y_i - \mathbf{X}_i \hat{\beta})' \mathbf{V}_i^{-1} (y_i - \mathbf{X}_i \hat{\beta}) \quad ,$$

e obtemos

$$\ell = n \ln(2\pi \hat{\sigma}^2) + \sum_i \ln(|\mathbf{V}_i|) + n \quad .$$

Efetua-se este procedimento para diversos valores iniciais dos parâmetros das matrizes \mathbf{B} e \mathbf{W}_i , até que ℓ seja minimizada. Então, nesse instante, teremos obtido as estimativas de máxima verossimilhança para β e elementos da estrutura de covariância do modelo de Laird-Ware.

Supondo conhecidas as matrizes \mathbf{V}_i , o estimador de $\hat{\beta}$, (2.18), obtido pelo método de máxima verossimilhança coincide com o estimador de mínimos quadrados generalizados de um vetor de parâmetros do modelo de Aitken, correspondente ao modelo de Laird-Ware. Ou seja, supondo \mathbf{V}_i conhecidas, $i = 1, \dots, m$, onde m é o número total de unidades, o modelo de Laird-Ware corresponde ao seguinte modelo de Aitken:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{bmatrix} \beta + \begin{bmatrix} \mathbf{Z}_1 \gamma_1 + \epsilon_1 \\ \vdots \\ \mathbf{Z}_1 \gamma_m + \epsilon_m \end{bmatrix} \quad , \quad (2.21)$$

onde

$$E \begin{bmatrix} \mathbf{Z}_1 \gamma_1 + \epsilon_1 \\ \vdots \\ \mathbf{Z}_1 \gamma_m + \epsilon_m \end{bmatrix} = \mathbf{0}$$

e

$$V \begin{bmatrix} \mathbf{Z}_1 \gamma_1 + \epsilon_1 \\ \vdots \\ \mathbf{Z}_1 \gamma_m + \epsilon_m \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{V}_m \end{bmatrix} = \sigma^2 \mathbf{V} \quad ,$$

notando que o modelo de Aitken não requer normalidade. O seu sistema de equações normais é:

$$\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{bmatrix}' \mathbf{V}^{-1} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{bmatrix} \underset{\sim}{\beta} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{bmatrix}' \mathbf{V}^{-1} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

ou

$$\left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right) \underset{\sim}{\beta} = \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} y_i \right) \quad ,$$

com solução

$$\underset{\sim}{\hat{\beta}} = \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} y_i \right) \quad (2.22)$$

supondo a existência da inversa.

O estimador (2.22) tem propriedades ótimas, como o de variância mínima entre os estimadores lineares, não viciados, de $\underset{\sim}{\beta}$. Também, a existência da inversa de $\left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)$ garante a estimabilidade de qualquer combinação linear dos elementos de $\underset{\sim}{\beta}$, sendo o correspondente estimador de mínimos quadrados generalizados a mesma combinação linear aplicada ao vetor $\left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} y_i \right)$. No entanto, em muitas situações onde é utilizado o modelo de Aitken a matriz de covariâncias não é conhecida. Segundo Harville (1985), o método usual é estimar essa matriz de alguma forma e proceder como se a mesma fosse conhecida. Neste trabalho, a estimação dos parâmetros relacionados à estrutura de covariâncias é feita pelo método de máxima verossimilhança, a partir da escolha de valores, o que é bem simples, especialmente na presença de poucos parâmetros.

Dessa forma, para qualquer combinação linear dos elementos de $\underset{\sim}{\beta}$, como por exemplo, $\underset{\sim}{x}_i' \underset{\sim}{\beta}$, onde $\underset{\sim}{x}_i'$ é uma linha de alguma matriz \mathbf{X}_i , o estimador é definido por

$$\underset{\sim}{x}_i' \underset{\sim}{\hat{\beta}} = \underset{\sim}{x}_i' \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} y_i \right) \quad , \quad (2.23)$$

com variância estimada

$$\hat{\sigma}^2 \underset{\sim}{x}_i' \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \underset{\sim}{x}_i \quad , \quad (2.24)$$

onde no lugar de V_i usamos os valores da última passagem do algoritmo de maximização da função de verossimilhança.

2.4 Seleção do Modelo

Um dos principais objetivos em análise de dados longitudinais é encontrar um modelo que represente adequadamente um conjunto de dados, entre inúmeros possíveis modelos. Uma técnica para selecionar um entre dois modelos é o teste assintótico da razão de verossimilhança. Neste teste, comparam-se dois modelos estimados por máxima verossimilhança, onde um dos modelos é uma versão restrita do outro, ou seja, um modelo tem r parâmetros adicionais. Este teste irá verificar se a adição de tais parâmetros à versão restrita melhorou significativamente o ajuste. Denominando ℓ_1 o valor de ℓ ($-2\ln$ da função de verossimilhança) para o primeiro modelo e ℓ_2 para o segundo modelo, com r parâmetros extras, o teste da razão de verossimilhança estabelece que, sob a hipótese nula de que os dois modelos são os mesmos (ou seja, os parâmetros extras são iguais a zero, logo, não melhoram o ajuste), a diferença entre os valores ℓ_1 e ℓ_2 é assintoticamente distribuída como uma Qui-Quadrada com r graus de liberdade,

$$\ell_1 - \ell_2 \sim \chi_r^2 \quad . \quad (2.25)$$

Embora este teste seja bastante eficaz, tem a desvantagem de só poder ser usado para comparar dois modelos de cada vez, sendo um destes modelos versão reduzida do outro.

O teste da razão de verossimilhança é equivalente ao teste das hipóteses

$$H_0 : \underset{\sim}{\mathbf{C}} \underset{\sim}{\boldsymbol{\beta}} = \underset{\sim}{\mathbf{0}} \quad \text{versus} \quad H_A : \underset{\sim}{\mathbf{C}} \underset{\sim}{\boldsymbol{\beta}} \neq \underset{\sim}{\mathbf{0}} \quad ,$$

onde $\underset{\sim}{\boldsymbol{\beta}}$ é o vetor de parâmetros do modelo com maior número de parâmetros e $\underset{\sim}{\mathbf{C}}$ é matriz $r \times p$, onde cada linha é composta de zeros e um único elemento igual a 1, de modo que

a hipótese nula defina os tais r componentes de $\hat{\beta}$ iguais a zero. A estatística do teste

$$(\mathbf{C} \hat{\beta})' [\mathbf{C} \{Cov(\hat{\beta})\} \mathbf{C}']^{-1} (\mathbf{C} \hat{\beta})$$

tem distribuição $F_{r, (n-p)}$, ou assintoticamente χ_r^2 , sob a hipótese nula. Assim, por (2.20),

$$\hat{\sigma}^2 (\mathbf{C} \hat{\beta})' \left[\mathbf{C} \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{C}' \right]^{-1} (\mathbf{C} \hat{\beta}) \sim \chi_r^2 \quad (\text{assintoticamente}) \quad .$$

Observe que este teste equivale a r testes independentes, correspondentes às hipóteses de que cada um dos tais r componentes de $\hat{\beta}$ é igual a zero. Em geral, poderíamos testar a hipótese

$$H_0 : \mathbf{C} \hat{\beta} = c_0 \quad \text{versus} \quad H_A : \mathbf{C} \hat{\beta} \neq c_0 \quad ,$$

para qualquer matriz \mathbf{C} , de dimensões $c \times p$, com posto igual a c . Neste caso, a estatística do teste é

$$(\mathbf{C} \hat{\beta} - c_0)' [\mathbf{C} \{Cov(\hat{\beta})\} \mathbf{C}']^{-1} (\mathbf{C} \hat{\beta} - c_0) \quad ,$$

com distribuição assintótica χ_c^2 , sob a hipótese nula.

Outro procedimento de seleção de modelo é a utilização do Critério de Informação de Akaike (AIC), definido como o valor de ℓ acrescentado de duas vezes o número de parâmetros no modelo,

$$AIC = \ell + 2 \times (\text{n}^\circ \text{ de parâmetros}) \quad , \quad (2.26)$$

O modelo com o menor AIC é o escolhido. A medida AIC penaliza os modelos com número grande de parâmetros.

2.5 Uma Breve Ilustração

Um exemplo de estudo observacional envolvendo dados longitudinais é analisado por Jones e Boadi-Boateng (1991). Este estudo envolveu 619 indivíduos (unidades); alguns com doença hereditária renal, alguns com hipertensão. Observações foram tomadas de cada indivíduo em idades variadas, sendo que o número de observações por pessoa variou de 1 até 22 observações. A variável resposta de interesse foi o recíproco da quantidade de uma substância, denotada SCR, que decresce à medida que a atuação do rim piora. Quatro grupos foram compostos da seguinte maneira: indivíduos sem doença renal e sem hipertensão, indivíduos só com doença renal, indivíduos só com hipertensão e indivíduos com ambas, doença renal e hipertensão. Foi definido um modelo de uma reta para cada grupo, totalizando 8 coeficientes, e efeitos aleatórios bivariados. Assim, queremos que a reta esperada para a população de indivíduos sem doença renal e hipertensão seja dada por

$$y_{t_i} = \beta_{11} + (\beta_{12} \times t_i) \quad ,$$

a reta esperada para a população de indivíduos apenas com doença renal seja

$$y_{t_i} = (\beta_{11} + \beta_{21}) + ((\beta_{12} + \beta_{22}) \times t_i) \quad ,$$

a reta esperada para a população de indivíduos apenas com hipertensão seja

$$y_{t_i} = (\beta_{11} + \beta_{31}) + ((\beta_{12} + \beta_{32}) \times t_i) \quad ,$$

e para a população de indivíduos com ambos os problemas seja

$$y_{t_i} = (\beta_{11} + \beta_{21} + \beta_{31} + \beta_{41}) + ((\beta_{12} + \beta_{22} + \beta_{32} + \beta_{42}) \times t_i) \quad .$$

Assim, por exemplo, para um indivíduo com 4 observações, sem doença renal e sem

hipertensão, temos a seguinte equação de modelo:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{bmatrix} = \begin{bmatrix} 1 & t_{i1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & t_{i2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & t_{i3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & t_{i4} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{42} \end{bmatrix} + \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ 1 & t_{i3} \\ 1 & t_{i4} \end{bmatrix} \begin{bmatrix} \gamma_{i1} \\ \gamma_{i2} \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \end{bmatrix},$$

onde t_{i1} , t_{i2} , t_{i3} e t_{i4} seriam as idades deste indivíduo por ocasião das observações. A parcela $\mathbf{X}_i \beta$ define a reta esperada da população para o grupo sem doença renal e com pressão do sangue normal. Cada indivíduo tem a sua própria reta que é assumida estar aleatoriamente distribuída em torno da reta média do grupo. Os dois coeficientes aleatórios, γ_{i1} e γ_{i2} , que acompanham a matriz \mathbf{Z}_i , representam os desvios do intercepto e da inclinação para este indivíduo, em relação ao intercepto e à inclinação do grupo. Assim, a reta esperada para este indivíduo é $(\beta_{11} + \hat{\gamma}_{i1}) + (\beta_{12} + \hat{\gamma}_{i2}) \times (idade)$. As matrizes de desenho para um indivíduo observado 3 vezes, com doença renal e sem hipertensão são

$$\mathbf{X}_i = \begin{bmatrix} 1 & t_{i1} & 1 & t_{i1} & 0 & 0 & 0 & 0 \\ 1 & t_{i2} & 1 & t_{i2} & 0 & 0 & 0 & 0 \\ 1 & t_{i3} & 1 & t_{i3} & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{Z}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ 1 & t_{i3} \end{bmatrix}$$

Isto permite que a reta esperada para este indivíduo seja $(\beta_{11} + \beta_{21} + \hat{\gamma}_{i1}) + (\beta_{12} + \beta_{22} + \hat{\gamma}_{i2}) \times (idade)$. Para um indivíduo com apenas uma observação, sem doença renal e com hipertensão, temos

$$\mathbf{X}_i = \begin{bmatrix} 1 & t_{i1} & 0 & 0 & 1 & t_{i1} & 0 & 0 \end{bmatrix}, \quad \mathbf{Z}_i = \begin{bmatrix} 1 & t_{i1} \end{bmatrix}$$

e para um indivíduo com duas observações e com doença renal e hipertensão

$$\mathbf{X}_i = \begin{bmatrix} 1 & t_{i1} & 1 & t_{i1} & 1 & t_{i1} & 1 & t_{i1} \\ 1 & t_{i2} & 1 & t_{i2} & 1 & t_{i2} & 1 & t_{i2} \end{bmatrix}, \quad \mathbf{Z}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \end{bmatrix}$$

Desta forma, para a geração das matrizes \mathbf{X}_i , o programa computacional utilizado requer a definição de 3 variáveis indicadoras agrupadoras, uma indicadora de doença renal (0-sim, 1-não), outra indicadora de hipertensão (0-sim, 1-não) e mais uma para indicação de ambas as doenças (1-unidade tem doença renal e hipertensão, 0- caso contrário). Usando a notação introduzida na seção (2.2.2), temos $p_1 = p_2 = p_3 = 2$.

Neste estudo, 288 pessoas foram observadas uma única vez, e mesmo assim é possível ajustar uma reta para cada uma delas. Segundo Jones e Boadi-Boateng, na reta ajustada para estes indivíduos, a inclinação estimada será a inclinação do grupo. No entanto, a reta não irá passar pelo único ponto observado, mas passará ao redor da reta do grupo. Note que estes indivíduos com apenas uma informação coletada são importantes para se obter um 'padrão' de comportamento da variável em estudo ao longo do tempo. Se estes fossem eliminados da análise, muita informação seria perdida.

Foi assumida estrutura de correlação dos resíduos seguindo um processo auto-regressivo de 1ª ordem (AR1), com erros observacionais, ou seja, com matriz de correlação dos resíduos dada por:

$$\mathbf{W}_i = \begin{bmatrix} 1 + \sigma_0^2 & \exp\{-\alpha_0(t_2 - t_1)\} & \cdots & \exp\{-\alpha_0(t_{n_i} - t_1)\} \\ \exp\{-\alpha_0(t_2 - t_1)\} & 1 + \sigma_0^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \exp\{-\alpha_0(t_{n_i} - t_1)\} & \cdots & \cdots & 1 + \sigma_0^2 \end{bmatrix}$$

Capítulo 3

Os Efeitos Aleatórios do Modelo Laird-Ware: Predição e Adequação do Modelo

3.1 Efeitos Aleatórios γ_i

Considere o modelo de Laird-Ware, apresentado no capítulo anterior,

$$y_i = \mathbf{X}_i \beta + \mathbf{Z}_i \gamma_i + \epsilon_i \quad , \quad (3.1)$$

onde $\gamma_i \sim N(0, \sigma^2 \mathbf{B})$ é independente de $\epsilon_i \sim N(0, \sigma^2 \mathbf{W}_i)$ e

$$Cov(y_i) = \sigma^2 \mathbf{V}_i = \sigma^2 (\mathbf{Z}_i \mathbf{B} \mathbf{Z}_i' + \mathbf{W}_i) \quad . \quad (3.2)$$

No contexto do modelo de Laird-Ware, γ_i e ϵ_i são variáveis aleatórias não observáveis, e, muitas vezes surge a necessidade de ‘estimar’ as variáveis γ_i . Robinson (1991) discute o uso do termo ‘estimação’ ou ‘predição’ de uma variável aleatória, dizendo preferir o primeiro, mas que usualmente o segundo termo é mais utilizado. Laird e Ware (1982) mostraram que os efeitos aleatórios podem ser preditos através da metodologia de Bayes empírico, ou seja, ‘estimar’ γ_i como sendo a média da distribuição a posteriori de γ_i dadas

às observações \tilde{y}_i ,

$$E[\tilde{\gamma}_i / \tilde{y}_i] = \mathbf{B}\mathbf{Z}'_i\mathbf{V}_i^{-1}(\tilde{y}_i - \mathbf{X}_i\tilde{\beta}) \quad . \quad (3.3)$$

Para verificarmos este resultado, basta mostrar que, como

$$\tilde{\gamma}_i \sim N(0, \sigma^2\mathbf{B}) \quad , \quad \tilde{y}_i \sim N(\mathbf{X}_i\tilde{\beta}, \sigma^2\mathbf{V}_i)$$

e ainda

$$Cov(\tilde{\gamma}_i, \tilde{y}_i) = Cov(\tilde{\gamma}_i, \mathbf{Z}_i\tilde{\gamma}_i + \epsilon_i) = Var(\tilde{\gamma}_i)\mathbf{Z}'_i = \sigma^2\mathbf{B}\mathbf{Z}'_i \quad ,$$

então, usando um conhecido resultado de distribuição normal, teremos

$$\begin{aligned} E[\tilde{\gamma}_i / \tilde{y}_i] &= E(\tilde{\gamma}_i) + Cov(\tilde{\gamma}_i, \tilde{y}_i) \cdot [Var(\tilde{y}_i)]^{-1} \cdot (\tilde{y}_i - E(\tilde{y}_i)) \\ &= \mathbf{B}\mathbf{Z}'_i\mathbf{V}_i^{-1}(\tilde{y}_i - \mathbf{X}_i\tilde{\beta}) \quad . \end{aligned}$$

Uma outra maneira de produzir o mesmo preditor para $\tilde{\gamma}_i$ é apresentada abaixo, logo após um resultado necessário para seu desenvolvimento.

Resultado 3.1:

A inversa da matriz de covariâncias (3.2) das observações para uma certa unidade i é dada por:

$$\mathbf{V}_i^{-1} = \mathbf{W}_i^{-1} - \mathbf{W}_i^{-1}\mathbf{Z}_i[\mathbf{B}^{-1} + \mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i]^{-1}\mathbf{Z}'_i\mathbf{W}_i^{-1} \quad . \quad (3.4)$$

Prova:

Isto pode ser mostrado verificando que $\mathbf{V}_i\mathbf{V}_i^{-1} = \mathbf{I}$

$$\begin{aligned} \mathbf{V}_i\mathbf{V}_i^{-1} &= (\mathbf{W}_i + \mathbf{Z}_i\mathbf{B}\mathbf{Z}'_i)(\mathbf{W}_i^{-1} - \mathbf{W}_i^{-1}\mathbf{Z}_i[\mathbf{B}^{-1} + \mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i]^{-1}\mathbf{Z}'_i\mathbf{W}_i^{-1}) = \\ &= \mathbf{I} - \mathbf{Z}_i[\mathbf{B}^{-1} + \mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i]^{-1}\mathbf{Z}'_i\mathbf{W}_i^{-1} + \mathbf{Z}_i\mathbf{B}\mathbf{Z}'_i\mathbf{W}_i^{-1} - \\ &\quad - \mathbf{Z}_i\mathbf{B}\mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i[\mathbf{B}^{-1} + \mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i]^{-1}\mathbf{Z}'_i\mathbf{W}_i^{-1} = \\ &= \mathbf{I} - \mathbf{Z}_i[\mathbf{I} - \mathbf{B}\mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i][\mathbf{B}^{-1} + \mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i]^{-1}\mathbf{Z}'_i\mathbf{W}_i^{-1} + \mathbf{Z}_i\mathbf{B}\mathbf{Z}'_i\mathbf{W}_i^{-1} = \\ &= \mathbf{I} - \mathbf{Z}_i\mathbf{B}[\mathbf{B}^{-1} - \mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i][\mathbf{B}^{-1} + \mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i]^{-1}\mathbf{Z}'_i\mathbf{W}_i^{-1} + \mathbf{Z}_i\mathbf{B}\mathbf{Z}'_i\mathbf{W}_i^{-1} = \\ &= \mathbf{I} - \mathbf{Z}_i\mathbf{B}\mathbf{Z}'_i\mathbf{W}_i^{-1} + \mathbf{Z}_i\mathbf{B}\mathbf{Z}'_i\mathbf{W}_i^{-1} = \mathbf{I} \quad . \end{aligned}$$

□

Resultado 3.2:

Usando a suposição de normalidade, o preditor de $\tilde{\gamma}_i$, $\mathbf{BZ}'_i\mathbf{V}_i^{-1}(y_i - \mathbf{X}_i \tilde{\beta})$ é dado pela solução da segunda equação do sistema

$$\begin{cases} \mathbf{X}'_i\mathbf{W}_i^{-1}\mathbf{X}_i \tilde{\beta} + \mathbf{X}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i \tilde{\gamma}_i = \mathbf{X}'_i\mathbf{W}_i^{-1} y_i \\ \mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{X}_i \tilde{\beta} + (\mathbf{B}^{-1} + \mathbf{Z}'_i\mathbf{W}_i^{-1}\mathbf{Z}_i) \tilde{\gamma}_i = \mathbf{Z}'_i\mathbf{W}_i^{-1} y_i \end{cases} \quad (3.5)$$

(Este sistema é conhecido como sistema de equações normais do modelo misto, porque ele pode ser derivado a partir de definição do modelo linear misto, sem a condição de normalidade.)

Prova:

A função de densidade conjunta de $\tilde{\epsilon}_i$ e $\tilde{\gamma}_i$ é dada por

$$\begin{aligned} f(\tilde{\epsilon}_i, \tilde{\gamma}_i) &= (2\pi\sigma^2)^{-\frac{1}{2}(n_i+g)} \left(\det \begin{bmatrix} B & 0 \\ 0 & W \end{bmatrix} \right)^{-\frac{1}{2}} \\ &\times \exp \left\{ -\frac{1}{2\sigma^2} \begin{pmatrix} \tilde{\gamma}_i \\ y_i - \mathbf{X}_i \tilde{\beta} - \mathbf{Z}_i \tilde{\gamma}_i \end{pmatrix}' \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix}^{-1} \begin{pmatrix} \tilde{\gamma}_i \\ y_i - \mathbf{X}_i \tilde{\beta} - \mathbf{Z}_i \tilde{\gamma}_i \end{pmatrix} \right\}, \end{aligned} \quad (3.6)$$

onde g é o número de efeitos aleatórios $\tilde{\gamma}_i$ e n_i é o número de observações para a unidade i . Considerando a função acima como uma função matemática com argumentos $\tilde{\beta}$ e $\tilde{\gamma}_i$, maximizar a função acima com relação a $\tilde{\beta}$ e $\tilde{\gamma}_i$ é o mesmo que minimizar

$$\begin{aligned} \Theta &= \begin{pmatrix} \tilde{\gamma}_i \\ y_i - \mathbf{X}_i \tilde{\beta} - \mathbf{Z}_i \tilde{\gamma}_i \end{pmatrix}' \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix}^{-1} \begin{pmatrix} \tilde{\gamma}_i \\ y_i - \mathbf{X}_i \tilde{\beta} - \mathbf{Z}_i \tilde{\gamma}_i \end{pmatrix} = \\ &= \tilde{\gamma}'_i \mathbf{B}^{-1} \tilde{\gamma}_i + (y_i - \mathbf{X}_i \tilde{\beta} - \mathbf{Z}_i \tilde{\gamma}_i)' \mathbf{W}_i^{-1} (y_i - \mathbf{X}_i \tilde{\beta} - \mathbf{Z}_i \tilde{\gamma}_i) \end{aligned}$$

Diferenciando Θ , com respeito a $\tilde{\beta}$ e $\tilde{\gamma}_i$, e igualando a zero, teremos

$$\begin{aligned}\frac{\partial \Theta}{\partial \tilde{\beta}} &= -2\mathbf{X}'_i \mathbf{W}_i^{-1} \tilde{y}_i + 2\mathbf{X}'_i \mathbf{W}_i^{-1} \mathbf{X}_i \tilde{\beta} + 2\mathbf{X}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i \tilde{\gamma}_i = 0 \\ &\Rightarrow \mathbf{X}'_i \mathbf{W}_i^{-1} \mathbf{X}_i \tilde{\beta} + \mathbf{X}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i \tilde{\gamma}_i = \mathbf{X}'_i \mathbf{W}_i^{-1} \tilde{y}_i\end{aligned}$$

$$\begin{aligned}\frac{\partial \Theta}{\partial \tilde{\gamma}_i} &= 2\mathbf{B}^{-1} \tilde{\gamma}_i - 2\mathbf{Z}'_i \mathbf{W}_i^{-1} \tilde{y}_i + 2\mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{X}_i \tilde{\beta} + 2\mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i \tilde{\gamma}_i = 0 \\ &\Rightarrow \mathbf{B}^{-1} \tilde{\gamma}_i + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{X}_i \tilde{\beta} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i \tilde{\gamma}_i = \mathbf{Z}'_i \mathbf{W}_i^{-1} \tilde{y}_i \\ &\Rightarrow \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{X}_i \tilde{\beta} + (\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i) \tilde{\gamma}_i = \mathbf{Z}'_i \mathbf{W}_i^{-1} \tilde{y}_i,\end{aligned}$$

que são as equações do sistema (3.5).

