

Modelos Não-Lineares para Análise de Dados Longitudinais

Ana Lúcia de Souza Brandão

Orientador: Prof. Dr. Dalton Francisco de Andrade (USP)

Co-orientador: Prof. Dr. Mauro Sérgio de Freitas Marques (Unicamp)

Dissertação apresentada ao Instituto de Matemática Estatística e Computação Científica (IMECC) da Universidade Estadual de Campinas, como requisito parcial para obtenção do grau de Mestre em Estatística.

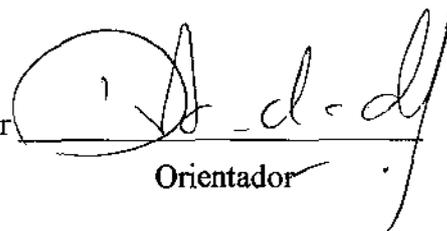
Campinas, dezembro/96

Modelos Não-Lineares para Análise de Dados Longitudinais

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Ana Lúcia de Souza Brandão e aprovada pela Comissão Julgadora.

Campinas, 13 de dezembro de 1996

Prof. Dr.



Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em ESTATÍSTICA.

UNIDADE	B3C
N.º CHAMADA:	11 UNICAMP
	B733m
V.	
TÍTULO	30565
PREÇO	281,97
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	24/05/97
N.º CPD	

CM-00098842-1

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP

Brandão, Ana Lúcia de Souza

B733m Modelos não-lineares para análise de dados longitudinais / Ana Lúcia de Souza Brandão -- Campinas, [S.P. :s.n.], 1996.

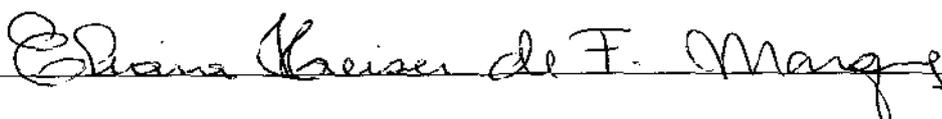
Orientadores : Dalton Francisco de Andrade, Mauro Sérgio de Freitas Marques

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

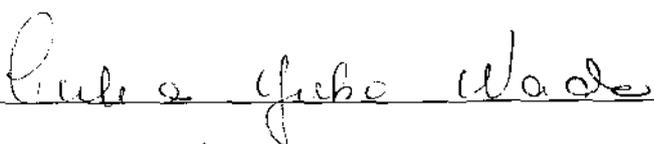
1. Estudos longitudinais. 2. Modelos não lineares (Estatística). 3. Análise de regressão. I. Andrade, Dalton Francisco de. II. Marques, Mauro Sérgio de Freitas. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Dissertação de Mestrado defendida e aprovada em 13 de dezembro de 1996

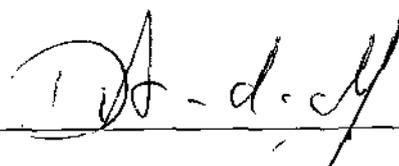
pela Banca Examinadora composta pelos Profs. Drs.



Prof (a). Dr (a). ELIANA HEISER DE FREITAS MARQUES



Prof (a). Dr (a). CÍCÍLIA YUKO WADA



Prof (a). Dr (a). DALTON FRANCISCO DE ANDRADE

Aos meus pais, Nilcinha e Brandão

Agradecimentos

Antes de mais ninguém e de forma muito especial, aos meus orientadores Profs. Dalton Andrade e Mauro Marques, pelo constante incentivo, dedicação e paciência!

Aos amigos do mestrado, que na maioria já defenderam suas teses e me fizeram acreditar que seria possível. Em especial, aos “puxões de orelha” do Carlos, ao apoio incondicional da Ana, à amizade e carinho dos “afilhados” Carlos Henrique e Ana Paula, aos votos de confiança e companheirismo do Sérgio. À Sílvia, Hildete, Aluizio, Aloísio, Bernadete e tantos outros amigos que contribuíram na dissertação através de dicas e “ajudas” técnicas. A estes e a todos os amigos da Unicamp com quem convivi e compartilhei bons (e maus) momentos ao longo do curso, credito parte deste trabalho.

Aos professores da ENCE e Unicamp que colaboraram com a minha formação acadêmica, incentivaram a acertada decisão de fazer o mestrado e contribuíram de alguma forma para que esta fosse de fato uma etapa valiosa na minha vida profissional. Em especial, aos professores dos Departamentos de Matemática e de Estatística da Unicamp que contribuíram de forma direta na dissertação com críticas e sugestões.

À USP, ao Prof. Emanuel (Unicamp) e à Prof^a Estela (UFSCar), por me cederem os dados utilizados nos exemplos.

À Nielsen e à CBPA, empresas onde trabalhei durante a elaboração da dissertação, pela compreensão e ajuda ao me concederem tempo e permitirem a utilização de seus equipamentos. Aos amigos do trabalho, pelas palavras de apoio.

À minha família, aos amigos de Campos, do Rio e de São Paulo, pelo gesto de carinho e compreensão demonstrado ao deixarem de me perguntar sobre como estava a tese. Finalmente posso responder: acabei!!!

Índice

	Página
Capítulo I - Dados Longitudinais	8
1.1. Introdução	8
1.2. Análise de Dados Longitudinais através do Ajuste de Curvas	10
1.3. Revisão da Literatura	12
Capítulo II - Modelos Não-Lineares para o Vetor de Médias com Estruturas Especiais para a Matriz de Covariância	 15
2.1. O Modelo Geral	15
2.2. Modelos Não-Lineares para o Vetor de Médias	16
2.3. Estruturas para a Matriz de Covariância	25
Capítulo III - Inferência em Modelos Não-Lineares com Estrutura de Covariância Geral	 30
3.1. Método de Mínimos Quadrados Generalizados	31
3.1.1. Procedimentos Iterativos para Estimação por Mínimos Quadrados.	36
3.1.2. Propriedades dos Estimadores de Mínimos Quadrados.....	41
3.2. Testes de Hipóteses e Intervalos de Confiança	47
3.2.1. Testes de Hipóteses	47
3.2.2. Intervalos de Confiança	49
3.3. Generalização para M Tratamentos	52
3.3.1. Testes de Hipóteses sobre Parâmetros de Diferentes Tratamentos..	59

	Página
Capítulo IV - Inferência em Modelos Não-Lineares Considerando Estruturas Especiais de Covariância	62
4.1. Estrutura de Independência	62
4.2. Estrutura Uniforme	65
4.3. Estrutura de Modelos de Efeitos Aleatórios	69
Capítulo V - Aplicação Prática	87
Capítulo VI - Considerações Finais	101
Apêndice: Programas	102
Referências Bibliográficas	133

Modelos Não-Lineares para Análise de Dados Longitudinais

Capítulo I - Dados Longitudinais

1.1. Introdução

Entende-se por dados longitudinais no sentido amplo, conjuntos de observações onde uma ou mais variáveis são medidas mais de uma vez em cada uma das unidades amostrais (por exemplo, pessoas, animais, ou plantas). As variáveis de interesse são medidas ao longo de uma dimensão específica, como por exemplo o tempo. Esta definição abrange dados provenientes de diversos tipos de estudos. Alguns exemplos são: registros de sinais captados em aparelhos de eletrocardiograma; ensaios clínicos para investigar a reação de pacientes contaminados por alguma doença sob diferentes dosagens de um medicamento; comparação da influência de diferentes tipos de ração no ganho de peso de animais; estudos do crescimento humano, medido pela altura em várias idades para comparação da evolução do crescimento de pessoas de diferentes países.

Embora as variáveis possam ser medidas ao longo de uma dimensão qualquer, como distância de um ponto no espaço ou dosagens de uma vacina, usaremos, sem perda de generalidade, o tempo como dimensão de interesse. Há algumas especificidades no tipo de dado que será considerado neste trabalho:

- i.* a origem do tempo é definida em relação a algum evento, como por exemplo, a hospitalização do paciente, nascimento, etc.;
- ii.* os tempos onde as características serão medidas são fixados antecipadamente. Exemplo: em estudos de crescimento, medir a altura uma vez por ano na data do aniversário;
- iii.* em geral, os dados deverão ser balanceados e completos com relação ao tempo, isto é, todas as unidades de investigação deverão ser planejadas e efetivamente observadas nas mesmas ocasiões. Exemplo: pesar todos os animais a cada 15 dias, medir a altura de cada criança nas respectivas datas de aniversário;

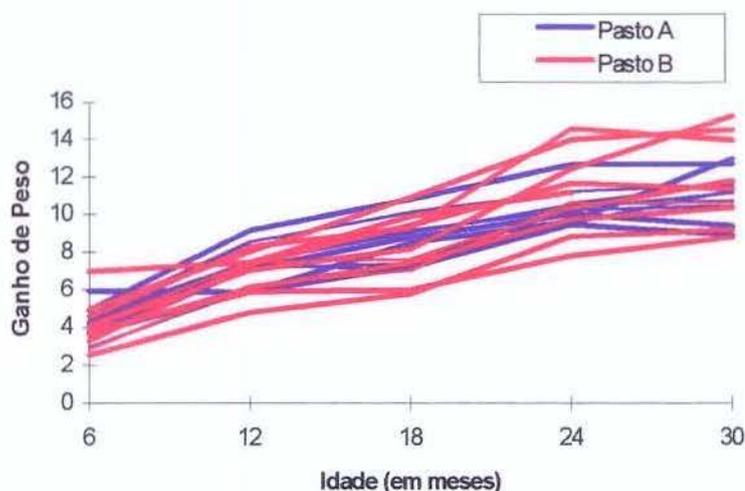
iv. apenas uma variável resposta será considerada em cada vez (respostas unidimensionais); a variável resposta deve ser contínua (por exemplo, pressão arterial, peso, altura, etc.) e assume-se, em geral, que as observações realizadas em uma mesma unidade amostral têm distribuição Normal Multivariada. A seguir apresentamos um exemplo onde os dados são como especificados em (i-iv).

Exemplo 1: Na EMBRAPA (São Carlos -SP) foi realizado um estudo com o objetivo de avaliar dois tipos de pasto como sendo adequados à alimentação de vacas da raça Charolesa. No total, fizeram parte da amostra 20 animais subdivididos em dois grupos de 10, sendo que cada grupo foi alimentado com um tipo de pasto (A ou B). A variável observada foi o ganho de peso com relação ao peso inicial (em kg), sendo que todos os animais foram pesados nas mesmas idades (em meses). Os resultados são apresentados na tabela e no gráfico a seguir.

Tabela 1.1 Ganho de peso relativo de vacas Charolesas (peso/ peso ao nascer) por idade (em meses).

Unid. Exp.	Tratamento (Pasto A=0;B=1)	Peso ao nascido	Idades (em meses)				
			6	12	18	24	30
1	0	44	2.93	6.00	7.07	9.52	11.27
2	0	40	3.32	7.10	7.35	10.32	11.10
3	0	42	5.93	5.79	8.43	9.62	12.95
4	0	39	3.67	7.26	7.00	10.59	11.61
5	0	46	4.33	7.50	9.11	10.20	10.43
6	0	45	4.53	7.40	8.95	10.47	10.73
7	0	35	4.89	9.11	10.77	12.63	12.69
8	0	40	4.60	8.52	10.10	11.20	11.57
9	0	56	3.68	5.84	7.20	9.43	8.89
10	0	48	4.19	7.08	8.75	10.00	9.35
11	1	30	4.50	8.27	9.27	14.60	13.93
12	1	36	2.86	6.17	7.00	10.61	10.61
13	1	53	2.53	4.72	5.75	8.77	9.15
14	1	47	3.64	6.08	7.36	9.70	10.34
15	1	40	4.10	7.42	7.55	10.30	11.80
16	1	50	3.74	5.84	5.94	7.74	8.74
17	1	40	4.20	6.90	7.40	9.90	11.15
18	1	35	4.03	7.00	8.14	12.43	15.26
19	1	40	3.27	7.27	9.40	11.65	11.32
20	1	40	4.92	8.00	10.90	13.97	14.50

Gráfico 1.1 Ganho de peso relativo (peso/ peso ao nascer) por idade



Conforme se pode observar, no exemplo acima os dados são balanceados e completos. O estudo foi planejado para que todos os bois fossem pesados com as mesmas idades e todas as observações foram efetivamente coletadas de acordo com este planejamento. Apesar do modelo geral proposto neste estudo ser adequado a dados com estas características, conforme será visto no **Capítulo IV**, há casos em que serão permitidos dados não-balanceados e/ou incompletos.

Em geral, o objetivo da análise de dados longitudinais é entender o comportamento das variáveis ao longo do tempo bem como comparar a influência de diferentes tipos de tratamentos (por exemplo: tipos de pasto) medida, frequentemente, sobre diferentes grupos (por exemplo: sexo, raça). Considerando estes objetivos e as características (*i-iv*), serão tratados aqui dados do tipo dos exemplos do ganho de peso de bois (**Exemplo 1**) e do crescimento humano, citado no primeiro parágrafo deste capítulo.

1.2. Análise de Dados Longitudinais através do Ajuste de Curvas

As técnicas de análise usualmente empregadas procuram descrever o comportamento da variável de interesse em função do tempo, através de curvas, e comparar efeitos de tratamentos através da comparação dos parâmetros das respectivas curvas. Assim, se na

realização de um experimento onde o objetivo é comparar dois tipos de pastos para animais ajusta-se uma reta para cada um dos grupos, a diferença entre os dois tratamentos pode ser avaliada através das diferenças entre os dois coeficientes angulares e lineares. Além disso, o ajuste de curvas possibilita fazer previsões para o comportamento médio da resposta ao longo do tempo, utilizando, em geral, poucos parâmetros. A metodologia de ajuste de curvas é conhecida como Análise de Curvas de Crescimento.

A análise pode ser feita comparando-se curvas individuais ou curvas médias. Normalmente considera-se as observações feitas em indivíduos sob o mesmo tratamento como repetições ajustando-se o modelo para a média. Este é o caso dos modelos propostos por Potthoff e Roy (1964) e Gosh, Grizzle e Sen (1973), que utilizam técnicas de Análise de Variância Multivariada. Ambos procuram descrever o comportamento da variável resposta através de modelos lineares. O primeiro é o principal ponto de referência para o ajuste de curvas de crescimento através de métodos multivariados (ver por exemplo, Singer (1977)) e o último utiliza métodos não-paramétricos.

O ajuste de curvas individuais, apesar de atraente pela possibilidade de acesso à dinâmica individual de crescimento, torna-se bastante mais complicado pela irregularidade das curvas. Neste caso os modelos adequados são freqüentemente não-lineares. No entanto, esta é apenas uma das situações em que o uso de modelos não-lineares em dados longitudinais torna-se apropriado. Mesmo no caso do ajuste de curvas médias, há situações em que o uso de funções não lineares dos parâmetros proporciona um ajuste mais adequado aos dados, tanto do ponto de vista descritivo, quanto no que diz respeito à interpretação dos parâmetros. Os modelos são construídos a partir de algum conhecimento que se tenha do fenômeno, de forma que geralmente não são polinômios e os parâmetros têm significado relacionado com o assunto estudado. Estas situações são encontradas com freqüência em biologia, engenharia, agronomia, economia, etc. Exemplos típicos de aplicação de modelos de crescimento não-lineares ocorrem em estudos do crescimento humano, de animais ou plantas.

Há outros tipos de modelos não-lineares encontrados na literatura, como por exemplo os chamados "Compartmental models" (ver por exemplo, Seber e Wild, 1989, pág.367). No entanto estes e outros modelos que não são em geral empregados em dados longitudinais, não serão de interesse para este trabalho.

De uma maneira geral, consideraremos que uma variável resposta de interesse pode ser explicada por uma função determinística comum a todos os indivíduos, e por uma parte aleatória. Neste estudo, a função determinística será um modelo de crescimento não-linear em relação aos parâmetros, conforme será descrito no **Capítulo II**.

A parte aleatória será alvo de interesse especial devido a sua importância na estimação dos parâmetros. A maioria dos trabalhos existentes sobre o ajuste de modelos não-lineares tem como hipótese básica resíduos independentes, normalmente distribuídos, com variâncias constantes. No entanto, a situação encontrada em dados longitudinais é aquela onde as medidas são feitas em um mesmo indivíduo, razão pela qual espera-se que haja correlação entre elas. Além disso, é comum a situação em que as variâncias não são constantes no tempo.

Uma das razões para que a maioria dos trabalhos ignore esta possível dependência é simples: a estimação dos parâmetros torna-se mais fácil, podendo ser feita pelo Método de Mínimos Quadrados Ordinários, através de subrotinas disponíveis em pacotes estatísticos.

No caso de se utilizar outras estruturas de covariância, os métodos de estimação mais utilizados são os de Mínimos Quadrados Generalizados e o de Máxima Verossimilhança. Devido à necessidade de se estimar a matriz de covariância, estes métodos tornam-se mais complicados computacionalmente que o de Mínimos Quadrados Ordinários.

1.3. Revisão da Literatura

Conforme dito anteriormente, a maioria dos trabalhos encontrados na literatura sobre o ajuste de modelos de crescimento a dados longitudinais supõe observações não-correlacionadas e variâncias constantes com relação ao tempo. Alguns exemplos são Nelder (1961), Day (1966) e Preece e Baines (1978).

Day e Nelder descrevem o ajuste de famílias de curvas exponenciais. Ambos utilizam o Método de Mínimos Quadrados Ordinários. Nelder utiliza ainda um método que se baseia em aproximações gráficas para obter estimativas iniciais para os parâmetros. Entre as hipóteses necessárias à aplicação do método está a suposição de que as observações medidas em diferentes tempos tenham sido feitas em unidades experimentais diferentes. Para

observações tomadas sobre uma mesma unidade experimental, a única observação feita pelo autor é que se a variância permanecer constante, o método fornece bons resultados. Não há comentário sobre a eficiência do método na presença de correlação entre as observações. O método utilizado por Nelder permite que os tempos em que as observações são medidas não sejam igualmente espaçados.

Preece e Baines (1978) desenvolveram um trabalho voltado para o ajuste de curvas que explicassem o crescimento humano, medido através da altura. Embora os autores tenham tomado como base dados provenientes de um estudo longitudinal, supõem independência entre as observações. Segundo Berkey (1982), esta suposição é aceitável devido ao grande espaço de tempo decorrido entre as medidas.

Ainda tratando do ajuste de curvas a dados de crescimento humano, alguns autores assumem resíduos com uma estrutura de processos auto-regressivos de primeira ordem. Destacamos em particular, os trabalhos de Glasbey (1979) e Bock e Thiessen (1980). Ambos têm a preocupação com o ajuste de curvas individuais. No entanto, a metodologia utilizada supõe séries individuais mais longas do que as normalmente encontradas em dados longitudinais. No trabalho de Glasbey, por exemplo, o número de observações por indivíduo é bem grande (da ordem de 40).

Glasbey emprega o modelo logístico. Os parâmetros são estimados por máxima verossimilhança, sendo que os métodos iterativos utilizados são os de Newton e Gauss. No mesmo trabalho, Glasbey apresenta uma generalização para o caso em que as medidas não são equiespaçadas no tempo.

No trabalho de Bock e Thiessen, o modelo ajustado é o triplo-logístico, soma de três funções logísticas. Os autores discutem o procedimento de ajuste sob três hipóteses: resíduos não correlacionados, estrutura geral de covariância e resíduos com estrutura de covariância auto-regressiva de primeira ordem. Ao final, apresentam um método para o caso de dados incompletos.

Utilizando estrutura de modelos de efeitos aleatórios para a matriz de covariância, há artigos recentes como os de Vonesh e Carter (1992) e Vonesh (1993), além de outros autores como Lindstrom e Bates (1990) e Sheiner e Beal (1980), que serão tratados com mais detalhe na Seção 4.3. Há ainda outros artigos na literatura que não serão discutidos por

abordarem o assunto sob o enfoque Bayesiano (ver, por exemplo, Berkey, 1982, ou Racine-Poon, 1985).

Neste trabalho, trataremos do ajuste de modelos não-lineares a dados longitudinais, concentrando em alguns modelos de crescimento encontrados na literatura e levando em conta algumas das estruturas de covariância mais comumente utilizadas nesta área. Ressaltamos, no entanto, que a teoria descrita aplica-se a qualquer modelo não-linear que satisfaça determinadas condições de regularidade (ver **Seção 3.1.2**). Além disso, embora o modelo geral proposto seja adequado a dados balanceados, completos e igualmente espaçados, este modelo poderá ser estendido a outros tipos de dados, dependendo da estrutura de covariância assumida.

No **Capítulo II** há uma breve descrição dos modelos utilizados para a média e para a matriz de covariância.

O método de estimação dos parâmetros utilizado é o de Mínimos Quadrados. Como as equações resultantes são não-lineares, são necessários procedimentos iterativos para solucioná-las. No caso, utiliza-se os algoritmos de Newton e Gauss-Newton. Estes procedimentos, bem como os métodos de estimação e testes de hipóteses estão descritos no **Capítulo III**.

No **Capítulo IV** especificamos os métodos de estimação para os casos de estruturas especiais de covariância.

No **Capítulo V** é feita uma aplicação dos modelos estudados em um exemplo prático.

Capítulo II - Modelos Não-Lineares para o Vetor de Médias com Estruturas Especiais para a Matriz de Covariância

2.1. O Modelo Geral

Considere a situação geral em que o objetivo do estudo é analisar o comportamento de uma variável ao longo do tempo. Considere ainda que se dispõe de uma amostra de n indivíduos submetidos a M tratamentos, onde a variável de interesse é medida nos mesmos p pontos no tempo para todos os indivíduos.

O modelo a ser ajustado terá a seguinte forma:

$$y_i = f(\mathbf{X}, \mathbf{A}_i, \boldsymbol{\alpha}) + \varepsilon_i, \quad i=1, \dots, n, \quad (2.1)$$

onde $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ip})^t$ é o vetor ($p \times 1$) das observações do i -ésimo indivíduo; $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_r)^t$ é o vetor ($r \times 1$) de parâmetros de locação; \mathbf{A}_i é a matriz ($k \times r$) que define o tratamento a que pertence o i -ésimo indivíduo com k sendo o número de parâmetros para cada tratamento e r o número total de parâmetros; $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)^t$ é uma matriz ($p \times w$) de variáveis regressoras sendo w o número de variáveis regressoras; f é uma função não-linear de $\boldsymbol{\alpha}$ com forma conhecida; $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{ip})^t$, $\boldsymbol{\varepsilon}_i$'s vetores aleatórios independentes com média $\mathbf{0}$ e covariância $\boldsymbol{\Sigma}$.

Em geral, trabalharemos somente com uma variável regressora ($w=1$). Assim, em lugar da matriz \mathbf{X} estaremos considerando um vetor $\mathbf{x}_{(p \times 1)}$ de tempos. Como em geral o modelo será adequado a dados balanceados e completos, a matriz $\boldsymbol{\Sigma}$ é definida como sendo a mesma para todos os indivíduos. Nos casos em que for possível a aplicação a dados não-balanceados, este fato será ressaltado e a notação será adequada. Ainda com relação à matriz de covariância $\boldsymbol{\Sigma}$, esta poderá assumir qualquer das estruturas especiais que serão descritas na Seção 2.3, dependendo da sua adequação ao problema.

Este modelo, bem como a teoria de estimação e testes de hipóteses utilizados neste trabalho são adaptações de modelos de regressão não-linear. Neste contexto, da mesma forma que em regressão linear, a literatura (ver por exemplo, Seber e Wild, 1989) trata de modelos de regressão univariados e multivariados.

Neste trabalho, em geral estaremos trabalhando com uma adaptação dos modelos de regressão não-linear multivariados. Isto significa que, em lugar de termos p variáveis resposta para cada indivíduo, estaremos trabalhando com p medidas da mesma variável ao longo do tempo. Além disso, como o modelo é balanceado, os p valores assumidos pelo vetor $\mathbf{x}_{(px1)}$ são os mesmos para todos os indivíduos, o que em geral não ocorre nos modelos de regressão. Em regressão não-linear, estas n repetições dos mesmos p pontos seriam replicações de um mesmo desenho experimental.

Outra diferença entre alguns modelos de regressão não-linear e (2.1) diz respeito à função f . Neste trabalho, f é a mesma para todos os indivíduos, podendo haver diferença apenas nos valores dos parâmetros (por exemplo, para indivíduos submetidos a diferentes tratamentos). Em regressão não-linear, as p variáveis resposta podem ser descritas por diferentes funções f .

2.2 Modelos Não-Lineares para o Vetor de Médias

Considerando-se a modelagem de dados de crescimento, pode-se dizer que existem dois enfoques sob os quais os autores buscam o modelo mais adequado aos dados: a modelagem empírica e a modelagem na qual os modelos são deduzidos a partir do fenômeno que gera os dados. Os modelos polinomiais são em geral propostos de forma empírica com o objetivo de descrever as observações, sem compromisso com o fenômeno que gera os dados. Estes modelos podem ser uma boa opção quando as informações forem coletadas durante um período limitado do processo de crescimento. No entanto, não incorporam informações previamente conhecidas, como por exemplo, em estudos de crescimento humano a estabilização da taxa de crescimento a partir de uma certa idade.

Em contrapartida, os modelos não-lineares costumam ser propostos sob o enfoque que incorpora informações sobre o processo de crescimento. Além de possibilitarem uma

melhor caracterização da variável resposta, modelos não-lineares têm como vantagem um menor número de parâmetros a estimar, se comparados com os modelos polinomiais. O leitor interessado em maiores detalhes sobre a origem dos modelos não-lineares de crescimento pode consultar Sandland (1983) ou Ratkowsky (1983).

A seguir apresentaremos alguns modelos não-lineares de crescimento que são utilizados em dados longitudinais.

Suponha uma situação em que a taxa de mudança no tamanho f de um organismo em função do tempo seja proporcional ao seu valor. Neste caso temos a seguinte equação diferencial:

$$\frac{df}{dx} = \kappa f, \quad (2.2)$$

que resulta no modelo

$$f(x) = \exp[\kappa(x-\lambda)], \quad (2.3)$$

onde x = tempo, e κ é a constante de proporcionalidade que estabelece a relação entre a taxa de crescimento e f . Observe que para $\kappa > 0$ o crescimento é ilimitado, o que não é a situação mais usual em estudos de crescimento. Em geral, espera-se que a taxa de crescimento estabilize com o tempo. Uma forma de satisfazer esta condição seria incorporar na equação a existência de um tamanho final β e supor a taxa de crescimento proporcional à diferença entre β e o tamanho atual:

$$\frac{df}{dx} = k(\beta - f), \quad (2.4)$$

para algum $\kappa > 0$. Neste caso, (2.4) resulta no modelo

$$f(x) = \alpha - (\alpha - \beta)\exp(-\kappa x), \quad (2.5)$$

onde $\alpha > \beta > 0$, para que a curva descreva o aumento no crescimento.

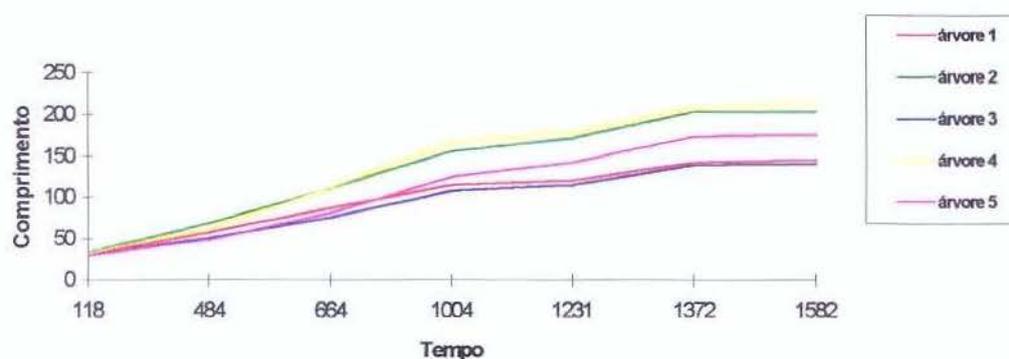
Em alguns tipos de crescimento, a taxa apresenta um comportamento diferente do que foi descrito em (2.4). Em lugar de decrescer sempre, tendendo a zero com o tempo, ela cresce até um certo ponto de máximo antes de iniciar a queda até zero. Este tipo de curva tem formato sigmoidal (formato de S).

Exemplo 2 (Draper e Smith, (1981)): Um estudo em Riverside, Califórnia, durante o período de 1969-1973 teve como objetivo estudar o crescimento de laranjeiras. A variável resposta (y) observada foi a medida da circunferência do tronco (em milímetros). A variável preditora (x) era o tempo (em dias) com origem arbitrariamente fixada em 31/12/68. Foram medidas 5 árvores em 7 ocasiões. Os dados são apresentados na Tabela 2.1 e representados no Gráfico 2.1:

Tabela 2.1. Comprimento da circunferência das laranjeiras (mm) por tempo (dias)

unid.	tempo (em dias)						
	118	484	664	1004	1231	1372	1582
1	30	58	87	115	120	142	145
2	33	69	111	156	172	203	203
3	30	51	75	108	115	139	140
4	32	62	112	167	179	209	214
5	30	49	81	125	142	174	177

Gráfico 2.1. Comprimento da circunferência das laranjeiras (mm) por tempo (dias)



Suponha que a partir de conhecimentos anteriores sobre a forma de crescimento das laranjeiras, tenha-se decidido por uma curva sigmóide, a Logística. Voltando ao modelo geral (2.1), tendo como objetivo descrever o comportamento de y , esta seria a forma funcional assumida para f .

A curva logística tem por base situações em que a taxa de crescimento em um particular tempo x é proporcional ao produto do tamanho atual pela futura quantidade de crescimento. Assim, temos a equação diferencial

$$\frac{df}{dx} = \frac{\kappa f (\beta - f)}{\beta} \quad (2.6)$$

onde, no exemplo, κ é a constante de proporcionalidade, f é o comprimento esperado da circunferência no tempo x e β é o comprimento final. Integrando (2.6), chega-se à curva logística

$$f(x) = \frac{\beta}{1 + \exp[-\kappa (x - \lambda)]}, \quad (2.7)$$

que neste exemplo será utilizada para descrever a parte determinística do modelo (2.1).

Vários modelos matemáticos foram propostos na tentativa de descrever curvas sigmóides. Além do Logístico, temos: Gompertz, Richards, Von Bertalanffy, entre outros. Alguns destes modelos são apresentados a seguir.

Logístico: A curva logística é simétrica em relação ao ponto de inflexão, que ocorre no ponto $x=\lambda$, quando $f= \beta/2$. Neste ponto a taxa de crescimento atinge o seu valor máximo, que é igual a $\kappa \cdot \beta/4$. Nos Gráficos 2.2 e 2.3 há alguns exemplos de como a curva logística varia em função dos parâmetros κ e λ (β é a assíntota e está fixo em ambos os gráficos).

Gráfico 2.2: Modelo logístico (2.7) com $\beta=1$, $\lambda=1$ e alguns valores para κ .

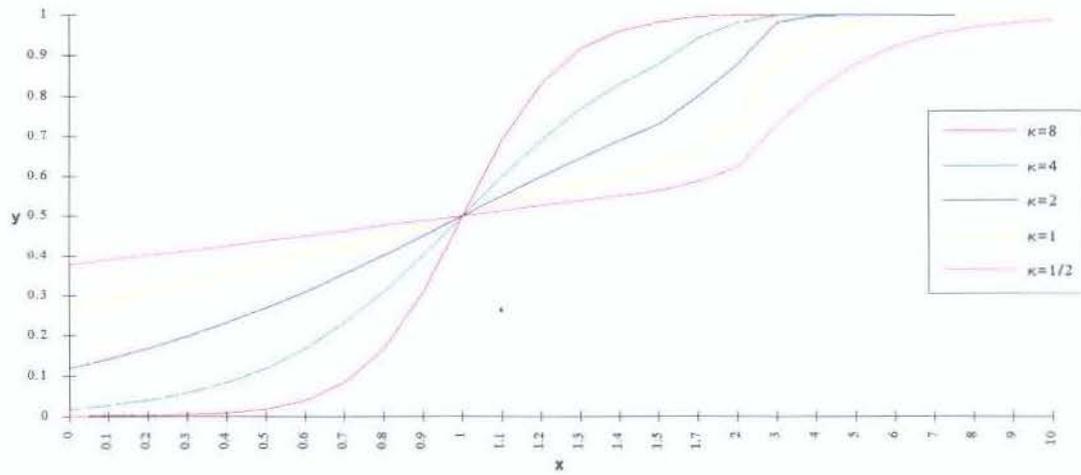
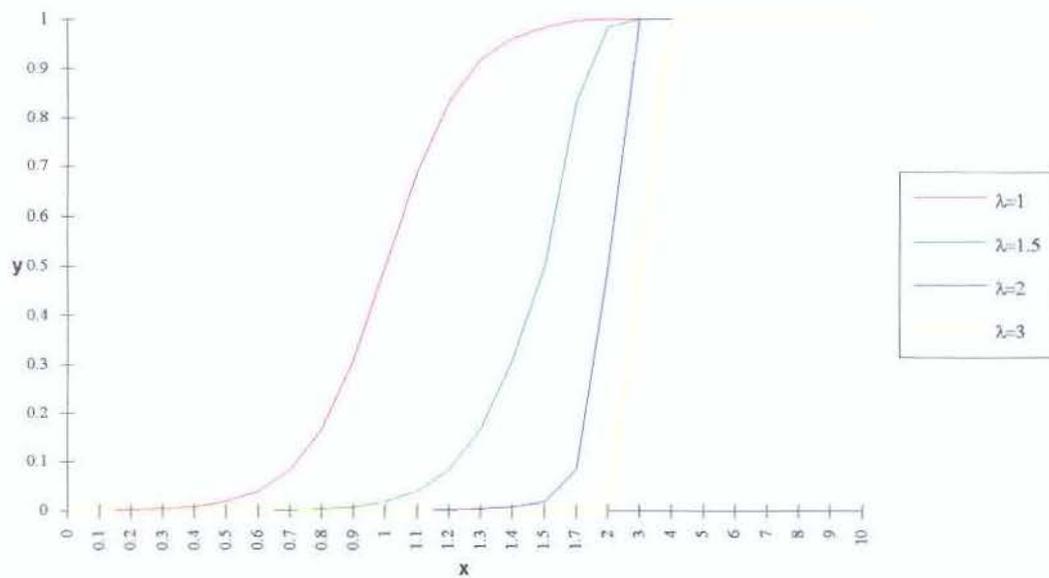


Gráfico 2.3: Modelo logístico (2.7) com $\beta=1$, $\kappa=8$ e alguns valores para λ .