□

Pela segunda equação do sistema (3.5),

$$\begin{aligned}\mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{X}_i \tilde{\beta} + (\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i) \tilde{\gamma}_i &= \mathbf{Z}'_i \mathbf{W}_i^{-1} \tilde{y}_i \\ \Rightarrow (\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i) \tilde{\gamma}_i &= \mathbf{Z}'_i \mathbf{W}_i^{-1} \tilde{y}_i - \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{X}_i \tilde{\beta} \\ \Rightarrow (\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i) \tilde{\gamma}_i &= \mathbf{Z}'_i \mathbf{W}_i^{-1} (\tilde{y}_i - \mathbf{X}_i \tilde{\beta}) \\ \Rightarrow \tilde{\gamma}_i &= (\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{W}_i^{-1} (\tilde{y}_i - \mathbf{X}_i \tilde{\beta})\end{aligned}\tag{3.7}$$

Usando o resultado 3.1, podemos reescrever a expressão acima como

$$\begin{aligned}\tilde{\gamma}_i &= (\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{W}_i^{-1} (\tilde{y}_i - \mathbf{X}_i \tilde{\beta}) = \\ &= (\mathbf{B}[\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i] - \mathbf{BZ}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i) [\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i]^{-1} \mathbf{Z}'_i \mathbf{W}_i^{-1} (\tilde{y}_i - \mathbf{X}_i \tilde{\beta}) \\ &= \mathbf{BZ}'_i \mathbf{W}_i^{-1} (\tilde{y}_i - \mathbf{X}_i \tilde{\beta}) - \mathbf{BZ}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i [\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i]^{-1} \mathbf{Z}'_i \mathbf{W}_i^{-1} (\tilde{y}_i - \mathbf{X}_i \tilde{\beta}) = \\ &= \mathbf{BZ}'_i [\mathbf{W}_i^{-1} - \mathbf{W}_i^{-1} \mathbf{Z}_i [\mathbf{B}^{-1} + \mathbf{Z}'_i \mathbf{W}_i^{-1} \mathbf{Z}_i]^{-1} \mathbf{Z}'_i \mathbf{W}_i^{-1}] (\tilde{y}_i - \mathbf{X}_i \tilde{\beta}) = \\ &= \mathbf{BZ}'_i \mathbf{V}_i^{-1} (\tilde{y}_i - \mathbf{X}_i \tilde{\beta})\end{aligned}$$

□

Segundo Harville (1977) é natural considerar $\hat{\gamma}_i = \mathbf{BZ}'_i \mathbf{V}_i^{-1} (y_i - \mathbf{X}_i \hat{\beta})$ e a sua matriz de covariâncias será dada pelo resultado abaixo:

Resultado 3.3:

$$Var(\hat{\gamma}_i) = \hat{\sigma}^2 [\mathbf{BZ}'_i \{ \mathbf{V}_i^{-1} - \mathbf{V}_i^{-1} \mathbf{X}_i (\sum_j \mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j) \mathbf{X}'_i \mathbf{V}_i^{-1} \} \mathbf{Z}_i \mathbf{B}]$$

Prova:

$$\begin{aligned} Var(\hat{\gamma}_i) &= Var(\mathbf{BZ}'_i \mathbf{V}_i^{-1} (y_i - \mathbf{X}_i \hat{\beta})) = \mathbf{BZ}'_i \mathbf{V}_i^{-1} Var(y_i - \mathbf{X}_i \hat{\beta}) \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{B} = \\ &= \mathbf{BZ}'_i \mathbf{V}_i^{-1} [Var(y_i) + Var(\mathbf{X}_i \hat{\beta}) - 2Cov(y_i, \mathbf{X}_i \hat{\beta})] \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{B} \quad . \end{aligned} \quad (3.8)$$

Mas considerando a independência das unidades, desenvolvemos a matriz de covariâncias

$$\begin{aligned} Cov(y_i, \mathbf{X}_i \hat{\beta}) &= Cov \left(y_i, \mathbf{X}_i \left(\sum_j \mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \sum_j \mathbf{X}'_j \mathbf{V}_j^{-1} y_j \right) = \\ &= Cov \left(y_i, \mathbf{X}_i \left(\sum_j \mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \mathbf{X}'_i \mathbf{V}_i^{-1} y_i \right) \quad , \end{aligned}$$

e desta forma,

$$\begin{aligned} Cov(y_i, \mathbf{X}_i \hat{\beta}) &= \mathbf{X}_i \left(\sum_j \mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \mathbf{X}'_i \mathbf{V}_i^{-1} E \left[y_i y_i \right] - \\ &- \mathbf{X}_i \left(\sum_j \mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \mathbf{X}'_i \mathbf{V}_i^{-1} E \left[y_i \right] E \left[y_i \right] = \mathbf{X}_i \left(\sum_j \mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \mathbf{X}'_i \mathbf{V}_i^{-1} Var(y_i) = \\ &= \hat{\sigma}^2 \mathbf{X}_i \left(\sum_j \mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \mathbf{X}'_i = Var(\mathbf{X}_i \hat{\beta}) \quad . \end{aligned} \quad (3.9)$$

Assim, voltando em (3.8)

$$\begin{aligned}
\text{Var}(\hat{\gamma}_i) &= \mathbf{B}\mathbf{Z}'_i\mathbf{V}_i^{-1}[\text{Var}(y_i) + \text{Var}(\mathbf{X}_i\hat{\beta}) - 2\text{Cov}(y_i, \mathbf{X}_i\hat{\beta})]\mathbf{V}_i^{-1}\mathbf{Z}_i\mathbf{B} = \\
&= \mathbf{B}\mathbf{Z}'_i\mathbf{V}_i^{-1}[\text{Var}(y_i) - \text{Cov}(y_i, \mathbf{X}_i\hat{\beta})]\mathbf{V}_i^{-1}\mathbf{Z}_i\mathbf{B} = \\
&= \mathbf{B}\mathbf{Z}'_i\mathbf{V}_i^{-1}[\hat{\sigma}^2\mathbf{V}_i - \mathbf{X}_i\hat{\sigma}^2(\sum_j \mathbf{X}'_j\mathbf{V}_j^{-1}\mathbf{X}_j)^{-1}\mathbf{X}'_i]\mathbf{V}_i^{-1}\mathbf{Z}_i\mathbf{B} = \\
&= \hat{\sigma}^2[\mathbf{B}\mathbf{Z}'_i\{\mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{X}_i(\sum_j \mathbf{X}'_j\mathbf{V}_j^{-1}\mathbf{X}_j)^{-1}\mathbf{X}'_i\mathbf{V}_i^{-1}\}\mathbf{Z}_i\mathbf{B}] .
\end{aligned}$$

□

3.2 Suposições Distribucionais

Quando utilizamos o modelo misto linear de Laird-Ware (3.1) com observações contínuas, é assumido que $\gamma_i \sim N(\mathbf{0}, \sigma^2\mathbf{B})$. Portanto é necessário que verifiquemos a adequação desta suposição. Até há pouco tempo atrás, métodos para verificar o ajuste de efeitos aleatórios eram escassos. Ryan e Dempster (1985) propuseram um gráfico de probabilidade normal ('normal plot') ponderado, como uma solução gráfica de verificação da normalidade. Lange e Ryan (1989) descreveram um método para verificar a adequação do modelo, em particular a suposição de normalidade para uma classe de modelos de efeitos aleatórios, generalizando o método ponderado de Ryan e Dempster (1985), quando fizeram uso de estimadores bayesianos empíricos padronizados.

Devemos salientar que falhas na suposição de normalidade no modelo Laird-Ware não invalidam as estimativas dos parâmetros fixos, já que os estimadores de máxima verossimilhança coincidem com os estimadores de mínimos quadrados ponderados. Entretanto, esta falha invalidará os testes usuais e intervalos de confiança baseados na suposição de normalidade.

3.2.1 Análise dos efeitos aleatórios

Dois recursos gráficos para verificar as suposições distribucionais do vetor $\tilde{\gamma}_i$ são o gráfico normal ponderado e o não ponderado. Embora o gráfico não ponderado seja o mais comumente usado e bastante disponível em pacotes estatísticos, o gráfico ponderado é uma solução mais adequada quando utilizamos o modelo de Laird-Ware, principalmente em conjuntos de dados não balanceados.

Os gráficos normais ponderados ou não ponderados são construídos obtendo-se primeiramente os efeitos aleatórios padronizados para cada indivíduo. Sejam

$$z_i = \frac{\tilde{c}' \tilde{\hat{\gamma}}_i}{[\tilde{c}' \text{Cov}(\tilde{\hat{\gamma}}_i) \tilde{c}]^{\frac{1}{2}}} \text{ para } i = 1, 2, \dots, n \quad , \quad (3.10)$$

onde \tilde{c} é um vetor que indica o efeito aleatório que está sendo verificado. Por exemplo, se o modelo assume dois efeitos aleatórios para cada indivíduo, intercepto aleatório e inclinação aleatória, podemos destacar o primeiro componente fazendo $\tilde{c}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e o

segundo, fazendo $\tilde{c}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Portanto, nesse instante, temos os efeitos aleatórios $\tilde{\hat{\gamma}}_i$'s padronizados, com média zero e variância 1. O nosso objetivo, então, é verificar a normalidade desses $\tilde{\hat{\gamma}}_i$'s a partir das variáveis transformadas z_i 's.

A distribuição acumulada empírica de z_1, z_2, \dots, z_n é dada pela função escada

$$F_n(x) = \frac{\sum_{i=1}^n I(x - z_i)}{n} \quad ,$$

onde

$$I(x - y) = \begin{cases} 1, & \text{se } x \geq y \\ 0, & \text{c.c.} \end{cases} \quad .$$

Verificaremos, agora, as propriedades desta função não ponderada quanto à sua média e variância.

Resultado 3.4:

Sob suposição de normalidade das variáveis $\tilde{\gamma}_i$,

$$E[F_n(x)] = \Phi(x) \quad ,$$

onde $\Phi(x)$ é o valor para x , na função de distribuição acumulada da normal padrão, e, para $y \geq x$,

$$Cov[F_n(x), F_n(y)] = \frac{1}{n} \Phi(x)[1 - \Phi(y)] \quad .$$

Prova:

Dado que $z_i, i = 1, \dots, n$ são variáveis aleatórias independentes normais padrão, então

$$E[F_n(x)] = \frac{1}{n} \sum_{i=1}^n E[\mathbf{I}(x - z_i)] = \Phi(x) \quad .$$

Além disso,

$$\begin{aligned} Cov[F_n(x), F_n(y)] &= E[F_n(x)F_n(y)] - E(F_n(x))E(F_n(y)) = \\ &= \frac{1}{n^2} E[\sum_{i=1}^n \mathbf{I}(x - z_i) \sum_{i=1}^n \mathbf{I}(y - z_i)] - \Phi(x)\Phi(y) = \\ &= \frac{1}{n^2} E[(\mathbf{I}(x - z_1)\mathbf{I}(y - z_1)) + (\mathbf{I}(x - z_1)\mathbf{I}(y - z_2)) + \dots \\ &+ (\mathbf{I}(x - z_1)\mathbf{I}(y - z_n)) + \dots + (\mathbf{I}(x - z_n)\mathbf{I}(y - z_1)) + (\mathbf{I}(x - z_n)\mathbf{I}(y - z_2)) + \dots \\ &+ (\mathbf{I}(x - z_n)\mathbf{I}(y - z_n))] - \Phi(x)\Phi(y) = \\ &= \frac{1}{n^2} E[(\mathbf{I}(x - z_1)\mathbf{I}(y - z_1))] + E[(\mathbf{I}(x - z_1)\mathbf{I}(y - z_2))] + \dots \\ &+ E[(\mathbf{I}(x - z_1)\mathbf{I}(y - z_n))] + \dots + E[(\mathbf{I}(x - z_n)\mathbf{I}(y - z_1))] + E[(\mathbf{I}(x - z_n)\mathbf{I}(y - z_2))] + \dots \\ &+ E[(\mathbf{I}(x - z_n)\mathbf{I}(y - z_n))] - \Phi(x)\Phi(y) \quad . \end{aligned}$$

Mas sabemos que $x \leq y$ e portanto,

$$\mathbf{I}(x - z_i) = 1 \Rightarrow \mathbf{I}(y - z_i) = 1 \quad .$$

Com isto, temos que

$$E[(\mathbf{I}(x - z_i)\mathbf{I}(y - z_i))] = \Phi(x) \quad ,$$

e para $i \neq j$

$$E[(\mathbf{I}(x - z_i)\mathbf{I}(y - z_j))] = \Phi(x)\Phi(y) \quad .$$

Voltando à demonstração, teremos

$$\begin{aligned} Cov[F_n(x), F_n(y)] &= \frac{1}{n^2}[n\Phi(x) + n(n-1)\Phi(x)\Phi(y)] - \Phi(x)\Phi(y) = \\ &= \frac{1}{n}\Phi(x) + \frac{1}{n}(n-1)\Phi(x)\Phi(y) - \Phi(x)\Phi(y) = \frac{1}{n}\Phi(x) - \frac{1}{n}\Phi(x)\Phi(y) = \\ &= \frac{1}{n}\Phi(x)[1 - \Phi(y)] \quad . \end{aligned}$$

□

Cabe aqui, ainda, enunciar o seguinte resultado:

Resultado 3.5:

Usando expansão em série de Taylor da função $\Phi^{-1}(x)$ sob $E(x)$, segue que, sob a hipótese de normalidade das variáveis $\tilde{\gamma}_i$,

$$E\{\Phi^{-1}[F_n(x)]\} \approx x$$

e

$$Cov\{\Phi^{-1}[F_n(x)], \Phi^{-1}[F_n(y)]\} \approx \frac{1}{n} \frac{\Phi(x)[1 - \Phi(y)]}{\phi(x)\phi(y)} \quad ,$$

onde $\phi(x)$ é a função de densidade normal padrão.

□

Portanto, o gráfico dos pares $(z_i, \Phi^{-1}[F_n(z_i)])$ deve ser aproximadamente linear, com inclinação 1, passando pela origem, sob suposição de normalidade dos efeitos aleatórios. Além disso, uma região crítica aproximada para $\Phi^{-1}[F_n(z_i)]$ é dada por

$$z_i \pm 1.96 \left\{ (1/n)\Phi(z_i)[1 - \Phi(z_i)] \right\}^{\frac{1}{2}} / \phi(z_i) \quad .$$

Dempster e Ryan (1985) propuseram uma modificação no gráfico de probabilidade

normal usual, atribuindo pesos aos estimadores dos efeitos aleatórios de cada unidade. Desta forma, será dado maior destaque (maior peso) àquele efeito aleatório que foi estimado com mais observações. Então, em vez de $F_n(x)$, consideremos $F_n^*(x)$ para denotar a distribuição acumulada empírica ponderada. Para obtermos $F_n^*(x)$ procedemos da seguinte maneira:

Consideremos os pesos

$$w_i = \underset{\sim}{c}' Cov(\underset{\sim}{\hat{\gamma}}_i) \underset{\sim}{c}, i = 1, \dots, n. \quad (3.11)$$

Então,

$$F_n^*(x) = \frac{\sum_{i=1}^n w_i \mathbf{I}(x - z_i)}{W}, \quad (3.12)$$

onde

$$W = \sum_{i=1}^n w_i$$

e

$$\mathbf{I}(x - y) = \begin{cases} 1, & \text{se } x \geq y \\ 0, & \text{c.c.} \end{cases} .$$

Verificaremos, agora, as propriedades da função distribuição acumulada empírica ponderada quanto à sua média e variância.

Resultado 3.6:

Sob suposição de normalidade das variáveis $\underset{\sim}{\gamma}_i$,

$$E[F_n^*(x)] = \Phi(x) ,$$

onde $\Phi(x)$ é o valor para x , na distribuição normal padrão acumulada, e, para $y \geq x$,

$$Cov[F_n^*(x), F_n^*(y)] = \frac{1}{n} \left(1 + \frac{v}{m^2}\right) \Phi(x) [1 - \Phi(y)] ,$$

onde

$$m = \frac{1}{n} \sum_{i=1}^n w_i$$

e

$$v = \left[\frac{1}{n} \sum_{i=1}^n w_i^2 \right] - m^2$$

são a média e a variância dos pesos definidos em (3.11).

Prova:

Dado que $z_i, i = 1, \dots, n$ são variáveis aleatórias independentes normais padrão, então

$$E[F_n^*(x)] = \frac{1}{W} \sum_{i=1}^n w_i E[\mathbf{I}(x - z_i)] = \Phi(x) \quad ,$$

e ainda, seguindo o mesmo procedimento usado anteriormente,

$$\begin{aligned} Cov[F_n^*(x), F_n^*(y)] &= \frac{1}{W^2} E \left[\sum_{i=1}^n w_i \mathbf{I}(x - z_i) \sum_{i=1}^n w_i \mathbf{I}(y - z_i) \right] - \Phi(x)\Phi(y) = \\ &= \frac{1}{W^2} \{ E[(\mathbf{I}(x - z_1)\mathbf{I}(y - z_1))w_1^2] + E[(\mathbf{I}(x - z_1)\mathbf{I}(y - z_2))w_1w_2] + \dots + \\ &E[(\mathbf{I}(x - z_1)\mathbf{I}(y - z_n))w_1w_n] + \dots + E[(\mathbf{I}(x - z_n)\mathbf{I}(y - z_1))w_nw_1] + \\ &+ E[(\mathbf{I}(x - z_n)\mathbf{I}(y - z_2))w_nw_2] + \dots + E[(\mathbf{I}(x - z_n)\mathbf{I}(y - z_n))w_n^2] \} - \Phi(x)\Phi(y) = \\ &= \frac{1}{W^2} [\Phi(x) \sum_i w_i^2 + \Phi(x)\Phi(y) \sum_{i \neq j} w_i w_j] - \Phi(x)\Phi(y) \quad . \end{aligned}$$

Mas sabemos que

$$\begin{aligned} \sum_{i=1}^n w_i^2 &= (m^2 + v)n \quad e \\ W^2 = n^2 m^2 &\Rightarrow \left[\sum_{i \neq j} w_i w_j + \sum_i w_i^2 \right] = n^2 m^2 \Rightarrow \\ \sum_{i \neq j} w_i w_j &= n^2 m^2 - (v + m^2)n \quad . \end{aligned}$$

Portanto,

$$\begin{aligned} Cov[F_n^*(x), F_n^*(y)] &= \frac{1}{n^2 m^2} [\Phi(x)(m^2 + v)n + \Phi(x)\Phi(y)(n^2 m^2 - (v + m^2)n)] - \Phi(x)\Phi(y) = \\ &= \frac{(m^2 + v)}{n m^2} \Phi(x) - \Phi(x)\Phi(y) \left[1 - \frac{(n^2 m^2 - (v + m^2)n)}{n^2 m^2} \right] = \\ &= \frac{1}{n} \Phi(x) \left[1 + \frac{v}{m^2} \right] - \Phi(x)\Phi(y) \left[\frac{1}{n} + \frac{v}{n m^2} \right] = \frac{1}{n} \left(1 + \frac{v}{m^2} \right) \Phi(x) [1 - \Phi(y)] \quad . \end{aligned}$$

Logo,

$$\begin{aligned} Cov[F_n^*(x), F_n^*(y)] &= (1 + \frac{v}{m^2})Cov[F_n(x), F_n(y)] = \\ &= \frac{1}{n}(1 + \frac{v}{m^2})\Phi(x)[1 - \Phi(y)] \quad . \end{aligned} \quad (3.13)$$

□

Resta ainda enunciarmos:

Resultado 3.7:

Usando expansão em série de Taylor da função $\Phi^{-1}(x)$ sob $E(x)$, segue que

$$E\{\Phi^{-1}[F_n^*(x)]\} \approx x$$

e

$$Cov\{\Phi^{-1}[F_n^*(x)], \Phi^{-1}[F_n^*(y)]\} \approx \frac{1}{n} \left(1 + \frac{v}{m^2}\right) \frac{\Phi(x)[1 - \Phi(y)]}{\phi(x)\phi(y)} \quad ,$$

onde $\phi(x)$ é a função densidade normal padrão. Portanto, uma região crítica aproximada para $\Phi^{-1}\{F_n^*(z_i)\}$ é dada por

$$z_i \pm 1.96 \left(1 + \frac{v}{m^2}\right)^{\frac{1}{2}} \left\{ \left(\frac{1}{n}\right)\Phi(z_i)[1 - \Phi(z_i)] \right\}^{\frac{1}{2}} / \phi(z_i) \quad .$$

□

Note, em (3.13), que a distribuição empírica ponderada apresenta uma maior variabilidade em relação à distribuição empírica não ponderada. Isto proporciona uma maior sensibilidade do gráfico normal ponderado para detectar possíveis violações das suposições distribucionais.

3.2.2 Ajustes finais

O gráfico de probabilidade normal ponderado precisa ainda ser ajustado para os pontos finais, pois, caso contrário, $\Phi^{-1}[F_n^*(x)]$ não pode ser calculado para os maiores z_i 's ($\Phi^{-1}[1] \approx \infty$). Dempster e Ryan (1985) sugerem o procedimento descrito a seguir.

Seja $w_{(i)}$ o peso que corresponde à i -ésima estatística de ordem $z_{(i)}$. Então, devemos

fazer

$$F_n^*(x) = \frac{\frac{5}{8}w(1)}{\frac{9}{8}w(1) + \dots + w(i) + \dots + \frac{9}{8}w(n)}$$

para $x \leq z(1)$, e

$$F_n^*(x) = \frac{\frac{9}{8}w(1) + w(2) + \dots + \frac{1}{2}w(i)}{\frac{9}{8}w(1) + \dots + w(i) + \dots + \frac{9}{8}w(n)}$$

para $x \in (z_{(i-1)}, z_{(i)}]$.

Para os gráficos de probabilidade normal ponderados deve-se considerar a variância de $\hat{\gamma}_i$, dada por Lange e Ryan (1989), que assume que todos os parâmetros do vetor $\underline{\beta}$ são conhecidos, ou seja:

$$Var(\hat{\gamma}_i) = \sigma^2 \mathbf{BZ}'_i \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{B} \quad . \quad (3.14)$$

Iremos considerar ainda, nos conjuntos de dados analisados neste trabalho, uma região crítica mais rigorosa, (mais estreita) que a apresentada no resultado 3.7,

$$z_i \pm \left(1 + \frac{v}{m^2}\right)^{\frac{1}{2}} \{(1/n)\Phi(z_i)[1 - \Phi(z_i)]\}^{\frac{1}{2}} / \phi(z_i) \quad ,$$

com o valor 1 em vez de 1.96. Lange e Ryan (1989) sugerem ainda uma correção na região crítica para compensar pelas estimativas usadas, ao invés dos parâmetros. Com este ajuste, para valores mais altos ou mais baixos de z_i 's as regiões críticas são aproximadamente iguais às regiões críticas sem ajuste. Nos valores centrais de z_i 's, entretanto, o intervalo é mais estreito, com o ajuste. Neste trabalho, não utilizaremos esta correção, pois é de difícil implementação computacional. Entretanto, temos interesse em, futuramente, desenvolver ainda mais os gráficos de probabilidade normal ponderados fazendo uso do ajuste sugerido.

3.3 Exemplos

3.3.1 Potthoff e Roy (1964)

Consideremos os dados indicados na tabela 3.1, extraídos de Little e Rubin (1987), referentes a um experimento analisado por Potthoff e Roy (1964). Neste conjunto de dados, as observações foram obtidas ao longo das idades de meninos e meninas para uma certa variável de interesse. Algumas observações (entre parênteses) foram retiradas com o objetivo de se estudar a metodologia na presença de dados não balanceados.

Meninas Unidades	Idade (em anos)				Meninos Unidades	Idade (em anos)			
	8	10	12	14		8	10	12	14
1	21	20	21.5	23	1	26	25	29	31
2	21	21.5	24	25.5	2	21.5	(22.5)	23	26.5
3	20.5	(24)	24.5	26	3	23	22.5	24	27.5
4	23.5	24.5	25	26.5	4	25.5	27.5	26.5	27
5	21.5	23	22.5	23.5	5	20	(23.5)	22.5	26
6	20	(21)	21	22.5	6	24.5	25.5	27	28.5
7	21.5	22.5	23	25	7	22	22	24.5	26.5
8	23	23	23.5	24	8	24	21.5	24.5	25.5
9	20	(21)	22	21.5	9	23	20.5	31	26
10	16.5	(19)	19	19.5	10	27.5	28	31	31.5
11	24.5	25	28	28	11	23	23	23.5	25
					12	21.5	(23.5)	24	28
					13	17	(24.5)	26	29.5
					14	22.5	25.5	25.5	26
					15	23	24.5	26	30
					16	22	(21.5)	23.5	(25)

Tabela 3.1: Dados extraídos de Little e Rubin

O modelo que melhor se ajustou aos dados define uma reta para cada grupo, com um efeito aleatório (inclinação individual) e considera os erros totalmente independentes, ou seja,

$$V(\epsilon_i) = \sigma^2 \mathbf{I}_i \quad .$$

Relacionamos, nas tabelas 3.2 e 3.3, as inclinações individuais estimadas (e.a) com seus desvios padrões (s.d), para as 11 meninas e os 16 meninos, respectivamente.

Meninas		
i	e.a	s.d
1	-1.051	0.6605
2	0.3420	0.6605
3	0.7386	0.7452
4	1.9490	0.6605
5	0.0205	0.6605
6	-1.307	0.7452
7	0.3420	0.6605
8	0.6634	0.6605
9	-1.307	0.7452
10	-3.626	0.7452
11	3.235	0.6605

Tabela 3.2: Efeitos aleatórios estimados para as meninas

Meninos		
i	e.a	s.d
1	2.372	0.6605
2	-1.294	0.7452
3	-0.628	0.6605
4	1.4080	0.6605
5	-1.976	0.7452
6	1.194	0.6605
7	-1.057	0.6605
8	-0.949	0.6605
9	0.1222	0.6605
10	3.873	0.6605
11	-1.164	0.6605
12	-0.612	0.7452
13	-0.885	0.7452
14	-0.092	0.6605
15	0.7651	0.6605
16	-1.076	0.8739

Tabcla 3.3: Efeitos aleatórios estimados para os meninos

Com estes efeitos aleatórios padronizados segundo (3.10), foi obtida a distribuição acumulada empírica ponderada F_n^* . Desta forma, seguindo o procedimento da seção anterior, obtivemos o seguinte gráfico:

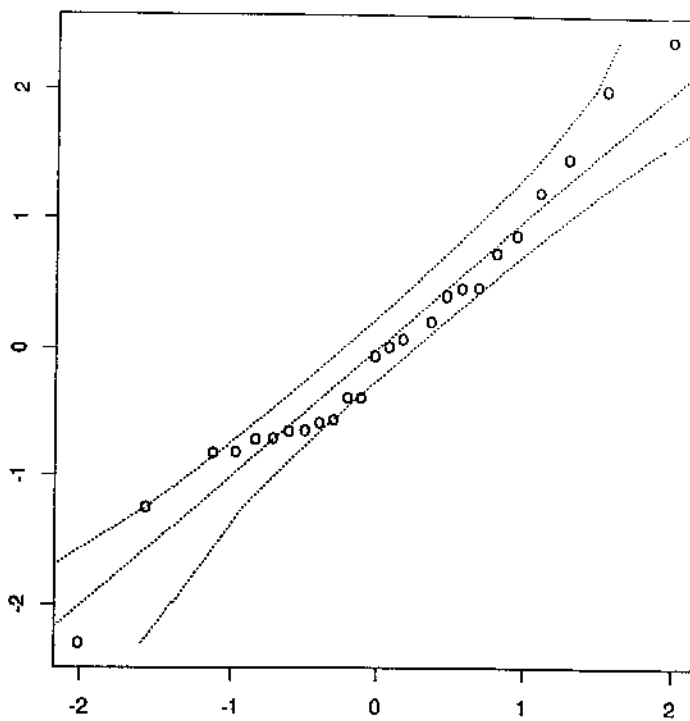


Figura 3-1: Gráfico de probabilidade normal ponderado do efeito aleatório

Podemos notar, pela análise do gráfico, que o modelo se ajusta satisfatoriamente aos dados.

3.3.2 Jones (1990)

Para o conjunto de dados, não balanceado, analisado em Jones (1990), foi ajustado um modelo considerando um vetor de efeitos aleatórios $\tilde{\gamma}_i$ tridimensional. Podemos constatar a boa adequação do modelo pela análise dos 3 gráficos de probabilidade normal ponderados, abaixo:

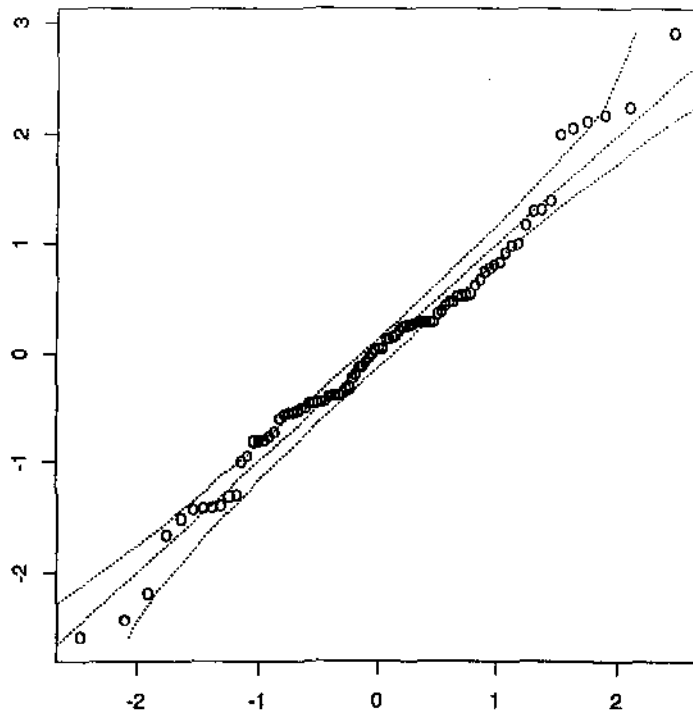


Figura 3-2: Gráfico de probabilidade normal ponderado do efeito aleatório γ_1

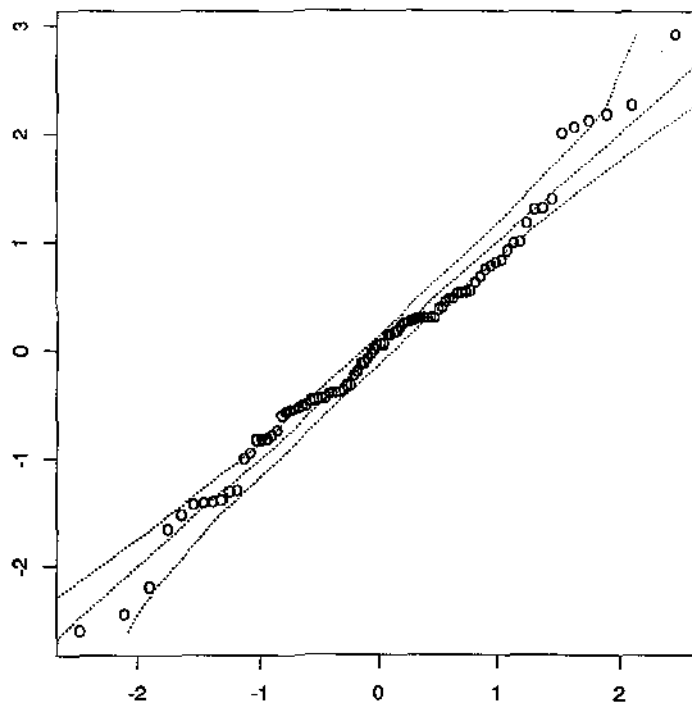


Figura 3-3: Gráfico de probabilidade normal ponderado do efeito aleatório γ_2

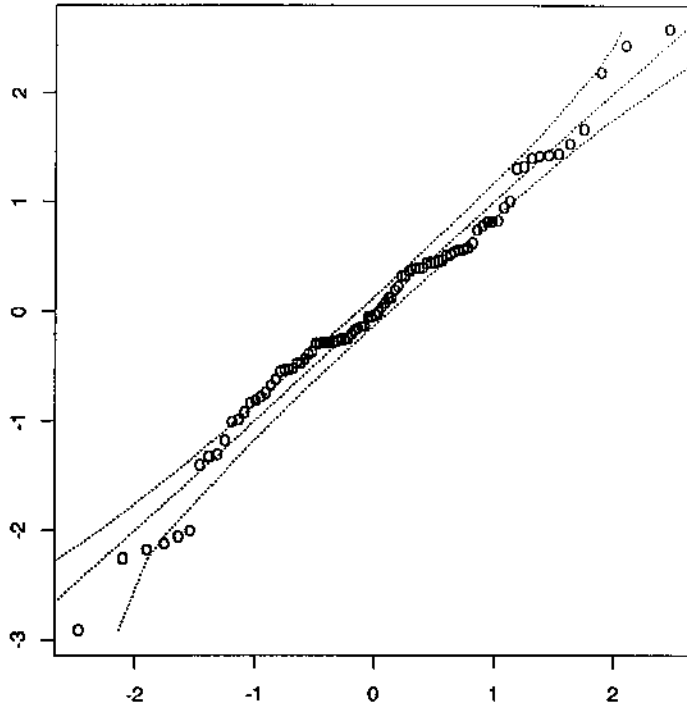


Figura 3-4: Gráfico de probabilidade normal ponderado do efeito aleatório γ_3

Capítulo 4

O Filtro de Kalman na Estimação e Predição do Modelo Laird-Ware

4.1 Introdução

O filtro de Kalman (FK) é um processo comumente empregado em áreas da Engenharia e outras ciências físicas desde a década de 60. Kalman (1960) introduziu esta nova abordagem ao problema clássico de estimação, desenvolvendo um método recursivo de estimação e predição de séries temporais, executando um algoritmo para cada ponto no tempo. Trabalhando com Bucy (1961) esses resultados foram generalizados a processos aleatórios e a implementação computacional foi facilitada. A praticidade da abordagem de Kalman logo tornou-se imensamente popular entre engenheiros.

Esta metodologia foi inicialmente utilizada na indústria aeroespacial para estimar a posição de foguetes. Supondo que as observações estão sujeitas a erros aleatórios, estimativas aperfeiçoadas da posição do foguete são obtidas combinando informações sobre a posição estimada por ocasião do tempo anterior e a posição observada no tempo atual. Este procedimento permite que as estimativas da posição do foguete sejam continuamente atualizadas, assim que os dados são coletados. O termo ‘filtro’ é usado quando a estimação da posição do foguete é executada no tempo da mais recente observação. Muitos trabalhos subsequentes ao artigo de Kalman apresentaram aplicações deste método a usos práticos como controle estatístico de qualidade (Phadke, 1981), navegação aérea e

submarina (Titus, 1977) e outras aplicações espaciais (Bucy e Joseph, 1968).

Schweppe (1965) mostrou que, se os erros tiverem distribuição normal, o processo FK pode ser usado para calcular funções de verossimilhança, como veremos mais adiante. Neste momento, foram abertas as portas para a aplicação do FK no ambiente estatístico, pois a possibilidade de calcular verossimilhanças nos permite obter estimativas de quantidades desconhecidas.

Devido à dificuldade de interpretação, pelos estatísticos, da literatura envolvendo FK, que utiliza notação, linguagem e estilo diferentes, somente mais tarde, a partir da década de 70, é que se deu a devida atenção a este método. Na literatura estatística, o primeiro tratamento ao FK foi dado por Duncan e Horn (1972) baseado em teoria de regressão com coeficientes aleatórios, ou seja, considerando as quantidades a serem estimadas como sendo variáveis aleatórias. A partir daí, muitos pesquisadores se esforçaram no sentido de obter derivações do procedimento recursivo FK de formas mais simplificadas. Desde a derivação original de Kalman (1960), fazendo uso do teorema da projeção ortogonal, até Jones (1993), que se baseou em esperanças condicionais, apareceram outras interessantes derivações. Wegman (1982) propõe uma discussão mais matemática do FK enfatizando uma abordagem de equação diferencial estocástica. Diderich (1985) deriva as equações do FK a partir de um estimador misto de Goldberger-Theil. Seguindo a abordagem bayesiana à teoria de regressão, temos o excelente artigo de Meinhold e Singpurwalla (1983), além dos trabalhos de Broemeling (1985) e Welch (1986). Atualmente, o FK é um método amplamente divulgado na literatura, pois percebeu-se que ele é potencialmente de grande utilidade na Estatística, principalmente nas áreas de Modelos Lineares de Regressão e de Séries Temporais.

Na seção 4.2 desenvolveremos a teoria do Filtro de Kalman como um procedimento de predição recursiva. Na seção 4.3 introduziremos uma forma de representar modelos em espaço de estados e enunciaremos os passos da recursão de Kalman utilizando o embasamento teórico desenvolvido na seção 4.2. Na seção 4.4 discutiremos duas formas de utilização do Filtro de Kalman ao modelo de Laird-Ware: a predição de efeitos aleatórios

e a estimação de parâmetros, abordando a associação entre a recursão de Kalman e a decomposição de Cholesky da matriz da estrutura de covariância do modelo de Laird-Ware.

4.2 Um Procedimento de Predição Recursiva

Sejam y_t e β_t variáveis aleatórias, associadas ao tempo t , sendo y_t observável e β_t , denominada estado da natureza, variável não observável. Uma equação define a relação entre o observado e o estado da natureza no tempo t :

$$y_t = H_t \beta_t + \omega_t \quad , \quad (4.1)$$

onde H_t é constante conhecida e ω_t é variável aleatória, com distribuição $N(0, R_t)$. Uma outra equação estabelece a relação entre o estado da natureza no tempo t e no tempo anterior, $t - \delta t$:

$$\beta_t = \Phi_t \beta_{t-\delta t} + v_t \quad , \quad (4.2)$$

onde Φ_t é constante conhecida e v_t é variável aleatória, com distribuição $N(0, \Sigma_t)$. Considera-se v_t independente de ω_t e v_t independente de $v_{t'}$ para $t' < t$.

Seja $Y_t = (y_1, y_2, \dots, y_t)'$, o vetor de observações até o tempo t . Num ponto de tempo t consideramos a distribuição de probabilidade de β_t , dado $Y_{t-\delta t}$, como a distribuição a priori, e dado Y_t , como a distribuição a posteriori.

No que segue, vamos definir a distribuição a posteriori de β_t , supondo conhecimento da distribuição a posteriori de $\beta_{t-\delta t}$ (ou seja, a partir da distribuição de probabilidade de $\beta_{t-\delta t}$, dado $Y_{t-\delta t}$). Este será o nosso objetivo.

Portanto, considere conhecida a distribuição a posteriori de $\beta_{t-\delta t}$:

$$(\beta_{t-\delta t} / Y_{t-\delta t}) \sim N(\hat{\beta}_{t-\delta t}, S_{t-\delta t}) \quad . \quad (4.3)$$

Usando (4.2), temos que

$$E[\beta_t / Y_{\sim t-\delta t}] = \Phi_t \hat{\beta}_{t-\delta t}$$

e

$$V[\beta_t / Y_{\sim t-\delta t}] = \Phi_t S_{t-\delta t} \Phi_t' + \sum_t \quad ,$$

e desta forma, temos definida a distribuição de probabilidade a priori de β_t :

$$(\beta_t / Y_{\sim t-\delta t}) \sim N(\Phi_t \hat{\beta}_{t-\delta t}, \Phi_t S_{t-\delta t} \Phi_t' + \sum_t) \quad . \quad (1) \quad (4.4)$$

Usando a esperança $\Phi_t \hat{\beta}_{t-\delta t}$, da distribuição a priori de β_t e usando (4.1), $H_t \Phi_t \hat{\beta}_{t-\delta t}$ é uma predição para a observação no tempo t , prévia à sua observação. Assim, o erro desta predição é definido por:

$$e_t = y_t - H_t \Phi_t \hat{\beta}_{t-\delta t} \quad , \quad (4.5)$$

ou seja, a diferença entre a observação no tempo t e a predição, com base em $Y_{\sim t-\delta t}$. Dado que H_t e Φ_t são constantes conhecidas e que $\hat{\beta}_{t-\delta t}$ já está definido por ocasião do tempo t , observar y_t é equivalente a observar e_t . Podemos reexpressar e_t :

$$\begin{aligned} e_t &= y_t - H_t \Phi_t \hat{\beta}_{t-\delta t} \\ &= (H_t \beta_t + \omega_t) - H_t \Phi_t \hat{\beta}_{t-\delta t} \quad , \text{ por (4.1)} \\ &= H_t (\beta_t - \Phi_t \hat{\beta}_{t-\delta t}) + \omega_t \quad . \end{aligned}$$

Desta forma, a distribuição de e_t , dados β_t e $Y_{\sim t-\delta t}$, é normal, com esperança $H_t (\beta_t - \Phi_t \hat{\beta}_{t-\delta t})$ e variância R_t .

¹Utilizamos $\Phi_t S_{t-\delta t} \Phi_t'$ ao invés de $\Phi_t^2 S_{t-\delta t}$, propositalmente, porque desta forma os resultados serão adequados a situações onde tenhamos vetores e matrizes, ao invés de escalares.

Até este ponto, temos que

$$(\beta_t / \underset{\sim}{Y}_{t-\delta t}) \sim N(\Phi_t \hat{\beta}_{t-\delta t}, \Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t)$$

$$e$$

$$(e_t / \beta_t, \underset{\sim}{Y}_{t-\delta t}) \sim N(H_t(\beta_t - \Phi_t \hat{\beta}_{t-\delta t}), R_t) .$$

Precisamos definir, a seguir, primeiramente, a distribuição conjunta de $(\beta_t / \underset{\sim}{Y}_{t-\delta t})$ e $(e_t / \underset{\sim}{Y}_{t-\delta t})$, e, finalmente, a distribuição de β_t , condicionada a e_t e $\underset{\sim}{Y}_{t-\delta t}$. Esta última é a distribuição de β_t , condicionada a y_t e $\underset{\sim}{Y}_{t-\delta t}$, ou seja, a posteriori de β_t . Um conhecido resultado sobre normais multivariadas é destacado abaixo, sem prova, para a obtenção das distribuições mencionadas.

Resultado 4.1:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right] \Leftrightarrow$$

$$\Leftrightarrow (X_1 / X_2 = x_2) \sim N(\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2), \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}) \quad e \quad X_2 \sim N(\mu_2, \Sigma_{22}) .$$

□

Considere estes outros resultados:

Resultado 4.2:

A distribuição conjunta de β_t e e_t , dado $\underset{\sim}{Y}_{t-\delta t}$ é:

$$\begin{bmatrix} \beta_t / \underset{\sim}{Y}_{t-\delta t} \\ e_t / \underset{\sim}{Y}_{t-\delta t} \end{bmatrix} \sim N \left[\begin{pmatrix} \Phi_t \hat{\beta}_{t-\delta t} \\ 0 \end{pmatrix}, \begin{pmatrix} \Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t & (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' \\ H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) & H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' + R_t \end{pmatrix} \right] .$$

Prova :

Temos

$$(\beta_t / \underset{\sim}{Y}_{t-\delta t}) \sim N(\Phi_t \hat{\beta}_{t-\delta t}, \Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t)$$

e

$$(e_t / \beta_t, \underset{\sim}{Y}_{t-\delta t}) \sim N(H_t(\beta_t - \Phi_t \hat{\beta}_{t-\delta t}), R_t) .$$

Usando o resultado 4.1, no sentido inverso, com $\beta_t / \underset{\sim}{Y}_{t-\delta t}$ no lugar de X_2 e $e_t / \beta_t, \underset{\sim}{Y}_{t-\delta t}$ no lugar de X_1 / X_2 , temos:

$$\mu_2 = \Phi_t \hat{\beta}_{t-\delta t} \quad e \quad \Sigma_{22} = \Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t . \quad (4.6)$$

Precisamos ainda definir μ_1, Σ_{12} e Σ_{11} :

$$\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2) = H_t (\beta_t - \Phi_t \hat{\beta}_{t-\delta t})$$

$$\Rightarrow \mu_1 = 0$$

$$\Rightarrow \Sigma_{12} \Sigma_{22}^{-1} = H_t$$

$$\Rightarrow \Sigma_{12} = H_t \Sigma_{22} = H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) \quad , \text{ por (4.6).}$$

$$\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = R_t$$

$$\Rightarrow \Sigma_{11} = (H_t \Sigma_{22}) \Sigma_{22}^{-1} (H_t \Sigma_{22})' + R_t \quad , \text{ usando } \Sigma_{12} = H_t \Sigma_{22},$$

$$= H_t \Sigma_{22} H_t' + R_t =$$

$$= H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' + R_t \quad , \text{ por (4.6).}$$

□

Resultado 4.3:

A distribuição a posteriori de β_t é normal com esperança

$$\Phi_t \hat{\beta}_{t-\delta t} + (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' [H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' + R_t]^{-1} e_t$$

e variância

$$(\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) - (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' [H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' + R_t]^{-1} H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t)$$

Prova:

A partir do argumento direto do resultado 4.1, com $(\beta_t / \underset{\sim}{Y}_{t-\delta t})$ no lugar de X_1 e $e_t / \underset{\sim}{Y}_{t-\delta t}$ no lugar de X_2 , temos sua conjunta dada pelo resultado 4.2 e a condicional $(\beta_t / e_t, \underset{\sim}{Y}_{t-\delta t})$ com esperança $\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$, na notação do resultado 4.1, e neste contexto:

$$\Phi_t \hat{\beta}_{t-\delta t} + (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' [H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' + R_t]^{-1} e_t$$

e variância $\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$, na notação do resultado 4.2, e neste contexto:

$$(\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) - (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' [H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t) H_t' + R_t]^{-1} H_t (\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t)$$

□

Podemos sumarizar os resultados apresentados, usando a notação $(S_t / Y_{t-\delta t})$ para a quantidade $\Phi_t S_{t-\delta t} \Phi_t' + \Sigma_t$:

$$(\beta_{t-\delta t} / \underset{\sim}{Y}_{t-\delta t}) \sim N(\hat{\beta}_{t-\delta t}, S_{t-\delta t}) \quad (\text{posteriori de } \beta_{t-\delta t})$$

$$(\beta_t / \underset{\sim}{Y}_{t-\delta t}) \sim N(\Phi_t \hat{\beta}_{t-\delta t}, (S_t / \underset{\sim}{Y}_{t-\delta t})) \quad (\text{priori de } \beta_t)$$

$$\begin{aligned} (\beta_t / \underset{\sim}{Y}_t) &\sim N(\Phi_t \hat{\beta}_{t-\delta t} + (S_t / \underset{\sim}{Y}_{t-\delta t}) H_t' [H_t (S_t / \underset{\sim}{Y}_{t-\delta t}) H_t' + R_t]^{-1} e_t, \\ (S_t / \underset{\sim}{Y}_{t-\delta t}) &- (S_t / \underset{\sim}{Y}_{t-\delta t}) H_t' [H_t (S_t / \underset{\sim}{Y}_{t-\delta t}) H_t' + R_t]^{-1} H_t (S_t / \underset{\sim}{Y}_{t-\delta t})) \quad (\text{posteriori de } \beta_t) \end{aligned} \quad (4.7)$$

Um procedimento recursivo começaria no tempo zero, com valores iniciais $\hat{\beta}_0$ e S_0 , que

seriam as melhores avaliações sobre a média e variância de β_0 , ou seja, estaria definido:

$$\beta_0 \sim N(\hat{\beta}_0, S_0) \quad .$$

As distribuições a priori e a posteriori de β_1 seriam definidas conforme (4.7), e assim por diante, ponto a ponto, obtém-se nova distribuição a posteriori. Este procedimento é denominado Filtro de Kalman. A recursão deve ser executada até a última observação. Ao final do processo a esperança da distribuição a posteriori do estado da natureza será a nossa predição do estado da natureza.

4.3 Representação em Espaço de Estados e o Filtro de Kalman

A situação vista na seção anterior será agora apresentada para o caso mais geral de vetores de observações e vetores de estados da natureza. Novamente são consideradas duas equações para a definição de um fenômeno e a obtenção da distribuição a posteriori dos estados da natureza é o objetivo da recursão.

Denomina-se representação em espaço de estados uma forma de representar um determinado fenômeno a partir de um sistema de duas equações, uma equação para um vetor de observações e outra equação para um vetor de quantidades aleatórias, que se deseja prever. Estas serão definidas a seguir.

Seja \underline{y}_t um vetor aleatório observável no tempo t , que depende de um vetor aleatório, não-observável, $\underline{\beta}_t$, conhecido como *estado da natureza*. A relação entre \underline{y}_t e $\underline{\beta}_t$, conhecida como *equação de observação* é dada por:

$$\underline{y}_t = \mathbf{H}_t \underline{\beta}_t + \underline{\omega}_t \quad , \quad (4.8)$$

onde:

- \underline{y}_t é vetor aleatório, $n \times 1$, de observações no tempo t ;
- $\underline{\beta}_t$ é vetor aleatório, $b \times 1$, de estados no tempo t ;

• \mathbf{H}_t é matriz, $n \times b$, conhecida, que determina quais elementos, ou combinações de elementos do vetor de estados, são destacados;

• ω_t é vetor aleatório, $n \times 1$, de erros observacionais, com distribuição gaussiana, com vetor de médias zero e matriz de covariância \mathbf{R}_t , ou seja,

$$\omega_t \sim N(0, \mathbf{R}_t) \quad .$$

A equação conhecida como *equação do sistema* é dada por

$$\beta_t = \Phi(t; t - \delta t) \beta_{t-\delta t} + \vartheta_t \quad ,$$

ou melhor,

$$\beta_t = \Phi(t; t - \delta t) \beta_{t-\delta t} + \mathbf{G}_t v_t \quad , \quad (4.9)$$

onde:

• $\Phi(t; t - \delta t)$ é matriz conhecida, $b \times b$, de transição de estados do tempo $t - \delta t$ para o tempo t ;

• v_t é vetor, $b \times 1$, de incrementos aleatórios ao estado no tempo t , com distribuição normal padrão, ou seja,

$$v_t \sim N(0, \mathbf{I}) \quad ;$$

• \mathbf{G}_t : matriz, $b \times b$, que incorpora as covariâncias dos elementos de ϑ_t , ou melhor, \mathbf{G}_t é tal que

$$\vartheta_t \sim N(0, \mathbf{G}_t \mathbf{G}_t') \quad .$$

Considera-se ω_t e v_t independentes. Uma eficiente determinação das matrizes $\Phi(t; t - \delta t)$, \mathbf{H}_t , \mathbf{R}_t e \mathbf{G}_t depende acima de tudo de um bom conhecimento do comportamento do sistema físico que rege o fenômeno em estudo.

Um exemplo freqüentemente citado na literatura é o rastreamento da órbita de um

satélite ao redor da Terra. O estado da natureza $\hat{\beta}_t$, que desejamos conhecer, pode ser a posição e a velocidade do satélite no tempo t , com respeito a um sistema de coordenadas esféricas com origem no centro da Terra. Estas quantidades não podem ser medidas diretamente. No entanto, através de estações de rastreamento ao redor da Terra, nós podemos obter medidas (com erros observacionais) da distância entre a Terra e o satélite e dos ângulos da posição do mesmo em relação ao centro da Terra; estes seriam os componentes do vetor y_t . Os princípios de Geometria e Física, etc, que relacionam y_t e $\hat{\beta}_t$, podem ser incorporados na matriz H_t , enquanto que ω_t refletiria o erro de mensuração; a matriz $\Phi(t; t - \delta t)$ representaria a forma como a posição e a velocidade mudam no tempo de acordo com as leis físicas que regem os corpos celestes orbitantes, enquanto que $G_t v_t$ representariam os desvios em relação a essas leis, devido a fatores como a não-uniformidade do campo gravitacional da Terra.

Antes de definir formalmente o método recursivo do FK, uma sucinta descrição é dada sobre seu uso com relação ao exemplo acima. Tendo como objetivo o conhecimento do vetor de estados da natureza, o processo recursivo do FK se inicia com uma definição dos valores iniciais de $\hat{\beta}_0$ e de S_0 , onde $\hat{\beta}_0$ representa a estimativa inicial do vetor de estados no tempo zero e S_0 representa a matriz com valores iniciais para as variâncias e covariâncias deste vetor de estados, no tempo zero. Engenheiros aeroespaciais envolvidos com o problema poderão proporcionar esses valores com base em experiências passadas ou então, se houver pouca informação, estabelecer uma estimativa inicial com grandes variâncias associadas, ou seja, valores altos para as variâncias na matriz S_0 . Na etapa seguinte (tempo 1), $\hat{\beta}_0$ e S_0 , juntamente com a nova observação, são usados para estabelecer uma estimativa atualizada, $\hat{\beta}_1$, e uma nova matriz de covariâncias, S_1 . Portanto, na recursão, ao atingirmos o tempo t , temos $\hat{\beta}_{t-\delta t}$ e $S_{t-\delta t}$ calculados, onde $\hat{\beta}_{t-\delta t}$ denotará a melhor estimativa do estado $\hat{\beta}_{t-\delta t}$, dadas as observações até o tempo $t - \delta t$; e $S_{t-\delta t}$ a sua respectiva matriz de covariâncias; e uma nova observação, combinando-se esses elementos para a obtenção de $\hat{\beta}_t$ e S_t . O processo segue até a última observação. Experiências têm mostrado que, se as estimativas iniciais são aproximadamente corretas, razoáveis

estimativas do vetor de estado podem ser obtidas junto com a matriz de covariâncias das estimativas.