O modelo logístico pode ser reparametrizado em outras formas que se ajustem melhor aos dados:

$$f(x) = \frac{\beta}{1 + \exp(\lambda - \kappa x)} \quad (2.8)$$

$$f(x) = \frac{1}{\beta + \lambda \exp(-\kappa x)} \quad (2.9)$$

$$f(x) = \frac{1}{\beta + \lambda \kappa^x} \quad (2.10)$$

$$f(x) = \frac{\beta}{1 + \exp(\lambda)\kappa^x} \quad (2.11)$$

$$f(x) = \frac{1}{\beta + \exp(\lambda)\kappa^x} \quad (2.12)$$

$$f(x) = \frac{\beta}{1 + \lambda \exp(-\kappa x)} \quad (2.13)$$

ou ainda, tomando $z = \ln(f)$,

$$z = \ln(\beta) - \ln(1 + \exp(\lambda - \kappa x)). \quad (2.14)$$

Gompertz: Utilizado em estudos onde o crescimento não é simétrico em relação ao ponto de inflexão. A taxa de crescimento relativa, $\frac{df}{dx} x \frac{1}{f}$, decresce de forma proporcional ao logaritmo do tamanho e é dada pela equação diferencial:

$$\frac{df}{dx} = \kappa f (\ln \beta - \ln f), \quad (2.15)$$

que resulta no modelo

$$f(x) = \beta \exp\{-\exp[-\kappa(x - \lambda)]\}. \quad (2.16)$$

O ponto de inflexão ocorre em $x = \lambda$ com $f = \beta/e$ e sendo a taxa de crescimento máxima igual a $\kappa \cdot \beta/e$. Outra forma em que é empregado:

$$f(x) = \exp(\beta - \lambda \kappa^x). \quad (2.17)$$

Nos Gráficos 2.4 e 2.5 são apresentados exemplos de curvas de Gompertz (2.16), variando os parâmetros κ e λ .

Gráfico 2.4: Curva de Gompertz (2.16) com $\beta=1$, $\lambda=1$ e alguns valores para κ .

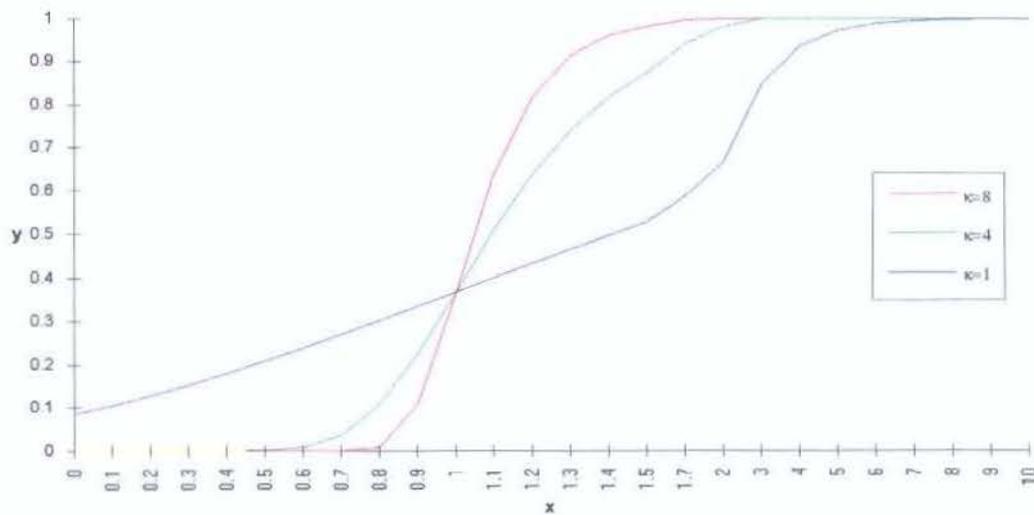
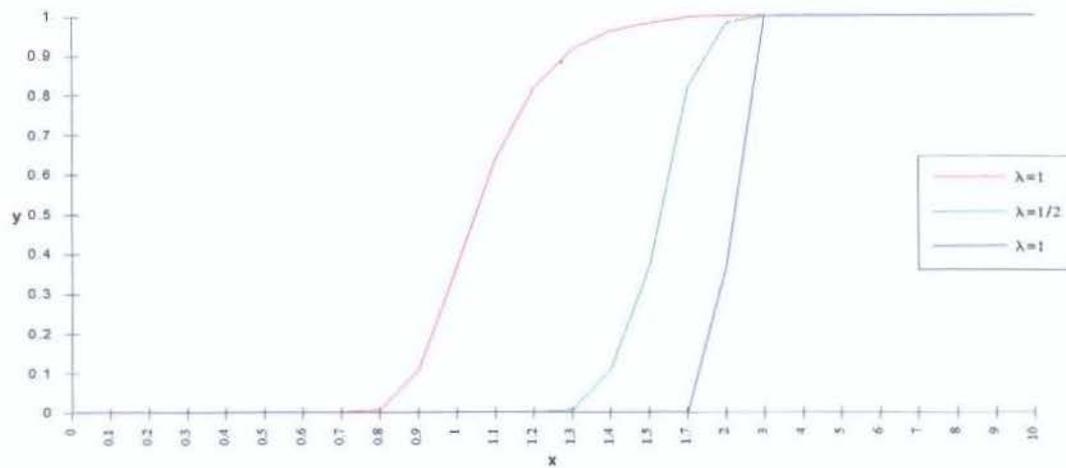


Gráfico 2.5: Curva de Gompertz (2.16) com $\beta=1$, $\kappa=8$ e alguns valores para λ .



Von Bertalanffy: Proposto de forma mecanicista para descrever crescimento animal. Segundo Von Bertalanffy (1957), o crescimento de um animal de peso f envolve a diferença entre o anabolismo e o catabolismo:

$$\frac{df}{dx} = \eta f^\delta - \xi f, \quad (2.18)$$

onde η é o parâmetro relacionado ao anabolismo e ξ ao catabolismo. Esta equação resulta no modelo:

$$f(x) = \left\{ \frac{\eta}{\xi} - \left[\frac{\eta}{\xi} - f_0^{(1-\delta)} \right] e^{-(1-\delta)\xi x} \right\}^{1/(1-\delta)}, \quad (2.19)$$

onde f_0 =peso no tempo $x=0$. Von Bertalanffy descobriu empiricamente que δ , a potência do peso relacionada ao anabolismo, poderia ser fixada em $2/3$ para grande parte das espécies de animais. Assim esta forma alternativa a (2.19) pode também ser encontrada na literatura. Abaixo, nos Gráficos 2.6 e 2.7, exemplos de como o modelo de Von Bertalanffy se comporta em relação aos parâmetros.

Gráfico 2.6: Curva de Von Bertalanffy (2.19) com $\delta=2/3$, $\eta=8$ e alguns valores para ξ .

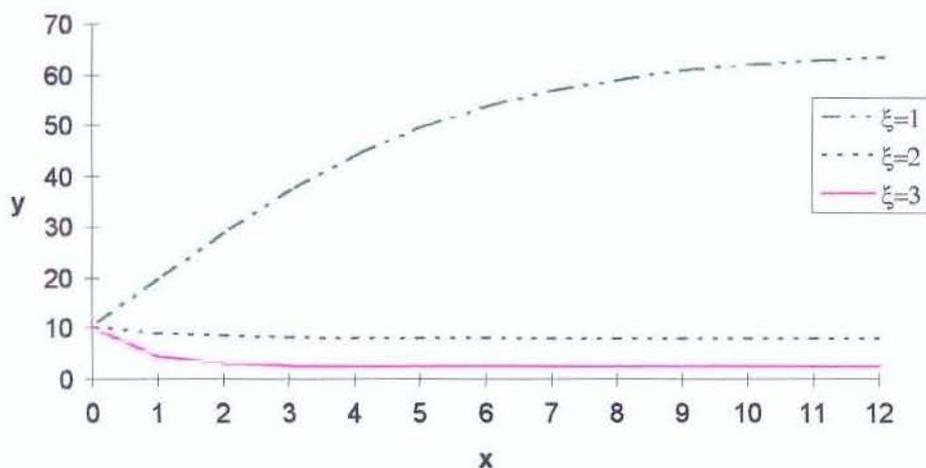
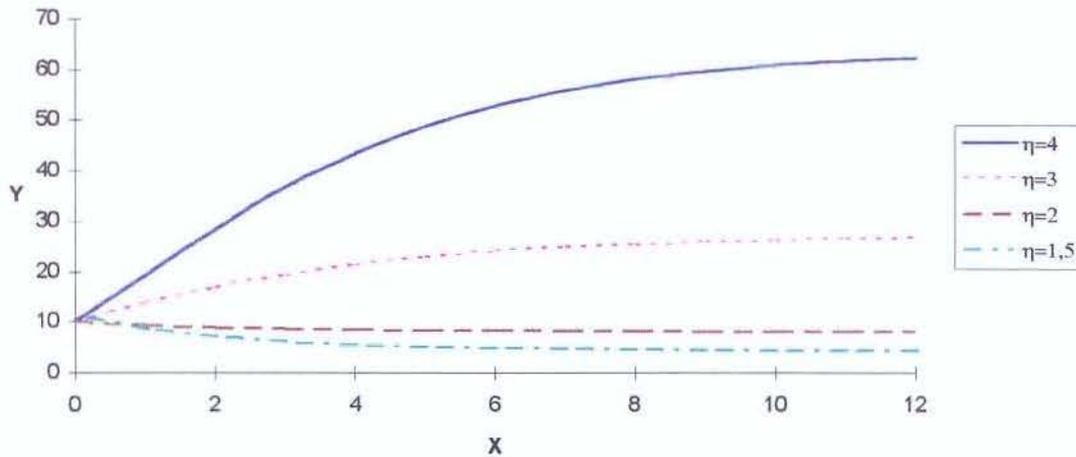


Gráfico 2.7: Curva de Von Bertalanffy (2.19) com $\delta=2/3$, $\xi=1$ e alguns valores para η .



Richards: Richards (1959) propôs um modelo mais geral que, aproveitando a formulação (2.18) de Von Bertalanffy, mantém δ como um parâmetro. Esta opção garante mais flexibilidade ao modelo, uma vez que δ incide sobre o ponto de inflexão, que ocorre quando $f = \beta \delta^{1/(\delta-1)}$.

$$\frac{df}{dx} = \frac{\kappa}{1-\delta} f \left[\left(\frac{f}{\beta} \right)^{\delta-1} - 1 \right], \quad (2.20)$$

onde $\delta \neq 1$ e $\kappa = \eta(1-\delta)\beta^{\delta-1}$. Através de algumas reparametrizações, este modelo inclui os modelos logístico ($\delta=2$), Von Bertalanffy ($\delta=2/3$) e Gompertz ($\delta \rightarrow 1$). Esta equação resulta no modelo (Seber e Wild, 1989, pág.333):

$$f(x) = \beta \left[1 + (\delta - 1) e^{-\kappa(x-\lambda)} \right]^{1/(1-\delta)}, \quad (2.21)$$

onde $\delta \neq 1$. De forma bastante semelhante, Nelder (1961) descreve o ajuste da família de curvas definidas pela equação diferencial

$$\frac{df}{dx} = \kappa f \left[1 - \left(\frac{f}{\beta} \right)^{1/\delta} \right]. \quad (2.22)$$

Integrando e restringindo $\delta > 0$ encontra-se a forma geral:

$$f(x) = \frac{\beta}{\{1 + \exp[-(\lambda + \kappa x) / \delta]\}^\delta}, \quad (2.23)$$

que é chamada de equação logística generalizada.

A partir da equação (2.23) é possível obter por exemplo a curva logística ($\delta = 1$); se $\delta \rightarrow \infty$, com β fixado e κ função linear de δ , chega-se à curva de Gompertz.

Os modelos descritos anteriormente são alguns exemplos de funções que podem ser utilizadas em (2.1) para descrever a parte determinística do modelo. Diversos outros modelos podem ser encontrados em estudos de crescimento. Por exemplo, em Bock et al. (1973) e em Bock e Thiessen (1980), os modelos duplo e triplo logístico resultam, respectivamente, da soma de duas e três funções logísticas. Estes modelos foram construídos com o objetivo de descrever o crescimento humano desde o nascimento até a fase adulta. Nesta linha há ainda os modelos de Preece e Baines (1978), que também envolvem exponenciais e foram desenvolvidos para ajustar os mesmos tipos de dados. Estes modelos de crescimento humano, bem como outros modelos encontrados na literatura descrevendo fenômenos específicos não serão definidos nesta seção por fugirem ao escopo deste trabalho.

2.3. Estruturas para a Matriz de Covariância

Várias estruturas podem ser adotadas para a matriz de covariância, dependendo do conhecimento que se tenha do fenômeno que gera os dados. A preferência por estruturas especiais de covariância em substituição à forma geral (ou não estruturada) é aconselhável por diversas razões. Uma escolha adequada pode tornar o modelo mais realista em relação aos dados, melhorar a eficiência das estimativas, além de trazer ganhos significativos no aspecto computacional. Em alguns casos, o número de parâmetros da matriz de covariância

a serem estimados diminui de $p(p + 1)/2$, onde p =número de ocasiões, para 2 parâmetros. É bom lembrar que esta redução torna-se especialmente importante quando p é grande.

Os tipos de estruturas de matrizes de covariância mais encontradas em dados longitudinais, segundo Andrade(1989), são:

2.3.1. Independência

$$\Sigma = \sigma^2 \mathbf{I}, \quad (2.24)$$

onde σ^2 = variância das observações e \mathbf{I}_p = matriz identidade de dimensão p .

Esta estrutura é adequada às situações em que é possível supor que as observações feitas em uma mesma unidade experimental são não-correlacionadas e possuem variância constante.

Cabe ressaltar que embora a condição de que as observações sejam não-correlacionadas não implique em independência, esta estrutura é assim chamada porque em geral, no contexto de dados longitudinais, assume-se que as observações seguem a distribuição Normal e, conseqüentemente, a ausência de correlação implica em independência entre as observações. Esta estrutura possui apenas um parâmetro a estimar.

2.3.2. Uniforme

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ & 1 & \cdots & \rho \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}, \quad (2.25)$$

onde σ^2 é a variância das observações medidas na mesma unidade experimental e ρ é o coeficiente de correlação entre duas observações medidas na mesma unidade experimental em diferentes tempos.

Esta estrutura assume que as variâncias são constantes no tempo e as correlações entre as observações feitas sobre a mesma unidade experimental são iguais. Possui dois parâmetros a estimar.

2.3.3. Efeitos aleatórios

$$\Sigma = \sigma_d^2 \mathbf{R} + \mathbf{ZDZ}^t, \quad (2.26)$$

onde \mathbf{Z} ($p \times v$) é uma matriz de especificação conhecida de posto $v \leq p$ e \mathbf{D} ($v \times v$) é a matriz de covariância dos efeitos aleatórios desconhecida, σ_d^2 é a variância do erros aleatórios e \mathbf{R} é uma matriz conhecida definida positiva. Frequentemente, $\mathbf{R} = \mathbf{I}_p$. Neste caso, o número de parâmetros a estimar é igual ao número de parâmetros distintos de \mathbf{D} mais um, um número em geral bem menor que $p(p+1)/2$.

Esta estrutura é característica dos modelos de efeitos aleatórios propostos por Laird e Ware (1982) e Ware (1985). Observe que esta representação permite a construção de outras estruturas a partir da especificação de \mathbf{Z} e \mathbf{D} . Por exemplo, fazendo $\mathbf{Z} = \mathbf{1}_p$ e $\mathbf{D} = \sigma_e^2$, temos a estrutura uniforme (2.25) com $\sigma^2 = \sigma_d^2 + \sigma_e^2$ e $\rho = \sigma_e^2 / (\sigma_d^2 + \sigma_e^2)$.

Embora tenha sido definida sem índices pelo fato de estarmos tratando de dados balanceados, a estrutura de covariância do modelo de efeitos aleatórios em geral é definida de forma mais ampla, sendo largamente utilizada em dados não balanceados e/ou incompletos. Neste caso:

$$\Sigma_i = \sigma^2 \mathbf{R}_i + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t, \quad (2.27)$$

2.3.4. Auto-regressiva de 1a. ordem

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{p-1} \\ & 1 & \rho & \dots & \rho^{p-2} \\ & & \ddots & \ddots & \vdots \\ & & & & 1 \end{bmatrix}, \quad (2.28)$$

onde σ^2 e ρ são como especificados em (2.25). Esta estrutura assume variâncias constantes enquanto que as correlações decrescem com o tempo. Possui dois parâmetros a estimar.

2.3.5. Markov

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho^{|t_1 - t_2|} & \rho^{|t_1 - t_3|} & \dots & \rho^{|t_1 - t_p|} \\ & 1 & \rho^{|t_2 - t_3|} & \dots & \rho^{|t_2 - t_p|} \\ & & 1 & \dots & \vdots \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix}, \quad (2.29)$$

onde σ^2 e ρ são também como especificados em (2.25). Esta estrutura é uma generalização do modelo anterior para o caso em que as observações não são igualmente espaçadas. Possui dois parâmetros a estimar.

2.3.6. Auto-regressivo geral

$$\Sigma = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \dots & \theta_p \\ & \theta_1 & \theta_2 & \dots & \theta_{p-1} \\ & & \ddots & \ddots & \vdots \\ & & & & \theta_1 \end{bmatrix}. \quad (2.30)$$

Neste caso, da mesma forma que no anterior, as variâncias são constantes ao longo do tempo enquanto as correlações variam de outra forma. Possui p parâmetros a estimar.

Embora as estruturas (2.28), (2.29) e (2.30) sejam utilizadas em dados longitudinais com bastante frequência, estas não serão consideradas neste estudo porque até o momento a metodologia desenvolvida para aplicação em modelos não-lineares abrange apenas situações em que se ajustam os modelos para curvas individuais bastante longas, com número de tempos da ordem de 30, por exemplo. No caso destas estruturas, as propriedades assintóticas dos estimadores estão ligadas ao tamanho da série (ver, por exemplo, Seber e Wild, 1989, página 18).

2.3.7. Forma geral ou não-estruturada

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1p} \\ & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2p} \\ & & & \ddots & \vdots \\ & & & & \sigma_p^2 \end{bmatrix}, \quad (2.31)$$

onde σ_{jl} é a covariância entre observações medidas nos j -ésimo e l -ésimo tempos e σ_j^2 é a variância das observações medidas no j -ésimo tempo, $j=1, \dots, p$. Possui $p(p+1)/2$ parâmetros a estimar.

Capítulo III. Inferência em Modelos Não-Lineares com Estrutura de Covariância Geral (ou Não-Estruturada)

No modelo (2.1), o valor esperado do vetor de respostas y é explicado por uma função que envolve o vetor de parâmetros desconhecidos α .

No contexto da inferência estatística, existem diferentes métodos para estimar este vetor de parâmetros. Neste trabalho será considerado um dos mais comumente utilizados em modelos de regressão não-linear: o método de Mínimos Quadrados. Para algumas estruturas de covariância será utilizado também o algoritmo EM como forma de obter uma aproximação para o estimador de Máxima Verossimilhança.

Por outro lado, assumimos que a parte aleatória, isto é, o vetor ε_i do modelo (2.1), possui distribuição normal com matriz de covariância Σ . A estimação dos elementos desta matriz é de fundamental importância para o ajuste de curvas. Ela contém toda a informação sobre as correlações existentes entre as observações feitas sobre uma mesma unidade experimental. É bom lembrar que as correlações, juntamente com as variâncias das observações em cada tempo, influenciam na variabilidade dos estimadores do vetor de médias. Por isso, utilizaremos métodos de estimação que levam em conta esta estrutura e onde serão necessárias estimativas dos parâmetros da referida matriz.

Neste trabalho, em geral será utilizado o teste de Wald para realização de testes de hipóteses. Em regressão não-linear, costuma-se utilizar além do teste de Wald, o da Razão de Verossimilhança e o teste de Rao (ou dos Multiplicadores de Lagrange). Segundo Seber e Wild (1989, pág. 228), os três são assintoticamente equivalentes, embora cada um apresente vantagens e desvantagens. No caso do teste de Wald, seu uso é bastante freqüente devido à facilidade de cálculo do valor da estatística-teste. Os outros dois testes exigem a estimação do vetor de parâmetros sob as restrições estabelecidas pela hipótese testada, o que os torna mais trabalhosos. Por outro lado, o teste de Wald não é invariante com relação a reparametrizações do modelo ou da hipótese testada, caso esta seja também não-

linear. Em outras palavras, diferentes formas de expressar o mesmo modelo podem levar a diferentes valores para o teste. No entanto, decidiu-se por utilizar este teste considerando-se que a discussão sobre qual dos três é o mais adequado é bastante ampla, e tem como base características específicas dos modelos e hipóteses testadas. Portanto, esta discussão foge ao escopo do trabalho, sendo que o leitor interessado pode consultar Seber e Wild (1989, pág.228) para uma revisão sobre o assunto.

Nas Seções 3.1 e 3.2, descreveremos os procedimentos de estimação e testes de hipóteses para o caso em que temos somente um grupo de indivíduos, isto é, $M=1$ e $A_i=I_r$ em (2.1). A generalização para $M>1$ será feita na Seção 3.3.

3.1. Método de Mínimos Quadrados Generalizados

O estimador de mínimos quadrados generalizados (EMQG) para α no modelo (2.1) é o vetor $\hat{\alpha}$ que minimiza

$$S(\alpha, \Sigma) = \sum_{i=1}^n [y_i - f(x, \alpha)]^t \Sigma^{-1} [y_i - f(x, \alpha)], \quad (3.1)$$

onde $y_i = (y_{i1}, \dots, y_{ip})^t$, $f(x, \alpha) = [f(x_1, \alpha), \dots, f(x_p, \alpha)]^t$, $\alpha = (\alpha_1, \dots, \alpha_r)^t$ e Σ^{-1} é a inversa da matriz $(p \times p)$ de covariância das observações.

O uso deste estimador supõe que a matriz de covariância seja conhecida, o que dificilmente ocorre, ou que dispomos de uma estimativa obtida de forma independente da amostra em que estamos trabalhando. Na prática, em geral, substitui-se o valor verdadeiro de Σ por uma estimativa calculada a partir dos dados.

Considerando a matriz de covariância geral ou não-estruturada (2.31), um estimador consistente para Σ é dado por (ver por exemplo, Malinvaud, 1970):

$$\hat{\Sigma} = (1/n) \sum_{i=1}^n [y_i - f(x, \tilde{\alpha})] [y_i - f(x, \tilde{\alpha})]^t, \quad (3.2)$$

onde $\tilde{\alpha}$ é o estimador de Mínimos Quadrados Ordinários (EMQO), que por sua vez é obtido a partir da minimização de

$$\sum_{i=1}^n [y_i - f(x, \alpha)]^t [y_i - f(x, \alpha)], \quad (3.3)$$

Conforme Malinvaud (1970), após obtermos uma primeira estimativa para Σ e em seguida, para α , podemos substituir o valor de $\hat{\alpha}$ em (3.2) e reiniciar o processo. Embora não haja resultados analíticos que comprovem, este procedimento pode fornecer estimativas mais precisas em amostras pequenas.

Na literatura encontramos em Gallant (1987) e Seber e Wild (1989), por exemplo, sugestões de estimadores para Σ quando esta tem as estruturas auto-regressivas (2.28) e (2.30). Vonesh e Carter (1992) e Vonesh (1992 e 1993) sugerem procedimento para o caso da matriz com estrutura de modelos de efeitos aleatórios (2.27), que será visto com maior detalhe em 4.3.

Sendo Σ conhecida, é possível calcular L , triangular superior, tal que através da decomposição de Cholesky, $\Sigma^{-1} = L^t L$. A partir desta matriz transformamos os dados, passando a tratar o problema como sendo de mínimos quadrados ordinários. Isto é,

$$\begin{aligned} S(\alpha, \Sigma) &= \sum_{i=1}^n [y_i - f(x, \alpha)]^t \Sigma^{-1} [y_i - f(x, \alpha)] = \\ &= \sum_{i=1}^n [y_i - f(x, \alpha)]^t L^t L [y_i - f(x, \alpha)] = \\ &= \sum_{i=1}^n [Ly_i - Lf(x, \alpha)]^t [Ly_i - Lf(x, \alpha)] = \\ &= \sum_{i=1}^n \sum_{j=1}^p [l_{ij} y_i - l_{ij} f(x, \alpha)]^2, \end{aligned} \quad (3.4)$$

onde l_{ij} = j-ésima linha de L .

Podemos então trabalhar com os dados transformados: $y_s^* = \mathbf{1}_{(0)}^t y_i$ e $f_s^*(\mathbf{x}, \alpha) = \mathbf{1}_{(0)}^t f(\mathbf{x}, \alpha)$, para $j=1, \dots, p$; $i=1, \dots, n$, e onde $s=p(i-1)+j$. Assim,

$$S(\alpha, \Sigma) = S(\alpha) = \sum_{s=1}^{np} [y_s^* - f_s^*(\mathbf{x}, \alpha)]^2 = [\mathbf{y}^* - \mathbf{f}^*(\mathbf{x}, \alpha)]^t [\mathbf{y}^* - \mathbf{f}^*(\mathbf{x}, \alpha)], \quad (3.5)$$

onde $\mathbf{y}^* = (y_1^*, y_2^*, \dots, y_{np}^*)^t$ é o vetor $(np \times 1)$ de respostas transformado, $\mathbf{f}^*(\mathbf{x}, \alpha) = [f_1^*(\mathbf{x}, \alpha), \dots, f_{np}^*(\mathbf{x}, \alpha)]^t$ é o vetor $(np \times 1)$ de valores calculados da função f transformada.

No **Exemplo 1**, suponha um único tratamento (tipo de pasto) para todos os animais, e que se tenha decidido utilizar a curva de Gompertz (2.16) para modelar o ganho de peso ao longo das idades. Neste caso, $n=20$ vacas; $p=5$ idades; $\mathbf{x} = (6, 12, 18, 24, 30)^t$; $\mathbf{y}_i = (y_{i1}, \dots, y_{i5})^t$, sendo y_{ij} = ganho de peso da i -ésima vaca na j -ésima idade, $j=1, \dots, 5$; $i=1, \dots, 20$. A estimativa inicial para a matriz de covariância (2.31) é dada por: $\hat{\Sigma} = (1/20) \sum_{i=1}^{20} [\mathbf{y}_i - \mathbf{f}(\mathbf{x}, \tilde{\alpha})] [\mathbf{y}_i - \mathbf{f}(\mathbf{x}, \tilde{\alpha})]^t$, onde $\mathbf{f}(\mathbf{x}, \tilde{\alpha})$ é um vetor de dimensão $p=5$, cuja j -ésima componente é dada por $f(x_j, \tilde{\alpha}) = \tilde{\alpha}_1 \exp\{-\exp[-\tilde{\alpha}_3(x_j - \tilde{\alpha}_2)]\}$, $j=1, \dots, 5$. Assim, a partir de $\tilde{\alpha} = (13.4206, 7.9875, 0.0838)^t$, obtém-se

$$\hat{\Sigma} = \begin{bmatrix} 0.6380 & 0.3724 & 0.8155 & 0.4392 & 0.7021 \\ & 1.2585 & 1.0319 & 1.4145 & 0.9434 \\ & & 2.3192 & 1.5606 & 1.4555 \\ & & & 2.7815 & 2.2888 \\ & & & & 3.0620 \end{bmatrix}$$

Na transformação dos dados (3.4) utiliza-se a matriz

$$\hat{\Sigma}^{-1} = \begin{bmatrix} 39866 & -10171 & -12663 & 15519 & -11589 \\ & 25000 & -01037 & -16920 & 07771 \\ & & 11884 & -0,5572 & 0,1739 \\ & & & 24254 & -13826 \\ & & & & 13037 \end{bmatrix},$$

decomposta no produto $L^t L$ onde

$$L = \begin{bmatrix} 19966 & -05094 & -06342 & 07772 & -05804 \\ 0 & 14968 & -02851 & -08659 & 03216 \\ 0 & 0 & 08396 & -03706 & -01220 \\ 0 & 0 & 0 & 09665 & -07224 \\ 0 & 0 & 0 & 0 & 05715 \end{bmatrix}.$$

Como exemplo, observe que a primeira componente do vetor y^* em (3.5) é igual a $y^*_1 = t_{(1)}^t y_1 = [1.9966 \ -0.5094 \ -0.6342 \ 0.7772 \ -0.5804] x [2.93 \ 6.00 \ 7.07 \ 9.52 \ 11.27]^t = -0,83$. A segunda componente é dada por: $y^*_2 = t_{(2)}^t y_1 = [0 \ 1.4968 \ -0.2851 \ -0,8659 \ 0,3216] x [2.93 \ 6.00 \ 7.07 \ 9.52 \ 11.27]^t = 2.34$. Da mesma forma obtém-se os valores da função f transformada. Por exemplo: $f^*_1(x, \tilde{\alpha}) = t_{(1)}^t f(x, \tilde{\alpha}) = [1.9966 \ -0.5094 \ -0.6342 \ 0.7772 \ -0.5804] x [f(6, \tilde{\alpha}) \ f(12, \tilde{\alpha}) \ f(18, \tilde{\alpha}) \ f(24, \tilde{\alpha}) \ f(30, \tilde{\alpha})]^t = [1.9966 \ -0.5094 \ -0.6342 \ 0.7772 \ -0.5804] x [4.12 \ 6.57 \ 8.71 \ 10.33 \ 11.46]^t = 0.7324$.

Para encontrar $\hat{\alpha}$ a partir de (3.5), derivamos $S(\alpha)$ e igualamos a zero:

$$\frac{\partial S(\alpha)}{\partial \alpha} = -2[y^* - f^*(x, \alpha)]^t J^*(\alpha) \quad (3.6)$$

onde $J^*(\alpha)_{(n \times r)} = (J^t(\alpha), J^t(\alpha), \dots, J^t(\alpha))^t$ e o jacobiano da função é dado por

$$J(\alpha) = \begin{bmatrix} \frac{\partial f_1^*(x, \alpha)}{\partial \alpha_1} & \dots & \frac{\partial f_1^*(x, \alpha)}{\partial \alpha_r} \\ \vdots & & \vdots \\ \frac{\partial f_p^*(x, \alpha)}{\partial \alpha_1} & \dots & \frac{\partial f_p^*(x, \alpha)}{\partial \alpha_r} \\ \vdots & & \vdots \end{bmatrix}_{(p \times r)} \quad (3.7)$$

Observe que $\frac{\partial f_j^*(X, \alpha)}{\partial \alpha_j} = \mathbf{1}_{(i)}^t (1 \times p) \times \left[\frac{\partial f(x_1, \alpha)}{\partial \alpha_j}, \dots, \frac{\partial f(x_p, \alpha)}{\partial \alpha_j} \right]^t_{(p \times 1)}$, onde $\mathbf{1}_{(i)}^t$ é

definida como em (3.4). Obtemos assim as equações normais

$$J^{*t}(\hat{\alpha}) [y^* - f^*(x, \hat{\alpha})] = 0, \quad (3.8)$$

que são não lineares em relação aos parâmetros, não possuindo, em geral, solução explícita. Para resolvê-las é necessário utilizar procedimentos iterativos.

A transformação dos dados através da decomposição de Cholesky visa aproveitar a metodologia desenvolvida para modelos não-lineares com observações independentes. Utilizando-se esta transformação, o procedimento de estimação torna-se computacionalmente mais simples devido à existência de pacotes que possuem subrotinas para o ajuste de modelos não lineares com esta estrutura de covariância (2.24).

Resumindo, o procedimento utilizado na estimação por mínimos quadrados generalizados segue as seguintes etapas:

- i.* Obtém-se uma estimativa inicial para α . Em geral utiliza-se o estimador de Mínimos Quadrados Ordinários.
- ii.* Estima-se a matriz de covariância Σ (o estimador vai variar de acordo com a estrutura da matriz). Por exemplo, a partir de (3.2) quando Σ é não estruturada.
- iii.* Obtém-se $\hat{\alpha}_{MQG}$, com a estimativa $\hat{\Sigma}$ obtida em *ii*.

Em *i* e *iii*, utilizamos procedimentos iterativos, como os de Newton e Gauss-Newton Modificado, que serão vistos na próxima Seção. No passo *ii*, em geral não é necessário utilizar procedimentos iterativos, em particular nas estruturas tratadas neste trabalho.

3.1.1. Procedimentos Iterativos para Estimação por Mínimos Quadrados

Os procedimentos iterativos utilizados na estimação dos parâmetros são comumente empregados em problemas matemáticos de minimização (ou maximização) de funções. Eles têm por base a expansão da função em série de Taylor (ver, por exemplo, Gallant (1987), pag. 8). Neste trabalho serão utilizados os algoritmos de Gauss-Newton Modificado de Hartley (Hartley, 1961), e Newton (ver por exemplo, Seber e Wild, 1989, pág. 620).

O método de Gauss-Newton Modificado é uma adaptação do método de Gauss-Newton, também chamado de linearização, onde expande-se primeiro a função não linear em série de Taylor até o 2º termo e depois minimiza-se a soma de quadrados.