Abaixo seguem, de forma compacta, os passos da recursão de Kalman, onde a notação \mathbf{Y}_t se refere à matriz cujas colunas são os vetores de observações, do primeiro tempo até o tempo t . A recursão começa com valores do estado no tempo $t - \delta t$, $\hat{\beta}_{t-\delta t}$, e sua matriz de covariâncias $\mathbf{S}_{t-\delta t}$:

1. Calcula-se a esperança a priori de β_t ,

$$(\beta_t / \mathbf{Y}_{t-\delta t}) = \Phi_t \hat{\beta}_{t-\delta t} \quad ;$$

2. Calcula-se a matriz de covariâncias a priori de β_t

$$(\mathbf{S}_t / \mathbf{Y}_{t-\delta t}) = \Phi_t \mathbf{S}_{t-\delta t} \Phi_t' + \mathbf{G}_t \mathbf{G}_t' \quad ;$$

3. Calcula-se a predição do vetor de observações no tempo t , prévia a sua observação,

$$\hat{y}_t = \mathbf{H}_t \Phi_t \hat{\beta}_{t-\delta t} \quad ;$$

4. Calcula-se a diferença entre o vetor de observações e a predição desse vetor, obtido acima,

$$e_t = y_t - \mathbf{H}_t \Phi_t \hat{\beta}_{t-\delta t} \quad .$$

Esta quantidade é denominada *Inovação*;

5. Constrói-se a matriz de covariâncias de e_t dado $\mathbf{Y}_{t-\delta t}$,

$$\mathbf{V}_t = \mathbf{R}_t + \mathbf{H}_t (\mathbf{S}_t / \mathbf{Y}_{t-\delta t}) \mathbf{H}_t' \quad ;$$

6. Atualiza-se o vetor de estados (que é a esperança da distribuição a posteriori de β_t),

$$\hat{\beta}_t = \Phi_t \hat{\beta}_{t-\delta t} + \mathbf{K}_t e_t \quad ;$$

onde K_t é o ganho de Kalman e é dado por $K_t = (S_t/Y_{t-\delta t})H_t'V_t^{-1}$.

7. Calcula-se a matriz de covariâncias desse vetor atualizado (que é a matriz de variâncias e covariâncias da distribuição a posteriori de $\tilde{\beta}_t$),

$$S_t = (S_t/Y_{t-\delta t}) - K_t H_t (S_t/Y_{t-\delta t}) \quad .$$

As justificativas para esses passos se encontram no desenvolvimento teórico da seção anterior, notando-se que $G_t G_t'$ substitui a notação Σ_t , na equação do sistema (4.2).

4.4 Aplicação ao Modelo de Laird-Ware

Numa análise de dados, para valores arbitrários da estrutura de covariância, primeiramente, são calculadas as estimativas de $\tilde{\beta}$, σ^2 , do modelo Laird-Ware, e o valor da função de verossimilhança, correspondente a esses valores. Após a estabilidade dos valores da função de verossimilhança, temos as estimativas (finais) de todos os parâmetros do modelo Laird-Ware.

O Filtro de Kalman pode ser utilizado com o propósito de obtenção destas estimativas, numa primeira fase. Em seguida, usando as estimativas de todos os parâmetros fixos, o FK pode ser utilizado, agora, com o objetivo específico de obtenção de predições dos efeitos aleatórios das unidades, caracterizando a segunda fase. As duas fases de utilização do FK têm definições distintas do vetor de observações.

Nesta seção, as duas utilizações do Filtro de Kalman serão discutidas, e, por considerarmos mais adequado, começaremos com a explicação da segunda fase, quando já supomos disponíveis as estimativas de $\tilde{\beta}$, σ^2 e B , do modelo de Laird-Ware. Os passos da recursão de Kalman serão apresentados considerando uma estrutura de correlação serial AR(1) e erros observacionais. Para outras estruturas, pequenas alterações seriam necessárias.

4.4.1 Predição de efeitos aleatórios

Jones (1993) propõe o emprego do FK para a predição dos efeitos aleatórios do modelo Laird-Ware, usando estimativas dos parâmetros do modelo. Estas estimativas já estariam disponíveis pois, cronologicamente, esta é a última das duas fases de utilização da recursão. Portanto, através dos passos da recursão de Kalman, são utilizadas as estimativas $\hat{\beta}$, $\hat{\sigma}^2$ e \hat{B} .

Seja o modelo de Laird-Ware com estrutura AR(1) e erro observacional. Assim, ϵ_i é descrito como uma série temporal, seguindo um processo contínuo auto-regressivo de 1ª ordem (AR(1)), com pontos de tempo desigualmente espaçados, possivelmente com observações perdidas. O elemento de ϵ_i correspondente à observação realizada no tempo t é descrito como

$$\epsilon_{it} = \phi_t \epsilon_{i(t-\delta t)} + \eta_{it} \quad , \quad (4.10)$$

onde

- ϕ_t , o coeficiente auto-regressivo é dado por

$$\phi_t = \exp[-\alpha_0 \delta t] \quad , \quad (4.11)$$

sendo α_0 um valor positivo;

- η_{it} é um incremento aleatório no tempo t com média zero e variância

$$\sigma^2 Q_{it} = \sigma^2 [1 - \exp(-2\alpha_0 \delta t)] \quad . \quad (4.12)$$

Então, para uma unidade, no tempo t , temos

$$y_{it} = \underset{\sim}{x}'_{it} \underset{\sim}{\beta} + \underset{\sim}{z}'_{it} \underset{\sim}{\gamma}_i + \epsilon_{it} \quad ,$$

ou

$$y_{it} - \underset{\sim}{x}'_{it} \underset{\sim}{\beta} = \underset{\sim}{z}'_{it} \underset{\sim}{\gamma}_i + \epsilon_{it} \quad ,$$

onde x'_{it} é a t -ésima linha da matriz \mathbf{X}_i , e podemos utilizar a representação de espaço de estados, definindo a equação de observação:

$$y_{it} - x'_{it}\hat{\beta} = \begin{bmatrix} 1 & z'_{it} \end{bmatrix} \begin{bmatrix} \epsilon_{it} \\ \gamma_i \end{bmatrix} + \omega_{it} \quad , \quad (4.13)$$

onde z'_{it} é a t -ésima linha da matriz \mathbf{Z}_i e $\omega_{it} \sim N(0, \sigma^2\sigma_0^2)$. Para completar a representação do modelo Laird-Ware em espaço de estados, a equação de estado é dada por:

$$\begin{bmatrix} \epsilon_{it} \\ \gamma_{it} \end{bmatrix} = \begin{bmatrix} \phi_{\delta t} & 0' \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \epsilon_{i(t-\delta t)} \\ \gamma_{i(t-\delta t)} \end{bmatrix} + \begin{bmatrix} \eta_{it} \\ 0 \end{bmatrix} \quad . \quad (4.14)$$

Nota-se que a primeira linha deste sistema de equações é a equação (4.10), informando a correlação serial AR(1), e que os vetores γ_{it} não se alteram, com respeito ao tempo t . O erro observacional incorporado na equação de observação é representado por η_{it} .

Relacionando à notação do caso geral apresentado na seção 4.3, temos:

$$\begin{aligned} \beta_t &\Leftrightarrow \begin{bmatrix} \epsilon_{it} \\ \gamma_{it} \end{bmatrix} \quad , \\ y_t &\Leftrightarrow y_{it} - x'_{it}\hat{\beta} \quad , \\ \mathbf{H}_t &\Leftrightarrow \begin{bmatrix} 1 & z'_{it} \end{bmatrix} \quad , \\ \mathbf{R}_t &\Leftrightarrow \hat{\sigma}^2\sigma_0^2 \quad , \\ \Phi(t; t - \delta t) &\Leftrightarrow \begin{bmatrix} \phi_{\delta t} & 0' \\ 0 & \mathbf{I} \end{bmatrix} \quad , \\ \mathbf{G}_t\mathbf{G}'_t &\Leftrightarrow \begin{bmatrix} \hat{\sigma}^2 Q_{i\delta t} & 0' \\ 0 & \mathbf{0} \end{bmatrix} \quad . \end{aligned}$$

O ponto chave desta fase da utilização do FK é que usamos como valor observado o

resíduo $y_{it} - x'_{it}\hat{\beta}$, na impossibilidade de observar diretamente γ_i ou ϵ_{it} .

Para iniciar a recursão de Kalman são necessários valores iniciais para o vetor de estados e sua matriz de covariâncias. Uma sugestão é usar as definições do modelo Laird-Ware, utilizando as esperanças e variâncias das suposições distribucionais:

$$\begin{bmatrix} \epsilon_{i0} \\ \gamma_{i0} \end{bmatrix} \underset{\sim}{=} 0$$

e

$$S_0 = \hat{\sigma}^2 \begin{bmatrix} 1 & 0' \\ 0 & \hat{B} \end{bmatrix}.$$

Desta forma, a recursão de Kalman deve ser utilizada conforme definida acima, com o objetivo de obter a distribuição a posteriori de γ_i e sua matriz de covariâncias. A recursão é inicializada novamente para cada unidade. No caso de utilização de outras estruturas de correlação, o procedimento deve ser o mesmo, com pequenas alterações.

Reenunciaremos abaixo os passos da recursão de Kalman, já adaptada para a predição dos efeitos aleatórios, a partir das estimativas no tempo $t - \delta t$, $\begin{bmatrix} \hat{\epsilon}_{i(t-\delta t)} \\ \hat{\gamma}_{i(t-\delta t)} \end{bmatrix}$ e $S_{t-\delta t}$:

1. Calcula-se a esperança a priori de $\begin{bmatrix} \epsilon_{it} \\ \gamma_{it} \end{bmatrix}$,

$$\begin{bmatrix} \epsilon_{it} \\ \gamma_{it} \end{bmatrix} / Y_{t-\delta t} = \begin{bmatrix} \phi_{\delta t} & 0' \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{\epsilon}_{i(t-\delta t)} \\ \hat{\gamma}_{i(t-\delta t)} \end{bmatrix};$$

2. Calcula-se a matriz de covariâncias a priori de $\begin{bmatrix} \epsilon_{it} \\ \gamma_{it} \end{bmatrix}$,

$$(S_t / Y_{t-\delta t}) = \begin{bmatrix} \phi_{\delta t} & 0' \\ 0 & I \end{bmatrix} S_{t-\delta t} \begin{bmatrix} \phi_{\delta t} & 0' \\ 0 & I \end{bmatrix} + \begin{bmatrix} \hat{\sigma}^2 Q_{i\delta t} & 0' \\ 0 & 0 \end{bmatrix};$$

3. Calcula-se a predição do vetor de observações no tempo t , prévia à sua observação,

$$\hat{y}_{it} = \begin{bmatrix} 1 & z'_{it} \end{bmatrix} \begin{bmatrix} \phi_{\delta t} & 0' \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\epsilon}_{i(t-\delta t)} \\ \hat{\gamma}_{i(t-\delta t)} \end{bmatrix} ;$$

4. Calcula-se a diferença entre o vetor de observações e a predição desse vetor, obtida acima,

$$e_{it} = (y_{it} - x'_{it}\hat{\beta}) - \hat{y}_{it} ;$$

5. Constrói-se a matriz de covariâncias de e_t dado $\mathbf{Y}_{t-\delta t}$,

$$\mathbf{V}_t = \hat{\sigma}^2 \sigma_0^2 + \begin{bmatrix} 1 & z'_{it} \end{bmatrix} (\mathbf{S}_t / \mathbf{Y}_{t-\delta t}) \begin{bmatrix} 1 \\ z_{it} \end{bmatrix} ;$$

6. Atualiza-se o vetor de estados (que é a esperança da distribuição a posteriori de $\begin{bmatrix} \epsilon_{it} \\ \gamma_{it} \end{bmatrix}$),

$$\begin{bmatrix} \hat{\epsilon}_{it} \\ \hat{\gamma}_{it} \end{bmatrix} = \begin{bmatrix} \phi_{\delta t} & 0' \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\epsilon}_{i(t-\delta t)} \\ \hat{\gamma}_{i(t-\delta t)} \end{bmatrix} + \mathbf{K}_t e_{it} ,$$

onde \mathbf{K}_t , o ganho de Kalman, é dado por $\mathbf{K}_t = (\mathbf{S}_t / \mathbf{Y}_{t-\delta t}) \begin{bmatrix} 1 \\ z_{it} \end{bmatrix} \mathbf{V}_t^{-1}$.

7. Calcula-se a matriz de covariâncias desse vetor atualizado (que é a matriz de variâncias e covariâncias da distribuição a posteriori de $\begin{bmatrix} \epsilon_{it} \\ \gamma_{it} \end{bmatrix}$),

$$\mathbf{S}_t = (\mathbf{S}_t / \mathbf{Y}_{t-\delta t}) - \mathbf{K}_t \begin{bmatrix} 1 & z'_{it} \end{bmatrix} (\mathbf{S}_t / \mathbf{Y}_{t-\delta t}) .$$

4.4.2 Estimação de parâmetros

Jones (1993) propõe o emprego do FK, numa fase inicial, quando a partir de valores arbitrários de alguns parâmetros da estrutura de covariância, se objetiva obter as quantidades necessárias para o cálculo das estimativas $\hat{\beta}$ e $\hat{\sigma}^2$, além do respectivo valor da função de verossimilhança. Ou seja, o objetivo desta fase é obter a matriz (1) e o vetor (2), necessários para estimar β ,

$$\hat{\beta} = \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{V}_i^{-1} y_i \right) ,$$

a soma de quadrados ponderada total (3), necessária para estimar σ^2 ,

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (y_i - \mathbf{X}_i \hat{\beta})' \mathbf{V}_i^{-1} (y_i - \mathbf{X}_i \hat{\beta}) ,$$

onde $n = \sum_i n_i$ é o número total de observações para todas as unidades e o termo determinante (4) em ℓ

$$\ell = n \ln(2\pi\hat{\sigma}^2) + \sum_i \ln |\mathbf{V}_i| + n .$$

No decorrer dos passos da recursão o valor 1 é usado em lugar de σ^2 , como se tivesse havido uma mudança de escala. A representação em espaço de estados é um pouco diferente daquela apresentada na subseção anterior (efeitos aleatórios).

Em primeiro lugar, vamos abordar a associação entre a recursão de Kalman e a decomposição de Cholesky da matriz da estrutura de covariância do modelo Laird-Ware, \mathbf{V}_i . Considere o modelo de Laird-Ware

$$y_i = \mathbf{X}_i \beta + \mathbf{Z}_i \gamma_i + \epsilon_i$$

$$\Rightarrow \text{Var}(\mathbf{Z}_i \gamma_i + \epsilon_i) = \sigma^2 \mathbf{V}_{i(n_i \times n_i)} .$$

Aplicando a fatorização de Cholesky,

$$\mathbf{V}_i = \mathbf{T}'_i \mathbf{T}_i \quad ,$$

onde \mathbf{T}_i é uma matriz triangular superior, teremos

$$\mathbf{V}_i^{-1} = \mathbf{T}_i^{-1} (\mathbf{T}'_i)^{-1} \quad ,$$

denominada fatorização de Cholesky reversa, pois é o produto da matriz triangular superior seguida de sua transposta.

Seja

$$\mathbf{L}_i = (\mathbf{T}'_i)^{-1} \quad . \quad (4.15)$$

Resultado 4.4:

Considere o modelo Laird-Ware

$$\tilde{y}_i = \tilde{\epsilon}_i \quad ,$$

com estrutura AR(1), sem erro observacional e observações igualmente espaçadas. Então, o vetor, $n_i \times 1$, das inovações divididas pela raiz quadrada das respectivas variâncias, da recursão de Kalman, é exatamente a pré-multiplicação do vetor \tilde{y}_i pela matriz \mathbf{L}_i (fatorização de Cholesky reversa de \mathbf{V}_i^{-1}). Além disso, $|\mathbf{L}_i|$ é igual ao produto dos inversos das raízes quadradas das variâncias das inovações.

Prova:

O modelo de Laird-Ware, com observações igualmente espaçadas, define

$$\tilde{y}_i = \tilde{\epsilon}_i \quad ,$$

onde,

$$\epsilon_{it} = \phi \epsilon_{i(t-1)} + \eta_{it}$$

e

$$\text{Var}(\eta_{it}) = 1 - \phi^2 \quad .$$

Desta forma,

$$\mathbf{V}_i = \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n_i-2} & \phi^{n_i-1} \\ & 1 & \phi & \phi^2 & & \phi^{n_i-2} \\ & & \ddots & \ddots & \ddots & \vdots \\ & & & \ddots & \ddots & \phi^2 \\ & & & & 1 & \phi \\ & & & & & 1 \end{bmatrix} \quad .$$

Seja a decomposição de Cholesky de \mathbf{V}_i : $\mathbf{V}_i = \mathbf{T}'_i \mathbf{T}_i$; então,

$$\mathbf{L}_i = (\mathbf{T}'_i)^{-1} = \frac{1}{\sqrt{1-\phi^2}} \begin{bmatrix} \sqrt{1-\phi^2} & 0 & 0 & \dots & 0 & 0 \\ -\phi & 1 & 0 & \dots & 0 & 0 \\ 0 & -\phi & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 & 0 \\ 0 & \dots & \dots & 0 & -\phi & 1 \end{bmatrix}$$

e

$$\mathbf{L}_i \tilde{\mathbf{y}}_i = \frac{1}{\sqrt{1-\phi^2}} \begin{bmatrix} \sqrt{1-\phi^2} y_{i1} \\ y_{i2} - \phi y_{i1} \\ y_{i3} - \phi y_{i2} \\ \vdots \\ y_{in_i} - \phi y_{i(n_i-1)} \end{bmatrix} \quad . \quad (4.16)$$

Os passos da recursão de Kalman para este modelo são:

1.

$$(\hat{\epsilon}_t / \mathbf{Y}_{t-1}) = \phi \hat{\epsilon}_{t-1} \quad ;$$

2.

$$(\mathbf{S}_t/\mathbf{Y}_{t-1}) = \phi^2 \mathbf{S}_{t-1} + (1 - \phi^2) \quad ;$$

3.

$$\hat{y}_t = \phi \hat{\epsilon}_{t-1} \quad ;$$

4.

$$e_t = y_t - \phi \hat{\epsilon}_{t-1} \quad ;$$

5.

$$\mathbf{V}_t = (\mathbf{S}_t/\mathbf{Y}_{t-1}) = \phi^2 \mathbf{S}_{t-1} + (1 - \phi^2) \quad ;$$

6.

$$\begin{aligned} \hat{\epsilon}_t &= \phi \hat{\epsilon}_{t-1} + \frac{\phi^2 \mathbf{S}_{t-1} + (1 - \phi^2)}{\phi^2 \mathbf{S}_{t-1} + (1 - \phi^2)} (y_t - \phi \hat{\epsilon}_{t-1}) \\ &= y_t \quad ; \end{aligned}$$

7.

$$\mathbf{S}_t = \phi^2 \mathbf{S}_{t-1} + (1 - \phi^2) - \frac{(\phi^2 \mathbf{S}_{t-1} + (1 - \phi^2))^2}{\phi^2 \mathbf{S}_{t-1} + (1 - \phi^2)} = 0 \quad .$$

Note que o valor constante zero para \mathbf{S}_t , no sétimo passo, implica na simplificação $\mathbf{V}_t = 1 - \phi^2$ no quinto passo. Assim, a primeira passagem da recursão, para este modelo, fornecerá os seguintes resultados: (considerando $\epsilon_0 = 0$ e $\mathbf{S}_0 = 1$)

1.

$$(\hat{\epsilon}_1/\mathbf{Y}_0) = 0 \quad ;$$

2.

$$(\mathbf{S}_1/\mathbf{Y}_0) = \phi^2 + (1 - \phi^2) = 1 \quad ;$$

3.

$$\hat{y}_1 = 0 \quad ;$$

4.

$$e_1 = y_1 \quad ;$$

5.

$$V_1 = 1 \quad ;$$

6.

$$\hat{\epsilon}_1 = y_1 \quad ;$$

7.

$$S_1 = 0 \quad .$$

A segunda passagem fornecerá

1.

$$(\hat{\epsilon}_2/Y_1) = \phi y_1 \quad ;$$

2.

$$(S_2/Y_1) = 1 - \phi^2 \quad ;$$

3.

$$\hat{y}_2 = \phi y_1 \quad ;$$

4.

$$e_2 = y_2 - \phi y_1 \quad ;$$

5.

$$V_2 = 1 - \phi^2 \quad ;$$

6.

$$\hat{\epsilon}_2 = y_2 \quad ;$$

7.

$$S_2 = 0 \quad .$$

As passagens seguintes fornecerão

$$e_j = y_j - \phi y_{j-1} \quad ,$$

$$\mathbf{V}_j = 1 - \phi^2 \quad ,$$

$$\hat{\epsilon}_j = y_j \quad ,$$

$$\mathbf{S}_j = 0 \quad ,$$

para $j = 2, \dots, n_i$.

Portanto, o vetor das inovações divididas pelas respectivas raízes quadradas das variâncias é:

$$\frac{1}{\sqrt{1 - \phi^2}} \begin{bmatrix} \sqrt{1 - \phi^2} y_{i1} \\ y_{i2} - \phi y_{i1} \\ y_{i3} - \phi y_{i2} \\ \vdots \\ y_{in_i} - \phi y_{i(n_i-1)} \end{bmatrix} \quad ,$$

que é igual ao vetor (4.16). Além disso, como podemos comprovar pelo exemplo acima, $|\mathbf{L}_i|$ é igual ao produto dos inversos da raiz quadrada das variâncias das inovações, isto é,

$$|\mathbf{L}_i| = 1 \times \frac{1}{\sqrt{1 - \phi^2}} \times \frac{1}{\sqrt{1 - \phi^2}} \times \dots \quad .$$

□

O resultado 4.4 pode ser um pouco mais abrangente, como vemos a seguir:

Resultado 4.5:

Sob as condições do resultado 4.4, se usarmos um vetor $(n_i \times 1)$, \underline{z} , como vetor de observações, no passo 4, a recursão de Kalman produzirá o produto $\mathbf{L}_i \underline{z}$, através das inovações e sua variâncias.

Prova:

É suficiente notar que a prova do resultado 4.4 seria idêntica, quando os elementos

do vetor \tilde{c} fossem utilizados ao invés dos elementos de \tilde{y}_i , no quarto passo da recursão de Kalman.

□

Considerando agora um modelo de Laird-Ware qualquer, temos um resultado direcionado aos nossos interesses.

Resultado 4.6:

Considere o modelo de Laird-Ware. Ao utilizarmos a recursão de Kalman, definida na subseção anterior (efeitos aleatórios), usando como valores observados, no passo 4, os elementos de \tilde{y}_i , o vetor das inovações divididas pelas raízes quadradas de suas variâncias é exatamente

$$\mathbf{L}_i \tilde{y}_i \quad ,$$

onde $\mathbf{L}_i = (\mathbf{T}'_i)^{-1}$ é a matriz triangular inferior da decomposição reversa de Cholesky de \mathbf{V}_i^{-1} . Repetindo a recursão de Kalman, usando como valores observados, qualquer coluna de \mathbf{X}_i , obtemos a pré-multiplicação desta coluna pela matriz \mathbf{L}_i .

□

Vamos, agora, analisar a associação entre a decomposição de Cholesky de \mathbf{V}_i e as estimativas dos parâmetros do modelo Laird-Ware.

Resultado 4.7:

As quantidades $\mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{X}_i$, $\mathbf{X}'_i \mathbf{V}_i^{-1} \tilde{y}_i$ e $\tilde{y}'_i \mathbf{V}_i^{-1} \tilde{y}_i$ podem ser expressas, respectivamente, como

$$\sum_{k=1}^{n_i} \left[l_{k1} \tilde{x}'_1 + \cdots + l_{kk} \tilde{x}'_k \right]' \left[l_{k1} \tilde{x}'_1 + \cdots + l_{kk} \tilde{x}'_k \right] \quad , \quad (4.17)$$

$$\sum_{k=1}^{n_i} \left[l_{k1} \tilde{x}'_1 + \cdots + l_{kk} \tilde{x}'_k \right]' \left[l_{k1} y_1 + \cdots + l_{kk} y_k \right] \quad , \quad (4.18)$$

e

$$\sum_{k=1}^{n_i} \left[l_{k1} y_1 + \cdots + l_{kk} y_k \right]' \left[l_{k1} y_1 + \cdots + l_{kk} y_k \right] \quad , \quad (4.19)$$

onde l_{kj} é o j -ésimo elemento da k -ésima linha, l'_k , da matriz \mathbf{L}_i , e \tilde{x}'_k é a k -ésima linha da matriz \mathbf{X}_i .

Prova:

L_i pode ser expressa como uma soma de matrizes:

$$\begin{bmatrix} l'_{11} \\ \tilde{0}' \\ \tilde{0}' \\ \vdots \\ \tilde{0}' \end{bmatrix} + \begin{bmatrix} \tilde{0}' \\ l'_{22} \\ \tilde{0}' \\ \vdots \\ \tilde{0}' \end{bmatrix} + \dots + \begin{bmatrix} \tilde{0}' \\ \tilde{0}' \\ \vdots \\ \tilde{0}' \\ l'_{nn} \end{bmatrix},$$

onde $l'_j = (l_{j1}, \dots, l_{jj}, 0, \dots, 0)$, porque L_i (4.15) é triangular inferior. Desta forma, $L_i X_i$ também pode ser expressa como uma soma:

$$\begin{aligned} L_i X_i &= \begin{bmatrix} l_{11} x'_1 \\ \tilde{0}' \\ \tilde{0}' \\ \vdots \\ \tilde{0}' \end{bmatrix} + \begin{bmatrix} \tilde{0}' \\ l_{21} x'_1 + l_{22} x'_2 \\ \tilde{0}' \\ \vdots \\ \tilde{0}' \end{bmatrix} + \dots + \begin{bmatrix} \tilde{0}' \\ \tilde{0}' \\ \vdots \\ \tilde{0}' \\ l_{n1} x'_1 + \dots + l_{nn} x'_n \end{bmatrix} \\ &= A_1 + \dots + A_n \quad . \end{aligned}$$

Por sua vez, o produto

$$\begin{aligned} (L_i X_i)' (L_i X_i) &= X_i' L_i' L_i X_i \\ &= X_i' T_i^{-1} (T_i')^{-1} X_i \\ &= X_i' V_i^{-1} X_i \end{aligned}$$

pode ser expresso como a soma

$$\sum_{k=1}^{n_i} \sum_{j=1}^{n_i} A_k' A_j = \sum_{k=1}^{n_i} A_k' A_k \quad ,$$

porque, para $k \neq j$, $\mathbf{A}'_k \mathbf{A}_j = \mathbf{0}$, sendo

$$\mathbf{A}'_k \mathbf{A}_k = \begin{bmatrix} \tilde{0}' \\ \vdots \\ \tilde{0}' \\ l_{k1} x'_1 + \dots + l_{kk} x'_k \\ \tilde{0}' \\ \vdots \\ \tilde{0}' \end{bmatrix}' \begin{bmatrix} \tilde{0}' \\ \vdots \\ \tilde{0}' \\ l_{k1} x'_1 + \dots + l_{kk} x'_k \\ \tilde{0}' \\ \vdots \\ \tilde{0}' \end{bmatrix}$$

$$= \left[l_{k1} x'_1 + \dots + l_{kk} x'_k \right]' \left[l_{k1} x'_1 + \dots + l_{kk} x'_k \right]$$

(matrizes de dimensão $n_i \times n_i$). De maneira análoga obtemos os outros dois resultados.

□

O método proposto por Jones (1993) sugere que a recursão de Kalman seja aplicada a cada coluna de \mathbf{X}_i e ao vetor de observações y_i independentemente, pois conforme o resultado 4.6, a recursão de Kalman tem meios de produzir a pré-multiplicação das colunas de \mathbf{X}_i e y_i pela matriz \mathbf{L}_i . Isto pode ser feito simultaneamente, definindo uma linha de observações no passo 4 da recursão de Kalman

$$\begin{bmatrix} x'_{ik} & y_{ik} \end{bmatrix} ,$$

de dimensão $b + 1$, onde x'_{ik} é a k -ésima linha de \mathbf{X}_i e replicando o vetor de estados em $b + 1$ colunas, ou seja, uma matriz de estados, denotada por $\mathbf{\Omega}$. Podemos notar que, a cada passagem da recursão, uma nova linha da matriz \mathbf{X}_i e a respectiva observação são inseridas. Assim, a cada passagem da recursão, as $b + 1$ inovações, divididas pela raiz quadrada da variância, correspondem à respectiva linha de $\mathbf{L}_i \begin{bmatrix} \mathbf{X}_i & y_i \end{bmatrix}$. Os vetores colunas compostos por tais $(b+1)$ elementos, multiplicados pelos seus transpostos, acumulados durante as passagens da recursão de Kalman, fornecem as quantidades $\mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{X}_i$,

$\mathbf{X}'_i \mathbf{V}_i^{-1} \tilde{y}_i$ e $\tilde{y}'_i \mathbf{V}_i^{-1} \tilde{y}_i$, conforme o resultado 4.7.

Apresentaremos ainda um último resultado, relacionando a decomposição de Cholesky com $\sum_i \ln |\mathbf{V}_i|$, uma das parcelas de $-2 \ln L$.

Resultado 4.8

A quantidade $|\mathbf{V}_i|$ é o inverso do produtório do quadrado dos elementos da diagonal de \mathbf{L}_i , isto é,

$$|\mathbf{V}_i| = \frac{1}{\left(\prod_{j=1}^{n_i} \ell_{jj}^2\right)} .$$

Portanto,

$$\ln |\mathbf{V}_i| = \sum_{j=1}^{n_i} \ln(\ell_{jj})^{-2} ,$$

onde ℓ_{jj} é o j -ésimo elemento da diagonal de \mathbf{L}_i .

Prova:

Sabemos que $\mathbf{V}_i^{-1} = \mathbf{L}'_i \mathbf{L}_i$, onde \mathbf{L}_i é dada por (4.15). Logo,

$$|\mathbf{V}_i^{-1}| = |\mathbf{L}'_i| |\mathbf{L}_i| = |\mathbf{L}_i|^2 = \prod_{j=1}^{n_i} \ell_{jj}^2$$

pois como \mathbf{L}_i é uma matriz triangular inferior, o seu determinante é o produto dos elementos da sua diagonal. Assim,

$$|\mathbf{V}_i| = |\mathbf{V}_i^{-1}|^{-1} = \frac{1}{\left(\prod_{j=1}^{n_i} \ell_{jj}^2\right)}$$

e

$$\ln |\mathbf{V}_i| = \ln \left[\frac{1}{\left(\prod_{j=1}^{n_i} \ell_{jj}^2\right)} \right] = \sum_{j=1}^{n_i} \ln(\ell_{jj})^{-2} .$$

□

As variâncias das inovações, no passo 5, são exatamente $(\ell_{jj})^{-2}$, como pudemos comprovar na segunda parte do resultado 4.4. Portanto, a quantidade $\sum_{i=1} \ln |\mathbf{V}_i|$ pode ser obtida através do FK, somando os logaritmos destas variâncias a cada passagem da recursão, para todas as unidades.

Agora, os passos da recursão são os seguintes:

1. Seja $\Omega_{t_j - \delta t}$ a estimativa da matriz de estados no tempo $t_j - \delta t$, dadas as observações até o tempo $t_j - \delta t$. Calculamos a predição da matriz de estados para o próximo tempo de observação:

$$(\Omega_{t_j / \mathbf{Y}_{t_j - \delta t}}) = \begin{bmatrix} \phi_{\delta t} & \tilde{0}' \\ \tilde{0} & \mathbf{I} \end{bmatrix} \cdot \Omega_{t_j - \delta t} \quad ;$$

2. Calcula-se a matriz de covariância desta predição,

$$(\mathbf{S}_{t_j / \mathbf{Y}_{t_j - \delta t}}) = \begin{bmatrix} \phi_{\delta t} & \tilde{0}' \\ \tilde{0} & \mathbf{I} \end{bmatrix} \cdot \mathbf{S}_{t_j - \delta t} \cdot \begin{bmatrix} \phi_{\delta t} & \tilde{0}' \\ \tilde{0} & \mathbf{I} \end{bmatrix} + \begin{bmatrix} Q_{i\delta t} & \tilde{0}' \\ \tilde{0} & \mathbf{0} \end{bmatrix} \quad ;$$

3. Calcula-se a predição do próximo vetor de observações, um vetor linha com $b + 1$ predições,

$$\left[\begin{array}{cc} (\tilde{x}'_{t_j} / \mathbf{Y}_{t_j - \delta t}) & (y_{t_j} / \mathbf{Y}_{t_j - \delta t}) \end{array} \right] = \left[\begin{array}{cc} 1 & \tilde{z}'_{it} \end{array} \right] \cdot (\Omega_{t_j / \mathbf{Y}_{t_j - \delta t}}) = \left[\begin{array}{cc} 1 & \tilde{z}'_{it} \end{array} \right] \cdot \begin{bmatrix} \phi_{\delta t} & \tilde{0}' \\ \tilde{0} & \mathbf{I} \end{bmatrix} \cdot \Omega_{t_j - \delta t} \quad ;$$

4. Calcula-se o vetor linha de $b + 1$ inovações,

$$e'_{t_j} = \left[\begin{array}{cc} \tilde{x}'_{t_j} & y_{t_j} \end{array} \right] - \left[\begin{array}{cc} (\tilde{x}'_{t_j} / \mathbf{Y}_{t_j - \delta t}) & (y_{t_j} / \mathbf{Y}_{t_j - \delta t}) \end{array} \right] \quad ;$$

5. Calcula-se a variância da inovação acima,

$$V_{t_j} = \left[\begin{array}{cc} 1 & \tilde{z}'_{it} \end{array} \right] \cdot (\mathbf{S}_{t_j / \mathbf{Y}_{t_j - \delta t}}) \cdot \left[\begin{array}{cc} 1 & \tilde{z}'_{it} \end{array} \right]' + \sigma_0^2 \quad .$$

Ainda, neste passo, acumulam-se os valores

$$e_{it_j} \cdot e'_{it_j} / V_{it_j} \quad (4.20)$$

e

$$\ln V_{it_j} \quad ; \quad (4.21)$$

6. Atualiza-se a estimativa da matriz de estados, com a entrada da j -ésima observação,

$$\Omega_{t_j} = (\Omega_{t_j} / \mathbf{Y}_{t_j - \delta t}) + \underset{\sim}{A}'_{t_j} \cdot \underset{\sim}{e}_{t_j} / V_{t_j} \quad ,$$

onde

$$\underset{\sim}{A}'_{t_j} = \begin{bmatrix} 1 & \underset{\sim}{z}'_{it} \end{bmatrix} \cdot (\mathbf{S}_{t_j} / \mathbf{Y}_{t_j - \delta t}) \quad ;$$

7. Calcula-se a matriz de covariância atualizada da matriz de estados calculada acima,

$$\mathbf{S}_{t_j} = (\mathbf{S}_{t_j} / \mathbf{Y}_{t_j - \delta t}) - \underset{\sim}{A}'_{t_j} \underset{\sim}{A}_{t_j} / V_{t_j} \quad .$$

Como estimativas iniciais, usamos

$$\Omega_0 = \mathbf{0}$$

e

$$\mathbf{S}_0 = \begin{bmatrix} 1 & \underset{\sim}{0}' \\ \underset{\sim}{0} & \hat{\mathbf{B}} \end{bmatrix} \quad .$$

Neste momento, retorna-se ao passo 1 e a recursão é reinicializada para 'filtrar' a próxima observação. Este procedimento deve ser repetido até a última observação de cada unidade. As quantidades acumuladas no passo 5, continuam a ser acumuladas para todas as unidades. No final da recursão, teremos uma matriz, $(b+1) \times (b+1)$, dada por

$$\begin{bmatrix} \sum_i \underset{\sim}{\mathbf{X}}_i' \underset{\sim}{\mathbf{V}}_i^{-1} \underset{\sim}{\mathbf{X}}_{i(b \times b)} & \sum_i \underset{\sim}{\mathbf{X}}_i' \underset{\sim}{\mathbf{V}}_i^{-1} \underset{\sim}{y}_i \\ (\sum_i \underset{\sim}{\mathbf{X}}_i' \underset{\sim}{\mathbf{V}}_i^{-1} \underset{\sim}{y}_i)'_{(1 \times b)} & \sum_i \underset{\sim}{y}_i' \underset{\sim}{\mathbf{V}}_i^{-1} \underset{\sim}{y}_i \end{bmatrix} \quad , \quad (4.22)$$

onde entre parênteses, no lado direito de cada bloco, está indicada a respectiva dimensão.

Os dois blocos superiores,

$$\left[\begin{array}{cc} \sum_i \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{X}_i & \sum_i \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{y}_i \\ \sum_i \mathbf{y}'_i \mathbf{V}_i^{-1} \mathbf{X}_i & \sum_i \mathbf{y}'_i \mathbf{V}_i^{-1} \mathbf{y}_i \end{array} \right] ,$$

forneem a matriz e o vetor necessários para obter a estimativa do vetor de parâmetros β e σ^2 . Para obtermos a estimativa de σ^2 devemos efetuar a fatorização de Cholesky na matriz do primeiro bloco:

$$\sum_i \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{X}_i = \mathbf{D}'_i \mathbf{D}_i .$$

Seja \underline{c} um vetor $b \times 1$ dado por

$$\underline{c} = (\mathbf{D}'_i)^{-1} \sum_i \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{y}_i ,$$

então

$$\hat{\sigma}^2 = \frac{1}{n} \left\{ \left(\sum_i \mathbf{y}'_i \mathbf{V}_i^{-1} \mathbf{y}_i \right) - \underline{c}' \underline{c} \right\} ,$$

onde a primeira parcela do lado direito é o bloco inferior direito da matriz (4.22). Finalmente, $\ell = -2 \ln L$ será

$$\ell = n \ln(2\pi\sigma^2) + \Delta + n .$$

onde Δ é a quantidade (4.21) obtida pelo passo 5 após serem acumuladas todas as observações.

Capítulo 5

Aplicações

5.1 Introdução

Neste capítulo iremos utilizar a metodologia descrita nos capítulos anteriores para analisar 3 conjuntos de dados longitudinais. O primeiro conjunto de dados se refere a um experimento desenvolvido na Pós-graduação de Engenharia de Alimentos da Unicamp, onde 30 ratos foram observados, ao longo do tempo, quanto a seu peso, com o objetivo de comparar 5 dietas. O segundo conjunto de dados é referente a um estudo observacional de corte transversal realizado pelo Departamento de Dança da Unicamp, envolvendo 209 meninas que praticam dança, em academias, pelo menos há seis meses, na faixa etária de 4 a 17 anos. Em cada bailarina foi obtida a medida do seu ângulo de rotação do quadril (média dos ângulos do quadril direito e esquerdo) e registrada a sua idade, na ocasião da observação. O objetivo desse estudo é analisar o comportamento do ângulo de rotação médio do quadril, ao longo das faixas etárias. O terceiro conjunto de dados é extraído de Grizzle e Allen (1969) e se refere a um experimento envolvendo 36 cães divididos em 4 grupos (tratamentos) e observados em 7 pontos de tempo equiespaçados. O experimento tem como objetivo comparar os quatro tratamentos.

Os conjuntos de dados deste trabalho foram analisados fazendo uso do programa FORTRAN CAR1, desenvolvido por Richard H. Jones & Francis Boadi-Boateng, em 1992. Constam anexos em apêndice: um disquete com o programa e um manual deste programa, desenvolvido por mim a partir do manual original e com base na minha experiência par-

ticular. Para a construção dos gráficos de probabilidade normal, para verificação das suposições distribucionais dos efeitos aleatórios γ_i , foram utilizados alguns programas em FORTRAN e S-PLUS.

Nas análises dos conjuntos de dados, consideramos importante:

- A elaboração de curvas individuais para possibilitar melhor visualização da variabilidade e de possíveis modelos.

- A consideração de modelos que incluíssem efeitos aleatórios e/ou possível correlação serial. A inclusão desses fatores poderia melhorar o ajuste dos dados.

5.2 Análise n°1

Cinco dietas foram alocadas aleatoriamente a 30 ratos(6 ratos por dieta), de mesmo sexo, idade e raça. Cada rato foi observado em 11 pontos de tempo durante 26 dias, e em cada ponto de tempo, o seu peso, em gramas, foi obtido. Um dos ratos, que recebeu a dieta 4, apresentou um comportamento bastante irregular e diferente dos demais ratos, segundo o experimentador, provavelmente devido a uma doença, e então, resolvemos excluí-lo da análise. Os gráficos com as observações coletadas para os 29 ratos nas 5 dietas se encontram a seguir:

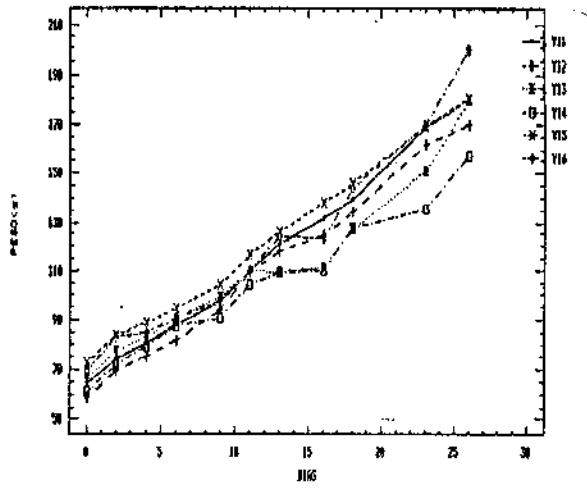


Figura 5-1: Peso observado dos ratos da dieta 1

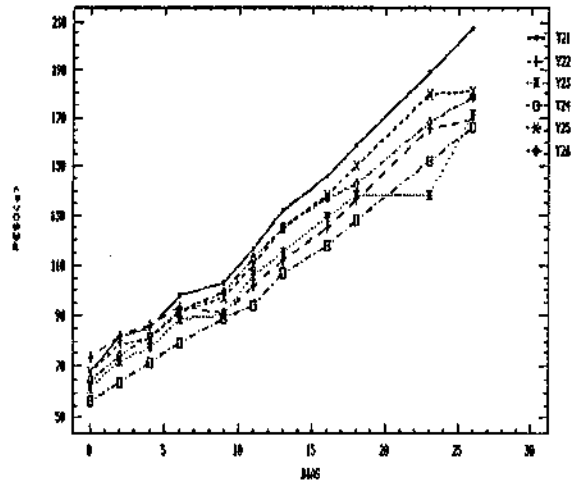


Figura 5-2: Peso observado dos ratos da dieta 2

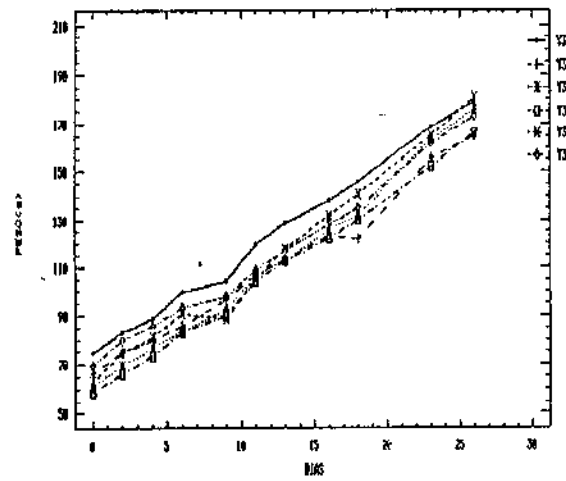


Figura 5-3: Peso observado dos ratos da dieta 3

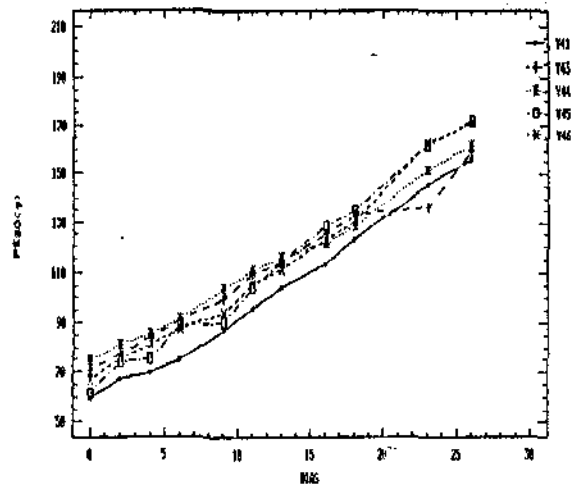


Figura 5-4: Peso observado dos ratos da dieta 4

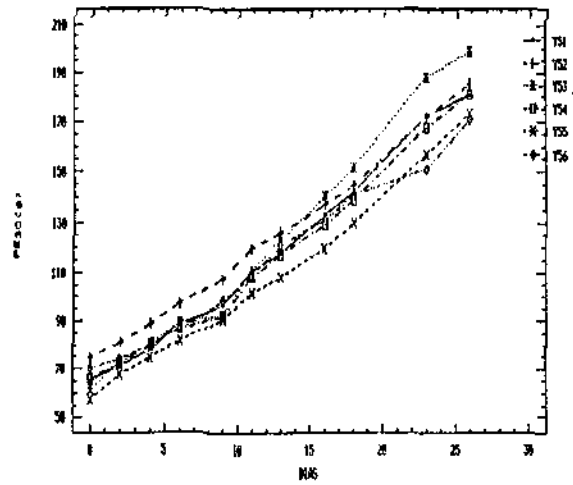


Figura 5-5: Peso observado dos ratos da dieta 5

Com base na análise dos gráficos deste conjunto de dados ao longo do tempo, definimos a matriz básica e o vetor de parâmetros $\tilde{\beta}$ da seguinte forma:

$$\mathbf{X}_y = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \\ 1 & 6 \\ 1 & 9 \\ 1 & 11 \\ 1 & 13 \\ 1 & 16 \\ 1 & 18 \\ 1 & 23 \\ 1 & 26 \end{bmatrix}, \tilde{\beta} = \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \\ \beta_{31} \\ \beta_{32} \\ \beta_{41} \\ \beta_{42} \\ \beta_{51} \\ \beta_{52} \end{bmatrix},$$

ou seja, definimos um ajuste linear para as 5 dietas. Assim, precisamos de 4 covariáveis agrupadoras para obter as matrizes de desenho \mathbf{X}_i . Os valores das covariáveis são os seguintes:

- 0 0 0 0 para a dieta 1,
- 1 0 0 0 para a dieta 2,
- 0 1 0 0 para a dieta 3,
- 0 0 1 0 para a dieta 4, e
- 0 0 0 1 para a dieta 5.

Assim, para as unidades que receberam a dieta 1, temos a seguinte matriz de

desenho:

$$X_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 13 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 16 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 18 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 23 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 26 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

Para as unidades que receberam a dieta 2, temos a seguinte matriz de desenho:

$$X_i = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 4 & 1 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 6 & 1 & 6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 9 & 1 & 9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 11 & 1 & 11 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 13 & 1 & 13 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 16 & 1 & 16 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 18 & 1 & 18 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 23 & 1 & 23 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 26 & 1 & 26 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} ,$$

e, assim por diante, até a dieta 5.

Consideramos, a princípio, os erros não correlacionados, ou seja,

$$\epsilon_i \sim N(0, \sigma^2 \mathbf{I}_i)$$

e ainda, definimos um intercepto aleatório e uma inclinação aleatória para cada rato fazendo

$$\mathbf{Z}_i = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \\ 1 & 6 \\ 1 & 9 \\ 1 & 11 \\ 1 & 13 \\ 1 & 16 \\ 1 & 18 \\ 1 & 23 \\ 1 & 26 \end{bmatrix} .$$

Este modelo, que chamamos modelo 1.1, ajustado ao conjunto de dados acima, com o auxílio do programa CAR1, forneceu os seguintes resultados:

Dieta	Estimativas	Desvio Padrão	t
1	64.19	2.172	29.55
	4.125	.1685	24.48
2	-1.512	3.072	-4.920
	.2539	.2382	1.066
3	-.5225	3.072	-1.701
	-.0159	.2382	-.0667
4	.5316	3.222	.1650
	-.4191	.2499	-1.677
5	-2.430	3.072	-7.912
	.3650	.2382	1.532

Tabela 5.1: Estimativas, desvios padrão e estatística t dos parâmetros do modelo 1.1

O valor de $-2 \ln L$, para este modelo foi de 1991.01, o que resultou em $AIC = 2019.01$ ($-2 \ln L + 2 \times 11$ parâmetros: 10 para β , 3 para \mathbf{B} e σ^2).

A figura 4-6 ilustra as retas ajustadas para as 5 dietas:

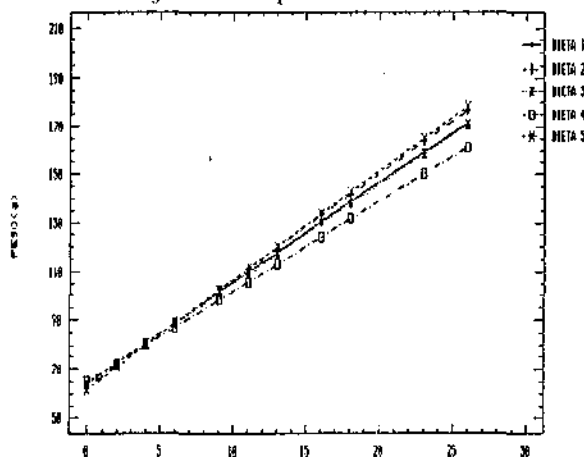


Figura 5-6: Retas ajustadas - Modelo 1.1

Note que, na tabela 5.1, com exceção das duas primeiras componentes de β , as demais não apresentaram um valor muito alto para a estatística t, ou seja, a significância estatística dos demais parâmetros não está totalmente comprovada. Neste sentido, de-

cidimos reduzir o modelo, ajustando somente uma reta para todas as 5 dietas, fazendo

$$\mathbf{X}_i = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \\ 1 & 6 \\ 1 & 9 \\ 1 & 11 \\ 1 & 13 \\ 1 & 16 \\ 1 & 18 \\ 1 & 23 \\ 1 & 26 \end{bmatrix} \quad e \quad \underset{\sim}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

para todos os 29 ratos. Isto equivale a testar se todas as dietas produzem o mesmo efeito, ou seja, proporcionam 'estatisticamente' o mesmo ganho de peso ao longo do tempo. Continuamos considerando os erros não correlacionados e o vetor de efeitos aleatórios γ_i bidimensional. Este modelo, 1.2, ajustado ao conjunto de observações, forneceu os seguintes resultados:

Estimativas	Desvio Padrão	t
63.36	1.006	62.97
4.177	.091	45.91

Tabela 5.2: Estimativas, desvios padrão e estatística t dos parâmetros do modelo 1.2

O valor de $-2 \ln L$ para este modelo foi de 2001.03, o que resultou em $AIC = 2013.03$. Portanto, o modelo 1.2, resultante da redução do modelo 1.1, nos parece mais apropriado. Verificamos, então, que não existe diferença estatisticamente significativa entre as dietas, quanto ao peso.

Continuamos ainda nossas tentativas de melhoria do ajuste. Tentamos a inclusão de correlação serial AR(1) no modelo 1.2, ao invés de considerarmos os erros independentes.

Com isso, o programa CAR1 forneceu os seguintes resultados:

Estimativas	Desvio Padrão	t
63.73	1.048	60.82
4.215	.0888	47.45

Tabela 5.3: Estimativas, desvios padrão e estatística t dos parâmetros do modelo 1.3

O valor de $-2 \ln L$ para este modelo foi de 1911.10, o que resultou em $AIC = 1925.10$. Portanto, o modelo 1.3, com a inclusão de correlação serial AR(1), melhora sensivelmente o ajuste.

E finalmente, consideramos o modelo 1.4, que ignora a covariância entre os dois efeitos aleatórios γ_{i1} e γ_{i2} . Desta forma, o modelo 1.4 tem um parâmetro a menos em relação ao modelo 1.3 e considera

$$B = \begin{bmatrix} Var(\gamma_{i1}) & 0 \\ 0 & Var(\gamma_{i2}) \end{bmatrix} .$$

Os resultados obtidos pelo programa CAR1 foram os seguintes:

Estimativas	Desvio Padrão	t
63.74	1.058	60.25
4.215	.0933	45.16

Tabela 5.4: Estimativas, desvios padrão e estatística t dos parâmetros do modelo 1.4

O valor de $-2 \ln L$ para este modelo foi de 1911.61, o que resultou em $AIC = 1923.61$. Portanto, o modelo 1.4 melhora um pouco mais o ajuste.