Considere a soma de quadrados (3.5) e suponha α^0 uma aproximação da estimativa de mínimos quadrados para α . Suponha ainda que a função f seja derivável. Pelo teorema de Taylor temos que, em uma vizinhança de α^0 , a função f pode ser aproximada da seguinte forma (ver por exemplo, Singer e Leite, 1990):

$$f^*(X, \alpha) = f^*(\alpha) \cong f^*(\alpha^0) + \left. \frac{\partial f^*(\alpha)}{\partial \alpha^t} \right|_{\alpha=\alpha^0} (\alpha - \alpha^0) + R_1(\alpha^0),$$

onde $R_1(\alpha^0)$ é o resto de f em torno de α^0 . Então,

$$f^*(\alpha) \cong f^*(\alpha^0) + J^*(\alpha^0) (\alpha - \alpha^0) \quad (3.9)$$

onde $J^*_{(np \times r)}(\alpha^0) = [J^t(\alpha^0), \dots, J^t(\alpha^0)]^t$ e $J^t_{(p \times r)}(\alpha^0)$ é como definido em (3.7). Substituindo este resultado em (3.5), podemos aproximar $S(\alpha)$ por:

$$\begin{aligned} S(\alpha) &= [y^* - f^*(\alpha^0) - J^*(\alpha^0)(\alpha - \alpha^0)]^t [y^* - f^*(\alpha^0) - J^*(\alpha^0)(\alpha - \alpha^0)] = \\ &= [y^* - f^*(\alpha^0)]^t [y^* - f^*(\alpha^0)] - 2[y^* - f^*(\alpha^0)]^t J^*(\alpha^0) (\alpha - \alpha^0) + \\ &\quad (\alpha - \alpha^0)^t J^{*t}(\alpha^0) J^*(\alpha^0) (\alpha - \alpha^0). \end{aligned} \quad (3.10)$$

Derivando $S(\alpha)$ e igualando a zero, chega-se à equação

$$\begin{aligned} -2 \mathbf{J}^{*t}(\alpha^0)[\mathbf{y}^* - \mathbf{f}^*(\alpha^0)] + 2 \mathbf{J}^{*t}(\alpha^0) \mathbf{J}^*(\alpha^0) (\hat{\alpha}^1 - \alpha^0) &= 0 \Rightarrow \\ \Rightarrow \mathbf{J}^{*t}(\alpha^0) \mathbf{J}^*(\alpha^0) (\hat{\alpha}^1 - \alpha^0) &= \mathbf{J}^{*t}(\alpha^0) [\mathbf{y}^* - \mathbf{f}^*(\alpha^0)] \Rightarrow \\ \Rightarrow \hat{\alpha}^1 &= \alpha^0 + [\mathbf{J}^{*t}(\alpha^0) \mathbf{J}^*(\alpha^0)]^{-1} \mathbf{J}^{*t}(\alpha^0) [\mathbf{y}^* - \mathbf{f}^*(\alpha^0)]. \end{aligned} \quad (3.11)$$

Assim, na iteração de ordem $(u+1)$, a aproximação de $\hat{\alpha}$ neste algoritmo é dada por

$$\hat{\alpha}^{(u+1)} = \hat{\alpha}^{(u)} + \delta^{(u)}(\text{GN}), \quad (3.12)$$

onde $\delta^{(u)}(\text{GN}) = [\mathbf{J}^{*t}(\hat{\alpha}^{(u)}) \mathbf{J}^*(\hat{\alpha}^{(u)})]^{-1} \mathbf{J}^{*t}(\hat{\alpha}^{(u)}) [\mathbf{y}^* - \mathbf{f}^*(\hat{\alpha}^{(u)})]$. Durante o procedimento iterativo, o valor aproximado de $\hat{\alpha}$ é atualizado a cada iteração, até que haja convergência. Em Seber e Wild (1989, pág. 619) ou Hartley (1961), o leitor interessado poderá encontrar maiores comentários sobre critérios de convergência e a velocidade obtida pelo método.

O algoritmo de Newton utiliza a expansão da função em série de Taylor até o terceiro termo, necessitando das segundas derivadas. Neste método, a soma de quadrados (3.5) é expandida diretamente.

Assim, de forma análoga a (3.9), aproximamos $S(\alpha)$:

$$\begin{aligned} S(\alpha) &\cong S(\alpha^0) + \left. \frac{\partial S(\alpha)}{\partial \alpha^t} \right|_{\alpha=\alpha^0} (\alpha - \alpha^0) + \frac{1}{2} (\alpha - \alpha^0)^t \left. \frac{\partial^2 S(\alpha)}{\partial \alpha \partial \alpha^t} \right|_{\alpha=\alpha^0} (\alpha - \alpha^0) = \\ &= S(\alpha^0) + \mathbf{g}^t(\alpha^0)(\alpha - \alpha^0) + \frac{1}{2} (\alpha - \alpha^0)^t \mathbf{H}(\alpha^0) (\alpha - \alpha^0), \end{aligned} \quad (3.13)$$

onde $\mathbf{g}(\alpha^0) = \left[\frac{\partial S(\alpha^0)}{\partial \alpha_1}, \dots, \frac{\partial S(\alpha^0)}{\partial \alpha_r} \right]^t$ é o gradiente de $S(\alpha)$ no ponto α^0 , e

$$H(\alpha^0) = \begin{bmatrix} \frac{\partial^2 S(\alpha^0)}{\partial \alpha_1^2} & \frac{\partial^2 S(\alpha^0)}{\partial \alpha_1 \partial \alpha_r} \\ \vdots & \vdots \\ & \frac{\partial^2 S(\alpha^0)}{\partial \alpha_r^2} \end{bmatrix} \text{ é a matriz Hessiana de } S(\alpha) \text{ no ponto } \alpha^0.$$

Derivando $S(\alpha)$ e igualando a zero:

$$\begin{aligned} \frac{\partial S(\alpha)}{\partial \alpha} &= g(\alpha^0) + H(\alpha^0)(\hat{\alpha} - \alpha^0) = 0 \\ H(\alpha^0)(\hat{\alpha} - \alpha^0) &= -g(\alpha^0) \\ \hat{\alpha} - \alpha^0 &= -H^{-1}(\alpha^0)g(\alpha^0) \\ \hat{\alpha} &= \alpha^0 - H^{-1}(\alpha^0)g(\alpha^0). \end{aligned} \quad (3.14)$$

Observe que

$$\begin{aligned} g^t(\alpha^0) &= \left. \frac{\partial S(\alpha)}{\partial \alpha^t} \right|_{\alpha=\alpha^0} (\alpha - \alpha^0) = 2 \sum_{s=1}^{np} [y_s^* - f_s^*(x, \alpha^0)] \cdot \left(\frac{-\partial f_s^*(x, \alpha^0)}{\partial \alpha^t} \right) = \\ &= -2 [y^* - f^*(x, \alpha^0)]^t J^*(\alpha^0) \end{aligned} \quad (3.15)$$

e

$$\begin{aligned} H(\alpha^0) &= \left. \frac{\partial^2 S(\alpha)}{\partial \alpha \partial \alpha^t} \right|_{\alpha=\alpha^0} = 2 \sum_{s=1}^{np} \left(-\frac{\partial f_s^*(x, \alpha)}{\partial \alpha} \right) \left(-\frac{\partial f_s^*(x, \alpha)}{\partial \alpha^t} \right) + \\ & \quad 2 \sum_{s=1}^{np} [y_s^* - f_s^*(x, \alpha^0)] \cdot \left(\frac{-\partial^2 f_s^*(x, \alpha^0)}{\partial \alpha \partial \alpha^t} \right) = \\ &= 2 \cdot [J^{*t}(\alpha^0) J^*(\alpha^0) - Q^*(\alpha^0)], \end{aligned} \quad (3.16)$$

onde $Q^*(\alpha^0) = \sum_{s=1}^{np} [y_s^* - f_s^*(x, \alpha^0)] \cdot \left(\frac{-\partial^2 f_s^*(x, \alpha^0)}{\partial \alpha \partial \alpha^t} \right)$

e

$$\frac{\partial^2 f_s^*(x, \alpha)}{\partial \alpha \partial \alpha^t} = \begin{bmatrix} \frac{\partial^2 f_s^*(x, \alpha)}{\alpha_1^2} & \frac{\partial^2 f_s^*(x, \alpha)}{\partial \alpha_1 \partial \alpha_2} & \dots & \frac{\partial^2 f_s^*(x, \alpha)}{\partial \alpha_1 \partial \alpha_r} \\ & \dots & & \\ & & \dots & \\ & & & \frac{\partial^2 f_s^*(x, \alpha)}{\alpha_r^2} \end{bmatrix}_{(r \times r)}$$

Substituindo (3.15) e (3.16) em (3.13), obtemos

$$S(\alpha) = [y^* - f^*(\alpha^0)]^t [y^* - f^*(\alpha^0)] - 2[y^* - f^*(\alpha^0)]^t J^*(\alpha^0) (\alpha - \alpha^0) + (\alpha - \alpha^0)^t [J^{*t}(\alpha^0) J^*(\alpha^0) - Q^*(\alpha^0)] (\alpha - \alpha^0). \quad (3.17)$$

Comparando (3.17) com (3.10), nota-se que no método de Gauss-Newton a matriz Hessiana é aproximada por $2[J^{*t}(\alpha^0) J^*(\alpha^0)]$. Observe que como

$$\frac{\partial^2 S(\alpha)}{\partial \alpha_i \partial \alpha_j} = 2 \sum_{s=1}^{np} \left\{ \left(-\frac{\partial f_s^*(x, \alpha)}{\partial \alpha_i} \right) \cdot \left(-\frac{\partial f_s^*(x, \alpha)}{\partial \alpha_j} \right) - [y_s^* - f_s^*(x, \alpha^0)] \cdot \left(\frac{-\partial f_s^{*2}(x, \alpha^0)}{\partial \alpha \partial \alpha^t} \right) \right\},$$

$$E \left(\frac{\partial^2 S(\alpha)}{\partial \alpha_i \partial \alpha_j} \right) = 2 J^{*t}(\alpha) J^*(\alpha), \text{ isto é, } H(\alpha) \text{ é aproximada pelo seu valor esperado}$$

no ponto α^0 . A aproximação de $\hat{\alpha}$ na iteração de ordem $(u+1)$ do algoritmo de Newton é dada por:

$$\hat{\alpha}^{(u+1)} = \alpha^{(u)} + \delta^{(u)}(N), \quad (3.18)$$

onde $\delta^{(u)}(N) = -H^{-1}(\alpha^{(u)})g(\alpha^{(u)}) = [J^{*t}(\alpha^{(u)}) J^*(\alpha^{(u)}) - Q^*(\alpha^{(u)})]^{-1} J^{*t}(\alpha^{(u)}) [y^* - f^*(\alpha^{(u)})]$.

O algoritmo de Gauss-Newton é uma adaptação de (3.18), conveniente por ser computacionalmente mais simples, não precisando das derivadas de 2ª ordem da função f . Em alguns casos, seu desempenho é bastante semelhante ao de Newton, notadamente quando $Q^*(\alpha)$ for pequeno em relação a $J^{*t}(\alpha) J^*(\alpha)$. Por (3.16),

observa-se que $Q^*(\alpha)$ pode ser ignorada quando os resíduos forem próximos de zero ou a função f for linear em uma vizinhança de $\alpha^{(u)}$.

Com o objetivo de melhorar o desempenho do algoritmo de Gauss-Newton, Hartley (1961) propôs uma pequena modificação. A cada iteração, o incremento $\delta^{(u)}(\text{GN})$ definido em (3.12) é multiplicado por uma constante $v^{(u)}$, $0 \leq v^{(u)} \leq 1$, sendo que o valor de $v^{(u)}$ é tal que a soma de quadrados $S(\alpha^{(u+1)})$ é mínimo. Desta forma, a soma de quadrados $S(\alpha^{(u+1)})$ resultante na iteração de ordem $(u+1)$ do método de Gauss-Newton Modificado será menor ou igual à resultante do mesmo método sem a modificação proposta por Hartley.

Nesta mesma linha, há outras adaptações do método de Newton e de Gauss-Newton, como o de Marquardt (1963), Quasi-Newton, em Broyden (1967), entre outros. Além de procurar tornar os algoritmos computacionalmente mais simples e melhorar a velocidade de convergência, as diversas propostas de procedimentos iterativos existentes na literatura buscam contornar problemas comuns no método de mínimos quadrados para modelos não-lineares. Um dos problemas que podem alterar o desempenho dos procedimentos é a possível existência de mínimos locais. Este fato torna maior a influência do valor inicial atribuído aos parâmetros na performance do procedimento, podendo mesmo causar a não convergência. Outro problema que costuma ser abordado pelos autores é o fato da matriz hessiana, ou sua aproximação, poder ser definida negativa em algumas iterações. Ela deve ser definida positiva para que o $\hat{\alpha}$ encontrado seja um mínimo local. Para maiores detalhes sobre problemas na utilização de procedimentos iterativos em mínimos quadrados, ver Draper e Smith (1981) ou Seber e Wild (1989), por exemplo.

Há vários pacotes que possuem procedimentos iterativos para mínimos quadrados ordinários em modelos não-lineares, como por exemplo, SAS, BMDP, SPSS, etc. Em geral, há diversas opções de procedimentos iterativos, sendo necessário fornecer a função, dependendo do procedimento escolhido, as derivadas e um valor inicial para os parâmetros. Neste trabalho utilizaremos o SAS.

Para ilustrar a utilização de procedimentos iterativos em mínimos quadrados, voltemos ao **Exemplo 1**, onde se procurava ajustar a curva de Gompertz (2.16) aos dados. Para simplificar, vamos supor um único tratamento para todos os animais. Atribuindo como valor inicial para α , $\alpha_0=(10,7,0.1)^t$ e utilizando o algoritmo de Gauss-Newton Modificado de Hartley, obteve-se os seguintes resultados através do procedimento NLIN do SAS:

Tabela 3.1.: Valores estimados para α através de MQG

Iteração	\hat{a}_1	\hat{a}_2	\hat{a}_3	SQR=S(\hat{a})
0	10.0000	7.0000	0.1000	143.4012
1	13.0187	7.3009	0.0796	99.1719
2	13.2215	7.4737	0.0830	97.8019
3	13.2273	7.4690	0.0829	97.8009
4	13.2273	7.4690	0.0829	97.8009

Então, o valor estimado para α foi $\hat{\alpha}=(13.2273, 7.4690, 0.0829)^t$.

3.1.2. Propriedades dos Estimadores de Mínimos Quadrados

Em modelos de regressão linear, é sabido que assumindo normalidade das observações, os estimadores obtidos por mínimos quadrados são não-tendenciosos, normalmente distribuídos e de variância mínima.

Em regressão não-linear, em geral as propriedades dos estimadores são assintóticas. É desejável que os estimadores obtidos sejam consistentes e assintoticamente normais (AN), isto é, $\sqrt{n}(\hat{\alpha} - \alpha^\#) \sim AN, (\mathbf{0}, \Omega^{-1})$, onde $\alpha^\#$ é o valor verdadeiro do parâmetro α e Ω^{-1} é uma matriz definida positiva.

Para tanto, há uma série de condições de regularidade que devem ser satisfeitas. Vários autores procuraram estabelecer estas condições e demonstrar alguns resultados importantes, principalmente para estimadores obtidos pelo método de Mínimos Quadrados.

Assumindo a estrutura de independência (2.24) para a matriz de covariância das observações, aproveitamos a teoria já bastante desenvolvida para o ajuste de modelos de regressão não-lineares. Jenrich (1969) estabeleceu condições de regularidade sob as quais os estimadores de mínimos quadrados são fortemente consistentes e assintoticamente normais. Outros trabalhos importantes foram Wu (1981), Amemiya (1983) e Gallant (1987). Este último estabelece condições de regularidade um pouco diferentes das discutidas pelos demais autores. Em Seber e Wild (1989, pág. 563), há uma boa revisão do que foi publicado na área.

Assumindo estrutura de covariância geral (2.31) para as observações, podemos utilizar resultados da teoria de regressão não-linear multivariada. Malinvaud (1970) e Phillips (1976) estabeleceram, respectivamente, condições para consistência fraca e forte dos estimadores de Mínimos Quadrados Generalizados. Phillips mostra ainda a equivalência entre os estimadores de Mínimos Quadrados Generalizados Iterativos (MQGI) e Quasi-Máxima Verossimilhança.

Para a estrutura de modelos de efeitos aleatórios ((2.26) e (2.27)), Vonesh e Carter (1992) provam consistência forte dos estimadores propostos (ver Seção 4.3) e normalidade assintótica. Para a estrutura uniforme (2.25), as mesmas propriedades podem ser provadas, considerando-se esta estrutura caso particular da de efeitos aleatórios.

Quanto às estruturas auto-regressivas ((2.28), (2.29) e (2.30)), não foram encontrados trabalhos que tratassem das propriedades dos estimadores de forma que pudessem ser utilizados no caso específico de dados longitudinais, com poucas observações por indivíduo. Apenas no caso de curvas individuais com muitas observações (da ordem de 40, por exemplo), é possível aproveitar a teoria de séries temporais (ver, por exemplo, Seber e Wild (1989), pág. 271).

De uma maneira um pouco mais geral, o estimador de Mínimos Quadrados, algumas vezes tratado como estimador de distância mínima, é definido como o vetor $\hat{\alpha}$ que minimiza:

$$\sum_{i=1}^n [y_i - f_i(\alpha)]^t S_n [y_i - f_i(\alpha)], \quad (3.19)$$

onde y_i e f_i são como definidos em (2.1), S_n é uma matriz definida positiva que pode depender da amostra. Na prática, em geral S_n é a inversa da matriz de covariância das observações.

Considere que assumimos para a matriz de covariância das observações a forma não estruturada (2.31). Sejam $\alpha^\#$ e $\Sigma^\#$ os valores verdadeiros de α e Σ , respectivamente. A seguir, a título de ilustração, apresentamos as condições de regularidade suficientes para que $\hat{\alpha}_{MQC}$ seja fortemente consistente (Seber e Wild, 1989, pág. 582), supondo $S_n = S$, S uma matriz definida positiva.

i. ϵ_i 's são iid(0, Σ). (3.20)

ii. Para cada i , os elementos de $f_i(\alpha)$ são funções contínuas de α , $\forall \alpha \in \Theta \subset \mathfrak{R}^r$. (3.21)

iii. Θ é um subconjunto fechado, limitado (compacto) de \mathfrak{R}^r . (3.22)

$$iv. \lim \mathbf{n}^{-1} \sum_{i=1}^n \mathbf{f}_i(\boldsymbol{\alpha}) \mathbf{f}_i(\boldsymbol{\alpha})^t \text{ existe e a convergência é uniforme } \forall \boldsymbol{\alpha}, \boldsymbol{\alpha}_1 \in \Theta; \text{ e} \quad (3.23)$$

$$\lim \mathbf{n}^{-1} \sum_{i=1}^n [\mathbf{f}_i(\boldsymbol{\alpha}) - \mathbf{f}_i(\boldsymbol{\alpha}^\#)] [\mathbf{f}_i(\boldsymbol{\alpha}) - \mathbf{f}_i(\boldsymbol{\alpha}^\#)]^t \text{ é definido positivo } \forall \boldsymbol{\alpha} (\neq \boldsymbol{\alpha}^\#) \in \Theta. \quad (3.24)$$

Neste caso, o estimador de Σ é dado por:

$$\mathbf{n}^{-1} \sum_{i=1}^n [\mathbf{y}_i - \mathbf{f}_i(\hat{\boldsymbol{\alpha}}(\mathbf{S}))] [\mathbf{y}_i - \mathbf{f}_i(\hat{\boldsymbol{\alpha}}(\mathbf{S}))]^t. \quad (3.25)$$

Se $\mathbf{S}_n = \mathbf{S}$, onde \mathbf{S} é uma matriz definida positiva, Phillips mostrou que $\hat{\boldsymbol{\alpha}}_{\text{MOG}}$ e $\hat{\Sigma}_n(\mathbf{S})$ são fortemente consistentes para $\boldsymbol{\alpha}$ e Σ , respectivamente. No caso mais geral em que \mathbf{S}_n depende de n , adicionamos a suposição:

$$\mathbf{S}_n \rightarrow \mathbf{S} \text{ quase certamente, sendo } \mathbf{S} \text{ definida positiva.} \quad (3.26)$$

Considere agora as seguintes suposições adicionais:

$$v. \boldsymbol{\alpha}^\# \text{ é um ponto interior de } \Theta. \quad (3.27)$$

$$vi. \text{ As derivadas de } \mathbf{f}_i(\boldsymbol{\alpha}) \text{ até a } 2^{\text{a}} \text{ ordem existem e são contínuas } \forall \boldsymbol{\alpha} \in \Theta; \quad (3.28)$$

vii. Os produtos de cauda matriciais envolvendo $f_i(\alpha)$, $\frac{\partial f_i(\alpha)}{\partial \alpha}$ e $\frac{\partial f_i(\alpha)}{\partial \alpha \partial \alpha^t}$ existem $\forall i$,
 $i=1, \dots, n$. (3.29)

viii. Sendo $M_n(\alpha, S)$ uma matriz com elemento de ordem (r, s) igual a

$$n^{-1} \sum_{i=1}^n \left[\frac{\partial f_i(\alpha)}{\partial \alpha_r} \right]^t S \left[\frac{\partial f_i(\alpha)}{\partial \alpha_{sr}} \right], \quad (3.30)$$

existe limite d.p. para $M_n(\alpha, S)$ quando $n \rightarrow \infty$.

Phillips compara os métodos de Mínimos Quadrados Generalizados Iterativos e Quasi-Máxima Verossimilhança. Dado que é necessário estimar sucessivamente α e Σ , duas questões são levantadas: se o processo converge e, caso convirja, se o ponto limite independe do valor inicial de Σ . Phillips mostrou que sob *i-viii* e para n grande, as duas respostas são positivas. Além disso, sob as mesmas suposições, $\hat{\alpha}_{MQG}$ e $\hat{\alpha}_{MLE}$ são equivalentes. Quanto à distribuição dos estimadores, Malinvaud (1970) prova a normalidade assintótica a partir de condições de regularidade bastante semelhantes.

Em amostras finitas, as propriedades dos EMQG's para regressão não-linear são desconhecidas. Alguns autores, como por exemplo, Vonesh e Carter (1992), afirmam que a utilização de Mínimos Quadrados Generalizados Iterativos pode melhorar as estimativas. Voltando ao procedimento descrito na **Seção 3.1**, isto significaria repetir iterativamente os passos *ii* e *iii*. No entanto, não há resultados analíticos sobre o assunto, sendo necessários estudos adicionais que comprovem essa afirmação.

Há ainda um aspecto específico de modelos não-lineares a ser considerado. A validade das propriedades assintóticas dos estimadores de mínimos quadrados para α varia bastante com o modelo para f . Em alguns casos, conforme Ratkowsky (1983), as propriedades são mantidas em amostras pequenas. No entanto, para alguns modelos não-lineares elas podem não ser satisfeitas até mesmo em amostras relativamente grandes. Alguns autores propuseram medidas através das quais analisam o quanto a não linearidade da parte determinística do modelo influencia as propriedades dos estimadores de mínimos quadrados. Estas medidas não serão tratadas neste trabalho. O leitor interessado poderá encontrar maiores detalhes sobre o assunto em Beale (1960), Bates e Watts (1980) ou Ratkowsky (1983).

3.2. Testes de Hipóteses e Intervalos de Confiança

3.2.1. Testes de Hipóteses

Conforme dito no início deste capítulo, neste trabalho utilizaremos uma das estatísticas comumente empregadas em testes de hipóteses sobre parâmetros de modelos não-lineares: a estatística de Wald. Além disso, consideraremos que seja de interesse testar somente hipóteses sobre os parâmetros de locação. Testes de hipóteses sobre a matriz de covariância podem ser encontrados em Mardia, Kent e Bibby (1979), ou Morrison (1976).

Considere que seja de interesse testar:

$$H_0: \mathbf{h}(\boldsymbol{\alpha}) = \mathbf{0} \quad H_1: \mathbf{h}(\boldsymbol{\alpha}) \neq \mathbf{0}, \quad (3.31)$$

onde $\mathbf{h}(\boldsymbol{\alpha})$ é uma função com primeira derivada contínua e $\mathbf{h}(\boldsymbol{\alpha}): \mathbb{R}^r \rightarrow \mathbb{R}^q$. Por definição, a estatística-teste de Wald é dada por:

$$W = \mathbf{h}'(\hat{\boldsymbol{\alpha}}) (\hat{\mathbf{H}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{H}}')^{-1} \mathbf{h}(\hat{\boldsymbol{\alpha}}) \quad (3.32)$$

onde $\hat{\boldsymbol{\Omega}}$ é uma estimativa da matriz de covariância dos estimadores $\hat{\boldsymbol{\alpha}}$, dada por

$$\hat{\boldsymbol{\Omega}} = \left[\sum_{i=1}^n \mathbf{J}'(\hat{\boldsymbol{\alpha}}) \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{J}(\hat{\boldsymbol{\alpha}}) \right]^{-1}, \text{ sendo que } \mathbf{J}(\hat{\boldsymbol{\alpha}}) \text{ é o jacobiano da função definido em (3.7)}$$

e $\hat{H} = H(\hat{\alpha}) = \left. \frac{\partial h(\alpha)}{\partial \alpha'} \right|_{\alpha = \hat{\alpha}}$. Assume-se que $H(\hat{\alpha})$ e $\hat{\Omega}$ são matrizes de posto completo, isto é, $\text{posto}(H(\hat{\alpha}))=q$ e $\text{posto}(\hat{\Omega})=r$.

Assintoticamente, sob H_0 , W tem distribuição qui-quadrado com q graus de liberdade.

Gallant (1987) sugere duas correções no teste. A primeira, para melhorar a precisão em amostras de tamanho moderado:

$$W_1 = W/qs^2, \quad (3.33)$$

onde $s^2 = \frac{S(\hat{\alpha}, \hat{\Sigma})}{np - r}$ e $W_1 \sim F_{(q, (np-r))}$. Em seguida, com base em simulações,

Gallant observa que a estimação dos parâmetros de Σ , juntamente com o uso de W_1 , pode aumentar o erro do tipo I. Daí vem a sugestão de se utilizar

$$W_2 = W/q, \quad (3.34)$$

e a comparação deste valor com o tabelado pela distribuição F com $(q, (n-r))$ graus de liberdade. A justificativa para este procedimento é que, dada uma estimativa para os parâmetros de locação, restam $(n-r)$ graus de liberdade para a estimativa de Σ . Como W_2 tem distribuição aproximada $F_{(q, \infty)}$, a utilização da distribuição $F_{(q, n-r)}$ deve fornecer uma boa aproximação em amostras pequenas.

Continuando a análise dos dados do **Exemplo 1**, considere ainda um único tratamento para todos os animais. Suponha que houvesse interesse em testar a hipótese: $H_0: \alpha_1 = 10$; $H_1: \alpha_1 \neq 10$. Neste caso, $h(\alpha) = \alpha_1 - 10$. Utilizando os valores obtidos para $\hat{\alpha}$ ao final da **Seção 3.1.1**, ou seja, $\hat{\alpha}=(13.2273, 7.4690, 0.0829)^t$, obtemos: $h(\hat{\alpha}) = 13,2273 - 10 = 3,2273$. Além disso,

$$\hat{\Omega} = \begin{bmatrix} 0.8449 & 0.7018 & -0.0087 \\ & 0.7495 & -0.0084 \\ (\text{sim.}) & & 0.0001 \end{bmatrix}; \mathbf{H}(\hat{\alpha})=(1 \ 0 \ 0); (\hat{\mathbf{H}}\hat{\Omega}\hat{\mathbf{H}}^t)^{-1}=\hat{\Omega}_{11}^{-1}=1/0.8449=1.1836 \text{ e}$$

$q=1$. Utilizando a estatística de Wald (3.32), obtemos $W=3.2273 \times 1.1836 \times 3.2273 = 12.33$. Como $P(\chi^2_1 > 12.33) = 0.0004$, rejeita-se H_0 , isto é, pode-se dizer que há evidências estatísticas de que α_1 seja diferente de 10.

Utilizando agora a primeira correção (3.33), obtém-se $s^2 = 97.8009/(20 \times 5 - 3) = 1.008$ e a partir deste valor: $W_1 = 12,33/1,008 = 12,23$. Comparando com o valor da distribuição $F(1,97)$ obtém-se $P(F > 12.23) = 0.0007$. Da mesma forma, a partir de (3.34) obtém-se $W_2=12.33/1 = 12.33$ e $P(F_{(1,17)} > 12.33) = 0.0037$.

3.2.2. Intervalos de Confiança

Considere uma função dos parâmetros $\Psi(\alpha)$ que tenha as duas primeiras derivadas contínuas. Para obtermos um intervalo de confiança de nível $100 \times (1-\phi)\%$ para $\Psi(\alpha)$, ou região de confiança no caso de $\Psi(\alpha)$ multidimensional, basta invertermos os testes descritos anteriormente. Considere a hipótese descrita em

(3.31), com $h(\alpha) = \Psi(\alpha) - \Psi_0$. Incluiremos no intervalo todos os valores de Ψ_0 para os quais a hipótese nula é aceita ao nível de significância ϕ .

No caso do teste de Wald (3.32), para $\Psi(\alpha)$ unidimensional, H_0 é aceita quando

$$\frac{|\psi(\hat{\alpha}) - \psi_0|}{(\hat{H} \hat{\Omega} \hat{H}^t)^{1/2}} \leq t_{\phi/2}, \quad (3.35)$$

onde $\hat{H} = \frac{\partial h(\hat{\alpha})}{\partial \alpha^t} = \frac{\partial |\psi(\hat{\alpha}) - \psi_0|}{\partial \alpha^t} = \frac{\partial \psi(\hat{\alpha})}{\partial \alpha^t}$, $t_{\phi/2}$ = ponto crítico de nível $\phi/2$

da distribuição t-Student com $(np-r)$ graus de liberdade.

Dessa forma, o intervalo de confiança de $100 \times (1 - \phi)\%$ para $\Psi(\alpha) = \Psi_0$ é dado por

$$\psi(\hat{\alpha}) - t_{\phi/2} (\hat{H} \hat{\Omega} \hat{H}^t)^{1/2} \leq \psi_0 \leq \psi(\hat{\alpha}) + t_{\phi/2} (\hat{H} \hat{\Omega} \hat{H}^t)^{1/2}. \quad (3.36)$$

As situações em que $\Psi(\alpha)$ é unidimensional são as mais comuns e abrangem a construção de intervalos de confiança para cada componente do vetor de parâmetros α . Neste caso, o intervalo de confiança de $100 \times (1 - \phi)\%$ é dado por

$$\hat{\alpha}_i - t_{\phi/2} (\hat{\Omega}_{ii})^{1/2} \leq \alpha_i \leq \hat{\alpha}_i + t_{\phi/2} (\hat{\Omega}_{ii})^{1/2}, \quad (3.37)$$

onde $\hat{\Omega}_{ii}$ = i-ésimo elemento da diagonal de $\hat{\Omega} = \left[\sum_{j=1}^n J_j^t(\hat{\alpha}) \hat{\Sigma}^{-1} J_j(\hat{\alpha}) \right]^{-1}$.

No **Exemplo 1**, considerando o valor de $\hat{\alpha}=(13.2273, 7.4690, 0.0829)^t$, um intervalo de confiança para α_1 obtido a partir de (3.37) é dado por:

$$\hat{\alpha}_1 \pm t(97)_{0,025} \left(\hat{\Omega}_{11} \right)^{1/2}. \text{ Assim, obtém-se: } 13,2273 + 1,98 \times 0,9192 = 15,0473;$$

$$13,2273 - 1,98 \times 0,9192 = 11,4073 \therefore IC_{95\%}(\alpha_1) = (11,4073, 15,0473).$$

Para $\Psi(\alpha)$ q-dimensional, podemos obter a seguinte região de confiança a partir de (3.32):

$$|\psi(\hat{\alpha}) - \psi_0|^t (\hat{H} \hat{\Omega} \hat{H}^t)^{-1} |\psi(\hat{\alpha}) - \psi_0| \leq \chi_{q, \alpha}^2 \text{ g.l.} \quad (3.38)$$

No caso de se utilizar o teste modificado de Wald (3.34), basta dividir (3.38) por q.

Dessa forma a região de confiança para $\Psi(\alpha)$ é obtida por

$$\frac{|\psi(\hat{\alpha}) - \psi_0|^t (\hat{H} \hat{\Omega} \hat{H}^t)^{-1} |\psi(\hat{\alpha}) - \psi_0|}{q} \leq F_{(q, (n-r))} \text{ g.l.} \quad (3.39)$$

3.3. Generalização para M Tratamentos

Em dados longitudinais é freqüente o interesse em comparar curvas referentes à variável resposta de diferentes tratamentos. Por exemplo, ao estudar a altura dos seres humanos, o pesquisador pode ter interesse em investigar se as taxas de crescimento masculina e feminina são iguais.

Os métodos de estimação e testes de hipóteses descritos nas seções anteriores referem-se ao ajuste de curvas não-lineares a uma variável resposta de um determinado grupo de indivíduos. Nesta Seção será apresentada uma extensão para quando os n indivíduos são agrupados em M tratamentos (veja modelo descrito na **Seção 2.1**).

Neste caso existem dois tipos de informações que precisam ser incorporadas ao modelo. O primeiro tipo é formado de características que variam entre os indivíduos mas permanecem constantes ao longo do tempo (variáveis “entre”). Fazem parte deste tipo variáveis tais como sexo, raça, exposição a um determinado tratamento que não muda com o tempo, ou covariáveis como peso ao nascer, etc.. O outro tipo de variáveis é composto por características que modificam-se ao longo do tempo (variáveis “dentro”). São elas: peso, idade, etc.. Em (2.1), o primeiro tipo é incorporado ao modelo pela matriz A_i e o segundo pelo vetor $\mathbf{x}_{(px1)}$ (que poderia ser uma matriz se tivéssemos mais de uma variável regressora).

Voltemos ao **Exemplo 1**. Nas **Seções 3.1 e 3.2**, obtivemos estimativas para os parâmetros sem levar em conta que os animais foram submetidos a tratamentos diferentes. Consideremos agora que os animais 1 a 10 foram alimentados com pasto

tipo A e os demais, com o tipo B. Dessa forma, poderíamos escrever o valor esperado para j-ésima observação medida na i-ésima unidade experimental da seguinte forma:

$$E(y_{ij}) = f(x_{ij}, \mathbf{A}_i \boldsymbol{\alpha}), \quad (3.40)$$

onde $j=1, \dots, 5$; $i=1, \dots, 20$; $k=3$, $r=6$ e $\boldsymbol{\alpha}=(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \alpha_{31}, \alpha_{32})^t$. Assim,

$$A_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \text{ se } i=1, \dots, 10, \text{ e } A_i = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ se } i=11, \dots, 20.$$

Escrevendo o modelo Gompertz para este caso, teríamos:

$$E(y_{ij}) = \begin{cases} \alpha_{11} \exp\{-\exp[-\alpha_{31} (x_j - \alpha_{21})]\}; & \text{se } j=1, \dots, 5; i=1, \dots, 10. \\ \alpha_{12} \exp\{-\exp[-\alpha_{32} (x_j - \alpha_{22})]\}; & \text{se } j=1, \dots, 5; i=11, \dots, 20. \end{cases} \quad (3.41)$$

Freqüentemente $r=Mk$, isto é, teremos o mesmo número de parâmetros para cada tratamento. No entanto, é possível encontrar modelos onde parte-se do princípio de que o valor verdadeiro de algum parâmetro é o mesmo para mais de um tratamento. No **Exemplo 1**, poderíamos supor que os parâmetros α_{31} e α_{32} são iguais para ambos os adubos. Assim, teríamos o seguinte vetor de parâmetros a estimar: $\boldsymbol{\alpha}_{r \times 1}=[\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \alpha_3]^t$. Neste caso, $r=5$, e temos as seguintes matrizes

de delineamento para os parâmetros $A_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, se $i=1, \dots, 10$ e

$$A_i = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ se } i=11, \dots, 20.$$

Com relação ao procedimento de estimação, com a presença de unidades experimentais submetidas a diferentes tratamentos, a única alteração no que foi apresentado nas seções anteriores deste capítulo é que o valor esperado dos vetores de observações individuais será diferente para unidades experimentais submetidas a diferentes tratamentos, conforme exemplificado em (3.41). Assim, levando em conta a existência de dois tratamentos, repetimos o procedimento de estimação por Mínimos Quadrados Ordinários para o **Exemplo 1**. Os resultados de cada iteração são apresentados a seguir:

Tabela 3.2: Valores obtidos para α através de Mínimos Quadrados Ordinários - Procedimento de Gauss-Newton Modificado

Iteração	$\tilde{\alpha}_{11}$	$\tilde{\alpha}_{12}$	$\tilde{\alpha}_{21}$	$\tilde{\alpha}_{22}$	$\tilde{\alpha}_{31}$	$\tilde{\alpha}_{32}$	SQR=S($\tilde{\alpha}$)
0	10.0000	10.0000	8.0000	8.0000	0.0500	0.0500	1200.4640
1	13.6166	7.8027	9.6647	-4.5788	0.0933	0.1257	433.1860
2	13.8580	8.8456	9.6574	-2.6941	0.0883	0.0747	388.7350
3	14.2703	10.6524	9.6905	6.4866	0.0809	0.0885	285.0710
4	14.7954	12.2298	9.8352	6.4031	0.0738	0.0984	195.5070
5	14.8255	12.2528	9.8741	6.4424	0.0741	0.0972	195.4580
6	14.8274	12.2561	9.8755	6.4449	0.0741	0.0972	195.4580
7	14.8274	12.2562	9.8755	6.4450	0.0741	0.0972	195.4580

Nesta etapa, a estimativa de MQO para α e os erros-padrão correspondentes foram os que se seguem:

Tabela 3.3.: Estimativas de MQO para α com Erros-Padrão

Parâmetro	Estimativa	Erro-padrão
α_{11}	14.8274	2.4247
α_{12}	12.2562	1.2222
α_{21}	9.8755	2.3845
α_{22}	6.4450	1.1365
α_{31}	0.0741	0.0228
α_{32}	0.0972	0.0280

Após obter o vetor $\tilde{\alpha}_{(rx1)}$, estima-se Σ através de

$$\hat{\Sigma} = (1/n) \sum_{i=1}^n [y_i - f(x, A_i \tilde{\alpha})] [y_i - f(x, A_i \tilde{\alpha})]^t \quad (3.42)$$

que é o mesmo estimador apresentado em (3.2), diferindo apenas pela matriz A_i , que define o tratamento a que pertence cada unidade experimental. Vale lembrar que em (3.2), $A_i = I_r$.

Em seguida, o procedimento de transformação dos dados é o mesmo de (3.4). Transformados os dados, passamos a trabalhar com y_{ms}^* , onde $m=1, \dots, M$ e

$$s=1, \dots, n_{mp}, n = \sum_{m=1}^M n_m. \text{ Assim,}$$

$$\begin{aligned}
S(\alpha, \Sigma) &= \sum_{m=1}^M \sum_{s=1}^{n_m p} [y_{ms}^* - f_s^*(\mathbf{x}, \mathbf{A}_s, \alpha)]^2 = \\
&= [\mathbf{y}^* - \mathbf{f}^*(\mathbf{x}, \mathbf{A}, \alpha)]' [\mathbf{y}^* - \mathbf{f}^*(\mathbf{x}, \mathbf{A}, \alpha)]
\end{aligned} \tag{3.43}$$

onde \mathbf{y}^* é o vetor ($np \times 1$) de respostas transformado, $\mathbf{f}^*(\mathbf{x}, \alpha)$ é o vetor ($np \times 1$) de valores calculados da função f transformada. Esclarecendo melhor esta notação, temos que

$$\mathbf{y}^*_{(np \times 1)} = [y^*_{11}, \dots, y^*_{1n_1}, y^*_{21}, \dots, y^*_{2n_2}, \dots, y^*_{M1}, \dots, y^*_{Mn_M}]', \tag{3.44}$$

sendo que $\mathbf{y}^*_{ms} = [y^*_{ms1}, y^*_{ms2}, \dots, y^*_{msp}]'$ é o vetor de p observações para o s -ésimo indivíduo do m -ésimo tratamento. Da mesma forma, podemos escrever

$$\mathbf{f}^*(\mathbf{x}, \mathbf{A}, \alpha)_{(np \times 1)} = [f^*_1(\mathbf{x}, \mathbf{A}_1, \alpha), \dots, f^*_p(\mathbf{x}, \mathbf{A}_1, \alpha), \dots, f^*_1(\mathbf{x}, \mathbf{A}_{np}, \alpha), \dots, f^*_p(\mathbf{x}, \mathbf{A}_{np}, \alpha)]'. \tag{3.45}$$

Para encontrar $\hat{\alpha}$, derivamos $S(\alpha, \Sigma)$ e igualamos a zero:

$$\frac{\partial S(\alpha, \Sigma)}{\partial \alpha^t} = -2[\mathbf{y}^* - \mathbf{f}^*(\mathbf{x}, \mathbf{A}, \alpha)]'_{(1 \times np)} \mathbf{J}^*(\alpha)_{(np \times r)}, \tag{3.46}$$

onde \mathbf{y}^* e $\mathbf{f}^*(\mathbf{x}, \mathbf{A}, \alpha)$ são como descritos em (3.44) e (3.45), respectivamente e

$$\mathbf{J}^*(\alpha) = \begin{bmatrix} \frac{\partial f_1^*(\mathbf{x}, \mathbf{A}_1, \alpha)}{\partial \alpha_1} & \dots & \frac{\partial f_1^*(\mathbf{x}, \mathbf{A}_1, \alpha)}{\partial \alpha_r} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n_1}^*(\mathbf{x}, \mathbf{A}_{n_1}, \alpha)}{\partial \alpha_1} & \dots & \frac{\partial f_{n_1}^*(\mathbf{x}, \mathbf{A}_{n_1}, \alpha)}{\partial \alpha_r} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n_1+1}^*(\mathbf{x}, \mathbf{A}_{n_1+1}, \alpha)}{\partial \alpha_1} & \dots & \frac{\partial f_{n_1+1}^*(\mathbf{x}, \mathbf{A}_{n_1+1}, \alpha)}{\partial \alpha_r} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n_2}^*(\mathbf{x}, \mathbf{A}_{n_2}, \alpha)}{\partial \alpha_1} & \dots & \frac{\partial f_{n_2}^*(\mathbf{x}, \mathbf{A}_{n_2}, \alpha)}{\partial \alpha_r} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n_{(M-1)+1}}^*(\mathbf{x}, \mathbf{A}_{n_{(M-1)+1}}, \alpha)}{\partial \alpha_1} & \dots & \frac{\partial f_{n_{(M-1)+1}}^*(\mathbf{x}, \mathbf{A}_{n_{(M-1)+1}}, \alpha)}{\partial \alpha_r} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n_M}^*(\mathbf{x}, \mathbf{A}_{n_M}, \alpha)}{\partial \alpha_1} & \dots & \frac{\partial f_{n_M}^*(\mathbf{x}, \mathbf{A}_{n_M}, \alpha)}{\partial \alpha_r} \end{bmatrix}_{(np \times r)} \quad (3.47)$$

sendo $\frac{\partial f_s^*(\mathbf{x}, \mathbf{A}_s, \alpha)}{\partial \alpha_r} = \mathbf{L} * \begin{bmatrix} \frac{\partial f(\mathbf{x}_1, \mathbf{A}_s, \alpha)}{\partial \alpha_r} \\ \frac{\partial f(\mathbf{x}_2, \mathbf{A}_s, \alpha)}{\partial \alpha_r} \\ \vdots \\ \frac{\partial f(\mathbf{x}_p, \mathbf{A}_s, \alpha)}{\partial \alpha_r} \end{bmatrix}$, isto é, o vetor (p x 1) de derivadas da

função em relação ao r-ésimo parâmetro para o s-ésimo indivíduo, transformado através da matriz L, sendo L definida como em (3.4).

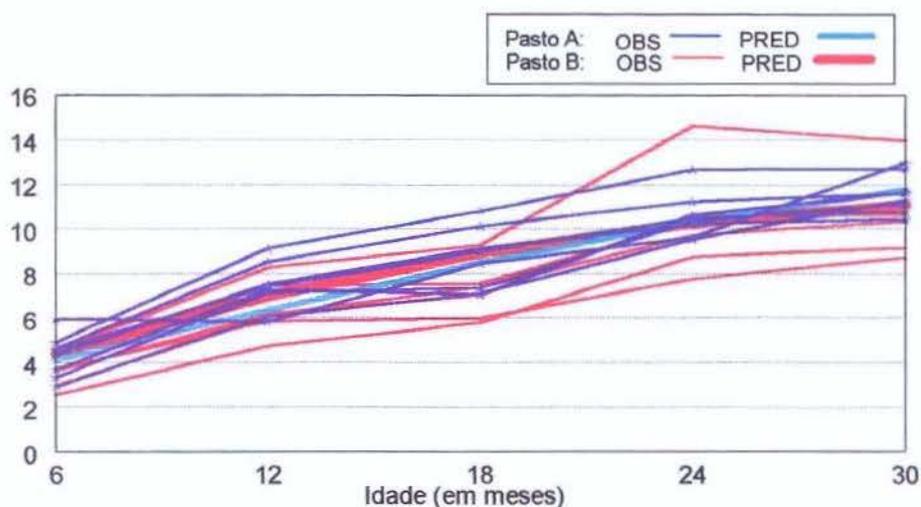
Este vetor de derivadas é igual a $\mathbf{0}_{(px1)}$ quando $\alpha_j \notin A_i \alpha$. Desta forma temos que $\mathbf{J}^*(\alpha)$ tem posto coluna completo, isto é, $\text{posto}(\mathbf{J}^*(\alpha))=r$. Assim, $(\mathbf{J}^{*t}(\alpha)\mathbf{J}^*(\alpha))$ é não singular.

Utilizando o mesmo procedimento das Seções anteriores para estimação dos parâmetros, obtivemos o valor final para $\hat{\alpha}$ no **Exemplo 1** com os erros-padrão correspondentes:

Tabela 3.4.: Estimativas de MQG para α com Erros-Padrão

Parâmetro	Estimativa	Erro-padrão
α_{11}	15.2747	1.9671
α_{12}	12.0735	0.8802
α_{21}	10.1569	2.1209
α_{22}	6.1371	0.6131
α_{31}	0.0671	0.0136
α_{32}	0.1010	0.0156

Gráfico 3.1.: Ganho de peso relativo (peso/peso ao nascer) por idade em meses



3.3.1. Testes de Hipóteses sobre Parâmetros de Diferentes Tratamentos

Para compararmos diferentes tratamentos, utilizaremos as mesmas estatísticas de teste descritas na **Seção 3.2**. Uma hipótese de interesse pode ser a igualdade entre as curvas, ou seja, testar se o vetor de parâmetros é igual para todos os tratamentos, ou ainda testar se o j -ésimo parâmetro de cada um dos M tratamentos tem o mesmo valor.

Suponha o vetor de parâmetros $\alpha = (\alpha_{11}, \dots, \alpha_{M1}, \dots, \alpha_{1k}, \dots, \alpha_{Mk})^t$, isto é $r=Mk$, onde k e M são como definidos em (2.1). Neste caso, se $M=2$ tratamentos, por exemplo, a hipótese de igualdade para todo o vetor de parâmetros poderia ser escrita da seguinte forma:

$$H_0 : [I_k \otimes (1,-1)] \alpha = 0 \quad \text{x} \quad H_1 : [I_k \otimes (1,-1)] \alpha \neq 0. \quad (3.48)$$

Neste exemplo, conforme (3.31), $h(\alpha)=[I_k \otimes (1,-1)]\alpha$, sendo que $h(\alpha):\mathbb{R}^r \rightarrow \mathbb{R}^q$ onde $q=k$.

Considerando novamente o **Exemplo 1**, $r=6$, $k=3$ e $M=2$ tratamentos. A partir dos valores obtidos para $\hat{\alpha}$ (ver Tabela 3.4), e utilizando a estatística modificada de Wald (3.34), temos: $h(\hat{\alpha}) = [3.2012 \quad 4.0198 \quad -0.0339]^t$; $H(\hat{\alpha})=[I_3 \otimes (1,-1)]$ e

$$\hat{\Omega} = \begin{bmatrix} 3.8694 & 0 & 3.9192 & 0 & -0.0246 & 0 \\ & 0.7748 & 0 & 0.3979 & 0 & -0.0105 \\ & & 4.4984 & 0 & -0.0275 & 0 \\ & & & 0.3759 & 0 & -0.0064 \\ & & & & 0.0002 & 0 \\ (\text{sim.}) & & & & & 0.0002 \end{bmatrix} \quad (3.49)$$

Obtemos então $W_2=1.29$. Pode-se tomar a decisão de não rejeitar H_0 uma vez que $P(F_{(3,14)}>1.29)=0.32$.

Caso houvesse $M=3$ tratamentos, a hipótese nula para este mesmo teste poderia ser escrita da seguinte forma:

$$H_0 : \left[\mathbf{I}_k \otimes \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \right] \alpha = 0, \quad (3.50)$$

e para $M=4$,

$$H_0 : \left[\mathbf{I}_k \otimes \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \right] \alpha = 0. \quad (3.51)$$

Para testar se um determinado parâmetro é igual para $M=2$ tratamentos, podemos escrever a hipótese nula na forma:

$$H_0 : [\mathbf{I}_{k(j)} \otimes (1,-1)] \boldsymbol{\alpha} = \mathbf{0}, \quad (3.52)$$

onde $\mathbf{I}_{k(j)}$ é a j -ésima linha da matriz \mathbf{I}_k . Neste caso, $h(\boldsymbol{\alpha}) = [\mathbf{I}_{k(j)} \otimes (1,-1)] \boldsymbol{\alpha}$, $\mathbf{H}(\boldsymbol{\alpha}) = [\mathbf{I}_{k(j)} \otimes (1,-1)]$ e $q=1$.

Da mesma forma, pode-se utilizar (3.32) e (3.33) como alternativas a (3.34) e também construir outras hipóteses que sejam de interesse para o problema.

Capítulo IV - Inferência em Modelos Não-Lineares Considerando Estruturas Especiais de Covariância

Neste capítulo, particularizamos a metodologia descrita no **Capítulo III**, considerando algumas das estruturas especiais para a matriz de covariância normalmente empregadas em dados longitudinais (ver **Seção 2.3**).

4.1. Estrutura de independência

Admitindo-se que a matriz de covariância das observações possui a estrutura de independência (2.24), estaremos supondo que as observações feitas sobre uma mesma unidade experimental são não correlacionadas e as variâncias constantes através do tempo.

Em geral, estas condições não são satisfeitas e esta estrutura é inadequada para dados longitudinais. No entanto em alguns casos a estrutura (2.24) é empregada, como por exemplo em estudos de crescimento onde há grande espaçamento entre duas medidas consecutivas. Nestes casos, alguns autores (ver por exemplo, Berkey, 1982) alegam que a correlação entre as medidas é tão pequena que pode ser ignorada. A estrutura de independência pode também ser empregada em estudos onde sua adequação seja indicada através do cálculo da matriz de covariância amostral.

A metodologia para o ajuste de modelos não lineares com erros não-correlacionados encontra-se bem descrita, por exemplo, em Seber e Wild (1989), Gallant (1987) ou Bard (1988). Esta é a situação mais simples para análise, pois o problema reduz-se ao ajuste de modelos não-lineares sem as especificidades de dados longitudinais, isto é, as p observações de cada indivíduo podem passar a compor um único vetor de respostas sem que seja necessário transformá-las na forma descrita no **Capítulo III**.

Neste caso, de forma diferente do que foi visto anteriormente, é possível fazer uma analogia com modelos de regressão univariados (Seber e Wild, 1989), e aplicar a metodologia desenvolvida para tais modelos.

O método de estimação dos parâmetros pode ser o de mínimos quadrados ordinários, uma vez que este é equivalente ao de mínimos quadrados generalizados para esta estrutura.

Observe por exemplo que no incremento de Gauss-Newton dado por (3.12), para $\Sigma = \sigma^2 \mathbf{I}$ temos:

$$\begin{aligned} \delta^{(u)}(\text{GN}) &= [\mathbf{J}^{*t}(\alpha^{(u)}) (\sigma^2 \mathbf{I})^{-1} \mathbf{J}^*(\alpha^{(u)})]^{-1} \mathbf{J}^{*t}(\alpha^{(u)}) (\sigma^2 \mathbf{I})^{-1} [\mathbf{y}^* - \mathbf{f}^*(\alpha^{(u)})] \\ &= [\mathbf{J}^{*t}(\alpha^{(u)}) \mathbf{J}^*(\alpha^{(u)})]^{-1} \mathbf{J}^{*t}(\alpha^{(u)}) [\mathbf{y}^* - \mathbf{f}^*(\alpha^{(u)})] \end{aligned} \quad (4.1)$$

Os testes de hipóteses e intervalos de confiança seguem também a teoria descrita no **Capítulo III**, apenas ressaltando que neste caso $\Sigma = \sigma^2 \mathbf{I}$. É o que se pode observar pelo teorema a seguir, transcrito da teoria de modelos não-lineares de regressão univariada (Seber e Wild, 1989, pág.24):

Seja $\alpha^\#$ o verdadeiro valor do vetor de parâmetros α . Dados $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ e condições de regularidade apropriadas, então para n grande temos, aproximadamente:

(i) $\hat{\alpha} - \alpha^\# \sim N_r(\mathbf{0}, \sigma^2 [\mathbf{J}^t(\alpha^\#) \mathbf{J}(\alpha^\#)]^{-1})$;

(ii) $(n-r)s^2 / \sigma^2 \cong \varepsilon^t (\mathbf{I}_n - \mathbf{P}_J) \varepsilon / \sigma^2 \sim \chi^2_{n-r}$; onde $s^2 = S(\hat{\alpha}) / (n-r)$ é o estimador de σ^2 ; $\mathbf{P}_J = \mathbf{J}(\mathbf{J}^t \mathbf{J})^{-1} \mathbf{J}^t$ e $\mathbf{I}_n - \mathbf{P}_J$ são simétricas e idempotentes.

(iii) $\hat{\alpha}$ é estatisticamente independente de s^2 ; e

(iv) $\frac{[S(\alpha^\#) - S(\hat{\alpha})] / r}{S(\hat{\alpha}) / (n-r)} \approx \frac{\varepsilon^t \mathbf{P}_J \varepsilon}{\varepsilon^t (\mathbf{I}_n - \mathbf{P}_J) \varepsilon} \times \frac{n-r}{n} \sim F_{r, n-r}$,

Conforme dito anteriormente, quando utilizado para dados longitudinais, no lugar de n observações, estaremos trabalhando com $n \times p$.

Como ilustração, vamos assumir esta estrutura e estimar os parâmetros do modelo proposto no **Exemplo 2**. Queremos estimar o vetor de parâmetros $\alpha = (\alpha_1, \alpha_2, \alpha_3)^t$, dado que assumimos o seguinte modelo para os dados: $y_{ij} = \alpha_1 / [1 + \alpha_2 \exp(-\alpha_3 x_{ij})] + \varepsilon_{ij}$,

$j=1,\dots,7$, $i=1,\dots,5$, onde $\mathbf{x}=(118, 484, 664, 1004, 1231, 1372, 1582)^t$ e $\varepsilon_i \sim N(0, \sigma^2 \mathbf{I}_5)$. Pelo método de mínimos quadrados, devemos encontrar $\hat{\alpha}$ que minimiza $S(\alpha) = \sum_{i=1}^5 \sum_{j=1}^7 \{y_{ij} - \alpha_1 / [1 + \alpha_2 \exp(-\alpha_3 x_{ij})]\}^2$. Apresentaremos apenas os resultados obtidos a partir do método de Gauss-Newton Modificado. O algoritmo de Newton forneceu resultados bastante semelhantes.

Tabela 4.1.: Valores estimados para α através de MQO

Iteração	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	SQR= $S(\hat{\alpha})$
0	150.0000	5.0000	0.002000	70109.1911
1	182.1164	7.3001	0.003077	17984.1316
2	191.0541	7.7528	0.002815	17523.1933
3	192.7011	7.8557	0.002828	17480.2338
4	192.6881	7.8565	0.002829	17480.2335
5	192.6876	7.8566	0.002829	17480.2335

Então, $\hat{\alpha}=(192.6876, 7.6876, 0.002829)^t$ e $s^2=17480.23/(35-3)=546.2573$. Para obtermos intervalos de confiança de $(1-\phi) \times 100\%$ para α_i , utilizamos:

$$\hat{\alpha}_i \pm t_{\phi/2} \left(\hat{\Omega}_{ii} \right)^{1/2}, \quad (4.2)$$

onde $\hat{\Omega} = s^2 \left[\sum_{j=1}^n \mathbf{J}^t(\hat{\alpha}) \mathbf{J}(\hat{\alpha}) \right]_{11}^{-1}$. No exemplo, temos que:

$$\hat{\Omega} = 546.2573 \times \begin{bmatrix} 0.7853 & -0.0418 & -2.2 \times 10^{-5} \\ & 0.0111 & 2.4 \times 10^{-5} \\ & & 8.2 \times 10^{-10} \end{bmatrix}, \quad (4.3)$$

e $t_{\phi/2}(np-r) = t_{(0,025)}(32) = 2.0369$. A partir de (4.2), obtemos $IC_{95\%}(\hat{\alpha}_1) = 192.6876 \pm 2.0369 \times \sqrt{546.2573 \times 0.7853} = [151.4520 ; 233.9233]$. Da mesma forma obtemos: $IC_{95\%}(\hat{\alpha}_2) = [2.9036 ; 12.8095]$ e $IC_{95\%}(\hat{\alpha}_3) = [0.0015 ; 0.0042]$.

Suponha que haja interesse em testar se $\alpha_3=0$. Utilizando a estatística modificada de Wald (3.34) obtemos: $h(\hat{\alpha}) = \hat{\alpha}_3 = 0.002829$. Neste caso: $q=1$,

$$\hat{H} = \left. \frac{\partial h(\alpha)}{\partial \alpha^t} \right|_{\alpha=\hat{\alpha}} = (0 \ 0 \ 1) \text{ e } \hat{H}\hat{\Omega}\hat{H}^t = \hat{\Omega}_{33} = 4.4 \times 10^{-7}. \text{ Daí calculamos}$$

$$W_2 = \frac{0.002829 \times 2232486.77 \times 0.002829}{1} = 17.86. \text{ Como } P(F_{(1,4)} > 17.86) = 0.0134,$$

rejeita-se H_0 ao nível de significância de 1 %.

O pacote estatístico utilizado foi o SAS, através do procedimento NLIN, e o algoritmo de Gauss-Newton Modificado (ver **Seção 3.1.1**). Através deste procedimento, o SAS fornece as estimativas para os parâmetros, os erros padrão, os intervalos de confiança e a matriz de correlação para os estimadores dos parâmetros. Com relação aos testes de hipóteses, os cálculos foram feitos em separado (o pacote não fornece). Ressaltamos que todos estes resultados são assintóticos.

4.2. Estrutura Uniforme

Conforme visto no **Capítulo II**, a estrutura uniforme (2.25) é adequada a situações em que as correlações entre as observações feitas sobre a mesma unidade experimental são iguais, independentemente do espaço de tempo decorrido entre elas. Além disso, estaremos supondo também que as variâncias são constantes no tempo.

Estas suposições são mais comuns no contexto de medidas repetidas, quando as condições experimentais sob as quais cada indivíduo será observado são aleatorizadas. Em geral, isto não ocorre em dados longitudinais, e espera-se que as correlações sejam menores quanto maior for o espaçamento de tempo entre duas observações. No entanto, em algumas situações esta estrutura pode ser adequada.

Neste caso, da mesma forma que na estrutura geral (2.31), pode-se utilizar o procedimento de Mínimos Quadrados Generalizados discutido no **Capítulo III**. Segundo Gennings et ali (1989), um estimador para a matriz de covariância quando esta segue o padrão (2.25) pode ser

$$\hat{\Sigma} = \hat{\sigma}^2(1 - \hat{\rho})I_p + \hat{\sigma}^2\hat{\rho}1_p1_p^t \quad (4.4)$$

onde $\hat{\sigma}^2 = \frac{1}{np} \sum_{j=1}^p \sum_{i=1}^n [y_{ij} - f(x_{ij}, A_i, \alpha)]^2$ e $\hat{\rho} = \{ \sum_{i=1}^n [\sum_{j=1}^p [y_{ij} - f(x_{ij}, A_i, \alpha)]]^2 - np\hat{\sigma}^2 \} / np(p-1)\hat{\sigma}^2$.

Conforme a **Seção 2.3**, a estrutura uniforme pode ser obtida ainda como caso particular dos modelos de efeitos aleatórios (2.27) quando se tem no modelo somente um efeito aleatório. Basta fazer a matriz $Z=1_p$ e $D=\sigma_e^2$. Neste caso, $\Sigma=ZDZ^t+\sigma_d^2I=1\sigma_e^21^t\sigma_d^2I=$

$$= \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ & 1 & \dots & \rho \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}, \text{ onde } \sigma^2 = \sigma_e^2 + \sigma_d^2 \text{ e } \rho = \sigma_e^2 / (\sigma_e^2 + \sigma_d^2).$$

Como caso particular dos modelos de efeitos aleatórios, será possível utilizar também para esta estrutura o procedimento de estimação que será apresentado na **Seção 4.3**. No entanto, no exemplo a seguir, utilizaremos o estimador proposto em (4.4).

Através do método de MQO obteve-se, no **Exemplo 2**, $\tilde{\alpha} = (192.6876, 7.6876, 0.002829)^t$. A partir de (4.4) estima-se a matriz de covariância das observações. Em primeiro lugar, obtém-se $\hat{\sigma}^2 = 17480.2335/35 = 499.4352$ $\hat{\rho} = 0.6236$, o que resulta em

$$\hat{\Sigma} = 499.4352 \times \begin{bmatrix} 1 & .6236 & \dots & .6236 \\ & 1 & .6236 & .6236 \\ & & .. & .6236 \\ & & & 1 \end{bmatrix} = \begin{bmatrix} 499.4352 & 311.4478 & \dots & 311.4478 \\ & 499.4352 & 311.4478 & 311.4478 \\ & & .. & 311.4478 \\ & & & 499.4352 \end{bmatrix} e$$

$$\hat{\Sigma}^{-1} = \begin{bmatrix} .00460 & -.0007 & \dots & -.0007 \\ & .00460 & & -.0007 \\ & & \dots & -.0007 \\ & & & .00460 \end{bmatrix} \text{ Transformando-se os dados a partir desta}$$

matriz, e utilizando-se o procedimento iterativo de Gauss-Newton, obtém-se os seguintes resultados a cada iteração:

Tabela 4.2.: Valores estimados para α através de MQG

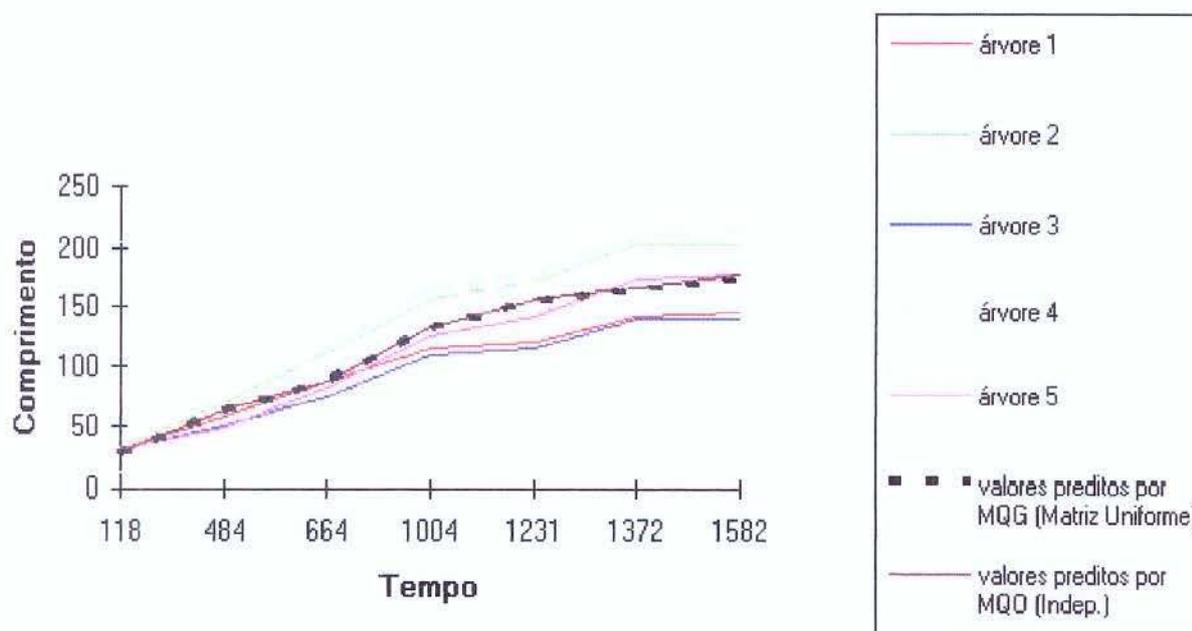
Iteração	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	SQR=S($\hat{\alpha}$)
0	100.0000	5.0000	0.0020	265.1231
1	181.5743	8.5133	0.0036	43.4620
2	185.5589	7.3476	0.0028	36.2896
3	192.2717	7.9161	0.0028	34.9987
4	192.2526	7.9223	0.0028	34.9987
5	192.2527	7.9222	0.0028	34.9987

Ao final, temos a estimativa de MQG para α , e obtemos da mesma forma que em (4.3) os erros padrão para cada componente do vetor. Comparando-se com os resultados obtidos na Seção 4.1, onde assume-se que as observações medidas na mesma árvore são não-correlacionadas, nota-se que as estimativas dos parâmetros são praticamente as mesmas, parecendo haver uma pequena diminuição nos erros:

Tabela 4.3.: Estimativa de MQG para α com Erros-Padrão

Parâmetro	Estrutura Uniforme		Estrutura de Independência	
	Estimativa	Erro-padrão	Estimativa	Erro-padrão
α_1	192.2527	16.3130	192.6876	20.7117
α_2	7.9222	2.2565	7.8566	2.4624
α_3	0.0028	0.0004	0.0028	0.0007

Gráfico 4.1: Comprimento da circunferência de laranjeiras (em mm) por tempo (em dias)



4.3. Estrutura de Modelos de Efeitos Aleatórios

Modelos de efeitos aleatórios, algumas vezes chamados na literatura de modelos de efeitos mistos, vêm sendo utilizados com bastante frequência em medidas repetidas e em particular, em dados longitudinais (Lindstrom e Bates, 1990). Além da vantagem de reduzir o número de parâmetros da matriz de covariância, esta estrutura pode ser empregada em casos de dados não balanceados ou incompletos (ver por exemplo, Jenrich e Schluchter, 1986, ou Laird e Ware, 1982). Para tanto, nesta Seção trabalharemos com uma generalização do modelo (2.1), mais flexível com relação a estas características.

Considerando que em grande parte dos trabalhos que tratam dos modelos de crescimento descritos na Seção 2.1 há o interesse no ajuste de curvas individuais, a estrutura de modelos de efeitos aleatórios oferece um atrativo especial. Ela parte da idéia de que as respostas individuais seguem uma mesma forma funcional com parâmetros que variam entre os indivíduos.

O uso de modelos de efeitos aleatórios no caso em que a variável resposta pode ser descrita por uma função linear é discutido por Laird e Ware (1982), Ware (1985) e Vonesh e Carter (1987), entre outros.

No modelo proposto por Laird e Ware (1982), a variável resposta para cada unidade experimental pode ser representada da seguinte forma

$$y_i = X_i \alpha + Z_i b_i + e_i, \quad i = 1, \dots, n \quad (4.5)$$

onde y_i ($p_i \times 1$) é o vetor de respostas, X_i ($p_i \times r$) é a matriz conhecida de delineamento de posto coluna completo, α ($r \times 1$) é o vetor de parâmetros populacionais desconhecidos, b_i ($v \times 1$) é o vetor de efeitos aleatórios individuais, Z_i ($p_i \times v$) é a matriz conhecida de especificação de posto completo que liga b_i a y_i e e_i ($p_i \times 1$) é o vetor de erros aleatórios. Além disso, $e_i \sim N(0, \sigma^2 R_i)$ e $b_i \sim N(0, D)$, onde R_i ($p_i \times p_i$) e D ($v \times v$) são matrizes definidas positivas e R_i é uma matriz conhecida que depende de i somente pela dimensão; b_i e e_i são independentes. Frequentemente assume-se $R_i = I_{p_i}$.

Laird e Ware consideram o modelo em dois estágios: no primeiro, α e b_i são fixos e a resposta y_i é condicional a b_i . Os erros e_i são independentes com $e_i \sim N(0, \sigma^2 R_i)$. No segundo supõe-se que os b_i 's são independentes com $b_i \sim N(0, D)$ e b_i 's independentes dos e_i 's, $i=1, \dots, n$. Dessa forma, a distribuição marginal dos y_i 's é $N(X_i \alpha, (Z_i D Z_i^t + \sigma^2 R_i))$.

Na adaptação destas idéias aos modelos não-lineares, embora a princípio os efeitos aleatórios possam ser ligados à variável resposta através de uma função não-linear, as soluções freqüentemente envolvem a linearização do modelo com relação a estes efeitos. Neste sentido, um dos trabalhos mais recentes encontrados na literatura, publicado por Vonesh e Carter (1992), propõe um modelo bastante geral, onde os efeitos aleatórios são ligados à variável resposta de forma linear através de uma matriz Z_i , enquanto a esperança da variável resposta é uma função não linear. Modelos anteriores com resposta linear em relação aos parâmetros (por exemplo, Vonesh e Carter, 1987, Laird e Ware, 1982;) e não linear (por exemplo, Lindstrom e Bates, 1990; Sheiner e Beal, 1980) podem ser considerados casos particulares deste modelo.

4.3.1. O Modelo de Vonesh e Carter

A forma geral do modelo é dada por:

$$y_i = f(X_i, a_i, \alpha) + Z_i b_i + e_i, \quad i = 1, \dots, n \quad (4.6)$$

onde y_i , Z_i e α são como definidos em (4.5); f é uma função vetorial não-linear; X_i ($p_i \times w$) é a matriz de variáveis "dentro" de indivíduos; a_i ($M \times 1$), vetor de valores de M variáveis "entre" indivíduos; b_i ($v \times 1$) é o vetor de efeitos aleatórios para o i -ésimo indivíduo, e_i ($p_i \times 1$) é o vetor de erros aleatórios. Assume-se que $e_i \sim iid(0, \sigma^2 I_{p_i})$ e $b_i \sim iid(0, D)$, onde D ($v \times v$) é uma matriz definida positiva; b_i e e_i são mutuamente independentes. Dependendo do método de estimação utilizado, não é necessário supor distribuição Normal para as observações. Conforme será visto mais tarde, basta que algumas condições de regularidade sejam satisfeitas para que os estimadores sejam consistentes e assintoticamente normais.