Na figura 5-7 ilustramos o modelo final ajustado:

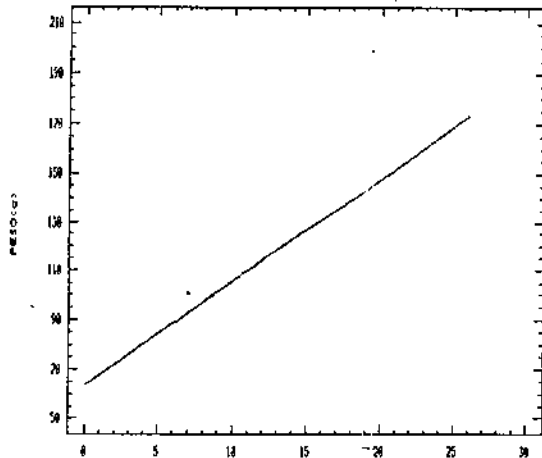


Figura 5-7: Modelo final do ganho de peso dos ratos

Os gráficos de probabilidade normal ponderados para verificar a suposição distribucional dos efeitos aleatórios γ_{i1} e γ_{i2} são os seguintes:

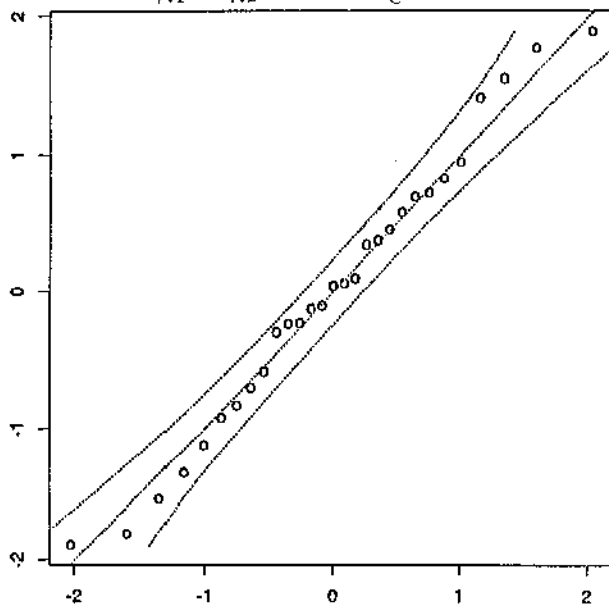


Figura 5-8: Gráfico de probabilidade normal ponderado para γ_{i1}

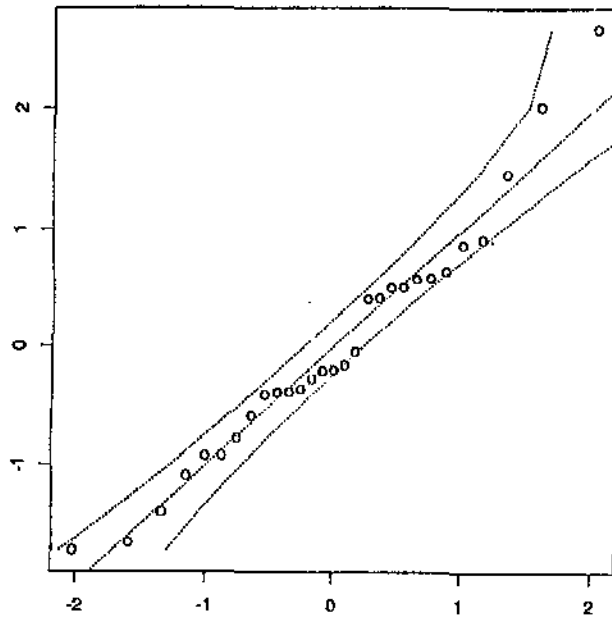


Figura 5-9: Gráfico de probabilidade normal ponderado para γ_{12}

Podemos notar que as suposições distribucionais estão satisfeitas. Portanto, podemos finalizar a análise deste conjunto de dados concluindo que não houve evidência de diferenças entre as 5 dietas testadas. O peso médio dos ratos no início do experimento foi de 63.74 g e o ganho de peso diário desses ratos submetidos às dietas foi de 4.215 g.

5.3 Análise nº 2

Este conjunto de dados se refere a um plano de corte transversal onde cada uma dentre 209 bailarinas foi observada em apenas uma oportunidade. Para cada uma das bailarinas foi medido o seu ângulo de rotação dos quadris direito e esquerdo, sendo obtida a média dos ângulos, e relacionado com a idade na ocasião da observação.

A figura 5-10 ilustra as observações tomadas das 209 bailarinas:

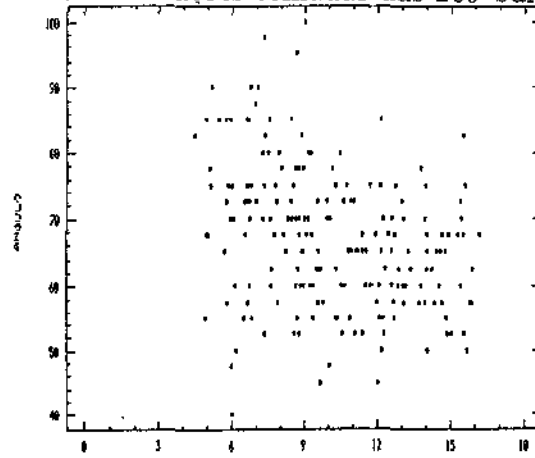


Figura 5-10: Observações das 209 bailarinas

A partir da figura acima, resolvemos ajustar uma reta aos dados. Com isto, cada bailarina terá uma matriz X_i (vetor) dada simplesmente por

$$X_i = \begin{bmatrix} 1 & t_i \end{bmatrix} \quad e \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} ,$$

onde t_i é a idade da bailarina i por ocasião da observação. Definimos um intercepto aleatório e uma inclinação aleatória para cada bailarina fazendo também

$$Z_i = \begin{bmatrix} 1 & t_i \end{bmatrix} ,$$

além de considerar os resíduos independentes. O programa CAR1 forneceu para este modelo ajustado, as seguintes estimativas:

Estimativas	Desvio Padrão	t
77.15	2.568	30.04
-1.018	0.2241	-4.541

Tabela 5.5: Estimativas, desvios padrão e estatística t dos parâmetros do modelo 2.1

A figura 5-11 ilustra o modelo final ajustado:

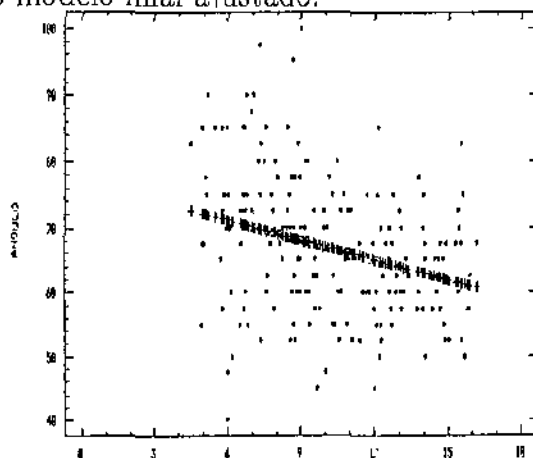


Figura 5-11: Modelo final ajustado

O modelo final forneceu como estimativa da matriz $\sigma^2\mathbf{B}$ de covariâncias dos efeitos aleatórios,

$$\hat{\sigma}^2\hat{\mathbf{B}} = \begin{bmatrix} 305.3 & -21.63 \\ -21.63 & 1.532 \end{bmatrix}$$

e os gráficos de probabilidade normal ponderados dos efeitos aleatórios:

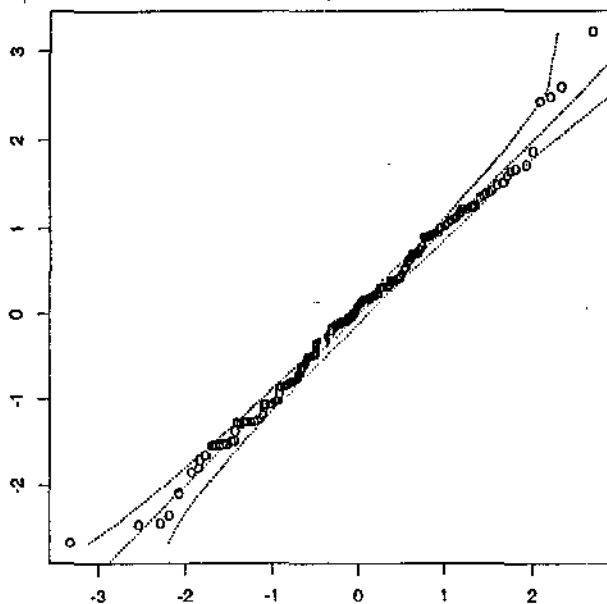


Figura 5-12: Gráfico de probabilidade normal ponderado para γ_{11}

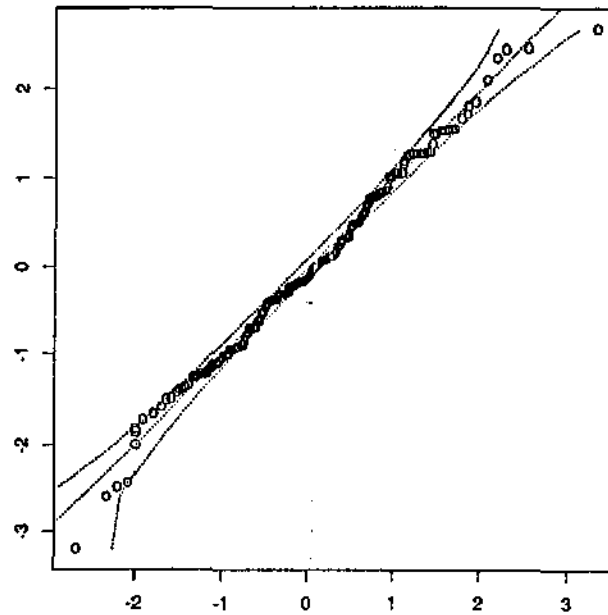


Figura 5-13: Gráfico de probabilidade normal ponderado para γ_{i2}

Assim, a suposição de normalidade dos efeitos aleatórios está satisfeita. Podemos concluir, então, que o ângulo médio de rotação dos quadris das bailarinas decresce levemente ao longo das idades. Podemos observar, também, o valor alto para a variância do efeito aleatório γ_1 . Ele indica, pela suposição de normalidade, que o intercepto aleatório se encontra aproximadamente, entre 42.15 e 102.15 graus com 95% de probabilidade. Esta grande variabilidade pode ser explicada pela própria natureza do fenômeno, e ainda, talvez, pela própria estrutura do plano, que observa cada unidade apenas uma vez, número pequeno para se conseguir um padrão de comportamento da variável em estudo.

5.4 Análise nº 3

Um conjunto de 36 cães foi dividido em 4 grupos (sendo o primeiro, um grupo de controle) e observados em 7 pontos de tempo (minutos após a oclusão coronária), quanto à variável *Coronary Sinus Potassium* (mil equivalentes/litro).

Decidimos inicialmente ajustar um polinômio de ordem 4 (completo de grau 3)

para cada grupo, considerando os erros independentes e o vetor de efeitos aleatórios, de dimensão 3. Assim, teremos uma matriz básica \mathbf{X}_y de dimensão 7×4 , e um vetor $\tilde{\beta}$, 16×1 , dados por

$$\mathbf{X}_y = \begin{bmatrix} 1 & 1 & 1^2 & 1^3 \\ 1 & 3 & 3^2 & 3^3 \\ 1 & 5 & 5^2 & 5^3 \\ 1 & 7 & 7^2 & 7^3 \\ 1 & 9 & 9^2 & 9^3 \\ 1 & 11 & 11^2 & 11^3 \\ 1 & 13 & 13^2 & 13^3 \end{bmatrix} \quad e \quad \tilde{\beta} = \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{14} \\ \beta_{21} \\ \vdots \\ \beta_{24} \\ \beta_{31} \\ \vdots \\ \beta_{34} \\ \beta_{41} \\ \vdots \\ \beta_{44} \end{bmatrix} .$$

Ainda,

$$\mathbf{Z}_i = \begin{bmatrix} 1 & 1 & 1^2 \\ 1 & 3 & 3^2 \\ 1 & 5 & 5^2 \\ 1 & 7 & 7^2 \\ 1 & 9 & 9^2 \\ 1 & 11 & 11^2 \\ 1 & 13 & 13^2 \end{bmatrix} ,$$

isto é, ajustamos um polinômio do segundo grau para os efeitos aleatórios individuais γ_i .

Dado isso, os resultados fornecidos pelo programa CAR1 foram os seguintes:

Grupo	Estimativas	Desvio Padrão	t
1	4.330	.1981	21.86
	-.2536	.1135	-2.234
	.7171E-01	.1752E-01	4.092
	-.3819E-02	.8073E-03	-4.731
2	-.7911	.2730	-2.898
	.2778	.1564	1.776
	-.7446E-01	.2415E-01	-3.083
	.3854E-02	.1113E-02	3.463
3	-.9305	.2888	-3.223
	.3971	.1654	2.400
	-.6572E-01	.2554E-01	-2.573
	.2865E-02	.1177E-02	2.434
4	-.8603	.2801	-3.071
	.4232	.1605	2.636
	-.8838E-01	.2478E-01	-3.566
	.4360E-02	.1142E-02	3.818

Tabela 5.6: Estimativas, desvios padrão e estatística t dos parâmetros do modelo 3.1

O modelo 3.1 forneceu um valor para $-2 \ln L = 263.61$ e para a medida $AIC = 309.61$. Em seguida ajustamos, no modelo 3.2, uma estrutura de covariância AR(1), no lugar da estrutura independente do modelo anterior. Ajustamos, também, um vetor de efeitos aleatórios γ bidimensional.

Os resultados foram os seguintes:

Grupo	Estimativas	Desvio Padrão	t
1	4.288	.2185	19.62
	-.2397	.1169	-2.049
	.7134E-01	.1951E-01	3.657
	-.3870E-02	.9129E-03	-4.239
2	-.7878	.3012	-2.615
	.2874	.1612	1.783
	-.7749E-01	.2689E-01	-2.882
	.4042E-02	.1258E-02	3.213
3	-.8120	.3185	-2.549
	.3485	.1705	2.045
	-.6214E-01	.2844E-01	-2.185
	.2856E-02	.1331E-02	2.146
4	-.7743	.3090	-2.506
	.3802	.1654	2.299
	-.8358E-01	.2759E-01	-3.030
	.4221E-02	.1291E-02	3.269

Tabela 5.7: Estimativas, desvios padrão e estatística t dos parâmetros do modelo 3.2

Este modelo apresentou $-2 \ln L = 238.86$ e $AIC = 280.86$ e está representado na figura abaixo:

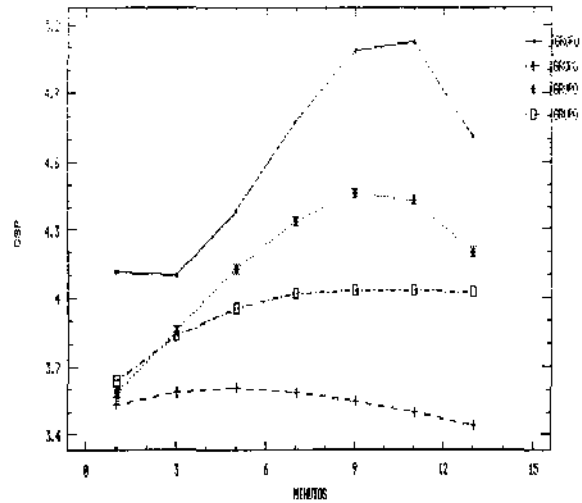


Figura 5-14: Ajuste do modelo 3.2

Devemos notar, pela tabela acima, que para os parâmetros do vetor $\tilde{\beta}$ o valor da estatística t é bastante razoável, parecendo demonstrar significância estatística. No entanto, como estas estatísticas estão relacionadas sempre com o grupo controle (primeiro grupo), elas demonstram diferença significativa com relação a este grupo, mas não entre os demais grupos. Pelo gráfico, nos parece que os 3 últimos grupos podem ser considerados como um só. Desta forma, no modelo 3.3, ajustamos um polinômio de grau 3 para os cães de controle e outro polinômio de grau 3 para os demais cães.

Os resultados obtidos foram os seguintes:

Grupo	Estimativas	Desvio Padrão	t
1	4.287	.2218	19.33
	-.2392	.1184	-2.020
	.7132E-01	.1972E-01	3.616
	-.3871E-02	.9228E-03	-4.195
2,3 e 4	-.7889	.2561	-3.081
	.3357	.1368	2.455
	-.7493E-01	.2278E-01	-3.290
	.3751E-02	.1066E-02	3.520

Tabela 5.8: Estimativas, desvios padrão e estatística t dos parâmetros do modelo 3.3

Este modelo apresentou $-2 \ln L = 251.58$ e $AIC = 277.58$. Com isto, chegamos

ao modelo final, ilustrado abaixo, que considera os cães dos últimos 3 grupos originais, como fazendo parte de um grupo só, ou seja, não existe diferença significativa entre estes grupos, quanto à variável em estudo. No entanto, estes cães diferem significativamente do grupo controle quanto à variável *Coronary Sinus Potassium*.

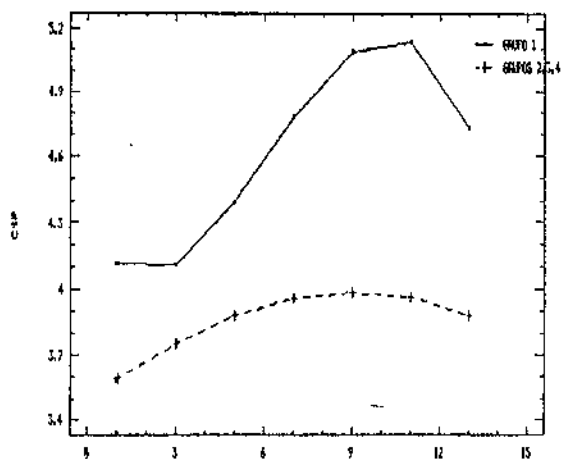


Figura 5-15: Modelo final ajustado

Os gráficos, abaixo, de probabilidade normal para os efeitos aleatórios, comprovam a suposição de normalidade do modelo final:

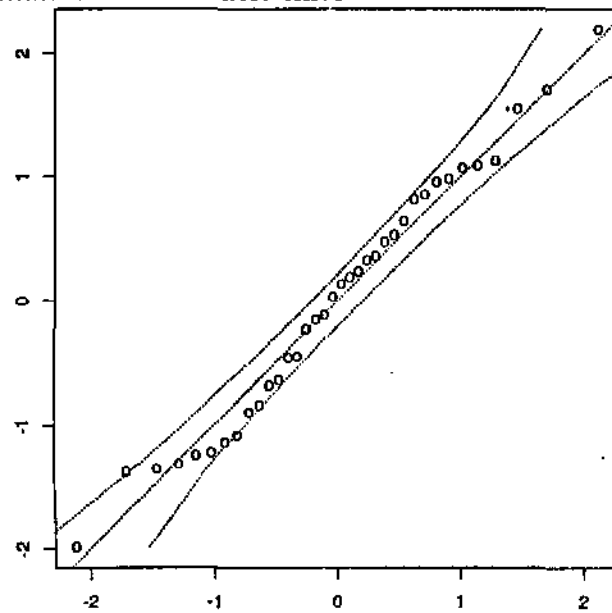


Figura 5-16: Gráfico de probabilidade normal ponderado para γ_{i1}

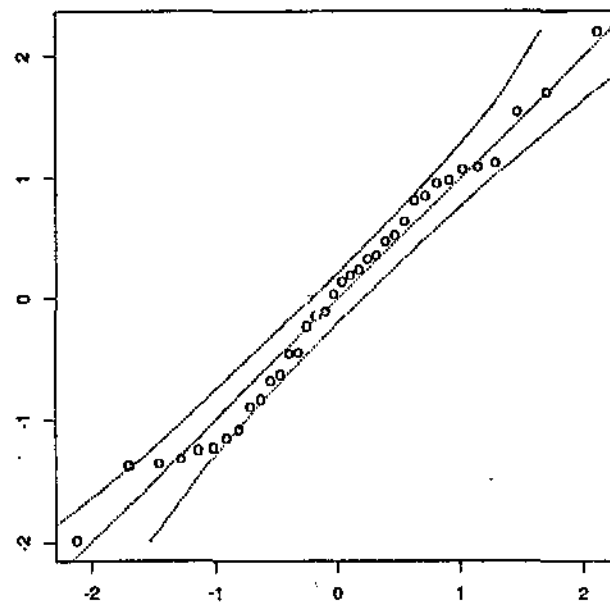


Figura 5-17: Gráfico de probabilidade normal ponderado para γ_{i2}

Um modelo considerado no início da análise, com efeito aleatório unidimensional, e logo descartado por apresentar um alto valor da medida AIC apresentou o seguinte gráfico de probabilidade normal dos efeitos aleatórios γ :

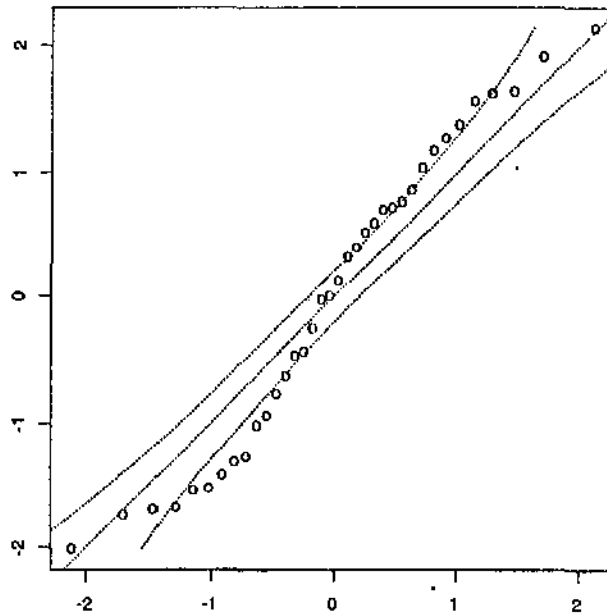


Figura 5-18: Gráfico de probabilidade normal ponderado para γ

Como podemos notar, vários pontos se apresentam fora da região crítica, indicando a falta de ajuste do modelo.

Apêndice A

Simetria Composta

A.1 Descrição do Modelo

Neste apêndice, descreveremos os detalhes da estimação de máxima verossimilhança para um modelo com estrutura de covariância de Simetria Composta, que poderíamos considerar como sendo uma das estruturas mais simples de covariância. Isto porque este modelo considera que a correlação entre quaisquer duas observações tomadas de uma mesma unidade, em qualquer ponto de tempo, é constante. Esta propriedade pode vir a ser adequada para situações onde várias observações são tomadas num mesmo ponto de tempo e a numeração dessas observações não tenha nenhum significado.

Para ilustrar este caso particular, consideremos um exemplo simples de comparação entre dois tratamentos, com m_1 unidades recebendo o tratamento 1 e m_2 unidades recebendo o tratamento 2. Usaremos o índice i para denotar a unidade, o índice j para denotar a observação e (1) ou (2) para referenciar o primeiro e o segundo grupo, respectivamente. Sejam $n_{i(1)}$, $i = 1, \dots, m_1$, os números de observações para as unidades do primeiro grupo e $n_{i(2)}$, $i = 1, \dots, m_2$, os números de observações para as unidades do segundo grupo. O modelo Laird-Ware para a i -ésima unidade do primeiro grupo, é dado por

$$y_{i(1)j} = \beta_1 + \gamma_{i(1)} + \epsilon_{i(1)j} \quad ,$$

para $j = 1, \dots, n_{i(1)}$ e $i = 1, \dots, m_1$ e ainda, para a i -ésima unidade no segundo grupo

$$y_{i(2)j} = \beta_1 + \beta_2 + \gamma_{i(2)} + \epsilon_{i(2)j} \quad ,$$

para $j = 1, \dots, n_{i(2)}$ e $i = 1, \dots, m_2$ com as seguintes suposições:

1. As variáveis $\gamma_{1(1)}, \dots, \gamma_{m_1(1)}, \gamma_{1(2)}, \dots, \gamma_{m_2(2)}$ são variáveis aleatórias independentes, identicamente distribuídas conforme a distribuição normal com média zero e variância σ_γ^2 ;
2. As variáveis $\epsilon_{i(1)j}, \epsilon_{i(2)j}$, para todos os valores de i e j definidos, são variáveis aleatórias independentes e identicamente distribuídas conforme a distribuição normal com média zero e variância σ_ϵ^2 ;
3. As variáveis $\epsilon_{i(1)j}$ são independentes das variáveis $\gamma_{i(1)}$, assim como as variáveis $\epsilon_{i(2)j}$ são independentes dos $\gamma_{i(2)}$, para quaisquer valores de i e j .

Os efeitos aleatórios da i -ésima unidade do grupo 1 e grupo 2, denotados por $\gamma_{i(1)}$ e $\gamma_{i(2)}$, respectivamente, são chamados efeitos aleatórios entre unidades. As variáveis aleatórias $\epsilon_{i(1)j}$ e $\epsilon_{i(2)j}$, para quaisquer valores de i e j , representam os fatores não controláveis e são conhecidos como efeitos aleatórios dentro das unidades.