É fácil ver que $V(y_i) = \Sigma_i = Z_i D Z_i^t + \sigma^2 I_{p_i}$. Para dados balanceados, $p_i = p$ e $X_i = X$. Além disso, em muitos casos, $r = Mk$.

Observe que se $f(\mathbf{X}_i, \boldsymbol{\alpha}) = \mathbf{X}_i\boldsymbol{\alpha}$, temos o modelo de efeitos aleatórios proposto por Laird e Ware (1982). Se além disso, $f(\mathbf{X}_i, \mathbf{a}_i, \boldsymbol{\alpha}) = \mathbf{X}_i(\mathbf{I}_w \otimes \mathbf{a}_i')\boldsymbol{\alpha}$ e $\mathbf{Z}_i = \mathbf{X}_i$, então temos o modelo de curva de crescimento de coeficientes aleatórios discutido por Vonesh e Carter (1987). O modelo (4.6) permite também o uso de matrizes de covariância alternativas. Por exemplo, se temos somente um efeito aleatório e fazemos $\mathbf{Z}_i = \mathbf{1}(p_i \times 1)$, obtemos um modelo com coeficiente linear aleatório. A matriz de covariância resultante segue a estrutura uniforme (2.25).

A princípio, (4.6) foge ao modelo geral (2.1) devido à presença do vetor de efeitos aleatórios ligado à variável resposta de forma linear. No entanto, veremos a seguir que com pequenas alterações esse modelo pode ser escrito sob a forma proposta anteriormente.

Considere o modelo em dois estágios:

$$1^{\text{a}}) \mathbf{y}_i = \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_i) + \mathbf{e}_i, \quad i = 1, \dots, n, \quad (4.7)$$

onde $\boldsymbol{\alpha}_i$ ($k \times 1$) é o vetor de parâmetro para o i -ésimo indivíduo, \mathbf{y}_i , \mathbf{f} e \mathbf{X}_i são definidos como em (4.6) e $\mathbf{e}_i \sim \text{NID}(0, \sigma^2 \mathbf{I}_{p_i})$.

$$2^{\text{a}}) \boldsymbol{\alpha}_i = \mathbf{g}(\mathbf{a}_i, \boldsymbol{\alpha}) + \mathbf{B}_i \boldsymbol{\xi}_i, \quad i = 1, \dots, n, \quad (4.8)$$

onde \mathbf{a}_i e $\boldsymbol{\alpha}$ são definidos como em (4.6), $\mathbf{g}(\mathbf{a}_i, \boldsymbol{\alpha})$ é uma função vetorial ($k \times 1$) que pode também ser não-linear; \mathbf{B}_i ($k \times v$) é uma matriz de valores conhecidos de posto completo; $\boldsymbol{\xi}_i$, $i=1, \dots, n$ são vetores aleatórios $\text{NID}(0, \mathbf{D})$ e assume-se que são independentes dos \mathbf{e}_i 's.

Combinando (4.7) e (4.8) e expandindo em série de Taylor de 1ª ordem sobre $\boldsymbol{\xi}_i=0$, obtemos o modelo:

$$\mathbf{y}_i = \mathbf{f}[\mathbf{X}_i, \mathbf{g}(\mathbf{a}_i, \boldsymbol{\alpha})] + \mathbf{Z}_i^* \mathbf{b}_i + \mathbf{e}_i, \quad i = 1, \dots, n, \quad (4.9)$$

onde $\mathbf{Z}_i^* = \left[\frac{\partial \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_i)}{\partial \boldsymbol{\alpha}_i'} \right]_{\boldsymbol{\alpha}_i = \mathbf{g}(\mathbf{a}_i, \boldsymbol{\alpha})} \mathbf{x} \mathbf{B}_i$ ($k \times v$) e $\mathbf{b}_i = \boldsymbol{\xi}_i$. Assim, como em (4.6), as

suposições de normalidade para \mathbf{b}_i e \mathbf{e}_i podem ser desconsideradas. Desta forma vemos que o

modelo de Vonesh e Carter pode ser visto como uma modificação de (2.1), onde a definição das variáveis “entre” indivíduos e os efeitos aleatórios constituem um segundo estágio do modelo.

Comparando agora (4.6) com a forma obtida em (4.9), observamos que esta última é ainda mais geral, dado que a matriz Z_i pode depender de α . Verificamos então que quando os efeitos aleatórios forem também não-lineares em relação à variável resposta, estaremos trabalhando com o seguinte modelo:

$$y_i = f(X_i, \alpha) + Z_i(\alpha) b_i + e_i, \quad i = 1, \dots, n, \quad (4.10)$$

onde a matriz de covariância ($\Sigma_i = Z_i D Z_i^t + \sigma^2 I$) vai depender também dos parâmetros α . Na prática o que se faz é fixar Z_i condicionando com relação a α , que será substituído por uma estimativa consistente (provado que existe). Em seguida, o procedimento de estimação é o mesmo que para o modelo (4.6), e será discutido posteriormente.

Como ilustração, suponha que no **Exemplo 2** um pesquisador decida ajustar a curva logística com um componente aleatório na assíntota (α_1). Neste caso, $v=1$. Como os dados são balanceados, $X_i = X_{(7 \times 1)}$ e $Z_i = Z_{(7 \times 1)}$. Podemos escrever o modelo em dois estágios da seguinte forma:

$$1^a) y_{ij} = \frac{\alpha_{11}}{1 + \alpha_{12} \exp(-\alpha_{13} x_j)} + e_{ij}, \quad i = 1, \dots, 5; j = 1, \dots, 7;$$

$$2^a) \alpha_i = I_3 \alpha + B_i^t \xi_i, \quad i = 1, \dots, 5,$$

onde $B_i^t = [1 \ 0 \ 0]$ e $\xi_i^t = [\xi_{i1} \ \xi_{i2} \ \xi_{i3}]$, isto é $\alpha_i = \begin{bmatrix} \alpha_1 + \xi_{i1} \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$. Em outras palavras, estaremos

trabalhando com o seguinte modelo:

$$y_{ij} = \frac{\alpha_1 + \xi_{i1}}{1 + \alpha_2 \exp(-\alpha_3 x_j)} + e_{ij}, \quad i = 1, \dots, 5; j = 1, \dots, 7 \quad (4.11)$$

Linearizando-se o efeito aleatório através de expansão de Taylor de 1ª ordem sobre $\xi_{i1}=0$, chega-se à forma proposta em (4.10):

$$y_{ij} = \frac{\alpha_i}{1 + \alpha_2 \exp(-\alpha_3 x_j)} + z_j(\alpha) b_{i1} + e_{ij}, \quad i = 1, \dots, 5; j = 1, \dots, 7, \quad (4.12)$$

onde $b_i = \xi_{i1}$ tem dimensão $v=1$ e z_j é o j -ésimo elemento de

$$Z_{(7 \times 1)}(\alpha) = \left[\frac{\partial f(x, \alpha_i)}{\partial \xi_i^t} \right] = \left[\frac{\partial f(x, \alpha_i)}{\partial \alpha_i^t} \right] \cdot \left[\frac{\partial \alpha_i}{\partial \xi_i^t} \right] = \begin{bmatrix} \frac{1}{1 + \alpha_2 \exp(-\alpha_3 x_1)} \\ \vdots \\ \frac{1}{1 + \alpha_2 \exp(-\alpha_3 x_7)} \end{bmatrix}. \quad (4.13)$$

O modelo em dois estágios descrito em (4.7), (4.8) e (4.9) é similar ao que foi proposto por Sheiner e Beal (1980). Estes autores desenvolveram um programa (NOMEM - Nonlinear Mixed Effect Model) para o ajuste de tais modelos, em que os efeitos aleatórios aparecem também na forma não-linear. Sheiner e Beal trabalharam com modelos aplicados a problemas de farmacocinética. Com base na suposição de normalidade dos b_i 's e dos e_i 's, estimaram os parâmetros por máxima verossimilhança. Vonesh e Carter (1992) criticam o método NOMEM, afirmando que além de ser computacionalmente intensivo, pode não fornecer estimativas consistentes.

Outro modelo anterior ao de Vonesh e Carter que pode ser aproximado por (4.10) é o de Lindstrom e Bates (1990). Estes autores utilizam também expansão de Taylor de 1ª ordem, mas a expansão é feita sobre os valores estimados de b_i , diferente de Sheiner e Beal, que expandem sobre a média. Isto significa ter em (4.9), $b_i = \xi_i - \hat{\xi}_i$. Dessa forma, segundo Jones (1993), a linearização fornece uma melhor aproximação para a função não-linear. Nesse modelo, a função g é sempre linear, isto é, $g = A_i \alpha$ e $b_i \sim N(0, D)$.

Lindstrom e Bates propõem um procedimento de estimação em dois passos, ambos iterativos, que se sucedem até que haja convergência. No primeiro, chamado “Pseudo-Data Step”, utilizam um procedimento comum de mínimos quadrados não-lineares aplicados aos dados transformados. No segundo passo, chamado “Linear Mixed Effects Step”, aproveitam a teoria desenvolvida para modelos de efeitos aleatórios com função resposta linear e adaptam um algoritmo para estimação dos parâmetros através dos métodos de máxima verossimilhança ou máxima verossimilhança restrita. O modelo proposto tem como desvantagens a necessidade de suposição de normalidade para os efeitos aleatórios, além do fato de ter um procedimento de estimação computacionalmente intensivo.

4.3.2. Procedimentos de Estimação

Com relação aos modelos (4.6) e (4.10), serão apresentados dois procedimentos alternativos de estimação. O primeiro é o de mínimos quadrados generalizados, na forma descrita na **Seção 3.1**. A estimativa consistente da matriz de covariância Σ , necessária à aplicação do método, pode ser obtida a partir de resultados apresentados por Reinsel (1985) e Vonesh e Carter (1987), não necessitando de procedimentos iterativos. O outro método de estimação que será visto em seguida fornece estimativas de máxima verossimilhança ou quasi-verossimilhança através de uma adaptação do algoritmo EM (ver, por exemplo, Laird e Ware, 1982) para modelos não-lineares.

(i) Método de Mínimos Quadrados Generalizados

Caso a matriz Z_i dependa de α (modelo (4.10)), o primeiro passo será obter estimativas iniciais consistentes para este vetor e substituí-las em $Z_i(\alpha)$. O estimador utilizado será o de mínimos quadrados ordinários, obtido a partir do 1º estágio do modelo (4.7), considerando, no entanto, estimativas diferentes para cada grupo de tratamento. Isto significa tratar os dados inicialmente como se fossem não-correlacionados, ignorando a presença de efeitos aleatórios estabelecida no 2º estágio (4.8). A matriz Z_i deve ter posto coluna completo e $p_i > v$, $\forall i$. Os estimadores α_{MQO} são consistentes sob as condições de regularidade apresentadas por Vonesh e Carter (1992).

O próximo passo será estimar a matriz de covariância dos efeitos aleatórios, \mathbf{D} , e a variância “dentro” de indivíduos, σ^2 . Para tanto, é preciso primeiro obter $\hat{\mathbf{b}}_i$, $i=1, \dots, n$, da seguinte forma: a partir das estimativas de α obtidas por MQO, substituímos ϵ_i pelos resíduos estimados $\hat{\epsilon}_i = y_i - f(\mathbf{X}_i, \tilde{\alpha}_{\text{MQO}})$ no modelo de regressão (4.14). Dessa forma, os parâmetros a serem estimados são os coeficiente aleatórios \mathbf{b}_i :

$$\epsilon_i = \mathbf{Z}_i \mathbf{b}_i + e_i, \quad i = 1, \dots, n \quad (4.14)$$

Além da estimação dos \mathbf{b}_i 's por MQO, podemos obter uma estimativa para σ^2 a partir do quadrado médio do resíduo (QMR_{*i*}) deste modelo para cada indivíduo:

$$\hat{\mathbf{b}}_i = (\mathbf{Z}_i^t \mathbf{Z}_i)^{-1} \mathbf{Z}_i^t \hat{\epsilon}_i, \quad (4.15)$$

$$\hat{\sigma}^2 = \left(\sum_{i=1}^n p_i - n\nu \right)^{-1} \sum_{i=1}^n (p_i - \nu) \text{QMR}_i, \quad (4.16)$$

onde $\text{QMR}_i = \frac{\hat{\epsilon}_i^t [\mathbf{I}_{p_i} - \mathbf{Z}_i (\mathbf{Z}_i^t \mathbf{Z}_i)^{-1} \mathbf{Z}_i^t] \hat{\epsilon}_i}{(p_i - \nu)}$. A matriz de covariância dos efeitos aleatórios pode

ser estimada por

$$\hat{\mathbf{D}} = \begin{cases} \mathbf{S}_{\hat{\mathbf{b}}\hat{\mathbf{b}}} - \hat{\sigma}^2 (\mathbf{n} - \mathbf{M})^{-1} \sum_{i=1}^n [1 - \mathbf{a}_i^t (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{a}_i] (\mathbf{Z}_i^t \mathbf{Z}_i)^{-1}, & \text{se } \hat{\lambda} > \hat{\sigma}^2 \\ \mathbf{S}_{\hat{\mathbf{b}}\hat{\mathbf{b}}} - \hat{\lambda}^2 (\mathbf{n} - \mathbf{M})^{-1} \sum_{i=1}^n [1 - \mathbf{a}_i^t (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{a}_i] (\mathbf{Z}_i^t \mathbf{Z}_i)^{-1}, & \text{se } \hat{\lambda} \leq \hat{\sigma}^2 \end{cases} \quad (4.17)$$

onde $\mathbf{S}_{\hat{\mathbf{b}}\hat{\mathbf{b}}} = \frac{\hat{\mathbf{B}}^t [\mathbf{I}_n - \mathbf{A} (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t] \hat{\mathbf{B}}}{n - \mathbf{M}}$ é o estimador da matriz de covariância de $\hat{\mathbf{b}}_i$, corrigida dos efeitos “entre” indivíduos; $\hat{\mathbf{B}} = (\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_n)^t$ é o estimador da matriz ($n \times \nu$) de coeficientes de regressão, $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)^t$ é a matriz de delineamento que indica o tratamento a que cada indivíduo pertence, $\hat{\lambda}$ é a menor raiz de

$$\left| S_{\hat{\beta}\hat{\beta}} - \hat{\lambda}^2 (\mathbf{n} - \mathbf{M})^{-1} \sum_{i=1}^n [1 - \mathbf{a}_i^t (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{a}_i] (\mathbf{Z}_i^t \mathbf{Z}_i)^{-1} \right| = 0 \quad (4.18)$$

Segundo os autores, utilizando-se apenas a primeira equação de (4.17) não se pode garantir estimativas semidefinidas positivas para \mathbf{D} . Por isso propõem que se utilize a alteração dada pela segunda equação, caso $\hat{\lambda} \leq \hat{\sigma}^2$.

Estes estimadores de σ^2 (4.16) e \mathbf{D} (4.17) foram apresentados anteriormente por Reinsel (1985) e utilizados por Vonesh e Carter (1987) para um modelo de efeitos aleatórios linear. O uso destes estimadores também em modelos não lineares é possível, considerando-se que com a expansão de Taylor, o modelo torna-se linear em relação aos efeitos aleatórios. Assim, voltando a (4.14), Vonesh e Carter (1992) mostram que este modelo pode ser reescrito de forma análoga ao discutido por aqueles autores:

$$\varepsilon_i = \mathbf{Z}_i (\mathbf{I}_v \otimes \mathbf{a}_i^t) \mathbf{0} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i, \quad i = 1, \dots, n \quad (4.19)$$

A matriz de covariância do vetor \mathbf{y}_i é então estimada por:

$$\hat{\Sigma}_i = [\mathbf{Z}_i \hat{\mathbf{D}} \mathbf{Z}_i^t + \hat{\sigma}^2 \mathbf{I}_{p_i}] \quad (4.20)$$

sendo utilizada na 3ª etapa do procedimento de estimação descrito em 3.1 para obtenção de $\hat{\alpha}_{\text{MQG}}$. Para ilustrar este procedimento, voltemos ao **Exemplo 2**. Partindo das estimativas iniciais para α obtidas por MQO, $\tilde{\alpha} = (192.6876, 7.8566, 0.0028)^t$ e substituindo estes valores na função logística, obtemos os resíduos estimados $\hat{\varepsilon}_i = \mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \tilde{\alpha})$:

$$\hat{\varepsilon}_1 = \begin{bmatrix} 09238 \\ -62649 \\ -05448 \\ -170618 \\ -351980 \\ -238081 \\ -318581 \end{bmatrix}, \quad \hat{\varepsilon}_2 = \begin{bmatrix} 39238 \\ 47351 \\ 234552 \\ 239382 \\ 168020 \\ 3719191 \\ 261419 \end{bmatrix}, \quad \hat{\varepsilon}_3 = \begin{bmatrix} 09238 \\ -132649 \\ -125448 \\ -240618 \\ -401980 \\ -268081 \\ -368581 \end{bmatrix}, \quad \hat{\varepsilon}_4 = \begin{bmatrix} 29238 \\ -22649 \\ 244552 \\ 349382 \\ 238020 \\ 431919 \\ 371419 \end{bmatrix}, \quad \hat{\varepsilon}_5 = \begin{bmatrix} 09238 \\ -152649 \\ -65448 \\ -70618 \\ -131980 \\ 81919 \\ 01419 \end{bmatrix}.$$

A partir de (4.13), obtemos $Z(\hat{\alpha}) = \begin{bmatrix} 01505 \\ 03304 \\ 04496 \\ 06791 \\ 07999 \\ 08557 \\ 09144 \end{bmatrix}$. Daí temos por (4.15) que

$\hat{B} = [-30.4393, 32.6725, -38.2975, 41.4120, -5.3476]^t$. De (4.16) e (4.17) obtemos, respectivamente, $\hat{\sigma}^2 = 61.6592$ e $\hat{D} = 1280.5442$. A matriz de covariância das observações é então estimada por (4.20) (abaixo da diagonal estão as correlações):

$$\hat{\Sigma} = \begin{bmatrix} 90.65 & 63.67 & 86.63 & 130.86 & 154.11 & 164.88 & 176.18 \\ .47 & 201.49 & 190.26 & 287.38 & 338.47 & 362.10 & 386.98 \\ .51 & .75 & 320.55 & 391.03 & 460.54 & 482.71 & 526.47 \\ .54 & .79 & .85 & 652.29 & 695.62 & 744.20 & 795.20 \\ .54 & .80 & .87 & .92 & 880.93 & 876.49 & 936.56 \\ .55 & .81 & .87 & .92 & .93 & 999.86 & 1001.96 \\ .55 & .81 & .87 & .93 & .94 & .94 & 1132.30 \end{bmatrix}.$$

Seguindo o procedimento descrito em 3.1, obtemos então $\hat{\alpha} = (192.6868, 7.8566, 0.0028)^t$ e a matriz de covariância desses estimadores:

$$\hat{\Omega}(\hat{\alpha}) = \left[\sum_{i=1}^n J(\hat{\alpha}) \hat{\Sigma}^{-1} J(\hat{\alpha}) \right]^{-1} = \begin{bmatrix} 305.1511 & -2.4103 & -0.0013 \\ & 0.6587 & 0.0002 \\ & & 4.77E-8 \end{bmatrix}.$$

Observe que, assim como na **Seção 4.2**, as estimativas obtidas por MQO e MQG são as mesmas, parecendo haver apenas uma diminuição nas estimativas dos erros.

Vonesh e Carter (1992) demonstraram que os estimadores obtidos são fortemente consistentes e têm distribuição assintótica Normal. O uso de Mínimos Quadrados Ordinários Iterativos, isto é, a repetição do procedimento a partir da segunda etapa (estimação da matriz de covariância) até obter convergência pode melhorar as estimativas em amostras pequenas.

Para dados balanceados e assumindo normalidade das observações, o procedimento pode ser ligeiramente modificado para fornecer estimadores de máxima verossimilhança (ou uma aproximação). Para isso, basta substituir S_{bb} por $(n-M)n^{-1}S_{bb}$ e iterar a partir da segunda etapa até obter convergência sob algum critério. Segundo os autores, é possível utilizar os mesmos argumentos de Gallant (1987, Cap.5.) para demonstrar que os estimadores obtidos nesta Seção para α , D e σ^2 e os estimadores de máxima verossimilhança são equivalentes.

(i) Métodos de Máxima Verossimilhança e Máxima Verossimilhança Restrita

Uma outra forma de estimar os parâmetros baseia-se numa adaptação do Algoritmo EM (Laird, Lange e Stram, 1987) para modelos de efeitos aleatórios não-lineares, proposta por Hirst et ali (1991).

Seja f uma função não linear de α e dos efeitos aleatórios b_i . Expandindo a função f em série de Taylor de 1ª. ordem sobre valores iniciais α_0 e b_{i0} de α e b_i , obtém-se uma aproximação do modelo:

$$y_i - f(X_i, \alpha_0, b_{i0}) \cong U_i[\alpha - \alpha_0] + Z_i[b_i - b_{i0}] + e_i, \quad (4.21)$$

onde $U_i(p \times k)$ é uma matriz cuja j -ésima linha é o vetor $(1 \times k)$

$$u_{ij} = \left[\frac{\partial f(x_{ij}, \alpha, b_i)}{\partial \alpha} \right]_{\alpha=\alpha_0}^t, \quad (4.22)$$

e $\mathbf{Z}_i(p_i \times k)$ é a matriz cuja j -ésima linha é dada por

$$\mathbf{z}_{ij} = \left[\frac{\partial f(\mathbf{x}_{ij}, \boldsymbol{\alpha}, \mathbf{b}_i)}{\partial \mathbf{b}_i} \right]_{\mathbf{b}_i = \mathbf{b}_{i0}}^t \quad (4.23)$$

Considerando a expansão de Taylor sobre $E(\mathbf{b}_i) = \mathbf{0}$, isto é, fazendo $\mathbf{b}_{i0} = \mathbf{0}$, obtemos

$$\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_0, \mathbf{0}) \cong \mathbf{U}_i[\boldsymbol{\alpha} - \boldsymbol{\alpha}_0] + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i. \quad (4.24)$$

Uma outra opção é considerar a expansão de Taylor sobre o valor estimado $\hat{\mathbf{b}}_i$.

Quando a variável resposta pode ser escrita através de uma função linear de $\boldsymbol{\alpha}$ e \mathbf{b}_i , é possível utilizar o algoritmo EM da mesma forma que em Laird e Ware (1982), bastando que sejam feitas algumas adaptações.

Considerando a metodologia na qual se baseia o algoritmo EM, e trabalhando com o modelo (4.21), a distribuição conjunta do sistema de dados completos é dada por

$$\begin{bmatrix} \mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_0, \mathbf{b}_{i0}) \\ \mathbf{b}_i \\ \mathbf{e}_i \end{bmatrix} \sim \mathbf{N} \left\{ \begin{bmatrix} \mathbf{U}_i(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + \mathbf{Z}_i\mathbf{b}_{i0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^t + \sigma^2\mathbf{I}_{p_i} & \mathbf{Z}_i\mathbf{D} & \sigma^2\mathbf{I}_{p_i} \\ \mathbf{D}\mathbf{Z}_i^t & \mathbf{D} & \mathbf{0} \\ \sigma^2\mathbf{I}_{p_i} & \mathbf{0} & \sigma^2\mathbf{I}_{p_i} \end{bmatrix} \right\}, \quad (4.25)$$

para $i=1, \dots, n$. Note que em Laird e Ware (1982), as matrizes \mathbf{U}_i e \mathbf{Z}_i eram fixas, não dependiam de $\boldsymbol{\alpha}$. Aqui elas dependem do valor inicial $\boldsymbol{\alpha}$, embora sejam tratadas como se fossem fixas durante o procedimento. Há ainda outras diferenças. Nesta adaptação, o EM fornece como resultado uma estimativa para a diferença entre $\boldsymbol{\alpha}$ e $\boldsymbol{\alpha}_0$, $\Delta\boldsymbol{\alpha}$. Além disso, no lugar de \mathbf{y}_i , estaremos trabalhando com a distribuição de $\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_0, \mathbf{b}_{i0})$. A função de verossimilhança da amostra será dada por

$$\begin{aligned}
L[\mathbf{y}|\mathbf{f}(\mathbf{X}, \alpha_0, \mathbf{b}_0), \mathbf{b}, \mathbf{e}, \alpha, \theta] &= \prod_{i=1}^n g(\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \alpha_0, \mathbf{b}_{i0}) | \mathbf{e}_i, \mathbf{b}_i; \alpha, \theta) g(\mathbf{b}_i; \alpha, \theta) g(\mathbf{e}_i; \alpha, \theta) \\
&= \prod_{i=1}^n g(\mathbf{b}_i; \alpha, \theta) g(\mathbf{e}_i; \alpha, \theta) = (2\pi)^{-n} (\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \mathbf{e}_i^t \mathbf{e}_i\right] |\mathbf{D}|^{-n/2} \exp\left[-\frac{1}{2} \text{tr} \mathbf{D}^{-1} \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^t\right],
\end{aligned} \tag{4.26}$$

onde $N = \sum_{i=1}^n p_i$; θ é o vetor de parâmetros que compõem a matriz de covariância; e $\mathbf{b} = (\mathbf{b}_1^t, \dots, \mathbf{b}_n^t)^t$.

A partir da função de verossimilhança é possível ver que as estatísticas suficientes para σ^2 e \mathbf{D} são $\mathbf{t}_1 = \sum_{i=1}^n \mathbf{e}_i^t \mathbf{e}_i$ e $\mathbf{T}_2 = \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^t$, respectivamente. No passo E são calculadas as esperanças condicionais de \mathbf{t}_1 e \mathbf{T}_2 dado \mathbf{y} , para α e θ fixados. O leitor interessado no cálculo das esperanças condicionais deve ver, por exemplo, Andreoni (1989):

$$\begin{aligned}
\mathbf{E}[\mathbf{t}_1 | \mathbf{y} - \mathbf{f}(\mathbf{X}, \alpha_0, \mathbf{b}_0); \alpha, \theta] &= \mathbf{E}\left[\sum_{i=1}^n \mathbf{e}_i^t \mathbf{e}_i | \mathbf{y} - \mathbf{f}(\mathbf{X}, \alpha_0, \mathbf{b}_0); \alpha, \theta\right] \\
&= \sum_{i=1}^n \{[\mathbf{I}_{p_i} - \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t \Sigma_i^{-1}] [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \alpha_0, \mathbf{b}_{i0}) - \mathbf{U}_i \Delta \alpha - \mathbf{Z}_i \mathbf{b}_{i0}]\} \\
&= \{[\mathbf{I}_{p_i} - \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t \Sigma_i^{-1}] [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \alpha_0, \mathbf{b}_{i0}) - \mathbf{U}_i \Delta \alpha - \mathbf{Z}_i \mathbf{b}_{i0}]\} + \sigma^2 \sum_{i=1}^n \text{tr}[\mathbf{I}_{p_i} - \sigma^2 \Sigma_i^{-1}]
\end{aligned} \tag{4.27}$$

$$\begin{aligned}
\mathbf{E}[\mathbf{T}_2 | \mathbf{y} - \mathbf{f}(\mathbf{X}, \alpha_0, \mathbf{b}_0); \alpha, \theta] &= \mathbf{E}\left[\sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^t | \mathbf{y} - \mathbf{f}(\mathbf{X}, \alpha_0, \mathbf{b}_0); \alpha, \theta\right] \\
&= \mathbf{nD} + \sum_{i=1}^n \{\mathbf{D} \mathbf{Z}_i^t \Sigma_i^{-1} [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \alpha_0, \mathbf{b}_{i0}) - \mathbf{U}_i (\alpha - \alpha_0) - \mathbf{Z}_i \mathbf{b}_{i0}]\} \\
&= [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \alpha_0, \mathbf{b}_{i0}) - \mathbf{U}_i (\alpha - \alpha_0) - \mathbf{Z}_i \mathbf{b}_{i0}]^t \Sigma_i^{-1} \mathbf{Z}_i \mathbf{D} - \mathbf{D} \left[\sum_{i=1}^n \mathbf{Z}_i^t \Sigma_i^{-1} \mathbf{Z}_i \right] \mathbf{D}.
\end{aligned} \tag{4.28}$$

Na iteração de ordem (u+1), fazemos

$$\mathbf{E}[\mathbf{t}_1 | \mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\alpha}_0, \mathbf{b}_0); \boldsymbol{\alpha} = \boldsymbol{\alpha}^{(u)}; \boldsymbol{\theta} = \boldsymbol{\theta}^{(u)}] \quad (4.29)$$

e

$$\mathbf{E}[\mathbf{T}_2 | \mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\alpha}_0, \mathbf{b}_0); \boldsymbol{\alpha} = \boldsymbol{\alpha}^{(u)}; \boldsymbol{\theta} = \boldsymbol{\theta}^{(u)}]. \quad (4.30)$$

No passo M, onde o vetor $\boldsymbol{\theta}^{(u+1)}$ é determinado como sendo a solução das equações de verossimilhança, obtém-se

$$\hat{\boldsymbol{\sigma}}^2 = \frac{\sum_{i=1}^n \mathbf{e}_i^t \mathbf{e}_i}{N} = \frac{\mathbf{t}_1}{N}, \quad (4.31)$$

$$\hat{\mathbf{D}} = \frac{\sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^t}{n} = \frac{\mathbf{T}_2}{n}, \quad (4.32)$$

e ainda,

$$\Delta \hat{\boldsymbol{\alpha}} = \left[\sum_{i=1}^n \mathbf{U}_i^t \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{U}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{U}_i^t \hat{\boldsymbol{\Sigma}}_i^{-1} [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_0, \mathbf{b}_{i0}) - \mathbf{Z}_i \mathbf{b}_{i0}] \right] \quad (4.33)$$

Antes de iniciar o EM, é necessário fixar um valor para $\boldsymbol{\alpha}_0$, que será a estimativa de MQO para $\boldsymbol{\alpha}$. Caso seja de interesse trabalhar com $\mathbf{b}_{i0} = \hat{\mathbf{b}}_i$, pode-se obter este valor de (4.15). A partir destas estimativas obtém-se $\mathbf{U}_i(\tilde{\boldsymbol{\alpha}})$ e $\mathbf{Z}_i(\tilde{\boldsymbol{\alpha}})$. Como estimativas iniciais para $\boldsymbol{\sigma}^2$ e \mathbf{D} , Vonesh e Carter (1992) sugerem (4.16) e (4.17). Feito isso, cada iteração de ordem $(u+1)$ do EM é dada por:

$$\begin{aligned} \boldsymbol{\sigma}^{2^{(u+1)}} = & \frac{1}{N} \sum_{i=1}^n [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_0, \mathbf{b}_{i0}) - \mathbf{U}_i \Delta \boldsymbol{\alpha}^{(u)} - \mathbf{Z}_i (\mathbf{b}_i^{(u)} - \mathbf{b}_{i0})]^t \\ & \times [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_0, \mathbf{b}_{i0}) - \mathbf{U}_i \Delta \boldsymbol{\alpha}^{(u)} - \mathbf{Z}_i (\mathbf{b}_i^{(u)} - \mathbf{b}_{i0})] + \frac{\boldsymbol{\sigma}^{2^*}}{N} \sum_{i=1}^n \text{tr}[\mathbf{I}_{p_i} - \boldsymbol{\sigma}^{2^*} \boldsymbol{\Sigma}^{-1(u)}], \end{aligned} \quad (4.34)$$

$$\mathbf{D}^{(u+1)} = \mathbf{D}^{(u)} + \frac{1}{n} \sum_{i=1}^n \mathbf{b}_i^{(u)} \mathbf{b}_i^{(u)t} - \frac{\mathbf{D}^{(u)}}{n} \left[\sum_{i=1}^n \mathbf{Z}_i^t \Sigma^{-1(u)} \mathbf{Z}_i \right] \mathbf{D}^{(u)}, \quad (4.35)$$

e

$$\Delta \alpha^{(u+1)} = \left[\sum_{i=1}^n \mathbf{U}_i^t \hat{\Sigma}_i^{-1(u)} \mathbf{U}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{U}_i^t \Sigma_i^{-1(u)} [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \alpha_0, \mathbf{b}_{i0}) - \mathbf{Z}_i \mathbf{b}_{i0}] \right], \quad (4.36)$$

onde $\mathbf{N} = \sum_{i=1}^n \mathbf{p}_i$, $\Sigma_i^{-1(u)} = [\mathbf{Z}_i \mathbf{D}^{(u)} \mathbf{Z}_i^t + \sigma^2(u) \mathbf{I}]^{-1}$ e $\mathbf{b}_i^{(u)} = \mathbf{D}^{(u)} \mathbf{Z}_i^t \Sigma_i^{-1(u)} [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \alpha_0, \mathbf{b}_{i0}) - \mathbf{U}_i \Delta \alpha^{(u)} - \mathbf{Z}_i \mathbf{b}_{i0}]$.

A convergência ocorre quando a diferença entre a verossimilhança da iteração anterior e a atual é muito pequena (segundo algum critério).

Quando empregado em modelos lineares, um único “loop” do algoritmo EM é suficiente para fornecer as estimativas de máxima verossimilhança para α . Nesta adaptação para modelos não-lineares, o EM está inserido em um “loop” maior, que será descrito a seguir. Na primeira vez, fixa-se $\mathbf{U}_i(\alpha_0)$, $\mathbf{Z}_i(\alpha_0)$ e $\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \alpha_0, \mathbf{b}_{i0})$, com $\alpha_0 = \tilde{\alpha}_{\text{MQO}}$. Ao final do EM (adaptado conforme descrito anteriormente), obtém-se $\Delta \hat{\alpha}$ e conseqüentemente, a primeira estimativa para α , que chamaremos $\hat{\alpha}_1$, $\hat{\alpha}_1 = \alpha_0 + \Delta \hat{\alpha}$. Em seguida, reinicia-se todo o processo, utilizando-se $\hat{\alpha}_1$ como valor inicial. Assim, este grande “loop” onde o EM está inserido, é repetido com “ α_0 ” assumindo os valores $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3 \dots$ até que haja convergência. O critério de convergência baseia-se também na comparação dos valores inicial e final da função de verossimilhança. No caso de se estar trabalhando com os valores estimados de \mathbf{b}_i ($\hat{\mathbf{b}}_i$) em lugar de $E(\mathbf{b}_i)$, mantém-se fixos os valores de \mathbf{b}_{i0} durante o EM, sendo atualizados no final, da mesma forma que os valores de α_0 .