A.2 Construção da Estrutura de Covariância

Com base nas suposições, podemos construir a matriz de variâncias e covariâncias das observações de cada unidade:

$$Var(y_{i(k)j}) = Var(\gamma_{i(k)} + \epsilon_{i(k)j}) = \sigma_\gamma^2 + \sigma_\epsilon^2 \quad ,$$

já que $\gamma_{i(k)}$ e $\epsilon_{i(k)j}$ são não correlacionados, e, para dois pontos de tempo j e j° ,

$$\begin{aligned} Cov(y_{i(k)j}, y_{i(k)j^\circ}) &= Cov(\gamma_{i(k)} + \epsilon_{i(k)j}, \gamma_{i(k)} + \epsilon_{i(k)j^\circ}) = \\ &= E[(\gamma_{i(k)} + \epsilon_{i(k)j}) \cdot (\gamma_{i(k)} + \epsilon_{i(k)j^\circ})] = \sigma_\gamma^2 \quad . \end{aligned}$$

Logo, a matriz de variâncias e covariâncias de

$$\underline{y}_{i(k)} = \begin{bmatrix} y_{i(k)1} \\ y_{i(k)2} \\ \vdots \\ y_{i(k)n_{i(k)}} \end{bmatrix},$$

para quaisquer valores de i e k é dada por

$$\mathbf{V}_{i(k)} = \begin{bmatrix} \sigma_c^2 + \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_c^2 + \sigma_\gamma^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_\gamma^2 & \cdots & \cdots & \sigma_\gamma^2 \end{bmatrix}. \quad (\text{A.1})$$

Note que os tamanhos dos vetores $\underline{y}_{i(k)}$ podem ser diferentes; portanto, para uma unidade com $n_{i(k)}$ observações, a sua matriz de variâncias e covariâncias será de ordem $n_{i(k)} \times n_{i(k)}$. Podemos notar também que a covariância constante entre as mensurações tomadas da mesma unidade é produzida pelo efeito aleatório $\gamma_{i(k)}$. Com estes elementos, temos que a distribuição de $\underline{y}_{i(k)}$, é dada por

$$\underline{y}_{i(k)} \sim N(\mathbf{X}_i \underline{\beta}, \mathbf{V}_{i(k)}) .$$

A.3 Estimação

Descrito o modelo e a estrutura de covariância, podemos seguir com o desenvolvimento da estimação de máxima verossimilhança.

A função de densidade de probabilidade conjunta das observações para uma unidade

$y_{i(k)}$, é dada por:

$$f(y_{i(k)}) = \frac{1}{(2\pi)^{\frac{n_{i(k)}}{2}} |\mathbf{V}_{i(k)}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y_{i(k)} - \mu_k \mathbf{1}_{i(k)})' \mathbf{V}_{i(k)}^{-1} (y_{i(k)} - \mu_k \mathbf{1}_{i(k)})\right\} \quad , \quad (\text{A.2})$$

onde,

$$\mu_1 = \beta_1 \quad \text{e} \quad \mu_2 = \beta_1 + \beta_2 \quad ,$$

sendo β_1 e $\beta_1 + \beta_2$ as médias populacionais para os grupos 1 e 2, respectivamente. O vetor $\mathbf{1}_{i(k)}$, $n_{i(k)} \times 1$, é um vetor coluna de elementos iguais a um, $|\mathbf{V}_{i(k)}|$ denota o determinante da matriz $\mathbf{V}_{i(k)}$ e $\mathbf{V}_{i(k)}^{-1}$ denota sua inversa.

Para encontrar a função de verossimilhança, basta efetuarmos o produtório da função acima para todas as unidades:

$$L(\mu_k, \sigma_\epsilon^2, \sigma_\gamma^2) = \prod_{i(k)} \frac{1}{(2\pi)^{\frac{n_{i(k)}}{2}} |\mathbf{V}_{i(k)}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y_{i(k)} - \mu_k \mathbf{1}_{i(k)})' \mathbf{V}_{i(k)}^{-1} (y_{i(k)} - \mu_k \mathbf{1}_{i(k)})\right\} \quad . \quad (\text{A.3})$$

Com isto, para obter as estimativas de máxima verossimilhança dos parâmetros, devemos encontrar valores que maximizam a função acima, sujeito às restrições de que eles estejam dentro do espaço paramétrico ($\sigma_\gamma^2 \geq 0, \sigma_\epsilon^2 \geq 0$).

Procedendo com a transformação $\ell = -2 \ln L$, teremos que os valores que irão maximizar L são os mesmos que irão minimizar ℓ .

$$\ell = \sum_{i(k)} \left[n_{i(k)} \ln(2\pi) + \ln |\mathbf{V}_{i(k)}| + (y_{i(k)} - \mu_k \mathbf{1}_{i(k)})' \mathbf{V}_{i(k)}^{-1} (y_{i(k)} - \mu_k \mathbf{1}_{i(k)}) \right] \quad . \quad (\text{A.4})$$

Uma técnica para encontrar o mínimo desta função é derivá-la em relação a algum parâmetro, igualando o resultado a zero e em seguida obter uma expressão (em função dos outros parâmetros) que explique aquele determinado parâmetro. Em seguida, substitui-se esta expressão na função de verossimilhança, que agora, é expressa através de um

número menor de parâmetros. Este processo deve ser executado até que cada parâmetro seja representado por uma expressão que independa dos demais parâmetros e seja função apenas da amostra. Desta forma, para um grupo qualquer k^* ,

$$\begin{aligned}
\frac{\partial \ell}{\partial \mu_{k^*}} &= \sum_{i(k^*)} \frac{\partial}{\partial \mu_{k^*}} \left[\nu_{i(k^*)} \ln(2\pi) + \ln |\mathbf{V}_{i(k^*)}| + (\underline{y}_{i(k^*)} - \mu_{k^*} \underline{1}_{i(k^*)})' \mathbf{V}_{i(k^*)}^{-1} (\underline{y}_{i(k^*)} - \mu_{k^*} \underline{1}_{i(k^*)}) \right] \\
&= \sum_{i(k^*)} \frac{\partial}{\partial \mu_{k^*}} (\underline{y}_{i(k^*)} - \mu_{k^*} \underline{1}_{i(k^*)})' \mathbf{V}_{i(k^*)}^{-1} (\underline{y}_{i(k^*)} - \mu_{k^*} \underline{1}_{i(k^*)}) \\
&= \sum_{i(k^*)} \frac{\partial}{\partial \mu_{k^*}} [\underline{y}'_{i(k^*)} \mathbf{V}_{i(k^*)}^{-1} \underline{y}_{i(k^*)} - \mu_{k^*} \underline{y}'_{i(k^*)} \mathbf{V}_{i(k^*)}^{-1} \underline{1}_{i(k^*)} - \\
&\quad - \mu_{k^*} \underline{1}'_{i(k^*)} \mathbf{V}_{i(k^*)}^{-1} \underline{y}_{i(k^*)} + \mu_{k^*}^2 \underline{1}'_{i(k^*)} \mathbf{V}_{i(k^*)}^{-1} \underline{1}_{i(k^*)}] = 0 \\
&\Rightarrow \sum_{i(k^*)} [\underline{y}'_{i(k^*)} \mathbf{V}_{i(k^*)}^{-1} \underline{1}_{i(k^*)} - \underline{1}'_{i(k^*)} \mathbf{V}_{i(k^*)}^{-1} \underline{y}_{i(k^*)} + 2\mu_{k^*} \underline{1}'_{i(k^*)} \mathbf{V}_{i(k^*)}^{-1} \underline{1}_{i(k^*)}] = 0 \quad .
\end{aligned}$$

Agora usaremos o seguinte resultado:

Resultado A.1

$$\mathbf{V}_{i(k^*)}^{-1} = \frac{1}{\sigma_\epsilon^2} \left[\mathbf{I}_{i(k^*)} - \frac{\sigma_\gamma^2 \underline{1}_{i(k^*)} \underline{1}'_{i(k^*)}}{n_{i(k^*)} \sigma_\gamma^2 + \sigma_\epsilon^2} \right] \quad (\text{A.5})$$

□

Assim,

$$\begin{aligned}
\frac{\partial \ell}{\partial \mu_{k^*}} &= \sum_{i(k^*)} \left[-2 \underline{y}'_{i(k^*)} \frac{1}{\sigma_\epsilon^2} \left[\mathbf{I}_{i(k^*)} - \frac{\sigma_\gamma^2 \underline{1}_{i(k^*)} \underline{1}'_{i(k^*)}}{n_{i(k^*)} \sigma_\gamma^2 + \sigma_\epsilon^2} \right] \underline{1}_{i(k^*)} + \right. \\
&\quad \left. + 2\mu_{k^*} \underline{1}'_{i(k^*)} \frac{1}{\sigma_\epsilon^2} \left[\mathbf{I}_{i(k^*)} - \frac{\sigma_\gamma^2 \underline{1}_{i(k^*)} \underline{1}'_{i(k^*)}}{n_{i(k^*)} \sigma_\gamma^2 + \sigma_\epsilon^2} \right] \underline{1}_{i(k^*)} \right] = 0 \\
&\Rightarrow \sum_{i(k^*)} 2\mu_{k^*} \underline{1}'_{i(k^*)} \frac{1}{\sigma_\epsilon^2} \left[\mathbf{I}_{i(k^*)} - \frac{\sigma_\gamma^2 \underline{1}_{i(k^*)} \underline{1}'_{i(k^*)}}{n_{i(k^*)} \sigma_\gamma^2 + \sigma_\epsilon^2} \right] \underline{1}_{i(k^*)} = \sum_{i(k^*)} 2 \underline{y}'_{i(k^*)} \frac{1}{\sigma_\epsilon^2} \left[\mathbf{I}_{i(k^*)} - \frac{\sigma_\gamma^2 \underline{1}_{i(k^*)} \underline{1}'_{i(k^*)}}{n_{i(k^*)} \sigma_\gamma^2 + \sigma_\epsilon^2} \right] \underline{1}_{i(k^*)} \\
&\Rightarrow \mu_{k^*} \sum_{i(k^*)} \underline{1}'_{i(k^*)} \left[\mathbf{I}_{i(k^*)} - \frac{(\sigma_\gamma^2 / \sigma_\epsilon^2) \underline{1}_{i(k^*)} \underline{1}'_{i(k^*)}}{n_{i(k^*)} (\sigma_\gamma^2 / \sigma_\epsilon^2) + 1} \right] \underline{1}_{i(k^*)} = \sum_{i(k^*)} \underline{y}'_{i(k^*)} \left[\mathbf{I}_{i(k^*)} - \frac{(\sigma_\gamma^2 / \sigma_\epsilon^2) \underline{1}_{i(k^*)} \underline{1}'_{i(k^*)}}{n_{i(k^*)} (\sigma_\gamma^2 / \sigma_\epsilon^2) + 1} \right] \underline{1}_{i(k^*)}
\end{aligned}$$

Infelizmente, neste caso, será impossível obtermos um estimador de máxima verossimilhança.

milhança para μ_{k^*} que não dependa de σ_γ^2 e σ_ϵ^2 . Mas considerando a constante

$$c^2 = \frac{\sigma_\gamma^2}{\sigma_\epsilon^2} \quad (\text{A.6})$$

teremos

$$\hat{\mu}_{k^*} = \frac{\sum_{i(k^*)} y_{i(k^*)}' \left[\mathbf{I}_{i(k^*)} - \frac{c^2 \mathbf{1}_{i(k^*)} \mathbf{1}_{i(k^*)}'}{1+n_{i(k^*)}c^2} \right] \mathbf{1}_{i(k^*)}}{\sum_{i(k^*)} \mathbf{1}_{i(k^*)}' \left[\mathbf{I}_{i(k^*)} - \frac{c^2 \mathbf{1}_{i(k^*)} \mathbf{1}_{i(k^*)}'}{1+n_{i(k^*)}c^2} \right] \mathbf{1}_{i(k^*)}}$$

Agora, façamos $y_{i(k^*)}' \mathbf{1}_{i(k^*)} = y_{i(k^*)o} = \sum_{i(k^*)} y_{i(k^*)j} c \mathbf{1}_{i(k^*)}' \mathbf{1}_{i(k^*)} = n_{i(k^*)}$

$$\begin{aligned} \hat{\mu}_{k^*} &= \frac{\sum_{i(k^*)} \left[y_{i(k^*)o} - \frac{y_{i(k^*)o} n_{i(k^*)} c^2}{1+n_{i(k^*)}c^2} \right]}{\sum_{i(k^*)} \left[n_{i(k^*)} - \frac{n_{i(k^*)} n_{i(k^*)} c^2}{(1+n_{i(k^*)}c^2)} \right]} \\ \Rightarrow \hat{\mu}_{k^*} &= \frac{\sum_{i=1}^{m_{k^*}} \{ y_{i(k^*)o} / (1+n_{i(k^*)}c^2) \}}{\sum_{i=1}^{m_{k^*}} \{ n_{i(k^*)} / (1+n_{i(k^*)}c^2) \}} \end{aligned} \quad (\text{A.7})$$

Este é o estimador para a média populacional μ_{k^*} quando o experimento é não balanceado. Note que se o experimento for balanceado, ou seja, quando as unidades forem observadas em todos os pontos de tempo pré-determinados, o estimador se reduzirá para

$$\hat{\mu}_{k^*} = \frac{\sum_{i=1}^{m_{k^*}} y_{i(k^*)o}}{\sum_{i=1}^{m_{k^*}} n_{i(k^*)}} = \bar{y}_{i(k^*)} \quad (\text{A.8})$$

que é a média de todas as observações no grupo k^* . Para encontrarmos o estimador de máxima verossimilhança de σ_ϵ^2 , devemos definir o vetor de resíduos das médias dos grupos

$$\tilde{y}_{i(k)} = y_{i(k)} - \hat{\mu}_k \mathbf{1}_{i(k)} \quad (\text{A.9})$$

Substituindo em A.4. teremos

$$\ell(\sigma_\epsilon^2, \sigma_\gamma^2) = n \ln(2\pi) + \sum_{i(k)} \left\{ \ln |\mathbf{V}_{i(k)}| + \tilde{\mathbf{y}}'_{i(k)} \mathbf{V}_{i(k)}^{-1} \tilde{\mathbf{y}}_{i(k)} \right\} ,$$

onde $n = \sum_{i(k)} n_{i(k)}$.

Agora devemos usar A.5

$$\mathbf{V}_{i(k)}^{-1} = \frac{1}{\sigma_\epsilon^2} \left[\mathbf{I}_{i(k)} - \frac{\sigma_\gamma^2 \mathbf{1}_{i(k)} \mathbf{1}'_{i(k)}}{n_{i(k)} \sigma_\gamma^2 + \sigma_\epsilon^2} \right] = \frac{1}{\sigma_\epsilon^2} \left[\mathbf{I}_{i(k)} - \frac{c^2 \sigma_\epsilon^2 \mathbf{1}_{i(k)} \mathbf{1}'_{i(k)}}{\sigma_\epsilon^2 (1 + n_{i(k)} c^2)} \right]$$

e o seguinte resultado:

Resultado A.2

O determinante da matriz $\mathbf{V}_{i(k)}$ é dado por

$$|\mathbf{V}_{i(k)}| = (n_{i(k)} \sigma_\gamma^2 + \sigma_\epsilon^2) (\sigma_\epsilon^2)^{n_{i(k)} - 1} = (1 + n_{i(k)} c^2) (\sigma_\epsilon^2)^{n_{i(k)}} . \quad (\text{A.10})$$

□

Assim, teremos

$$\begin{aligned} \ell(c^2, \sigma_\epsilon^2) &= n \ln(2\pi) + \sum_{i(k)} \left\{ \ln(1 + n_{i(k)} c^2) + n_{i(k)} \ln(\sigma_\epsilon^2) + \tilde{\mathbf{y}}'_{i(k)} \frac{1}{\sigma_\epsilon^2} \left[\mathbf{I}_{i(k)} - \frac{c^2 \sigma_\epsilon^2 \mathbf{1}_{i(k)} \mathbf{1}'_{i(k)}}{\sigma_\epsilon^2 (1 + n_{i(k)} c^2)} \right] \tilde{\mathbf{y}}_{i(k)} \right\} \\ &= n \ln(2\pi) + n \ln \sigma_\epsilon^2 + \sum_{i(k)} \left\{ \ln(1 + n_{i(k)} c^2) + \tilde{\mathbf{y}}'_{i(k)} \frac{1}{\sigma_\epsilon^2} \left[\mathbf{I}_{i(k)} - \frac{c^2 \mathbf{1}_{i(k)} \mathbf{1}'_{i(k)}}{(1 + n_{i(k)} c^2)} \right] \tilde{\mathbf{y}}_{i(k)} \right\} \end{aligned}$$

que é a forma da verossimilhança reparametrizada em termos de σ_ϵ^2 e c^2 . Diferenciando com respeito a σ_ϵ^2 e igualando a zero teremos

$$\begin{aligned} \frac{\partial \ell}{\partial \sigma_\epsilon^2} &= \frac{n}{\sigma_\epsilon^2} - \frac{1}{(\sigma_\epsilon^2)^2} \sum_{i(k)} \left\{ \tilde{\mathbf{y}}'_{i(k)} \tilde{\mathbf{y}}_{i(k)} - (c \tilde{\mathbf{y}}_{i(k)})^2 / (1 + n_{i(k)} c^2) \right\} = 0 \\ \Rightarrow \frac{n}{\sigma_\epsilon^2} &= \frac{1}{(\sigma_\epsilon^2)^2} \sum_{i(k)} \left\{ \tilde{\mathbf{y}}'_{i(k)} \tilde{\mathbf{y}}_{i(k)} - (c \tilde{\mathbf{y}}_{i(k)})^2 / (1 + n_{i(k)} c^2) \right\} \end{aligned}$$

$$\Rightarrow \hat{\sigma}_c^2 = \frac{1}{n} \sum_{i(k)} \left\{ \tilde{y}'_{i(k)} \tilde{y}_{i(k)} - \frac{(c \tilde{y}_{i(k)})^2}{(1 + n_{i(k)} c^2)} \right\} \quad (\text{A.11})$$

E agora, substituindo em ℓ , a expressão $-2 \ln L$ depende apenas de um único parâmetro desconhecido, c^2 .

$$\ell(c^2) = n \ln(2\pi \hat{\sigma}_c^2) + \sum_{i(k)} \{\ln(1 + n_{i(k)} c^2)\} + n \quad (\text{A.12})$$

Encontrando o valor de c^2 que minimiza a equação $\ell(c^2)$, poderemos obter as estimativas de máxima verossimilhança para μ_k , $\hat{\sigma}_c^2$ e $\hat{\sigma}_\gamma^2$.

Para o caso balanceado, onde as unidades (divididas em m grupos) são observadas nos mesmos n pontos de tempo e não existem observações perdidas, teremos os seguintes estimadores de máxima verossimilhança (que não demonstraremos aqui), derivados por Herbach (1959):

-Se $(n-1)T_B \geq T_W$,

$$\hat{\sigma}_c^2 = \frac{1}{m \cdot (n-1)} T_W \quad ; \quad (\text{A.13})$$

$$\hat{\sigma}_\gamma^2 = \frac{1}{n \cdot m} \left(T_B - \frac{T_W}{n-1} \right) \quad ; \quad (\text{A.14})$$

-Se $(n-1)T_B < T_W$,

$$\hat{\sigma}_c^2 = \frac{1}{n \cdot m} T_T \quad ; \quad (\text{A.15})$$

$$\hat{\sigma}_\gamma^2 = 0 \quad ; \quad (\text{A.16})$$

onde

$$T_T = \sum_{i(k)j} (y_{i(k)j} - \mu_k)^2 \quad , \quad (\text{A.17})$$

$$T_B = \frac{1}{n} \sum_{i(k)} \left(\sum_{j=1}^{n_i} (y_{i(k)j} - \mu_k) \right)^2, \quad (\text{A.18})$$

e

$$T_W = T_T - T_B. \quad (\text{A.19})$$

Apêndice B

Manual do Programa CAR1

Richard H. Jones

Francis Boadi-Boateng

1992

B.1 Introdução

Este programa permite que ajustemos um modelo polinomial¹ a observações desigualmente espaçadas² que podem ter sido colhidas em diferentes pontos de tempo para diferentes unidades³. O conjunto de observações pode conter 'missing values'⁴ ou não.

Este programa possibilita que os erros sejam modelados como gaussianos e independentes (1) ou ainda com estrutura gaussiana AR1 (2). A matriz de covariância dos resíduos, nos dois casos, é a seguinte:

$$(1) \quad \sigma^2 \mathbf{W}_i = \sigma^2 \mathbf{I} \quad \text{I. matriz identidade;}$$

¹É um polinômio em função do tempo. Por ex., $y = a + b.t + c.t^2$, função quadrática, é um polinômio de grau 2.

²O intervalo de tempo entre duas observações não é sempre o mesmo.

³É muito difícil controlarmos o experimento de modo que todos os indivíduos sejam observados ao mesmo tempo. Entretanto, este programa não faz esta exigência.

⁴dados incompletos ou observações perdidas.

$$(2) \quad \sigma^2 \mathbf{W}_i = \sigma^2 \begin{bmatrix} 1 & \exp(-\alpha_0(t_2 - t_1)) & \cdots & \exp(-\alpha_0(t_{n_i} - t_1)) \\ \exp(-\alpha_0(t_2 - t_1)) & 1 & \cdots & \exp(-\alpha_0(t_{n_i} - t_2)) \\ \vdots & \vdots & \ddots & \vdots \\ \exp(-\alpha_0(t_{n_i} - t_1)) & \cdots & \cdots & 1 \end{bmatrix},$$

onde α_0 é um parâmetro positivo não linear desconhecido e t_1, t_2, \dots, t_{n_i} são os tempos de observação em ordem crescente;

obs: poderá ainda, na estrutura (2), ser incluída uma variação referente aos erros observacionais. Neste caso, aos elementos da diagonal da matriz (2) é acrescida mais uma componente de variação, σ_0^2 .

B.2 O Modelo

O modelo considerado neste programa é o modelo de Laird-Ware, desenvolvido a partir do trabalho de Harville:

$$\underset{\sim}{y}_i = \underset{\sim}{\mathbf{X}}_i \underset{\sim}{\beta} + \underset{\sim}{\mathbf{Z}}_i \underset{\sim}{\gamma}_i + \underset{\sim}{\epsilon}_i$$

onde,

- $\underset{\sim}{y}_i$ é o vetor de observações da unidade i ;
- $\underset{\sim}{\mathbf{X}}_i$ é a matriz de desenho (polinomial) da unidade i ;
- $\underset{\sim}{\beta}$ é o vetor de parâmetros;
- $\underset{\sim}{\mathbf{Z}}_i$ é a matriz de desenho dos efeitos aleatórios da unidade i ;
- $\underset{\sim}{\gamma}_i$ é o vetor de efeitos aleatórios da unidade i ;
- $\underset{\sim}{\epsilon}_i$ é o vetor de resíduos da unidade i .

Devemos considerar ainda as seguintes suposições:

- $\underset{\sim}{\gamma}_i \sim iN(\underset{\sim}{0}, \sigma^2 \mathbf{B}_i)$, onde σ^2 é a variância do resíduo (mean square error - mse) e \mathbf{B} é a matriz de correlação dos efeitos aleatórios. Observe que $\underset{\sim}{\gamma}_j$ é independente de $\underset{\sim}{\gamma}_l$, ou seja, os efeitos aleatórios são independentes para diferentes unidades ($j \neq l$).

$-\tilde{\epsilon}_i \sim iN(0, \sigma^2 \mathbf{W}_i)$, onde \mathbf{W}_i é a matriz de correlação dos resíduos. Note que os resíduos são independentes para diferentes unidades.

$-\tilde{\gamma}_i$ é independente de $\tilde{\epsilon}_i$.

B.3 Estrutura da Matriz \mathbf{X}_i

Como em Dados Longitudinais estamos interessados em estudar o comportamento de um ou mais tratamentos através do tempo, a construção da nossa matriz \mathbf{X}_i dependerá de algumas informações:

- a) n° de observações tomadas da unidade i ;
- b) n° de tratamentos (grupos) em estudo;
- c) ordem do polinômio nos diversos tratamentos.

Estas informações definirão a dimensão e a estrutura básica da matriz \mathbf{X}_i . Outra informação importante é a respeito de como serão definidas as covariáveis ou variáveis ‘dummy’ agrupadoras. Por exemplo, para um experimento envolvendo apenas um tratamento teremos a seguinte matriz \mathbf{X}_i ⁵:

$$\mathbf{X}_i = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & \cdots & t_{i1}^{p-1} \\ 1 & t_{i2} & t_{i2}^2 & \cdots & t_{i2}^{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 & \cdots & t_{in_i}^{p-1} \end{bmatrix}$$

Para experimentos com dois ou mais tratamentos deveremos fazer uso das covariáveis (variáveis ‘dummy’ agrupadoras). Considere o simples exemplo abaixo:

Suponha que dois tratamentos sejam comparados de tal forma que a cada unidade seja aplicado um desses tratamentos. Desta forma, teremos dois grupos de unidades com uma variável indicadora que assume o valor 0 para um grupo e o valor 1 para o outro grupo. Considere uma matriz que consiste de duas colunas, e define um polinômio de

⁵isto representa um modelo polinomial de ordem p , com n_i observações tomadas da unidade i .

primeiro grau

$$\mathbf{X}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{bmatrix} ;$$

chamaremos esta matriz de matriz básica⁶. Para construir as matrizes \mathbf{X}_i de cada unidade, devemos multiplicar as colunas da matriz básica pela variável indicadora (0 ou 1), aumentando esta mesma matriz com as colunas resultantes da multiplicação. Assim teremos uma matriz \mathbf{X}_i com quatro colunas. O modelo ajustado, então, será uma diferente reta para cada grupo:

grupo 1 - valor da variável indicadora = 0. A matriz resultante é dada por

$$\mathbf{X}_i = \begin{bmatrix} 1 & t_{i1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & 0 & 0 \end{bmatrix} .$$

grupo 2 - valor da variável indicadora = 1. A matriz resultante é dada por

$$\mathbf{X}_i = \begin{bmatrix} 1 & t_{i1} & 1 & t_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & 1 & t_{in_i} \end{bmatrix} .$$

E assim teremos uma reta ajustada para o primeiro grupo e outra para o segundo grupo. No entanto os parâmetros estimados para o segundo grupo (3° e 4° parâmetros) serão na verdade, desvios em relação aos parâmetros do primeiro grupo (1° e 2° parâmetros, respectivamente).