Após a convergência do procedimento, as variâncias das estimativas são obtidas a partir da diagonal de

$$\mathbf{V}(\hat{\alpha}) = \left[\sum_{i=1}^n \mathbf{U}_i^t \Sigma_i^{-1} \mathbf{U}_i \right]^{-1}. \quad (4.37)$$

Caso \mathbf{Z}_i não dependa de α (modelo (4.9)) e assumindo que \mathbf{b}_i 's e \mathbf{e}_i 's têm distribuição Normal, as estimativas obtidas são de máxima verossimilhança. Para os modelos em que os \mathbf{b}_i 's

aparecem de forma não-linear, a distribuição dos y_i 's induzida pelo modelo torna-se bastante complicada. Neste caso, as estimativas obtidas são de quasi-verossimilhança, com propriedades similares às dos EMQG's discutidos anteriormente.

Para obter estimativas de máxima verossimilhança restrita, utiliza-se também a teoria desenvolvida para modelos de efeitos aleatórios lineares (ver, por exemplo, Andreoni, 1989) com as devidas adaptações. Neste caso, a iteração de ordem $(u+1)$ do algoritmo EM é dada por

$$\begin{aligned} \sigma^{2^{(u+1)}} = & \frac{1}{N} \sum_{i=1}^n [y_i - f(X_i, \alpha_0, \mathbf{b}_{i0}) - U_i \Delta \alpha^{(u)} - Z_i (\mathbf{b}_i^{(u)} - \mathbf{b}_{i0})]^t \times \\ & \times [y_i - f(X_i, \alpha_0, \mathbf{b}_{i0}) - U_i \Delta \alpha^{(u)} - Z_i (\mathbf{b}_i^{(u)} - \mathbf{b}_{i0})] + \frac{\sigma^{2^u}}{N} \sum_{i=1}^n \text{tr}[\mathbf{I}_{p_i} - \sigma^{2^u} \mathbf{P}^{-1(u)}], \end{aligned} \quad (4.38)$$

$$\mathbf{D}^{(u+1)} = \mathbf{D}^{(u)} + \frac{1}{n} \sum_{i=1}^n \mathbf{b}_i^{(u)} \mathbf{b}_i^{(u)t} - \frac{\mathbf{D}^{(u)}}{n} \left[\sum_{i=1}^n \mathbf{Z}_i^t \mathbf{P}^{-1(u)} \mathbf{Z}_i \right] \mathbf{D}^{(u)} \quad (4.39)$$

e

$$\Delta \alpha^{(u+1)} = \left[\sum_{i=1}^n \mathbf{U}_i^t \hat{\Sigma}_i^{-1(u)} \mathbf{U}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{U}_i^t \Sigma_i^{-1(u)} [y_i - f(X_i, \alpha_0, \mathbf{b}_{i0}) - \mathbf{Z}_i \mathbf{b}_{i0}] \right] \quad (4.40)$$

onde N , $\Sigma_i^{-1(u)}$ e $\mathbf{b}_i^{-1(u)}$ são definidos como em (4.36), e $\mathbf{P}_i^{(u)} = \Sigma_i^{-1(u)} - \Sigma_i^{-1(u)} \mathbf{U}_i \left[\sum_{i=1}^n \mathbf{U}_i^t \Sigma_i^{-1(u)} \mathbf{U}_i \right]^{-1} \mathbf{U}_i^t \Sigma_i^{-1(u)}$.

4.3.3. Seleção da estrutura de covariância em modelos de efeitos aleatórios

Vonesh (1992) discute alguns critérios encontrados na literatura para escolha da estrutura de covariância mais adequada, entre possíveis modelos com efeitos aleatórios. São medidas de "bondade do ajuste", em alguns casos utilizadas também para comparar modelos onde são introduzidos diferentes conjuntos de efeitos aleatórios.

Vonesh sugere a utilização de um índice chamado *coeficiente de correlação de concordância*, que é uma modificação do coeficiente de determinação, R^2 , usualmente empregado em Regressão Linear (ver, por exemplo, Draper e Smith, 1981).

O coeficiente de concordância entre duas variáveis y_1 e y_2 é definido por

$$\rho_c = 1 - \frac{E[(y_1 - y_2)^2]}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2}, \quad (4.41)$$

onde $y_1 \sim (\mu_1, \sigma_1)$, $y_2 \sim (\mu_2, \sigma_2)$, e $\text{corr}(y_1, y_2) = \rho$. Este índice pode ser estimado substituindo os parâmetros que o compõem pelos usuais estimadores: \bar{y}_1 e \bar{y}_2 (médias amostrais), s_1^2 e s_2^2 (variâncias amostrais), e $\hat{\rho}$ (coeficiente de correlação de Pearson). Além disso, fazendo $y_1 = y_i$, valores observados para a variável resposta, $y_2 = \hat{y}_i$, os valores preditos, \bar{y} e \hat{y} as respectivas médias, obtém-se:

$$\hat{\rho}_{c1} = 1 - \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{n^{-1} \sum_{i=1}^n (y_i - \bar{y})^2 + n^{-1} \sum_{i=1}^n (\hat{y}_i - \hat{y})^2 + (\bar{y} - \hat{y})^2} = 1 - \frac{SQ_{\text{erros}}}{SQ_{\text{total}} + \sum_{i=1}^n (\hat{y}_i - \hat{y})^2 + n(\bar{y} - \hat{y})^2}. \quad (4.42)$$

Conforme se pode observar em (4.42), a diferença para o R^2 está nos dois últimos termos do denominador. Segundo o autor, esta é uma medida mais direta da proximidade entre y_i e \hat{y}_i . Observe que, neste caso, as medidas repetidas por indivíduo estariam sendo tratadas como observações independentes, em um único vetor. Para comparar diferentes especificações de um modelo de efeitos aleatórios e levando em conta os vetores de observações por indivíduo, Vonesh sugeriu uma modificação de (4.42) onde, em lugar de se trabalhar com y_i e \hat{y}_i , verifica-se a proximidade entre os resíduos observados, $\tilde{\epsilon}_i = y_i - f(x_i, \tilde{\alpha}_{MQO})$ e os resíduos “preditos” $\hat{\epsilon}_i = Z_i(\tilde{\alpha})\hat{b}_i$, sendo \hat{b}_i dado por (4.15) (ver Seção 4.3.2). Neste caso, supõe-se que $Z(\tilde{\alpha})$ seja diferente de $1_{(px1)}$, caso contrário este coeficiente não se aplica (o resultado será sempre zero). O coeficiente de concordância para a amostra é uma estimativa “combinada” dos coeficientes individuais, dados por

$$\hat{\rho}_{c2,i} = 1 - \frac{p_i^{-1} \sum_{j=1}^{p_i} (\tilde{\epsilon}_{ij} - \hat{\epsilon}_{ij})^2}{p_i^{-1} \sum_{j=1}^{p_i} (\tilde{\epsilon}_{ij} - \tilde{\epsilon}_i)^2 + p_i^{-1} \sum_{i=1}^n (\hat{\epsilon}_{ij} - \hat{\epsilon}_i)^2 + (\tilde{\epsilon}_i - \hat{\epsilon}_i)^2}, \quad (4.43)$$

onde $\tilde{\epsilon}_i$ e $\hat{\epsilon}_i$ são as médias dos resíduos para o indivíduo i . A estimativa “combinada” é então dada por :

$$\hat{\rho}_{c2} = \frac{\sum_{i=1}^n \pi_i^2 \hat{\rho}_{c2,i}}{\sum_{i=1}^n \pi_i^2}, \quad (4.44)$$

onde π_i^2 é o denominador de (4.43). Pode-se ainda empregar (4.44) com y_i e \hat{y}_i em lugar dos resíduos, o que corresponderia a (4.42), mas levando em conta que se tratam de medidas repetidas por indivíduo.

4.3.4. Testes de Hipóteses

Após a obtenção das estimativas para α e Σ , os testes de hipóteses utilizados são os de Wald, da mesma forma descrita no **Capítulo III**.

4.3.5. Aspectos Computacionais

A implementação dos procedimentos de estimação descritos neste capítulo pode ser feita através de pacotes que disponham de subrotinas para ajuste de modelos não-lineares e sejam adequados à manipulação de matrizes, como por exemplo SAS[®], S-Plus[®], Minitab[®], etc.

Vonesh e Carter (1993) desenvolveram o programa MIXNLIN (Mixed NonLINEar regression) para o ajuste de modelos de efeitos aleatórios não-lineares. Este programa foi desenvolvido a partir dos recursos do pacote estatístico SAS, mais especificamente, utilizando os módulos SAS-MACRO e SAS-IML. O MIXNLIN permite estimar os parâmetros de modelos de efeitos aleatórios através dos métodos de mínimos quadrados ordinários, mínimos quadrados

generalizados, mínimos quadrados generalizados iterativos, máxima verossimilhança e máxima verossimilhança restrita. Nos métodos de máxima verossimilhança e máxima verossimilhança restrita, utilizam a adaptação do algoritmo EM descrita na **Seção 4.3.2**. O procedimento iterativo adotado para estimação por mínimos quadrados é o de Gauss-Newton (ver **Seção 3.1.1**). Além disso, os estimadores são os mesmos utilizados por Vonesh e Carter (1992), apresentados anteriormente. Este programa possibilita ainda a realização de alguns testes de hipóteses (Wald). No **Apêndice**, apresentamos os programas SAS utilizados nos exemplos para estimação por MQG (não-iterativos), com e sem a utilização do MIXNLIN.

Capítulo V. Aplicação Prática

Este capítulo tem como objetivo exemplificar a utilização de modelos não-lineares na análise de dados longitudinais, levando em conta diversas estruturas para a matriz de covariância das observações. Desta forma, são apresentados os resultados obtidos no ajuste de um modelo não-linear a um conjunto real de dados, assumindo-se diferentes estruturas de covariância e alternando-se os procedimentos de estimação descritos neste trabalho.

5.1. Descrição do conjunto de dados

Os dados são provenientes de um estudo conduzido na Escola de Educação Física da USP (Cordani, L.K. e Ribeiro, R. O., 1993). A variável observada foi a *quantidade de lactato de sódio* (em mmol/l) presente no sangue de atletas durante corrida em pista, medida ao longo de diversas *velocidades* (em km/h).

A quantidade de lactato é utilizada como indicadora de condição física e potencial de atletas, sendo portanto fundamental na prescrição de um treinamento adequado. Um ponto importante para a prescrição de treinamento de atletas de alto nível é o limiar anaeróbico, definido como a transição do metabolismo aeróbico para o anaeróbico. Durante o metabolismo anaeróbico ocorre uma reação que faz aumentar a produção de lactato de sódio, sendo que o aumento excessivo de lactato acarreta “stress” no atleta.

Fazem parte da amostra 14 corredores de fundo e meio fundo, e 12 triatletas. O objetivo da análise foi comparar o condicionamento físico dos dois grupos (que são os *tratamentos*), através da análise da concentração de lactato ao longo da velocidade. Nesta comparação, é de interesse verificar se o lactato de sódio obtido no limiar anaeróbico dos fundistas é equivalente ao dos triatletas.

Conforme trabalhos específicos da área de Educação Física (ver por exemplo, Pompeu (1994)), a concentração de lactato no sangue durante teste ergométrico ocorre de forma exponencial. No entanto não se tomou como objetivo principal investigar uma função que fosse fiel ao “fenômeno” que gera os dados ou seja, à relação funcional entre as variáveis, uma vez que esta etapa exigiria maior dedicação ao tema e envolvimento de

profissionais da área. Tomando como base principalmente a forma aparente da distribuição dos dados nas velocidades em que foram observados, empregamos neste estudo o modelo exponencial (2.3). Na Tabela 5.1 e nos gráficos a seguir apresentamos os dados que foram observados.

Tabela 5.1: Lactato de sódio (mmol/l) por velocidade (km/h) para *fundistas* e *triatletas*

Atleta	Trat.: <i>fund.</i> =0; <i>trial.</i> =1	Velocidade (em km/h)									
		9.5	10.9	12	13.3	14.4	15.7	17.1	18	19	20
1	0	-	2.5	2.5	2.7	3.2	4.1	4.2	7.3	13.3	-
2	0	-	1.9	2.2	2.4	2.8	3.5	3.7	4.9	5.3	7.1
3	0	-	1.9	2.2	2.3	2.5	2.9	3.1	3.4	4.2	5.3
4	0	1.7	1.7	1.8	1.7	1.8	2.1	2.9	3.7	4.7	6.6
5	0	1.6	1.7	2.1	2.4	3.0	3.5	5.0	5.8	6.5	-
6	0	1.9	1.8	1.8	2.0	2.1	4.1	7.9	-	-	-
7	0	1.6	1.8	2	2.2	3.0	3.5	4.5	5.6	-	-
8	0	1.4	2.0	2.2	2.4	2.8	3.6	4.5	5.8	7.9	9.2
9	0	1.2	1.8	2.9	3.2	4.2	5.9	-	-	-	-
10	0	-	1.4	1.8	1.8	2.2	2.8	-	6.9	8.6	-
11	0	-	1.3	1.4	1.7	2.3	2.3	4.0	4.4	5.3	-
12	0	1.4	1.8	1.8	1.8	1.9	2.2	3.4	4.0	7.1	8.7
13	0	1.7	1.8	1.8	1.9	2.0	2.3	4.0	7.5	9.9	-
14	0	1.7	1.8	2.1	2.2	2.4	2.6	2.9	3.2	3.6	5.1
15	1	1.3	1.2	1.2	1.5	1.7	2.7	-	8.8	-	-
16	1	-	2.3	1.8	2.4	1.6	3.4	5.2	6.8	-	-
17	1	1.3	1.1	1.1	1.4	1.7	2.1	3.2	-	-	-
18	1	1.3	1.7	1.2	1.7	2.2	2.4	7.3	-	-	-
19	1	1.6	1.3	1.2	1.1	1.3	1.3	1.9	2.8	3.3	6.6
20	1	1.3	1.4	1.5	1.3	1.4	1.7	2.3	3.7	6.0	-
21	1	1.2	0.8	1.1	0.9	0.9	1.2	1.8	6.2	7.6	-
22	1	-	1.1	1.1	1.0	1.3	1.9	2.3	6.7	-	-
23	1	1.7	1.0	1.2	1.7	1.6	2.5	3.6	6.2	-	-
24	1	1.0	-	1.3	1.9	2.2	3.0	-	-	-	-
25	1	1.4	1.0	0.9	1.1	2.4	1.9	2.5	2.9	4.5	-
26	1	0.9	0.8	1.1	2.0	1.9	2.1	4.9	6.0	-	-

Gráfico 5.1a: Lactato (mmol/l) por Velocidade (km/h) - Fundistas

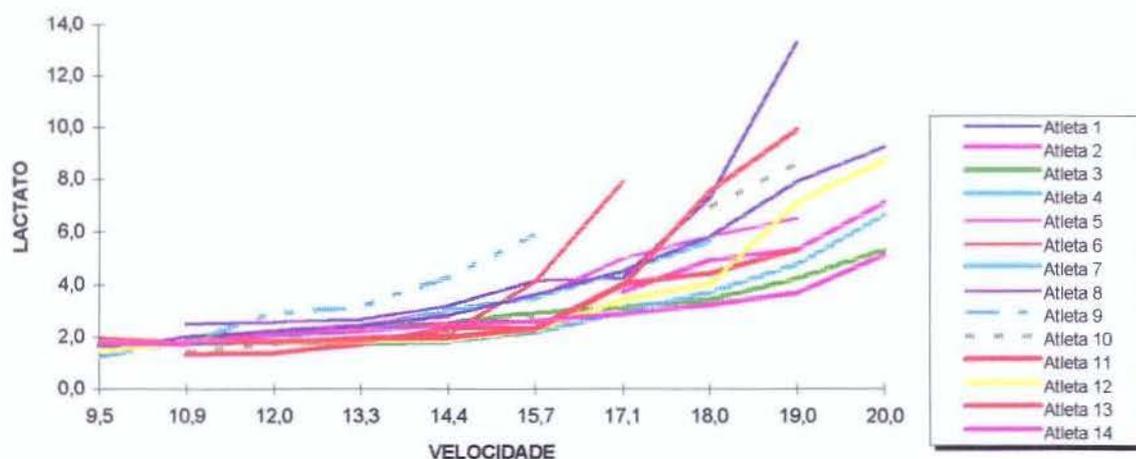
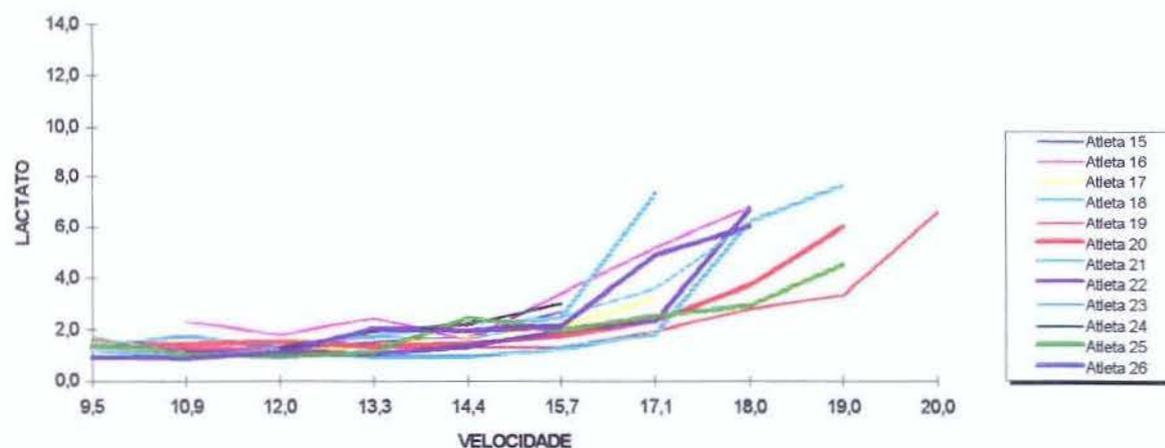


Gráfico 5.1b: Lactato (mmol/l) por Velocidade (km/h) - Triatletas



Algumas observações devem ser feitas com relação aos dados. Tanto as velocidades quanto os respectivos valores de lactato para os *fundistas* foram coletados considerando-se duas casas decimais. No entanto, foi feito o arredondamento para uma casa decimal de

forma que se pudesse ter exatamente os mesmos pontos fixados para a velocidade e tratar os dados como sendo de um experimento balanceado. Conforme visto no **Capítulo III**, na utilização de algumas estruturas de covariância esta suposição é uma condição necessária. Além disso, seguindo o mesmo procedimento adotado em trabalho anterior com este conjunto de dados (Cordani, L.K. e Ribeiro, R. O., 1993), o atleta nº 9 foi excluído da amostra. Seu desempenho era bastante superior aos demais, tendo os pesquisadores concluído que se tratava de um velocista.

5.2. Modelo para o vetor de médias e para a matriz de covariância

Como temos dois tratamentos (fundistas e triatletas), escrevemos o modelo (2.3) conforme descrito na **Seção 3.3.:**

$$E(y_{ij}) = \begin{cases} \exp[\alpha_{11}(x_j - \alpha_{21})], & \text{se } j = 1, \dots, 10; i = 1, \dots, 8, 10, \dots, 14 \\ \exp[\alpha_{12}(x_j - \alpha_{22})], & \text{se } j = 1, \dots, 10; i = 15, \dots, 26 \end{cases}$$

onde y_{ij} é a concentração de lactato de sódio medida em uma amostra de sangue do i -ésimo atleta, na j -ésima velocidade (x_j); $k=2$ e $r=4$, sendo $\alpha = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})^t$. Pode-se considerar o experimento balanceado com relação à condição de avaliação (sendo $p=10$) embora haja observações incompletas.

Com relação à interpretação dos parâmetros, α_{11} e α_{12} são constantes de proporcionalidade que estabelecem a relação entre a taxa de crescimento e a quantidade de lactato, para cada tratamento. Já α_{21} e α_{22} , além de interferirem também como constantes de proporcionalidade, definem o ponto onde o lactato é igual a 1 ($x = \alpha_{21}$ ou $x = \alpha_{22}$, conforme o tratamento). Segundo informações obtidas junto a pesquisadores, quando em repouso o lactato costuma ser aproximadamente igual a 1. De posse desta informação, procurou-se ajustar o modelo (2.3) supondo $\alpha_{21} = \alpha_{22} = 0$. No entanto, devido à não disponibilidade de dados em baixas velocidades, a imposição do ponto ($x=0, y=1$) acabou por pesar no ajuste do modelo, tornando-o menos adequado que o anteriormente adotado. Por esta razão,

optou-se por manter o segundo parâmetro no modelo, ainda que as estimativas possam ser incoerentes com a interpretação (por exemplo, valores de α_{21} ou α_{22} muito abaixo de 1).

O “ajuste” deste modelo foi repetido algumas vezes supondo-se, a cada vez, diferentes estruturas de covariância:

- I. Independência: estrutura (2.24), descrita anteriormente.
- II. Independência com variância proporcional à velocidade: $\sigma^2 \text{diag}(x_1, \dots, x_{10})$
- III. Uniforme: estrutura (2.25), descrita anteriormente.
- IV. “Uniforme”, com variância proporcional à velocidade: $\sigma_e^2 \mathbf{1}\mathbf{1}^t + \sigma_d^2 \text{diag}(x_1, \dots, x_{10})$.
- V. Efeitos aleatórios (2.27), com efeito não-linear em α_1 .
- VI. Efeitos aleatórios (2.27), com efeito não-linear em α_2 .
- VII. Efeitos aleatórios (2.27), com efeitos não-lineares em α_1 e α_2 .

Além disso, foram utilizados diferentes procedimentos de estimação: mínimos quadrados ordinários (MQO), mínimos quadrados generalizados (MQG), mínimos quadrados generalizados iterativos (MQGI) e máxima verossimilhança restrita (REML).

5.3. Análise dos resultados

Embora o objetivo principal deste trabalho não seja escolher a estrutura de covariância mais adequada para um conjunto de dados, foram feitas algumas considerações iniciais para melhor compreensão dos resultados obtidos nos ajustes supondo diferentes estruturas.

Devido ao tamanho reduzido da amostra, aliado ao número de observações incompletas, torna-se inviável o cálculo da matriz de covariância amostral na forma descrita na Seção 3.1, que auxiliaria na escolha de uma estrutura especial. Foram calculadas as variâncias e as correlações para os valores existentes em cada velocidade. Através destes resultados e dos Gráficos 5.1a e 5.1b, observa-se que as variâncias aumentam ao longo da velocidade:

Tabela 5.2: Estatísticas descritivas para a variável lactato, em cada velocidade

Velocidade	n	Média	Variância	Mínimo	Máximo
9.5	18	1.44	0.07	0.9	1.9
10.9	24	1.54	0.21	0.8	2.5
12	25	1.62	0.21	0.9	2.5
13.3	25	1.81	0.24	0.9	2.7
14.4	25	2.08	0.35	0.9	3.2
15.7	25	2.63	0.64	1.2	4.1
17.1	22	3.87	2.44	1.8	7.9
18	21	5.36	2.91	2.8	8.8
19	15	6.52	7.15	3.3	13.3
20	7	6.94	2.42	5.1	9.2

Tabela 5.3: Coeficientes de correlação de Pearson - () nº de obs.utilizadas no cálculo.

	VELOCIDADE									
	9.5	10.9	12	13.3	14.4	15.7	17.1	18	19	20
9.5	1	0.57	0.54	0.24	0.21	0.33	0.16	-0.20	-0.23	-0.94
10.9	(18)	(17)	(18)	(18)	(18)	(18)	(16)	(14)	(10)	(5)
12		(24)	(24)	(24)	(24)	(24)	(22)	(21)	(15)	(7)
13.3			(25)	(25)	(25)	(25)	(22)	(21)	(15)	(7)
14.4				(25)	(25)	(25)	(22)	(21)	(15)	(7)
15.7					(25)	(25)	(22)	(21)	(15)	(7)
17.1						(25)	(22)	(21)	(15)	(7)
18							(22)	(19)	(14)	(7)
19								(19)	(14)	(7)
20									(15)	(7)
										(7)

Conforme se pode observar, as variâncias crescem com a velocidade e, embora haja poucas observações, na última velocidade parece decrescer. As correlações não parecem obedecer a um padrão muito definido.

Apesar do reduzido tamanho de amostra, analisando-se as estatísticas descritivas para cada tratamento em separado (Tabelas 5.4 e 5.5), verifica-se que entre os *fundistas* é um pouco mais claro o comportamento das variâncias, que crescem com a média.

Tabela 5.4: Fundistas: estatísticas descritivas para a variável lactato, em cada velocidade

Velocidade	n	Média	Variância	Mínimo	Máximo
9.5	8	1.63	0.03	1.4	1.9
10.9	13	1.80	0.08	1.3	2.5
12	13	1.98	0.08	1.4	2.5
13.3	13	2.11	0.10	1.7	2.7
14.4	13	2.46	0.21	1.8	3.2
15.7	13	3.04	0.51	2.1	4.1
17.1	12	4.18	1.82	2.9	7.9
18	12	5.21	2.28	3.2	7.5
19	11	6.94	8.22	3.6	13.3
20	6	7.00	2.88	5.1	9.2

Tabela 5.5: Triatletas: estatísticas descritivas para a variável lactato, em cada velocidade

Velocidade	n	Média	Variância	Mínimo	Máximo
9.5	10	1.30	0.06	0.9	1.7
10.9	11	1.24	0.19	0.8	2.3
12	12	1.22	0.05	0.9	1.8
13.3	12	1.50	0.21	0.9	2.4
14.4	12	1.68	0.19	0.9	2.4
15.7	12	2.18	0.43	1.2	3.4
17.1	10	3.50	3.19	1.8	7.3
18	9	5.57	4.06	2.8	8.8
19	4	5.35	3.47	3.3	7.6
20	1	6.60	-	6.6	6.6

A primeira estrutura de covariância utilizada (Independência) supõe variâncias constantes e dados não correlacionados. Diante do comportamento dos dados, com variâncias que crescem com a velocidade, decidiu-se por utilizar também a estrutura II, mais flexível com relação a esta característica. No entanto, esta estrutura permanece sendo pouco indicada por não levar em conta as correlações.

A estrutura III (Uniforme), embora tenha sido utilizada a título de exemplificação, também não parece ser a mais adequada, uma vez que supõe variâncias e correlações constantes ao longo da velocidade. Foi então utilizada uma modificação (estrutura IV), com o objetivo de pelo menos torná-la mais adequada no que se refere às variâncias. Esta modificação, embora chamada de estrutura “Uniforme” com variância proporcional à velocidade, na realidade descaracterizou-se como estrutura uniforme, uma vez que além das variâncias, as correlações deixaram também de ser constantes ao longo da velocidade.

A estrutura de modelos de efeitos aleatórios parece ser a mais indicada ao problema por diversas razões. Em primeiro lugar não impõe que variâncias e correlações devam ser constantes ao longo da velocidade. Além disso, separa a variabilidade dentro e entre indivíduos, permitindo também que se tenha além da curva média, estimativas das curvas individuais.

Foram utilizadas todas as possíveis opções de inclusão de componente aleatório no modelo: acrescentando somente a α_1 (V), somente a α_2 (VI) e a ambos (VII), não esquecendo que a estrutura Uniforme foi também gerada a partir da utilização deste modelo. Neste caso, conforme visto na **Seção 4.3**, o componente aleatório não está ligado a um parâmetro especificamente e relaciona-se à variável resposta de forma linear.

Com relação ao procedimento de estimação, para as estruturas I e II utilizou-se MQO, através do procedimento NLIN, do SAS. Para as demais, utilizou-se mínimos quadrados generalizados iterativos (MQGI), tendo sido escolhida posteriormente uma única estrutura para exemplificar outros métodos (MQG e REML). Estes cálculos foram feitos através da macro MIXNLIN (Vonesh, 1993). Os programas utilizados encontram-se no **Apêndice**.

Na tabela a seguir são apresentados os resultados do ajuste com cada uma das estruturas, através de MQGI.

Tabela 5.6: Estimativas para os parâmetros e respectivos erros-padrão (ep)

Estruturas	$\hat{\alpha}_{11}$	ep($\hat{\alpha}_{11}$)	$\hat{\alpha}_{12}$	ep($\hat{\alpha}_{12}$)	$\hat{\alpha}_{21}$	ep($\hat{\alpha}_{21}$)	$\hat{\alpha}_{22}$	ep($\hat{\alpha}_{22}$)
I	0,1796	0,0124	0,2277	0,0210	8,8100	0,6255	11,3176	0,5700
II	0,1745	0,0110	0,2188	0,0186	8,5589	0,5689	11,0824	0,5201
III	0,1878	0,0138	0,2490	0,0247	9,2212	0,7439	11,8469	0,6732
IV	0,1798	0,0120	0,2330	0,0213	8,8443	0,6668	11,4806	0,6164
V	0,2101	0,0152	0,3058	0,0218	9,7642	0,3333	12,4299	0,2766
VI	0,1983	0,0089	0,2921	0,0169	9,4225	0,4851	12,2304	0,4350
VII	0,2105	0,0152	0,2928	0,0212	9,7789	0,3831	12,2612	0,3521

A estrutura V é a que apresenta menores erros para as estimativas de α_{21} e α_{22} . Com relação às estimativas para α_{11} e α_{12} , a estrutura VI parece apresentar menores erros. No entanto, considerando-se o significado dos parâmetros α_{21} e α_{22} no modelo como determinando o ponto onde o lactato é igual a 1 para cada um dos tratamentos, verifica-se através dos Gráficos 5.1a e 5.1b que na região de baixos valores da velocidade (onde o lactato aproxima-se de 1), a variância é pequena entre os indivíduos, não se justificando a inclusão de componente aleatório nestes parâmetros. Assim, entre as três estruturas de covariância com efeitos aleatórios, a mais indicada parece ser a V, que inclui efeito aleatório apenas no parâmetro que interfere na curvatura do modelo (α_{11} e α_{12}).

Outro critério de escolha entre as estruturas seria utilizar os indicadores sugeridos por Vonesh (1993) e apresentados na **Seção 4.3**: os *coeficientes de concordância* calculados a partir da comparação entre valores observados e valores preditos (ρ_{ei}) ou sobre os resíduos das estimativas de mínimos quadrados ordinários em comparação com os resíduos finais obtidos (ρ_{e2}). Conforme a Tabela 5.7, para ambos os indicadores, as estruturas V ou VII parecem ser as mais indicadas entre as testadas. Assim permanece a escolha da estrutura V, pela forma como se apresentam os dados (com variância maior apenas nos maiores valores de velocidade).

Tabela 5.7: Coeficientes de concordância para os modelos com diferentes estruturas

Estruturas	$\hat{\rho}_{e1}$	$\hat{\rho}_{e2}$
I	0,80	-
II	0,75	-
III	0,81	-
IV	0,75	-
V	0,82	0,70
VI	0,82	0,56
VII	0,82	0,75

Com o objetivo de exemplificar no mesmo modelo diferentes procedimentos de estimação, escolhemos a estrutura V e repetimos o ajuste utilizando MQG, REML e REML com expansão sobre valores estimados dos efeitos aleatórios (REML-SS). As estimativas dos parâmetros e respectivos erros são apresentados na Tabela 5.8..

Tabela 5.8: Estimativas para os parâmetros e respectivos erros-padrão (ep) com diferentes procedimentos iterativos - estrutura V

Proced.	$\hat{\alpha}_{11}$	$ep(\hat{\alpha}_{11})$	$\hat{\alpha}_{12}$	$ep(\hat{\alpha}_{12})$	$\hat{\alpha}_{21}$	$ep(\hat{\alpha}_{21})$	$\hat{\alpha}_{22}$	$ep(\hat{\alpha}_{22})$
MQGI	0,2101	0,0152	0,3058	0,0218	9,7642	0,3333	12,4299	0,2766
MQG	0,2112	0,0117	0,3021	0,0181	9,6431	0,3555	12,2755	0,2921
REML	0,2099	0,0148	0,3045	0,0216	9,7592	0,3366	12,4156	0,2800
REML-SS	0,2083	0,0148	0,2749	0,0204	9,8462	0,3273	11,9523	0,2992
ML	0,2097	0,0143	0,3031	0,0211	9,7533	0,3354	12,3991	0,2798

A seguir são apresentados alguns resultados do ajuste do modelo com a estrutura V, através do procedimento de estimação de MQGI. Em primeiro lugar, estimamos a matriz de covariância induzida no modelo, $\Sigma_1 = \mathbf{Z}_1(\alpha)\mathbf{D}\mathbf{Z}_1^t(\alpha) + \sigma^2\mathbf{I}_{p_1}$. Como só temos um efeito aleatório, a matriz \mathbf{D} tem dimensão 1x1, $\mathbf{Z}_1(\alpha)$ é um vetor coluna (10 x 1), obtido como em (4.9). Para cada tratamento, a matriz Σ_1 varia entre os indivíduos apenas com relação à dimensão. A seguir são apresentadas as matrizes estimadas para cada um dos tratamentos,

supondo um indivíduo com dados completos. Em ambas observa-se o crescimento das variâncias ao longo da velocidade, sendo este padrão mais claro entre os *fundistas*.

**Matriz de covariância (acima da diagonal) e correlação (abaixo da diagonal)
marginais estimadas para os *fundistas*, supondo dados completos**

$$\hat{\Sigma} = \begin{bmatrix} 0.5503 & -0.0007 & -0.0018 & -0.0037 & -0.0061 & -0.0103 & -0.0171 & -0.0231 & -0.032 & -0.0438 \\ -0.0013 & 0.5543 & 0.0103 & 0.0213 & 0.0353 & 0.0594 & 0.0984 & 0.1335 & 0.1847 & 0.2526 \\ -0.0032 & 0.0182 & 0.5756 & 0.0530 & 0.0875 & 0.1472 & 0.2441 & 0.3312 & 0.4582 & 0.6265 \\ -0.0061 & 0.0353 & 0.0859 & 0.6602 & 0.1818 & 0.3059 & 0.5074 & 0.6882 & 0.9522 & 1.3020 \\ -0.0089 & 0.0514 & 0.1250 & 0.2426 & 0.8505 & 0.5054 & 0.8382 & 1.1368 & 1.5729 & 2.1508 \\ -0.0117 & 0.0674 & 0.1640 & 0.3181 & 0.4631 & 1.4005 & 1.4102 & 1.9128 & 2.6465 & 3.6188 \\ -0.0135 & 0.0778 & 0.1893 & 0.3674 & 0.5347 & 0.7011 & 2.8890 & 3.1723 & 4.3892 & 6.0020 \\ -0.0142 & 0.0814 & 0.1981 & 0.3845 & 0.5596 & 0.7337 & 0.8472 & 4.8530 & 5.9533 & 8.1404 \\ -0.0146 & 0.0837 & 0.2037 & 0.3953 & 0.5754 & 0.7544 & 0.8711 & 0.9117 & 8.7872 & 11.2631 \\ -0.0148 & 0.0849 & 0.2068 & 0.4012 & 0.5839 & 0.7656 & 0.8841 & 0.9252 & 0.9513 & 15.9510 \end{bmatrix}$$

**Matriz de covariância (acima da diagonal) e correlação (abaixo da diagonal)
marginais estimadas para os *triatletas*, supondo dados completos**

$$\hat{\Sigma} = \begin{bmatrix} 0.5535 & 0.0026 & 0.0010 & -0.0029 & -0.0088 & -0.0210 & -0.0441 & -0.0675 & -0.105 \\ 0.0047 & 0.5522 & 0.0008 & -0.0022 & -0.0068 & -0.0161 & -0.0339 & -0.0519 & -0.0808 \\ 0.0018 & 0.0014 & 0.5505 & -0.0008 & -0.0026 & -0.0061 & -0.0129 & -0.0198 & -0.0308 \\ -0.0052 & -0.0040 & -0.0015 & 0.5526 & 0.0075 & 0.0178 & 0.0375 & 0.0573 & 0.0892 \\ -0.0157 & -0.0121 & -0.0046 & 0.0133 & 0.5732 & 0.0547 & 0.1151 & 0.1760 & 0.2738 \\ -0.0342 & -0.0263 & -0.0100 & 0.0291 & 0.0876 & 0.6803 & 0.2736 & 0.4187 & 0.6512 \\ -0.0554 & -0.0426 & -0.0163 & 0.0471 & 0.1419 & 0.3097 & 1.1476 & 0.8807 & 1.3698 \\ -0.0659 & -0.0507 & -0.0193 & 0.0560 & 0.1688 & 0.3685 & 0.5968 & 1.8975 & 2.0957 \\ -0.0723 & -0.0557 & -0.0212 & 0.0615 & 0.1853 & 0.4045 & 0.6551 & 0.7794 & 3.8098 \end{bmatrix}$$

Nos gráficos a seguir são apresentados os valores observados para os dois grupos de atletas com as respectivas curvas médias (Gráfico 5.2) e as curvas individuais (Gráfico 5.3a e 5.3b).

Gráfico 5.2.: Lactato (mmol/l) por Velocidade (km/h) - Fundistas e Triatletas
Valores observados e curvas médias - Estrutura V

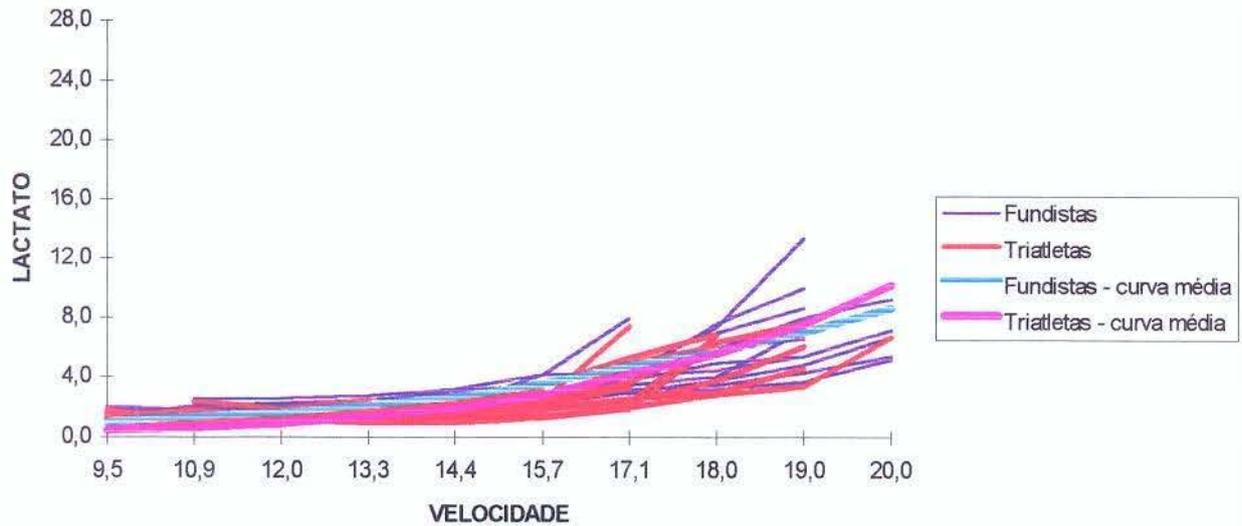


Gráfico 5.3a.: Lactato (mmol/l) por Velocidade (km/h)
Valores preditos individuais - FUNDISTAS

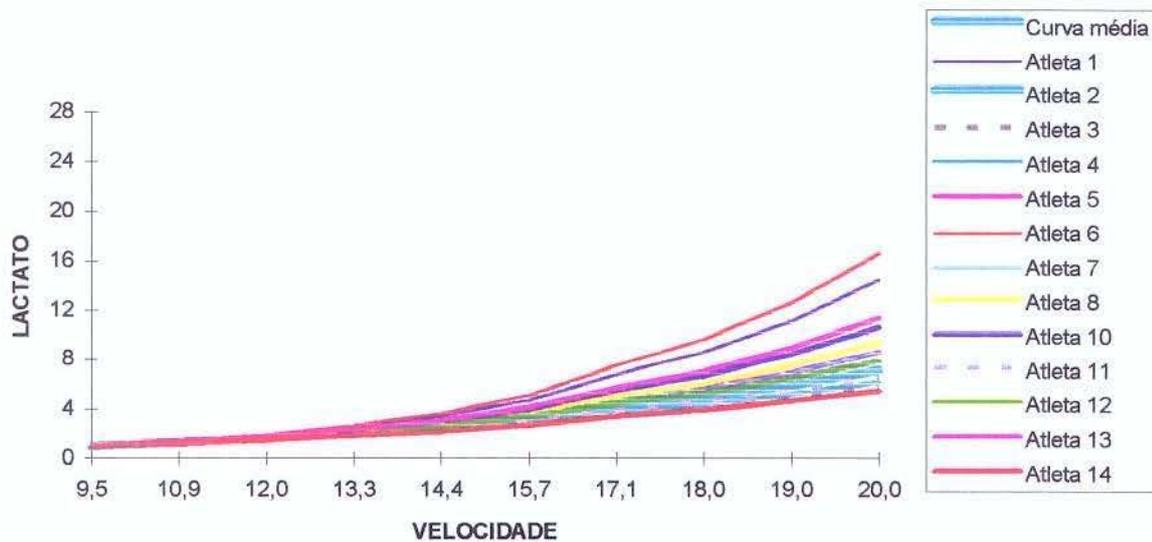
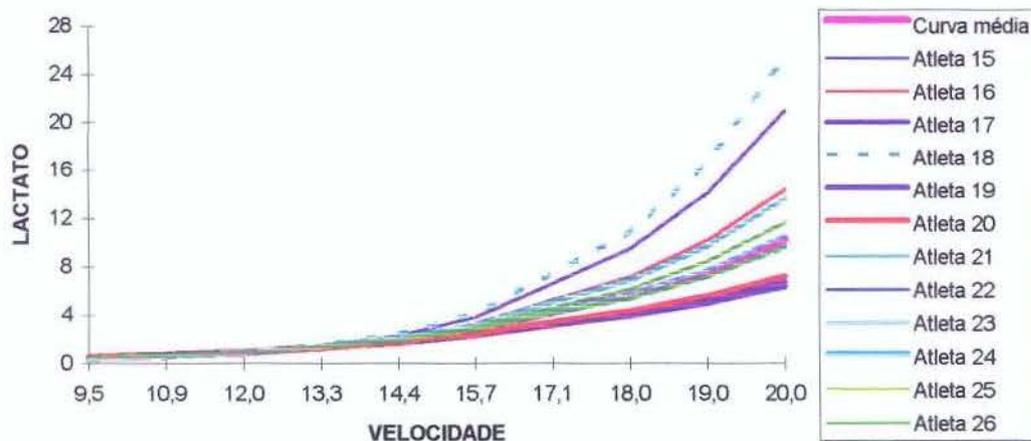


Gráfico 5.3b.: Lactato (mmol/l) por Velocidade (km/h)
Valores preditos individuais - TRIATLETAS



Os *fundistas* parecem apresentar maiores valores de lactato nas velocidades mais baixas. No entanto, a produção de lactato dos *triatletas* cresce a uma taxa superior, tendendo a superar os *fundistas*, embora haja poucas observações nos valores mais altos de velocidade. Este resultado pode ser observado também pela tabela a seguir, onde são apresentadas as taxas de crescimento estimadas para os dois grupos e percebe-se que as taxas dos triatletas superam os fundistas a partir da 6^a velocidade:

Tabela 5.8.: Taxas de crescimento de lactato para fundistas e triatletas

Velocidade	Fundistas		Triatletas	
	\hat{y}	taxa	\hat{y}	taxa
9.5	0.95	-	0.41	-
10.9	1.27	0.2311	0.63	0.1558
12	1.60	0.3000	0.88	0.2277
13.3	2.10	0.3864	1.30	0.3292
14.4	2.65	0.4968	1.83	0.4743
15.7	<u>3.48</u>	<u>0.6398</u>	<u>2.72</u>	<u>0.6859</u>
17.1	4.67	0.8085	4.17	1.0375
18	5.64	1.0800	5.49	1.4682
19	6.96	1.3191	7.46	1.9646
20	8.59	1.6274	10.12	2.6673

Podemos agora verificar se a diferença entre as taxas de crescimento de *fundistas* e *triatletas* é significativa, testando a hipótese de igualdade entre os parâmetros. Conforme os resultados apresentados no **Capítulo III**, para testar se as curvas coincidem, utilizamos por exemplo o teste modificado de Wald (3.34), onde a hipótese nula a ser testada é: $H_0: [I_2 \otimes (1, -1)]\alpha = 0$, $H_1: [I_2 \otimes (1, -1)]\alpha \neq 0$. Neste caso, $h(\hat{\alpha}) = [\hat{\alpha}_{11} - \hat{\alpha}_{12}, \hat{\alpha}_{21} - \hat{\alpha}_{22}] = [-0.0957$

$$-2.6658]; \mathbf{H}(\hat{\alpha}) = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, q=2 \text{ e a matriz de variância-covariância}$$

assintótica dos estimadores é dada por

$$\hat{\Omega} = \begin{bmatrix} 0.0002 & 0 & 0.0028 & 0 \\ & 0.0005 & 0 & 0.0045 \\ & & 0.1111 & 0 \\ & & & 0.0765 \end{bmatrix}$$

Obtemos então $W_2 = [h^1(\hat{\alpha}) (\hat{H} \hat{\Omega} \hat{H}^t)^{-1} h(\hat{\alpha})] / q = 19.02$. Como $P(F_{(2,21)} > 19.02) \approx 0.0001$, rejeita-se H_0 . Para verificar se a diferença entre os tratamentos ocorre para um ou para os dois parâmetros, testamos cada um em separado.

Testando se $\alpha_{11} = \alpha_{12}$, temos: $h(\hat{\alpha}) = -0.0957$; $\mathbf{H}(\hat{\alpha}) = [1 \ -1 \ 0 \ 0]$; neste caso, $q=1$ e $\hat{W}_2 = 12.95$. Sendo $P(F_{(1,21)} > 12.95) \approx 0.0017$, rejeita-se a hipótese de igualdade entre α_{11} e α_{12} . Testando se $\alpha_{21} = \alpha_{22}$ obtemos $h(\hat{\alpha}) = -2.6658$; $\mathbf{H}(\hat{\alpha}) = [0 \ 0 \ 1 \ -1]$; $q=1$ e $W_2 = 37.88$. Rejeita-se também a hipótese nula de igualdade entre α_{21} e α_{22} , uma vez que $P(F_{(1,21)} > 37.88) \approx 4.17E-06$.

Embora não se tenha determinado o ponto de limiar anaeróbico, verifica-se que há diferença entre os vetores de parâmetros dos dois grupos. Alguns pesquisadores (ver Pompeu, 1994) indicam que no limiar anaeróbico, a concentração de lactato é aproximadamente igual a 4 mmol/l. Assim, uma forma de determinar este ponto seria estimar o valor da velocidade para o qual $\hat{y} = 4$ em cada um dos tratamentos.

Capítulo VI. Considerações Finais

Neste trabalho, tratamos do ajuste de modelos não-lineares a dados longitudinais sob um enfoque específico. A abordagem foi paramétrica e em geral, consideramos que os conjuntos de dados poderiam ser representados através de uma parte determinística, uma função não-linear em relação aos parâmetros, e uma parte aleatória. Foram estudados basicamente quatro tipos de estrutura de covariância para as observações: independência (2.4), uniforme (2.5), efeitos aleatórios (2.27) e a forma geral (2.31).

Na análise de dados longitudinais através de modelos lineares, observa-se maior diversidade nas estruturas de covariância empregadas, sendo frequentes estudos onde se procuram estabelecer novas formas que representem mais adequadamente a relação de dependência entre os dados (ver por exemplo, Suyama, 1995). Quando os modelos são não-lineares, percebe-se que a maior ênfase é dada à busca da função não-linear que melhor descreva os dados. Assim, este trabalho têm como continuação natural a aplicação de outras estruturas de covariância em modelos não-lineares e a verificação das condições de regularidade correspondentes.

Ainda dentro da abordagem utilizada neste estudo, recentemente foram publicados dois livros que tratam especificamente do ajuste de modelos não-lineares para medidas repetidas (ver Davidian e Giltinan, 1995; Vonesh e Chinchilli, 1997 - em fase de publicação), mas que não chegaram a ser abordados neste trabalho.

Outras linhas de pesquisa complementares são a Bayesiana, tratada em Berkey (1982) e Racine-Poon (1985) ou Mazucheli (1995), por exemplo, e a abordagem não-paramétrica (Gasser et ali, 1984 ou Müller, 1989).

Apêndice : Programas

A seguir, apresentamos os programas SAS utilizados em cada um dos exemplos numéricos deste trabalho: o exemplo dos bois (**Exemplo 1**), o das laranjeiras (**Exemplo 2**) e o dos atletas, apresentado no **Capítulo V**. Em alguns programas utilizou-se, além dos recursos usuais oferecidos pelo pacote, a macro MIXNLIN (Vonesh, 1993), cujo programa foi gentilmente cedido pelo autor.

Exemplo 1: Peso dos bois

- Example1.dat: Arquivo onde estão armazenados os dados.

1	1	0	44	44	4	1	12	283	39	7	1	24	442	35
1	1	6	129	44	4	1	18	273	39	7	1	30	444	35
1	1	12	264	44	4	1	24	413	39	8	1	0	40	40
1	1	18	311	44	4	1	30	453	39	8	1	6	184	40
1	1	24	419	44	5	1	0	46	46	8	1	12	341	40
1	1	30	496	44	5	1	6	199	46	8	1	18	404	40
2	1	0	40	40	5	1	12	345	46	8	1	24	448	40
2	1	6	133	40	5	1	18	419	46	8	1	30	463	40
2	1	12	284	40	5	1	24	469	46	9	1	0	56	56
2	1	18	294	40	5	1	30	480	46	9	1	6	206	56
2	1	24	413	40	6	1	0	45	45	9	1	12	327	56
2	1	30	444	40	6	1	6	204	45	9	1	18	403	56
3	1	0	42	42	6	1	12	333	45	9	1	24	528	56
3	1	6	249	42	6	1	18	403	45	9	1	30	498	56
3	1	12	243	42	6	1	24	471	45	10	1	0	48	48
3	1	18	354	42	6	1	30	483	45	10	1	6	201	48
3	1	24	404	42	7	1	0	35	35	10	1	12	340	48
3	1	30	544	42	7	1	6	171	35	10	1	18	420	48
4	1	0	39	39	7	1	12	319	35	10	1	24	480	48
4	1	6	143	39	7	1	18	377	35	10	1	30	449	48

11	0	0	30	30	14	0	12	286	47	17	0	24	396	40
11	0	6	135	30	14	0	18	346	47	17	0	30	446	40
11	0	12	248	30	14	0	24	456	47	18	0	0	35	35
11	0	18	278	30	14	0	30	486	47	18	0	6	141	35
11	0	24	438	30	15	0	0	40	40	18	0	12	245	35
11	0	30	418	30	15	0	6	164	40	18	0	18	285	35
12	0	0	36	36	15	0	12	297	40	18	0	24	435	35
12	0	6	103	36	15	0	18	302	40	18	0	30	534	35
12	0	12	222	36	15	0	24	412	40	19	0	0	40	40
12	0	18	252	36	15	0	30	472	40	19	0	6	131	40
12	0	24	382	36	16	0	0	50	50	19	0	12	291	40
12	0	30	382	36	16	0	6	187	50	19	0	18	376	40
13	0	0	53	53	16	0	12	292	50	19	0	24	466	40
13	0	6	134	53	16	0	18	297	50	19	0	30	453	40
13	0	12	250	53	16	0	24	387	50	20	0	0	40	40
13	0	18	305	53	16	0	30	437	50	20	0	6	197	40
13	0	24	465	53	17	0	0	40	40	20	0	12	320	40
13	0	30	485	53	17	0	6	168	40	20	0	18	436	40
14	0	0	47	47	17	0	12	276	40	20	0	24	559	40
14	0	6	171	47	17	0	18	296	40	20	0	30	580	40

- **Boimqo1.sas:** Programa para estimar os parâmetros através do método de mínimos quadrados ordinários, supondo um único tratamento. Utiliza somente a procedure NLIN - ver **Capítulo III**.

```
options ps=60;
data um;
infile "c:\tese\Example1.dat";
input boi $ trat idade pesoi pzero ;
proc sort data=um;
```

```

by trat boi;
title ' EXEMPLO 1: PESO DE BOIS- USANDO MQO - CONSIDERANDO 1
TRATAMENTO';
data um;
  set um;
  peso = pesoi/pzero ;
  if idade=0 then delete;
proc nlin data=um iter=50;
  parms b1=10 b2=8 b3=.05 ;
  model PESO = b1*exp(-exp(-b3*(idade-b2)));
  der.b1=exp(-exp(-b3*(idade-b2)));
  der.b2=-b1*b3*exp(-(b3*(idade-b2) + exp(-b3*(idade-b2))));
  der.b3=b1*(idade-b2)*exp(-(b3*(idade-b2) + exp(-b3*(idade-b2))));
  id boi;
  output out = saida1 residual=e1 predicted=pred;
proc print data =saida1;
proc plot data=saida1;
  plot pred*idade='*' peso*idade/ overlay;
run;

```

- **Boimqg1.sas**: Programa para estimar os parâmetros através do método de mínimos quadrados generalizados, supondo um único tratamento. Deve ser rodado após o programa anterior, uma vez que utiliza o arquivo SAIDA1, onde estão armazenados os valores da variável resposta (Y), da variável explicativa (T) e dos resíduos (E). Matriz de covariância segue a estrutura geral (2.31) - ver **Capítulo III**.

```

OPTIONS PS=60;
PROC IML ;
  USE SAIDA1;
  READ ALL INTO M;
  T=M[,1];

```

```

Y=M[,4];
E=M[,6];
/**CALCULO DA MATRIZ DE COVARIANCIA COM ESTRUTURA GERAL**/;
/* N=NUMERO TOTAL DE INDIVÍDUOS */;
N=20;
/* p=número de ocasiões (tempos) */;
p=5;
Sigma= (E[1:5,]*'(E[1:5,]') + E[6:10,]*'(E[6:10,]') + E[11:15,]*'(E[11:15,]') +
E[16:20,]*'(E[16:20,]') + E[21:25,]*'(E[21:25,]') + E[26:30,]*'(E[26:30,]') +
E[31:35,]*'(E[31:35,]') + E[36:40,]*'(E[36:40,]') + E[41:45,]*'(E[41:45,]') +
E[46:50,]*'(E[46:50,]') + E[51:55,]*'(E[51:55,]') + E[56:60,]*'(E[56:60,]') +
E[61:65,]*'(E[61:65,]') + E[66:70,]*'(E[66:70,]') + E[71:75,]*'(E[71:75,]') +
E[76:80,]*'(E[76:80,]') + E[81:85,]*'(E[81:85,]') + E[86:90,]*'(E[86:90,]') +
E[91:95,]*'(E[91:95,]') + E[96:100,]*'(E[96:100,]'))/N;
PRINT sigma;

/* CALCULO DA DECOMPOSICAO DE CHOLESKY */;
WI = INV(SIGMA);
PRINT WI;
P=ROOT(WI);
PRINT P;
Z1=REPEAT(0,25,75);
Z2=REPEAT(0,75,25);
DIAPS=BLOCK(P,P,P,P,P,P,P,P,P,P,P,P,P,P,P);
DIAPI=BLOCK(P,P,P,P,P);
DIAP1=DIAPS//Z1;
DIAP2=Z2//DIAPI;
DIAP=DIAP1||DIAP2;
YT=DIAP*Y;
PT=P//P//P//P//P//P//P//P//P//P//P//P//P//P//P//P//P;
DADO=YT||PT;

```

```

CREATE DOIS FROM DADO;
APPEND FROM DADO;
CLOSE DOIS;
QUIT;
DATA DOIS ;
SET DOIS;
RENAME COL1=PESO COL2=P1 COL3=P2 COL4=P3 COL5=P4 COL6=P5 ;
T1=6; T2=12; T3=18; T4=24; T5=30;
TITLE 'EXEMPLO DOS BOIS USANDO MQG COM MATRIZ ESTRUTURADA' ;
PROC NLIN ITER=50 ;
PARMS B1=10 B2=7 B3=0.10 ;
model PESO = P1*b1*exp(-exp(-b3*(T1-b2))) + P2*b1*exp(-exp(-b3*(T2-b2))) +
P3*b1*exp(-exp(-b3*(T3-b2))) + P4*b1*exp(-exp(-b3*(T4-b2))) +
P5*b1*exp(-exp(-b3*(T5-b2)));
der.b1= P1*exp(-exp(-b3*(T1-b2))) + P2*exp(-exp(-b3*(T2-b2))) +
P3*exp(-exp(-b3*(T3-b2))) + P4*exp(-exp(-b3*(T4-b2))) +
P5*exp(-exp(-b3*(T5-b2))) ;

der.b2= -P1*b1*b3*exp(-(b3*(T1-b2) + exp(-b3*(T1-b2))))
-P2*b1*b3*exp(-(b3*(T2-b2) + exp(-b3*(T2-b2))))
-P3*b1*b3*exp(-(b3*(T3-b2) + exp(-b3*(T3-b2))))
-P4*b1*b3*exp(-(b3*(T4-b2) + exp(-b3*(T4-b2))))
-P5*b1*b3*exp(-(b3*(T5-b2) + exp(-b3*(T5-b2)))) ;

der.b3= P1*b1*(T1-b2)*exp(-(b3*(T1-b2) + exp(-b3*(T1-b2)))) +
P2*b1*(T2-b2)*exp(-(b3*(T2-b2) + exp(-b3*(T2-b2)))) +
P3*b1*(T3-b2)*exp(-(b3*(T3-b2) + exp(-b3*(T3-b2)))) +
P4*b1*(T4-b2)*exp(-(b3*(T4-b2) + exp(-b3*(T4-b2)))) +
P5*b1*(T5-b2)*exp(-(b3*(T5-b2) + exp(-b3*(T5-b2)))) ;
output out = saida3 p=PRED3 r=e3 PARMS=B1 B2 B3;
proc print data saida3;

```

```
proc plot data=saida3;
  plot e3*pred3;
run;
```

- **Boimqo2.sas:** Programa para estimar os parâmetros através do método de mínimos quadrados ordinários, supondo dois tratamentos. Utiliza somente a procedure NLIN - ver **Capítulo III**.

```
options ps=60;
data um;
infile "c:\tese\Example1.dat";
  input boi trat idade pesoi pzero ;
proc sort data=um;
  by trat boi;
data um;
  set um;
  peso=pesoi/pzero;
  if idade=0 then delete;
title 'EXEMPLO : PESO DE BOIS- USANDO MQO - considerando 2 tratamentos';
proc nlin data=UM iter=50;
  parms b11=10 b12=10 b21=8 b22=8 b31=.05 b32=.05 ;
  model PESO =(b11*(1-trat) + b12*trat)*exp(-exp(-(b31*(1-trat) +
    b32*trat)*(idade-(b21*(1-trat) + b22*trat))));
  der.b11=(1-trat)*exp(-exp(-b31*(idade-b21)));
  der.b12=trat*exp(-exp(-b32*(idade-b22)));
  der.b21=-(1-trat)*b11*b31*exp(-(b31*(idade-b21) + exp(-b31*(idade-b21))));
  der.b22=-trat*b12*b32*exp(-(b32*(idade-b22) + exp(-b32*(idade-b22))));
  der.b31=(1-trat)*b11*(idade-b21)*exp(-(b31*(idade-b21) + exp(-b31*(idade-b21))));
  der.b32=trat*b12*(idade-b22)*exp(-(b32*(idade-b22) + exp(-b32*(idade-b22))));
  id boi;
output out = sai1 p=p1 r=r1 ;
```

```
proc print;
run;
```

- Boimqg2.sas: Programa para estimar os parâmetros através do método de mínimos quadrados generalizados, supondo dois tratamentos. Deve ser rodado após o programa anterior, uma vez que utiliza o arquivo SAI1, onde estão armazenados os valores da variável resposta (Y), da variável explicativa (T) e dos resíduos (E). Matriz de covariância segue a estrutura geral (2.31) - ver **Capítulo III**.

```
OPTIONS PS=60;
PROC IML ;
  USE SAI1;
  READ ALL INTO M;
  TRAT=M[,2];
  T=M[,3];
  Y=M[,6];
  E=M[,8];
  PRINT TRAT T Y E ;
  /**CALCULO DA MATRIZ DE COVARIANCIA COM ESTRUTURA GERAL**/;

  /* N=NUMERO TOTAL DE INDIVÍDUOS */;
  N=20;
  /* p=número de ocasioes (tempos) */;
  p=5;
  Sigma= (E[1:5,]*'(E[1:5,]') + E[6:10,]*'(E[6:10,]') + E[11:15,]*'(E[11:15,]') +
    E[16:20,]*'(E[16:20,]') + E[21:25,]*'(E[21:25,]') + E[26:30,]*'(E[26:30,]') +
    E[31:35,]*'(E[31:35,]') + E[36:40,]*'(E[36:40,]') + E[41:45,]*'(E[41:45,]') +
    E[46:50,]*'(E[46:50,]') + E[51:55,]*'(E[51:55,]') + E[56:60,]*'(E[56:60,]') +
    E[61:65,]*'(E[61:65,]') + E[66:70,]*'(E[66:70,]') + E[71:75,]*'(E[71:75,]') +
    E[76:80,]*'(E[76:80,]') + E[81:85,]*'(E[81:85,]') + E[86:90,]*'(E[86:90,]') +
    E[91:95,]*'(E[91:95,]') + E[96:100,]*'(E[96:100,]'))/N;
```

```

PRINT sigma;

/* CALCULO DA DECOMPOSICAO DE CHOLESKY */;

WI = INV(SIGMA);
PRINT WI;
P=ROOT(WI);
PRINT P;
Z1=REPEAT(0,25,75) ;
Z2=REPEAT(0,75,25) ;
DIAPS=BLOCK(P,P,P,P,P,P,P,P,P,P,P,P,P);
DIAPI=BLOCK(P,P,P,P,P);
DIAP1=DIAPS//Z1;
DIAP2=Z2//DIAPI;
DIAP=DIAP1||DIAP2;
YT=DIAP*Y;
PT=P//P//P//P//P//P//P//P//P//P//P//P//P//P//P//P//P;
DADO=YT||PT||TRAT;
CREATE DOIS FROM DADO;
APPEND FROM DADO;
CLOSE DOIS;
QUIT;

DATA DOIS ;
SET DOIS;
    RENAME COL1=PESO COL2=P1 COL3=P2 COL4=P3 COL5=P4 COL6=P5
COL7=TRAT;
T1=6; T2=12; T3=18; T4=24; T5=30;
TITLE 'EXEMPLO DOS BOIS USANDO MQG COM MATRIZ ESTRUTURADA' ;
PROC NLIN ITER=50 ;

```

PARMS B11=10 B12=10 B21=7 B22=7 B31=0.10 B32=.10 ;

model PESO = P1*(b11*(1-TRAT) + B12*TRAT)*exp(-exp(-(b31*(1-TRAT) + B32*TRAT)*(T1- (b21*(1-TRAT) + B22*TRAT)))) +
P2*(b11*(1-TRAT) + B12*TRAT)*exp(-exp(-(b31*(1-TRAT) + B32*TRAT)*(T2- (b21*(1-TRAT) + B22*TRAT)))) +
P3*(b11*(1-TRAT) + B12*TRAT)*exp(-exp(-(b31*(1-TRAT) + B32*TRAT)*(T3- (b21*(1-TRAT) + B22*TRAT)))) +
P4*(b11*(1-TRAT) + B12*TRAT)*exp(-exp(-(b31*(1-TRAT) + B32*TRAT)*(T4-(b21*(1-TRAT) + B22*TRAT)))) +
P5*(b11*(1-TRAT) + B12*TRAT)*exp(-exp(-(b31*(1-TRAT) + B32*TRAT)*(T5- (b21*(1-TRAT) + B22*TRAT)))) ;

der.b11=(1-TRAT)*(P1*exp(-exp(-b31*(T1-b21))) + P2*exp(-exp(-b31*(T2-b21))) +
P3*exp(-exp(-b31*(T3-b21))) + P4*exp(-exp(-b31*(T4-b21))) +
P5*exp(-exp(-b31*(T5-b21)))) ;

der.b12=TRAT*(P1*exp(-exp(-b32*(T1-b22))) + P2*exp(-exp(-b32*(T2-b22))) +
P3*exp(-exp(-b32*(T3-b22))) + P4*exp(-exp(-b32*(T4-b22))) +
P5*exp(-exp(-b32*(T5-b22)))) ;

der.b21=(1-TRAT)*(-P1*b11*b31*exp(-(b31*(T1-b21) + exp(-b31*(T1-b21))))
-P2*b11*b31*exp(-(b31*(T2-b21) + exp(-b31*(T2-b21))))
-P3*b11*b31*exp(-(b31*(T3-b21) + exp(-b31*(T3-b21))))
-P4*b11*b31*exp(-(b31*(T4-b21) + exp(-b31*(T4-b21))))
-P5*b11*b31*exp(-(b31*(T5-b21) + exp(-b31*(T5-b21)))) ;

der.b22= TRAT*(-P1*b12*b32*exp(-(b32*(T1-b22) + exp(-b32*(T1-b22))))
-P2*b12*b32*exp(-(b32*(T2-b22) + exp(-b32*(T2-b22))))
-P3*b12*b32*exp(-(b32*(T3-b22) + exp(-b32*(T3-b22))))
-P4*b12*b32*exp(-(b32*(T4-b22) + exp(-b32*(T4-b22))))
-P5*b12*b32*exp(-(b32*(T5-b22) + exp(-b32*(T5-b22)))) ;

der.b31=(1-TRAT)*(P1*b11*(T1-b21)*exp(-(b31*(T1-b21) + exp(-b31*(T1-b21)))) +
P2*b11*(T2-b21)*exp(-(b31*(T2-b21) + exp(-b31*(T2-b21)))) +
P3*b11*(T3-b21)*exp(-(b31*(T3-b21) + exp(-b31*(T3-b21)))) +

```

P4*b11*(T4-b21)*exp(-(b31*(T4-b21) + exp(-b31*(T4-b21)))) +
P5*b11*(T5-b21)*exp(-(b31*(T5-b21) + exp(-b31*(T5-b21)))));
der.b32=TRAT*(P1*b12*(T1-b22)*exp(-(b32*(T1-b22) + exp(-b32*(T1-b22)))) +
P2*b12*(T2-b22)*exp(-(b32*(T2-b22) + exp(-b32*(T2-b22)))) +
P3*b12*(T3-b22)*exp(-(b32*(T3-b22) + exp(-b32*(T3-b22)))) +
P4*b12*(T4-b22)*exp(-(b32*(T4-b22) + exp(-b32*(T4-b22)))) +
P5*b12*(T5-b22)*exp(-(b32*(T5-b22) + exp(-b32*(T5-b22)))));
output out = saida3 p=pRED3 r=e3 PArms=b11 b12 b21 b22 b31 b32 ;

```

DATA SAIDA;

```
set saida3;
```

```
proc print;
```

```
run;
```

Exemplo 2: Circunferência das laranjeiras

- Example2.dat: Programa onde estão armazenados os dados.