Não é necessário, no entanto, multiplicar todas as colunas da matriz básica pela

⁶é uma matriz de desenho \mathbf{X}_i , para um tratamento de referência ou controle.

variável 'dummy'. Para efeito de 'redução do modelo', ou seja, eliminação dos parâmetros não significativos do modelo, podemos multiplicar a variável 'dummy' apenas pelas p_1 primeiras colunas da matriz básica ($p_1 \leq p$) onde p é a dimensão desta matriz. Uma outra alternativa para testar a redução de um modelo é através do teste de contrastes.

O método de estimação dos parâmetros pode ser máxima verossimilhança exata ou máxima verossimilhança restrita, sempre fazendo uso do filtro de Kalman⁷ para calcular a verossimilhança e obter as estimativas de parâmetros e do processo de otimização não-linear de Schnabel para estimar os demais parâmetros não lineares (σ^2 , elementos de \mathbf{B} e α_0).

B.4 Execução do Programa

Para executar o programa devemos, inicialmente, digitar o comando `exec CAR1`. A seguir será pedido o nome de 3 arquivos: o arquivo de dados, o arquivo de saída e o arquivo de parâmetros.

4.1-Arquivo de dados

A primeira linha do arquivo de dados deverá conter as seguintes informações:

$$NCV \quad NTVCV$$

onde NCV é o número de covariáveis ou variáveis agrupadoras ('dummy') para o modelo desejado. Este valor irá depender do número de tratamentos (ou grupos) do experimento. Ex: se o experimento possuir 3 tratamentos, então precisaremos de duas covariáveis para discriminar os tratamentos entre os indivíduos, isto é, assumiremos $NCV = 2$ e 0 0 para o tratamento 1, 1 0 para o tratamento 2, e 0 1 para o tratamento 3. Poderemos ainda, se estivermos interessados em verificar a interação entre o tratamento 2 e o tratamento 3, usar os valores 1 1 para as duas covariáveis. Por outro lado, se notarmos que não

⁷processo recursivo destinado a estimar os parâmetros fixos ou prever efeitos aleatórios.

existe diferença entre os tratamentos 2 e 3, poderemos reduzir o número de parâmetros, simplificando o modelo, e então deveremos fazer $NCV = 1$ e $\underline{0}$ para o tratamento 1, e $\underline{1}$ para os tratamentos 2 e 3.

Se para todas as unidades forem obtidas algumas informações adicionais, deveremos definir este número de informações como $NTVCV$. Ex: $NTVCV = 3 \implies$ Sexo (1-masculino, 0-feminino), Classe Sócio-econômica (1-baixa, 2-média, 3-alta) e Idade (18,19,20,...anos).

Nas linhas seguintes devemos proceder à entrada dos dados, sempre com um espaço em branco entre eles, preenchendo sempre 2 linhas para cada unidade envolvida no experimento. Na primeira linha pede-se $NOBS(i)$, o número de observações para a unidade i , e logo em seguida ($X(j), j = 1, \dots, NOBS(i)$), os $NOBS(i)$ pontos de tempo observados para a mesma unidade. Na segunda linha, colocamos os valores das NCV covariáveis para esta unidade ($CV(k), k = 1, \dots, NCV$) e em seguida, as $NOBS(i)$ observações ($y(j), j = 1, \dots, NOBS(i)$). Quando existir apenas um tratamento ou grupo para todas as unidades, deveremos fazer $NCV = 0$ e as informações ($CV(k), k = 1, NCV$) deverão ser desprezadas. Se $NTVCV > 0$, deveremos acrescentar mais $NTVCV$ linhas de dados para cada unidade, correspondentes aos valores de cada informação adicional sobre a unidade (obs: em cada uma das $NTVCV$ linhas, deve-se entrar com um valor para cada um dos $NOBS(i)$ pontos de tempo observados). Ex: $NTVCV = 2 \implies$ Sexo(0: feminino ou 1: masculino) e Idade(anos).

Indivíduo i - 6 observações, sexo masculino, 18 anos na ocasião da primeira observação:

(entrada dos valores)

1 1 1 | | |

18 21 23 24 26 33

4.2-Arquivo de parâmetros

A estrutura do arquivo de parâmetros deve ser a seguinte:

NQ, NX, NXCV(1), ..., NXCV(NCV), NXTVCV(1), ..., NXTVCV(NTVCV)

MODEL(1), MODEL(2)

P(1), P(2)

NOD

2 Índices de U *PR(1)*

2 Índices de U *PR(2)*

⋮

2 Índices de U *PR(NOD)*

ORIGIN SCALE

REML

IOPT

NCTRST

Número de linhas e título do contraste 1

matriz de contraste

Número de linhas e título do contraste 2

matriz de contraste

⋮

Número de linhas e título do contraste *NCTRST*

matriz de contraste

NFILES

IRES

A primeira linha do arquivo de parâmetros deve trazer as informações referentes a:

-*NQ*, o número de colunas da matriz Z_i . Ex: se $NQ = 2$, $Z_i = \begin{bmatrix} 1 & t \\ \sim & \sim \end{bmatrix}$. (obs: se não existem efeitos aleatórios devemos fazer $NQ = 1$ e $NOD = 0$);

-*NX*, o número de colunas da matriz básica, que irá definir a ordem do polinômio

que representa a curva média da população:

- $NXCV(1), \dots, NXCV(NCV)$, denotam o número de colunas da matriz básica que serão multiplicadas por cada uma das NCV covariáveis;

- $NXTVCV(1), \dots, NXTVCV(NTVCV)$, denotam o número de colunas da matriz de desenho da população básica que serão multiplicadas por cada uma das $NTVCV$ informações adicionais.

ATENÇÃO: $NCV + NTVCV$ deve ser ≤ 5 , senão este limite deverá ser alterado no programa principal, bem como nas sub-rotinas, e recompilado em Fortran.

Na segunda linha deveremos definir:

- $MODEL(1)$, que assumirá o valor 0 se estivermos supondo que os resíduos são independentes e 1 se estivermos supondo que os resíduos têm uma estrutura AR(1) de tempo contínuo;

- $MODEL(2)$, assumirá o valor 1 se erros observacionais forem incluídos no modelo e 0 se não forem incluídos.

Na terceira linha, se $MODEL(1) = 1$, deveremos indicar o valor inicial para $\ln \alpha_0$, logaritmo do coeficiente auto-regressivo α_0 de tempo contínuo do AR(1). Se ainda, $MODEL(2) = 1$, deveremos indicar em seguida o valor inicial para σ_0 , o desvio padrão dos erros observacionais. $P(1) = 0.0$ e $P(2) = 0.0$ são estimativas iniciais naturais para os parâmetros não lineares descritos acima. (obs: se $MODEL(1) = 0$, esta linha deverá ser desprezada).

Na quarta linha deveremos definir o valor da variável NOD . NOD é o número de elementos da matriz U a serem estimados. U , por sua vez é uma matriz triangular superior obtida por decomposição de Choleski⁸. A matriz B de covariâncias dos efeitos aleatórios é dada por $U'U$.

A seguir teremos NOD linhas onde deveremos indicar os índices para locação de cada elemento da matriz U (NOD elementos) seguido da sua respectiva estimativa inicial,

⁸decompõe a matriz B de forma que $B=U'U$, onde U é uma matriz triangular superior. Esta decomposição assegura que a matriz de covariância será não negativa definida.

sendo uma linha para cada um dos NOD elementos. Ex: $NOD = 3$

linha 1: 1 1 0.5 (elemento a_{11} da matriz U com estimativa inicial 0.5)

linha 2: 1 2 0.0 (elemento a_{12} da matriz U com estimativa inicial 0.0)

linha 3: 2 2 0.5 (elemento a_{22} da matriz U com estimativa inicial 0.5)

Atenção: estimativas iniciais dos elementos da diagonal não devem ser iguais a zero pois podem produzir um mínimo local na função de verossimilhança em vez de um mínimo global. É interessante que o programa seja rodado uma primeira vez apenas para se obter uma boa estimativa dos 'chutes iniciais' e em seguida rodado novamente com os chutes iniciais estimados anteriormente para se obterem as estimativas finais. Para grandes conjuntos de dados pode ocorrer que uma pequena alteração nas estimativas iniciais modifique as estimativas dos demais parâmetros.

Na linha seguinte devem ser definidos dois valores:

- *ORIGIN*, a origem do tempo;

- *SCALE*, a escala do tempo. 0.0 e 1.0 são os valores usuais. Esta opção permite que o tempo de origem seja alterado para a média geral dos tempos de observação com o objetivo de tornar a matriz de desenho melhor condicionada.

Na linha seguinte deveremos definir se deverá ser usado estimação por máxima verossimilhança (0) ou por máxima verossimilhança restrita (1). A estimação por máxima verossimilhança restrita fornecerá estimativas das variâncias menos viciadas que as estimativas de máxima verossimilhança. Deve ser usada quando o número de parâmetros fixos não for pequeno em relação ao número total de observações, ou seja, quando $\frac{\sum_i NOBS(i)}{\sum_i NOBS(i)}$ for bem menor que 1.

Na linha seguinte deveremos definir o valor de *IOPT* (0 ou 1). Se *IOPT* = 1, um programa de otimização não linear será executado para estimar os parâmetros não lineares desconhecidos. Estes parâmetros não lineares são: α_0 , o coeficiente de auto-regressão, os elementos da matriz B , e σ_0^2 , a variância dos erros observacionais (quando pedidos). Se *IOPT* = 0, a otimização não será executada. Então, o programa considerará as estimativas iniciais desses parâmetros como sendo estimativas obtidas de uma execução

anterior. Esta opção permite ao programa gerar testes para contrastes, estruturar dados para gráficos, etc, e assim poupar tempo computacional.

Na linha seguinte deveremos definir *NCTRST*, o número de contrastes de efeitos fixos que estamos desejando testar. Os contrastes possibilitam testar se um ou mais parâmetros são iguais a zero, ou seja, são desprezíveis para o modelo, ou ainda, testar se existe diferença entre um ou mais pares de parâmetros ou qualquer tipo de contraste possível. Se *NCTRST* = 0, nenhum contraste será testado. Se *NCTRST* > 0, as matrizes de contrastes devem entrar a seguir, em *NCTRST* blocos, da seguinte forma:

- Entra com o número de linhas (*c*) da matriz de contraste (2 primeiros espaços) e em seguida (a partir do quarto espaço) define um comentário a respeito do contraste. Então, na linha seguinte, entra com a matriz de contrastes linha por linha. Este processo deve ser executado para cada contraste, ou seja, *NCTRST* vezes.

Ex:

```
5
1 _ _ beta1=0
0 1 0 0 0 0 0 0 0 0
3 _ _ beta1=beta2=beta3=0
0 1 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0
2 _ _ alfa1=alfa2 e alfa3=alfa1
1 0 -1 0 0 0 0 0 0 0
0 0 0 0 1 0 1 0 0 0
1 _ _ beta1=média(beta2,3,4,5)
0 0 0 1 0 1 0 1 0 1
1 _ _ beta5=média(beta1,2,3,4)
0 0 0 1 0 1 0 1 0 1
```

⁹lembre-se que beta2,3,4 e 5 são desvios em relação a beta1

(obs: a dimensão da matriz de contrastes é $c \times b$, onde c é o número de linhas do contraste e b é o número de colunas de X_i .)

Na linha seguinte devemos definir o valor de *NFILES* que é o número de arquivos que devem ser criados para a saída das médias, erros padrões das médias, e desvios padrões das unidades para dados valores de t e de covariáveis. Estas saídas estruturam os dados para permitir a construção de gráficos em algum software apropriado. Se *NFILES* > 0, o programa irá ler a linha seguinte que define:

-*NPLOTS*, o número de gráficos a serem gerados (geralmente um para cada grupo ou código de covariável diferente);

-*NPTS*, o número de pontos para cada gráfico;

- $t(1), t(2), \dots, t(NPTS)$, a locação dos *NPTS* pontos. Se existem covariáveis o programa lê mais *NPLOTS* linhas, uma para cada código de valores das covariáveis. Este processo será executado *NFILES* vezes. Isto irá permitir a construção de diferentes intervalos de confiança para diferentes curvas médias. O programa irá solicitar o nome dos *NFILES* arquivos de saída. Ex:

1 : (um arquivo de médias, erros-padrões e desvios padrões)

4 6 20 30 40 50 60 70 : (estrutura de dados para 4 gráficos com 6 pontos de tempo: 20, 30, 40, 50, 60 e 70)

0 0 0 : (valores das covariáveis)

1 0 0

0 1 0

1 1 1

E finalmente na última linha deveremos definir o valor da variável *IRES* (0 ou 1). Se *IRES* = 1, o programa irá pedir o nome de outro arquivo de saída onde irão constar os dados referentes a todas as unidades em 5 colunas.

As colunas se referem a:

1. Tempo observado;

2. Observação;

3. Observação predita;
4. Desvio padrão da observação predita;
5. Resíduos padronizados (para investigação de normalidade, correlação serial, outliers, etc.).

Se $IRENS = 0$, não será pedido outro nome de arquivo de saída e conseqüentemente nenhuma das informações acima será salva.

4.3 - Arquivo de saída (análise)

Comentários e interpretações dos resultados salvos no arquivo de saída serão feitos no exemplo dado abaixo, com base no artigo de JONES (1990).

B.5 Exemplo

Consideremos o conjunto de dados do artigo de JONES (1990). Abaixo descrevemos resumidamente como se estrutura o arquivo de dados, seguido de alguns comentários a respeito de tal estrutura:

2 0

• Estamos considerando um modelo com 3 grupos, ou seja, um grupo referência (experimento 1) mais dois outros grupos (experimentos 2 e 3). Daí, $NCV = 2$. Como não temos informações adicionais, $NTVCV = 0$.

5 0 7 26 42 70

0 0 1.5 1.5 1.5 1.5 1.5

• A primeira unidade é observada em 5 pontos de tempo, são eles: 0, 7, 26, 42 e 70. Na segunda linha o código 0 0 indica que esta unidade pertence ao grupo referência. Na seqüência surgem as 5 observações.

⋮

4 0 26 42 70

0 0 .5 .56 .75 .8

• A sétima unidade, também do grupo referência, foi observada apenas 4 vezes pois

tem uma observação perdida no tempo 7.

⋮

4 0 7 14 18

1 0 .5 1.13 1.31 1.88

3 0 14 48

1 0 .56 .25 .75

• Os dados acima se referem às duas primeiras unidades do segundo grupo (experimento 2, código 1 0). A primeira foi observada em 4 ocasiões e a segunda foi observada em apenas 3 ocasiões (observação perdida no tempo 7).

⋮

4 0 8 14 29

0 1 1.56 1.55 2.25 4.6

• E finalmente a última unidade do terceiro grupo (experimento 3, código 0 1) com 4 observações nos tempos 0, 8, 14 e 29.

A seguir devemos definir o arquivo de parâmetros para o conjunto de dados acima:

3 4 4 4

0 0

3

1 1 2.6

1 2 0.13

1 3 -0.001

18.08847 1.0

0

1

0

0

0

A primeira linha indica que a matriz Z_i possui 3 colunas, ou seja,

$$Z_i = \begin{bmatrix} 1 & t & t^2 \end{bmatrix},$$

onde t é o vetor de tempos observados. A matriz X_i , $n_i \times 12$, é dada por:(considerando também as informações do arquivo de dados)

$$X_i = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & t_{i1}^3 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 & t_{in_i}^3 & 0 & \dots & 0 \end{bmatrix}$$

para o experimento 1 (grupo de referência).

$$X_i = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & t_{i1}^3 & 1 & t_{i1} & t_{i1}^2 & t_{i1}^3 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 & t_{in_i}^3 & 1 & t_{in_i} & t_{in_i}^2 & t_{in_i}^3 & 0 & 0 & 0 & 0 \end{bmatrix}$$

para o experimento 2, e

$$X_i = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & t_{i1}^3 & 0 & 0 & 0 & 0 & 1 & t_{i1} & t_{i1}^2 & t_{i1}^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 & t_{in_i}^3 & 0 & 0 & 0 & 0 & 1 & t_{in_i} & t_{in_i}^2 & t_{in_i}^3 \end{bmatrix}$$

para o experimento 3.

A segunda linha do arquivo de parâmetros, com dois zeros, indica que não existe correlação serial no modelo. A terceira linha indica que 3 elementos da matriz U deverão ser estimados, e a linha seguinte define os elementos desta matriz a serem estimados: a_{11}, a_{12}, a_{13} (somente os elementos da primeira linha de U pois aqui já estamos eliminando os demais elementos da matriz, que são não significativos para o modelo).

A linha seguinte define a origem do tempo, que no caso será o tempo médio para todas as observações, 18.08847, e a escala que será igual a 1.

As linhas seguintes definem, nesta ordem, que:

- (0) será usada estimação de máxima verossimilhança;
- (1) um programa de otimização não-linear será executado para estimar os parâmetros da matriz \mathbf{U} ;

- (0) não serão testados contrastes;
- (0) dados não serão estruturados para gráficos;
- (0) não serão salvas as informações (já descritas acima) em um arquivo de saída.

Com estas informações, podemos rodar o programa e os resultados serão salvos em um arquivo de saída. Este arquivo de saída apresentará o seguinte conteúdo:

- O primeiro bloco traz informações sobre o arquivo de parâmetros como: elementos da matriz \mathbf{U} que foram estimados, com respectivas estimativas iniciais, estrutura das matrizes \mathbf{X}_i , existência ou não de correlação serial, número de unidades e de observações, origem do tempo, escala e método de estimação.

- A seguir, temos os valores de $-2 \ln L$ e o respectivo valor de AIC ($-2 \ln L + (2 \times n^\circ$ de parâmetros)), o valor da variância do resíduo (mse) e seu respectivo desvio padrão.

- Em seguida surgem as estimativas dos parâmetros lineares acompanhados de seus respectivos erros padrões e valores da estatística 't de Student'.

- Após isto teremos as 3 estimativas pedidas dos parâmetros não-lineares (elementos da matriz \mathbf{U}), a matriz $\sigma \mathbf{U}$ obtida com estas estimativas e a matriz $\sigma^2 \mathbf{B}$.

- E finalmente, em seguida teremos para cada unidade os valores de seus respectivos efeitos aleatórios γ_i , desvios padrões e resíduos recursivos (padronizados).

Referências Bibliográficas

- [1] BERGER, M. P. F. (1986) *A Comparison of Efficiencies of Longitudinal, Mixed Longitudinal and Cross-Sectional Designs*. JOURNAL OF EDUCATIONAL STATISTICS, **11**, 3, 171-81.
- [2] BICKEL, P. J. & DOKSUM, A. M. (1977) *Mathematical Statistics: Basic Ideas and Selected Topics*. HOLDEN-DAY, San Francisco.
- [3] BOX, G. E. P. & JENKINS, G. M. (1976) *Time Series Analysis, Forecasting and Control*, revised edition. HOLDEN-DAY, San Francisco.
- [4] BROEMELING, L. (1985) *Bayesian Analysis of Linear Models*, MARCEL DEKKER, New York.
- [5] CHI, E. M. & REINSEL, G. C. (1989) *Models for Longitudinal Data with Random Effects and AR(1) Errors*. JASA, **84**, 406, 452-9.
- [6] CHRISTENSEN, R. A. (1987) *Plane Answers to Complex Questions*, SPRINGER-VERLAG, New York.
- [7] CROWDER, M. J. & HAND, D. J. (1990) *Analysis of Repeated Measures*, CHAPMAN AND HALL, London.
- [8] DIDERRICH, G. T. (1985) *The Kalman Filter from the Perspective of Goldberger-Theil Estimators*. THE AMERICAN STATISTICIAN, **39**, 3, 193-8.
- [9] DUNCAN, D. B. & HORN, S. D. (1972) *Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis*. JASA, **67**, 340, 815-21.

- [10] EFRON, B. & GONG, G. A. (1983) *A Liberally Look at the Bootstrap, the Jackknife and Cross-Validation*, THE AMERICAN STATISTICIAN, **37**, 170-74.
- [11] GABOW, P. A., JOHNSON, A. M., KAEHNY, W. D., KIMBERLING, W. J., LEZOTTE, D. C., DULEY, I. T., JONES, R. H. (1992) *Factors Affecting the Progression of Renal Disease in Autosomal-Dominant Polycystic Kidney Disease*, KIDNEY INTERNATIONAL, **41**, 1311-9.
- [12] GRIZZLE, J. E. & ALLEN, D. M. (1969) *Analysis of Growth and Dose Response Curves*, BIOMETRICS, **25**, 357-81.
- [13] HARVILLE, D. A. (1976) *Extension of the Gauss-Markov Theorem to include the Estimation of Random Effects*, THE ANNALS OF STATISTICS, **4**, 384-95.
- [14] HARVILLE, D. A. (1977) *Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems*, JASA, **72**, 358, 320-40.
- [15] HARVILLE, D. A. (1985) Notas de aula.
- [16] HENDERSON, C. R. (1950) *Estimation of Genetic Parameters*, ANNALS OF MATHEMATICAL STATISTICS, **21**, 309-16.
- [17] HERBACH, L. H. (1959) *Properties of Model II-type Analysis of Variance Tests*, ANNALS OF MATHEMATICAL STATISTICS, **30**, 939-59.
- [18] HUI, S. L. & BERGER, J. O. (1983) *Empirical Bayes Estimation of Rates in Longitudinal Studies*, JASA, **78**, 384, 753-60.
- [19] JENRICH, R. L. & SCHLUCHTER, M. D. (1986) *Unbalanced Repeated-Measures Models with Structured Covariances Matrices*, BIOMETRICS, **42**, 805-20.
- [20] JONES, R. H. (1990) *Serial Correlation or Random Subject Effects?*, COMMUNICATIONS OF STATISTICS SIMULATION, **19**, 1105-23.

- [21] JONES,R.H.&ACKERSON.L.M.(1990) *Serial Correlation in Unequally Spaced Longitudinal Data*, BIOMETRIKA, **77**, 1, 721-31.
- [22] JONES,R.H.&BOADI-BOATENG.F.(1991) *Unequally Spaced Longitudinal Data with AR(1) Serial Correlation*, BIOMETRICS, **47**, 161-75.
- [23] JONES,R.H.(1993) *Longitudinal Data with Serial Correlation: A State Space Approach*, CHAPMAN & HALL, London.
- [24] KALMAN,R.E.(1960) *A New Approach to Linear Filtering and Prediction Problems*, TRANSACTIONS ASME J. BASIC ENGINEERING, **82D**, 39-45.
- [25] KALMAN,R.E.&BUCKY,R.S.(1961) *New Results in Linear Filtering and Prediction Theory*, TRANSACTIONS ASME J. BASIC ENGINEERING, **83D**, 95-108.
- [26] LAIRD,N.M.&WARE,J.H.(1982) *Random-Effects Models for Longitudinal Data*, BIOMETRICS, **38**, 963-74.
- [27] LANGE,N.&RYAN,L.(1989) *Assessing Normality in Random Effects Models*, THE ANNALS OF STATISTICS, **17**, 2, 621-42.
- [28] LITTLE,R.J.A.&RUBIN,D.B.(1987) *Statistical Analysis with Missing Data*, WILEY, New York.
- [29] LOUIS,T.A.(1988) *General Methods for analysing Repeated Measures*, STATISTICS IN MEDICINE, **7**, 29-45.
- [30] MEINHOLD,D.R.J.&SINGPURWALLA,N.D.(1983) *Understanding the Kalman Filter*, THE AMERICAN STATISTICIAN, **37**, 2, 123-7.
- [31] PHADKE,M.S.(1981) *Quality Audit Using Adaptive Kalman Filtering*, ASQC QUALITY CONGRESS TRANSACTIONS , San Francisco, 1045-52.

- [32] POTTHOFF, R.F. & ROY, S.N. (1961) *A Generalized Multivariate Analysis of Variance Model useful especially for Growth Curve Problems*, *BIOMETRIKA*, **51**, 3 e 4, 313-26.
- [33] PURI, M.L. & SEN, P.K. (1971) *Nonparametrics Methods in Multivariate Analysis*, JOHN WILEY, New York.
- [34] ROBINSON, G.K. (1991) *That BLUP is a Good Thing: The Estimation of Random Effects*, *STATISTICAL SCIENCE*, **1**, 15-51.
- [35] SCHLUCHTER, M.D. (1988) *Analysis of Incomplete Multivariate Data using Linear Models with Structured Covariance Matrices*, *STATISTICS IN MEDICINE*, **7**, 317-24.
- [36] SINGER, B. (1985) *Longitudinal Data Analysis*, *ENCYCLOPEDIA OF STATISTICAL SCIENCES*, **5**, 142-55, WILEY, New York.
- [37] SINGER, J.M. & ANDRADE, D.F. (1986) *Análise de Dados Longitudinais*, SINAPE.
- [38] SOUZA, L. & IEMMA, A.F. (1990) *Curvas de Crescimento: Análise de Dados Incompletos com Resíduos Autocorrelacionados*, *REVISTA DE MATEMÁTICA E ESTATÍSTICA*, **8**, 45-55.
- [39] TITUS, H. (1977) *Advances in Passive Target Tracking*, Rep. No. NPS-62Ys-77071, Naval Postgraduate School, Monterey, California.
- [40] VACEK, P.M., MCKEY, R., BELL, D.V. (1989) *Application of a Two-stage Random Effects Model to Longitudinal Pulmonary Function Data From Sarcoidosis Patients*, *STATISTICS IN MEDICINE*, **8**, 189-200.
- [41] WARE, J.H. (1985) *Linear Models for the Analysis of Longitudinal Studies*, *THE AMERICAN STATISTICIAN*, **39**, 2, 95-101.
- [42] WATERNAUX, C., LAIRD, N.M., WARE, J.H. (1989) *Methods for Analysis of Longitudinal Data: Blood-lead Concentrations and Cognitive Development*, *JASA*, **84**, 405, 33-41.

- [43] WILSON,P.D.(1988) *Autoregressive Growth Curves and Kalman Filtering*, STATISTICS IN MEDICINE, **7**, 73-86.
- [44] WOLFINGER,R.(1993) *Covariance Structure Selection in General Mixed Models*, COMMUNICATIONS OF STATISTICS SIMULATION, **22**, 1079-106.
- [45] ZERBE,C. (1979) *Randomization Analysis of the Completely Randomized Design Extended to Growth and Response Curves*. JASA, **74**, 215-21.