1 118 30	2 118 33	3 118 30	4 118 32	5 118 30
1 484 58	2 484 69	3 484 51	4 484 62	5 484 49
1 664 87	2 664 111	3 664 75	4 664 112	5 664 81
1 1004 115	2 1004 156	3 1004 108	4 1004 167	5 1004 125
1 1231 120	2 1231 172	3 1231 115	4 1231 179	5 1231 142
1 1372 142	2 1372 203	3 1372 139	4 1372 209	5 1372 174
1 1582 145	2 1582 203	3 1582 140	4 1582 214	5 1582 177

- Arvmqo.sas: Programa para estimar os parâmetros através do método de mínimos quadrados ordinários. Utiliza somente a procedure NLIN - ver **Seção 4.1**.

```

OPTIONS PS=60;
DATA UM;
INFILE "c:\tese\example2.dat";
INPUT IND DAYS Y;
PROC NLIN DATA=UM METHOD=GAUSS ITER=50;
  PARMS ALFA=150 BETA=5 GAMA=0.002 ;
  D=BETA*EXP(-GAMA*DAYS);
  MODEL Y=ALFA/(1 + D);
  DER.ALFA=1/(1 + D);
  DER.BETA=-(ALFA*EXP(-GAMA*DAYS))/((1 + D)**2);
  DER.GAMA=(ALFA*DAYS*D)/((1 + D)**2);
  OUTPUT OUT=SAIDA1 RESIDUAL=E;
PROC PRINT;
RUN;

```

- Arvmqg.sas: Programa para estimar os parâmetros através do método de mínimos quadrados generalizados. Deve ser rodado após o programa anterior, uma vez que utiliza o arquivo SAIDA1, onde estão armazenados os valores da variável resposta (Y), da variável explicativa (DAYS) e dos resíduos (E). Matriz de covariância segue a estrutura de modelos de efeitos aleatórios (2.27). Neste caso, foi considerado somente um efeito aleatório, ligado ao parâmetro α_1 - ver **Seção 4.3**.

```

PROC IML ;
  USE SAIDA1;
  READ ALL INTO M;
  DAYS=M[,1];
  Y=M[,2];
  E=M[,3];

```

```

/** PRIMEIRO ESTAGIO: ENCONTRAR ESTIMATIVAS POR MQO PARA OS
PARAM.*;/
    ALFA=192.6876;
    BETA=7.8566;
    GAMA=0.0028;

/*N=numero de individuos*/;
    N=5;
/*RI=numero de condicoes de avaliacao*/;
    RI=7;
/*K2=numero de between-variables*/;
    K2=1;
/*NI=numero de coeficientes aleatorios*/;
    NI=1;
/*****CALCULO DA MATRIZ Z*****/;
/**no caso da matriz de covariancia seguir o padrao uniforme, Z=UNIRI.*;/
/**no caso do modelo utilizado nao ser o pseudo-linear (utiliza apro-*/;
/**ximacao de Taylor), Z={UNIRI, DAYS[1:RI,], DAYS[1:RI]##2}.*****/;
/**no caso do modelo pseudo-linear Z e calculada como se segue.*****/;
    UNIRI=VECDIAG(I(RI));
    D=BETA#EXP(-GAMA#DAYS[1:RI,]);
    DALFA=UNIRI/(UNIRI + D);
    DBETA=- (ALFA#D)/BETA#((UNIRI + D)##2);
    DGAMA=(ALFA#DAYS[1:RI,]#D)/((UNIRI + D)##2);
    Z=DALFA||DBETA||DGAMA;
    A=I(3);
    A=A[,1:NI];
    ZI=Z*A;
    PRINT ZI;

```

```
/* SEGUNDO ESTAGIO: ESTIMACAO DOS EFEITOS ALEATORIOS E DAS
COMPONENTES*/
```

```
/* SI DA VARIANCIA (ENTRE) */
```

```
B1=INV(ZI*ZI)*ZI*E[1:7,];
B2=INV(ZI*ZI)*ZI*E[8:14,];
B3=INV(ZI*ZI)*ZI*E[15:21,];
B4=INV(ZI*ZI)*ZI*E[22:28,];
B5=INV(ZI*ZI)*ZI*E[29:35,];
S1=E[1:7,]*(I(7) - ZI*INV(ZI*ZI)*ZI)*E[1:7,]/6;
S2=E[8:14,]*(I(7) - ZI*INV(ZI*ZI)*ZI)*E[8:14,]/6;
S3=E[15:21,]*(I(7) - ZI*INV(ZI*ZI)*ZI)*E[15:21,]/6;
S4=E[22:28,]*(I(7) - ZI*INV(ZI*ZI)*ZI)*E[22:28,]/6;
S5=E[29:35,]*(I(7) - ZI*INV(ZI*ZI)*ZI)*E[29:35,]/6;
```

```
/* TERCEIRO ESTAGIO: ESTIMACAO DE SIGMA2 E PSI */
```

```
SIGMA2=((RI-NI)/(N*(RI-NI)))*(S1+S2+S3+S4+S5);
PRINT SIGMA2;
B=B1||B2||B3||B4||B5;
B=B';
PRINT B;
X2I=1;
X2={1 1 1 1 1}';
SBB=B'*(I(N)-X2*INV(X2'*X2)*X2')*B/(N-K2);
PSI=SBB - SIGMA2*N*(1-X2I*INV(X2'*X2)*X2I)*INV(ZI*ZI)/(N-K2);
PRINT SBB;
PRINT PSI;
```

```
/* QUARTO ESTAGIO: ESTIMACAO DE ALFA ATRAVES DE MQG */
```

```

WI=INV(ZI*PSI*ZI + SIGMA2*I(RI));
PRINT WI;
P=ROOT(WI);
PRINT P;
DIAP=BLOCK(P,P,P,P,P);
YT=DIAP*Y;
PRINT YT;
P=P//P//P//P//P;
DADO=YT||T||P ;
CREATE DOIS FROM DADO;
APPEND FROM DADO;
CLOSE DOIS;
QUIT;
DATA DOIS ;
SET DOIS ;
RENAME COL1=YT COL2=T COL3=P1 COL4=P2 COL5=P3 COL6=P4 COL7=P5
      COL8=P6 COL9=P7;
T1=118; T2=484; T3=664; T4=1004; T5=1231; T6=1372; T7=1582;
PROC NLIN METHOD=NEWTON ;
PARMS ALFA=100 BETA=5 GAMA=0.002 ;
D1=BETA*EXP(-GAMA*T1);
D2=BETA*EXP(-GAMA*T2);
D3=BETA*EXP(-GAMA*T3);
D4=BETA*EXP(-GAMA*T4);
D5=BETA*EXP(-GAMA*T5);
D6=BETA*EXP(-GAMA*T6);
D7=BETA*EXP(-GAMA*T7);
MODEL YT=P1*ALFA/(1 + D1) + P2*ALFA/(1+D2) + P3*ALFA/(1+D3) +
      P4*ALFA/(1 + D4) + P5*ALFA/(1+D5) + P6*ALFA/(1+D6) +

```

$$P7*ALFA/(1 + D7) ;$$

$$DER.ALFA=P1/(1 + D1) + P2/(1+D2) + P3/(1+D3) +$$

$$P4/(1 + D4) + P5/(1+D5) + P6/(1+D6) +$$

$$P7/(1 + D7) ;$$

$$DER.BETA=-(P1*ALFA*D1/(BETA*((1+D1)**2)) +$$

$$P2*ALFA*D2/(BETA*((1+D2)**2)) +$$

$$P3*ALFA*D3/(BETA*((1+D3)**2)) +$$

$$P4*ALFA*D4/(BETA*((1+D4)**2)) +$$

$$P5*ALFA*D5/(BETA*((1+D5)**2)) +$$

$$P6*ALFA*D6/(BETA*((1+D6)**2)) +$$

$$P7*ALFA*D7/(BETA*((1+D7)**2)));$$

$$DER.GAMA=P1*ALFA*T1*D1/((1 + D1)**2) + P2*ALFA*T2*D2/((1+D2)**2) +$$

$$P3*ALFA*T3*D3/((1 + D3)**2) + P4*ALFA*T4*D4/((1+D4)**2) +$$

$$P5*ALFA*T5*D5/((1 + D5)**2) + P6*ALFA*T6*D6/((1+D6)**2) +$$

$$P7*ALFA*T7*D7/((1 + D7)**2) ;$$

$$DER.ALFA.ALFA=0;$$

$$DER.ALFA.BETA=DER.BETA/ALFA ;$$

$$DER.ALFA.GAMA=DER.GAMA/ALFA;$$

$$DER.BETA.BETA=P1*2*ALFA*(EXP(-2*GAMA*T1))/((1 + D1)**3) +$$

$$P2*2*ALFA*(EXP(-2*GAMA*T2))/((1 + D2)**3) +$$

$$P3*2*ALFA*(EXP(-2*GAMA*T3))/((1 + D3)**3) +$$

$$P4*2*ALFA*(EXP(-2*GAMA*T4))/((1 + D4)**3) +$$

$$P5*2*ALFA*(EXP(-2*GAMA*T5))/((1 + D5)**3) +$$

$$P6*2*ALFA*(EXP(-2*GAMA*T6))/((1 + D6)**3) +$$

$$P7*2*ALFA*(EXP(-2*GAMA*T7))/((1 + D7)**3) ;$$

$$DER.BETA.GAMA=P1*ALFA*T1*D1*(1 - D1)/(BETA*((1 + D1)**3)) +$$

$$P2*ALFA*T2*D2*(1 - D2)/(BETA*((1 + D2)**3)) +$$

$$P3*ALFA*T3*D3*(1 - D3)/(BETA*((1 + D3)**3)) +$$

$$P4*ALFA*T4*D4*(1 - D4)/(BETA*((1 + D4)**3)) +$$

```

P5*ALFA*T5*D5*(1 - D5)/(BETA*((1 + D5)**3)) +
P6*ALFA*T6*D6*(1 - D6)/(BETA*((1 + D6)**3)) +
P7*ALFA*T7*D7*(1 - D7)/(BETA*((1 + D7)**3));
DER.GAMA.GAMA=P1*ALFA*(T1**2)*D1*(D1 - 1)/((1 + D1)**3) +
P2*ALFA*(T2**2)*D2*(D2 - 1)/((1 + D2)**3) +
P3*ALFA*(T3**2)*D3*(D3 - 1)/((1 + D3)**3) +
P4*ALFA*(T4**2)*D4*(D4 - 1)/((1 + D4)**3) +
P5*ALFA*(T5**2)*D5*(D5 - 1)/((1 + D5)**3) +
P6*ALFA*(T6**2)*D6*(D6 - 1)/((1 + D6)**3) +
P7*ALFA*(T7**2)*D7*(D7 - 1)/((1 + D7)**3);
RUN;

```

- Arvmix.sas: Programa que utiliza a macro MIXNLIN para estimar os parâmetros por mínimos quadrados generalizados, supondo estrutura de efeitos aleatórios (2.27). Substitui o uso dos dois programas anteriores- ver **Seção 4.3**.

```

libname rmlib 'c:\TESE';
filename mixnlin 'c:\TESE\mixnlin.sas';
filename example2 'c:\tese\example2.dat';
%include mixnlin / nosource2;
options nodate pagesize=60 nocenter nonumber ls=80;

data example2;
infile example2;
input tree $ days Y ;
intercep=1;
group='grand mean' ;
proc sort data=example2;
by group tree;

```

```

title ' EXEMPLO 2 : DRAPER E SMITH - LARANJEIRAS-USANDO MQG';
%mixnlin(imllib=rmlib.storemod);
proc mixnlin data=example2 method=egls best=5 expand=average print=all
  outpred=results;
  subject tree;
  class group;
  vars 'intercep days';
  parms b1='150,100'
        b2='10,20'
        b3='-.001, -.01';
  design all='group' ;
  random 'b1';
  model 'y=b1*intercep/(intercep + b2*(exp(b3*days)))';
deriv
  der_b1='intercep/(intercep + b2*(exp(b3*days))'
  der_b2='(-b1/((intercep + b2*(exp(b3*days)))**2))*(exp(b3*days))'
  der_b3='(-b1/((intercep + b2*(exp(b3*days)))**2))*(b2*days*(exp(b3*days)))';
execute mixnlin;
proc print data=results;
run;

```

- Arvuni.sas: Programa para estimar os parâmetros através do método de mínimos quadrados generalizados. Deve ser rodado após o programa Arvmqo.sas, uma vez que utiliza o arquivo SAIDA1, onde estão armazenados os valores da variável resposta (Y), da variável explicativa (T) e dos resíduos (E). Matriz de covariância segue a estrutura uniforme (2.25). - ver **Seção 4.2**.

```

PROC IML ;
  USE SAIDA1;
  READ ALL INTO M;
  T=M[,1];

```

```

Y=M[,2];
E=M[,3];
PRINT T Y E;

/** PRIMEIRO ESTAGIO: ENCONTRAR ESTIMATIVAS POR MQO PARA OS
PARAM.*/;
ALFA=192.6876;
BETA=7.8566;
GAMA=0.0028;
SQR=17480.2335;

/*N=numero de individuos*/;
N=5;
/*RI=numero de condicoes de avaliacao*/;
RI=7;
/*K2=numero de between-variables*/;
K2=1;
/*NI=numero de coeficientes aleatorios*/;
NI=1;

/* TERCEIRO ESTAGIO: ESTIMACAO DE SIGMA2 E PSI */
SIGMA2=SQR/(N*RI);
PRINT SIGMA2;

S1=(SUM(E[1:7,]));
S2=SUM(E[8:14,]);
S3=SUM(E[15:21,]);
S4=SUM(E[22:28,]);
S5=SUM(E[29:35,]);
PRINT S1 S2 S3 S4 S5;

```

```
RO=(S1##2 + S2##2 + S3##2 + S4##2 + S5##2 - N#RI#SIGMA2)/(N#RI#(RI-1)#SIGMA2);
```

```
JOTA = J(RI);
```

```
SIG=SIGMA2#(1-RO)*I(RI) + SIGMA2#RO#JOTA;
```

```
PRINT RO JOTA SIG ;
```

```
/* QUARTO ESTAGIO: ESTIMACAO DE ALFA ATRAVES DE MQG */
```

```
WI=INV(SIG);
```

```
PRINT WI;
```

```
P=ROOT(WI);
```

```
PRINT P;
```

```
DIAP=BLOCK(P,P,P,P,P);
```

```
YT=DIAP*Y;
```

```
PRINT YT;
```

```
P=P//P//P//P//P;
```

```
DADO=YT||P ;
```

```
CREATE DOIS FROM DADO;
```

```
APPEND FROM DADO;
```

```
CLOSE DOIS;
```

```
QUIT;
```

```
DATA DOIS ;
```

```
SET DOIS ;
```

```
RENAME COL1=YT COL2=P1 COL3=P2 COL4=P3 COL5=P4 COL6=P5 COL7=P6
```

```
COL8=P7 ;
```

```
T1=118; T2=484; T3=664; T4=1004; T5=1231; T6=1372; T7=1582;
```

```
PROC NLIN ;
```

```
PARMS ALFA=100 BETA=5 GAMA=0.002 ;
```

```
D1=BETA*EXP(-GAMA*T1);
```

```
D2=BETA*EXP(-GAMA*T2);
```

```
D3=BETA*EXP(-GAMA*T3);
```

```
D4=BETA*EXP(-GAMA*T4);
```

$$D5=BETA*EXP(-GAMA*T5);$$

$$D6=BETA*EXP(-GAMA*T6);$$

$$D7=BETA*EXP(-GAMA*T7);$$

$$MODEL\ YT=P1*ALFA/(1 + D1) + P2*ALFA/(1+D2) + P3*ALFA/(1+D3) + \\ P4*ALFA/(1 + D4) + P5*ALFA/(1+D5) + P6*ALFA/(1+D6) + \\ P7*ALFA/(1 + D7) ;$$

$$DER.ALFA=P1/(1 + D1) + P2/(1+D2) + P3/(1+D3) + \\ P4/(1 + D4) + P5/(1+D5) + P6/(1+D6) + \\ P7/(1 + D7) ;$$

$$DER.BETA=-(P1*ALFA*D1/(BETA*((1+D1)**2)) + \\ P2*ALFA*D2/(BETA*((1+D2)**2)) + \\ P3*ALFA*D3/(BETA*((1+D3)**2)) + \\ P4*ALFA*D4/(BETA*((1+D4)**2)) + \\ P5*ALFA*D5/(BETA*((1+D5)**2)) + \\ P6*ALFA*D6/(BETA*((1+D6)**2)) + \\ P7*ALFA*D7/(BETA*((1+D7)**2)));$$

$$DER.GAMA=P1*ALFA*T1*D1/((1 + D1)**2) + P2*ALFA*T2*D2/((1+D2)**2) + \\ P3*ALFA*T3*D3/((1 + D3)**2) + P4*ALFA*T4*D4/((1+D4)**2) + \\ P5*ALFA*T5*D5/((1 + D5)**2) + P6*ALFA*T6*D6/((1+D6)**2) + \\ P7*ALFA*T7*D7/((1 + D7)**2) ;$$

RUN;

Aplicação prática: Concentração de lactato no sangue de atletas

- Pistat2.dat: Programa onde estão armazenados os dados.

1	0	10.9	2.5		1	0	15.7	4.1		2	0	10.9	1.9
1	0	12	2.5		1	0	17.1	4.2		2	0	12	2.2
1	0	13.3	2.7		1	0	18	7.3		2	0	13.3	2.4
1	0	14.4	3.2		1	0	19	13.3		2	0	14.4	2.8

2	0	15.7	3.5	5	0	18	5.8	9	0	14.4	4.2
2	0	17.1	3.7	5	0	19	6.5	9	0	15.7	5.9
2	0	18	4.9	6	0	9.5	1.9	10	0	10.9	1.4
2	0	19	5.3	6	0	10.9	1.8	10	0	12	1.8
2	0	20	7.1	6	0	12	1.8	10	0	13.3	1.8
3	0	10.9	1.9	6	0	13.3	2	10	0	14.4	2.2
3	0	12	2.2	6	0	14.4	2.1	10	0	15.7	2.8
3	0	13.3	2.3	6	0	15.7	4.1	10	0	18	6.9
3	0	14.4	2.5	6	0	17.1	7.9	10	0	19	8.6
3	0	15.7	2.9	7	0	9.5	1.6	11	0	10.9	1.3
3	0	17.1	3.1	7	0	10.9	1.8	11	0	12	1.4
3	0	18	3.4	7	0	12	2	11	0	13.3	1.7
3	0	19	4.2	7	0	13.3	2.2	11	0	14.4	2.3
3	0	20	5.3	7	0	14.4	3	11	0	15.7	2.3
4	0	9.5	1.7	7	0	15.7	3.5	11	0	17.1	4
4	0	10.9	1.7	7	0	17.1	4.5	11	0	18	4.4
4	0	12	1.8	7	0	18	5.6	11	0	19	5.3
4	0	13.3	1.7	8	0	9.5	1.4	12	0	9.5	1.4
4	0	14.4	1.8	8	0	10.9	2	12	0	10.9	1.8
4	0	15.7	2.1	8	0	12	2.2	12	0	12	1.8
4	0	17.1	2.9	8	0	13.3	2.4	12	0	13.3	1.8
4	0	18	3.7	8	0	14.4	2.8	12	0	14.4	1.9
4	0	19	4.7	8	0	15.7	3.6	12	0	15.7	2.2
4	0	20	6.6	8	0	17.1	4.5	12	0	17.1	3.4
5	0	9.5	1.6	8	0	18	5.8	12	0	18	4
5	0	10.9	1.7	8	0	19	7.9	12	0	19	7.1
5	0	12	2.1	8	0	20	9.2	12	0	20	8.7
5	0	13.3	2.4	9	0	9.5	1.2	13	0	9.5	1.7
5	0	14.4	3	9	0	10.9	1.8	13	0	10.9	1.8
5	0	15.7	3.5	9	0	12	2.9	13	0	12	1.8
5	0	17.1	5	9	0	13.3	3.2	13	0	13.3	1.9

13	0	14.4	2	23	1	12	1.1	27	1	9.5	1.2
13	0	15.7	2.3	23	1	13.3	1.4	27	1	10.9	0.8
13	0	17.1	4	23	1	14.4	1.7	27	1	12	1.1
13	0	18	7.5	23	1	15.7	2.1	27	1	13.3	0.9
13	0	19	9.9	23	1	17.1	3.2	27	1	14.4	0.9
14	0	9.5	1.7	24	1	9.5	1.3	27	1	15.7	1.2
14	0	10.9	1.8	24	1	10.9	1.7	27	1	17.1	1.8
14	0	12	2.1	24	1	12	1.2	27	1	18	6.2
14	0	13.3	2.2	24	1	13.3	1.7	27	1	19	7.6
14	0	14.4	2.4	24	1	14.4	2.2	28	1	10.9	1.1
14	0	15.7	2.6	24	1	15.7	2.4	28	1	12	1.1
14	0	17.1	2.9	24	1	17.1	7.3	28	1	13.3	1
14	0	18	3.2	25	1	9.5	1.6	28	1	14.4	1.3
14	0	19	3.6	25	1	10.9	1.3	28	1	15.7	1.9
14	0	20	5.1	25	1	12	1.2	28	1	17.1	2.3
21	1	9.5	1.3	25	1	13.3	1.1	28	1	18	6.7
21	1	10.9	1.2	25	1	14.4	1.3	29	1	9.5	1.7
21	1	12	1.2	25	1	15.7	1.3	29	1	10.9	1
21	1	13.3	1.5	25	1	17.1	1.9	29	1	12	1.2
21	1	14.4	1.7	25	1	18	2.8	29	1	13.3	1.7
21	1	15.7	2.7	25	1	19	3.3	29	1	14.4	1.6
21	1	18	8.8	25	1	20	6.6	29	1	15.7	2.5
22	1	10.9	2.3	26	1	9.5	1.3	29	1	17.1	3.6
22	1	12	1.8	26	1	10.9	1.4	29	1	18	6.2
22	1	13.3	2.4	26	1	12	1.5	30	1	9.5	1
22	1	14.4	1.6	26	1	13.3	1.3	30	1	12	1.3
22	1	15.7	3.4	26	1	14.4	1.4	30	1	13.3	1.9
22	1	17.1	5.2	26	1	15.7	1.7	30	1	14.4	2.2
22	1	18	6.8	26	1	17.1	2.3	30	1	15.7	3
23	1	9.5	1.3	26	1	18	3.7	31	1	9.5	1.4
23	1	10.9	1.1	26	1	19	6	31	1	10.9	1

31	1	12	0.9	31	1	18	2.9	32	1	13.3	2
31	1	13.3	1.1	31	1	19	4.5	32	1	14.4	1.9
31	1	14.4	2.4	32	1	9.5	0.9	32	1	15.7	2.1
31	1	15.7	1.9	32	1	10.9	0.8	32	1	17.1	4.9
31	1	17.1	2.5	32	1	12	1.1	32	1	18	6

- Atlestr1.sas: Programa que utiliza a procedure NLIN para estimar os parâmetros por mínimos quadrados ordinários, supondo a estrutura I (ver **Capítulo V**).

```
options ps=60;
data um;
infile "c:\tese\pistat2.dat";
input ind $ tipo veloc lactato;
if ind=9 then delete;
proc sort data=um;
by tipo ind;
title1 'EXEMPLO : LACTATO DE SÓDIO ATLETAS NA PISTA - USANDO MQO';
title2 'Modelo: Exponencial - f(x)=exp(b1*(x-b2))';
data um;
set um;
proc nlin data=UM iter=50;
parms b11=.5 b12=.5 b21=6 b22=12;
model lactato = exp((b11*(1-tipo) + b12*tipo)*(veloc-(b21*(1-tipo) + b22*tipo)));
der.b11=(1-tipo)*exp((b11*(1-tipo) + b12*tipo)*(veloc-(b21*(1-tipo)
+ b22*tipo)))*(veloc-(b21*(1-tipo) + b22*tipo)) ;
der.b12=tipo*exp((b11*(1-tipo) + b12*tipo)*(veloc-(b21*(1-tipo)
+ b22*tipo)))*(veloc-(b21*(1-tipo) + b22*tipo));
der.b21=-(1-tipo)*(b11*(1-tipo) + b12*tipo)*exp((b11*(1-tipo)
+ b12*tipo)*(veloc-(b21*(1-tipo) + b22*tipo)));
der.b22=-tipo*(b11*(1-tipo) + b12*tipo)*exp((b11*(1-tipo)
```

```
+ b12*tipo)*(veloc-(b21*(1-tipo) + b22*tipo)));
```

```
output out = saida1 p=p1 r=e1 ;
```

```
run;
```

- Atlestr2.sas: Programa que utiliza a procedure NLIN do SAS para estimar os parâmetros por mínimos quadrados ordinários, supondo a estrutura II (ver **Capítulo V**).

```
options ps=60;
```

```
data um;
```

```
infile "c:\tese\pistat2.dat";
```

```
input ind tipo veloc lactato;
```

```
LACPON=0;
```

```
IF ind=9 then delete;
```

```
proc sort data=um;
```

```
by tipo ind;
```

```
title1 ' EXEMPLO : LACTATO DE SÓDIO ATLETAS NA PISTA - USANDO MQO -  
PONDER.';
```

```
title2 ' Modelo: Exponencial -  $f(x)=\exp(b1*(x-b2))$ ';
```

```
data um;
```

```
set um;
```

```
LACPON=LACTATO/SQRT(VELOC);
```

```
proc nlin data=UM iter=50;
```

```
parms b11=.5 b12=.5 b21=6 b22=12;
```

```
model lacPON = exp((b11*(1-tipo) + b12*tipo)*(veloc-(b21*(1-tipo) +  
b22*tipo)))/SQRT(VELOC);
```

```
der.b11=(1-tipo)*exp((b11*(1-tipo) + b12*tipo)*(veloc-(b21*(1-tipo)  
+ b22*tipo)))*(veloc-(b21*(1-tipo) + b22*tipo))/ SQRT(VELOC);
```

```
der.b12=tipo*exp((b11*(1-tipo) + b12*tipo)*(veloc-(b21*(1-tipo)  
+ b22*tipo)))*(veloc-(b21*(1-tipo) + b22*tipo))/SQRT(VELOC);
```

```
der.b21=-(1-tipo)*(b11*(1-tipo) + b12*tipo)*exp((b11*(1-tipo)  
+ b12*tipo)*(veloc-(b21*(1-tipo) + b22*tipo)))/ SQRT(VELOC);
```

```

der.b22=-tipo*(b11*(1-tipo) + b12*tipo)*exp((b11*(1-tipo)
+ b12*tipo)*(veloc-(b21*(1-tipo) + b22*tipo)))/ SQRT(VELOC);

```

```
output out = saida1 p=p1 r=e1 ;
```

```
run;
```

- Atlestr3.sas: Programa que utiliza a macro MIXNLIN para estimar os parâmetros por mínimos quadrados generalizados, supondo a estrutura III (ver **Capítulo V**).

```
libname rmlib 'c:\TESE';
```

```
filename mixnlin 'c:\TESE\mixnlin.sas';
```

```
filename pista 'c:\TESE\PISTAT2.dat';
```

```
%include mixnlin / nosource2;
```

```
options nodate pagesize=60 nocenter nonumber ls=80;
```

```
data um;
```

```
infile PISTA;
```

```
input ind tipo veloc lactato;
```

```
intercep=1;
```

```
if ind=9 then delete;
```

```
proc sort data=um;
```

```
by tipo ind ;
```

```
title1 ' EXEMPLO : LACTATO DE SÓDIO - ATLETAS NA PISTA - MATRIZ
UNIFORME';
```

```
title2 ' 1 EFEITO ALEAT. LINEAR - IRGLS - AVERAGE ';
```

```
%mixnlin(imllib=rmlib.storemod);
```

```
proc mixnlin data=UM method=IRGLS best=5 expand=AVERAGE print=all
```

```
outpred=results;
```

```
subject ind;
```

```
class tipo;
```

```
vars 'intercep veloc';
```

```
parms b1=.5 .2,
```

```

        .05 .01,
        .15 .1'
    b2='6 12,
        12 15,
        12 6';
design all='tipo';
random 'intercep';
model 'lactato = exp(b1*(veloc-b2));
deriv
    der_b1='(veloc-b2)*exp(b1*(veloc-b2))'
    der_b2= '-b1*exp(b1*(veloc-b2))';
execute mixnlin;
proc print data=results;
run;

```

- Atlestr4.sas: Programa que utiliza a macro MIXNLIN para estimar os parâmetros por mínimos quadrados generalizados, supondo a estrutura IV (ver **Capítulo V**).

```

libname rmlib 'c:\TESE';
filename mixnlin 'c:\TESE\mixnlin.sas';
filename pista 'c:\TESE\PISTAT2.dat';
%include mixnlin / nosource2;
options nodate pagesize=60 nocenter nonumber ls=80;

data um;
infile PISTA;
    input ind tipo veloc lactato;
    lacpon = 0;
    intercep=1;
    if ind=9 then delete;
proc sort data=um;

```

```

by tipo ind ;
data um;
  set um;
  lacpon=lactato/sqrt(veloc);
title1 ' EXEMPLO : LACTATO DE SÓDIO - ATLETAS NA PISTA ';
title2 ' UNIF PONDERADO - IRGLS - AMOSTRA TOTAL';
%mixnlin(imllib=rmlib.storemod);
proc mixnlin data=UM method=irgls best=5 expand=AVERAGE print=all
  outpred=results;
  subject ind;
  class tipo;
  vars 'INTERCEP veloc';
  parms b1=.5 .2,
        .05 .01,
        .15 .1'
        b2='6 12,
        12 15';
  design all='tipo';
  random 'user_1';
  model 'lacpon = (exp(b1*(veloc-b2)))/ sqrt(veloc)';
deriv
  der_b1='((veloc-b2)*exp(b1*(veloc-b2)))/ sqrt(veloc)'
  der_b2='(-b1*exp(b1*(veloc-b2)))/ sqrt(veloc)';
  function user_1='intercep/sqrt(veloc)';
execute mixnlin;
proc print data=results;
run;

```

- Atlestr5.sas: Programa que utiliza a macro MIXNLIN para estimar os parâmetros por mínimos quadrados generalizados, supondo a estrutura V (ver **Capítulo V**).

```

libname rmlib 'c:\TESE';
filename mixnlin 'c:\TESE\mixnlin.sas';
filename pista 'c:\TESE\PISTAT2.dat';
%include mixnlin / nosource2;
options nodate pagesize=60 nocenter nonumber ls=80;

data um;
infile PISTA;
    input ind tipo veloc lactato;
    IF IND=9 THEN DELETE;
proc sort data=um;
    by tipo ind ;
title1 ' EXEMPLO : LACTATO DE SÓDIO - ATLETAS NA PISTA ';
title2 ' MODELO EXPONENCIAL - Efeito Aleat. no b1 - MQGI - AVERAGE';
%mixnlin(imllib=rmlib.storemod);
proc mixnlin data=UM method=IRGLS best=5 expand=AVERAGE print=all
    outpred=results;
    subject ind;
    class tipo;
    vars 'veloc';
    parms b1=.5 .2,
        .05 .01,
        .15 .1'
        b2='6 12,
        12 15';
    design all='tipo';
    random 'b1';
    model 'lactato = exp(b1*(veloc-b2))';
deriv
    der_b1='(veloc-b2)*exp(b1*(veloc-b2))'
    der_b2= '-b1*exp(b1*(veloc-b2)) ';

```

```
execute mixnlin;
proc print data=results;
run;
```

- Atlestr6.sas: Programa que utiliza a macro MIXNLIN para estimar os parâmetros por mínimos quadrados generalizados, supondo a estrutura VI (ver **Capítulo V**).

```
libname rmlib 'c:\TESE';
filename mixnlin 'c:\TESE\mixnlin.sas';
filename pista 'c:\TESE\PISTAT2.dat';
%include mixnlin / nosource2;
options nodate pagesize=60 nocenter nonumber ls=80;

data um;
infile PISTA;
    input ind tipo veloc lactato;
    IF IND=9 THEN DELETE;
proc sort data=um;
    by tipo ind ;
title1 ' EXEMPLO : LACTATO DE SÓDIO - ATLETAS NA PISTA ';
title2 ' MODELO EXPONENCIAL - Efeito Aleat. no b2';
%mixnlin(imllib=rmlib.storemod);
proc m mixnlin data=UM method=IRGLS best=5 expand=AVERAGE print=all
    outpred=results;
    subject ind;
    class tipo;
    vars 'veloc';
    parms b1=.5 .2,
        .05 .01,
        .15 .1'
        b2='6 12,
```

```

12 15';
design all='tipo';
random 'b2';
model 'lactato = exp(b1*(veloc-b2));
deriv
  der_b1='(veloc-b2)*exp(b1*(veloc-b2))'
  der_b2= '-b1*exp(b1*(veloc-b2))';
execute mixnlin;
proc print data=results;
run;

```

- Atlestr7.sas: Programa que utiliza a macro MIXNLIN para estimar os parâmetros por mínimos quadrados generalizados, supondo a estrutura VII (ver **Capítulo V**).

```

libname rmlib 'c:\TESE';
filename mixnlin 'c:\TESE\mixnlin.sas';
filename pista 'c:\TESE\PISTAT2.dat';
%include mixnlin / nosource2;
options nodate pagesize=60 nocenter nonumber ls=80;
data um;
infile PISTA;
  input ind tipo veloc lactato;
  IF IND=9 THEN DELETE;
proc sort data=um;
  by tipo ind ;
title1 ' EXEMPLO : LACTATO DE SÓDIO - ATLETAS NA PISTA ';
title2 ' MODELO EXPONENCIAL - 2 Ef. Al.';
%mixnlin(imllib=rmlib.storemod);
proc mixnlin data=UM method=IRGLS best=5 expand=AVERAGE print=all
  outpred=results;
  subject ind;

```

```

class tipo;
vars 'veloc';
parms b1=.5 .2,
      .05 .01,
      .15 .1'
      b2='6 12,
      12 15';
design all='tipo';
random 'b1 b2';
model 'lactato = exp(b1*(veloc-b2))';
deriv
  der_b1='(veloc-b2)*exp(b1*(veloc-b2))'
  der_b2='-b1*exp(b1*(veloc-b2))';
execute mixnlin;
proc print data=results;
run;

```

Referências Bibliográficas

- Amemyia, T. (1983) Non-linear regression models. Z. Griliches e M.D. Intriligator (Eds.), Handbook of Econometrics, Vol. I, 333-389. North-Holland: Amsterdam.
- Andrade, D.F. (1989) Uma visão dos modelos lineares longitudinais. Atas da 1ª Escola de Modelos Lineares. ABE-IME/USP.
- Andrade, D.F. e Singer, J. (1986) Análise de dados longitudinais. VI SINAPE.
- Andreoni, S. (1989) Modelos de efeitos aleatórios para análise de dados longitudinais não-balanceados em relação ao tempo. Dissertação de Mestrado. IME/USP.
- Bard, Y. (1974) Nonlinear parameter estimation. Academic Press: New York.
- Bates, D.M. e Watts, D. G. (1980) Relative curvatures measures of nonlinearity (with discussion). JRSS B, 42, 1-25.
- Bates, D.M. e Watts, D. G. (1988) Nonlinear regression analysis and its applications. New York: Wiley.
- Beale, E.M.L. (1960) Confidence regions in nonlinear estimation (with discussion). JRSS B, 22, 41-88.
- Berkey, C.S.(1982) Bayesian approach for nonlinear growth model. Biometrics 38, 953-961.
- Bock, R.D., Wainer, H., Petersen, A., Thiessen, D., Murray, J., e Roche, A. (1973) A parametrization for individual human growth curves. Human Biology 45, 63-80.
- Bock, R.D. e Thiessen, D. (1980) Statistical problems of fitting individual growth curves. Human Physical Growth and Maturation - Methodologies and Factors, 265-290
- Box, M.J. (1971) Bias in a nonlinear estimation (with Discussion). JRSS B, 33, 171-201.
- Broyden, C.G. (1967) Quasi-Newton methods and their application to function minimization. Math. Comp. 21, 368-381.
- Carter, R.L. e Yang, M.C.K. (1986) Large-sample inference in random coefficient regression models. Comm. Stat. - Theory and Methods 8, 2507-2526.
- Cordani, L.K. e Ribeiro, R.O. (1993) Validade do teste escalonado em pista para avaliação de "endurance" no atletismo. RAE-CEA9308-USP.
- Crowder, M.J. e Hand, D.J. (1990) Analysis of repeated measures. Chapman & Hall.

- Day, N.E. (1966) Fitting curves to longitudinal data. *Biometrics* 22, 276-291.
- Draper, N.R. e Smith, H. (1981). *Applied regression analysis*. Wiley: NY.
- Davidian, M e Giltinan (1995) *Nonlinear models for repeated measurement data*. Chapman & Hall, London: UK
- El Lozy, M. (1978) A critical analysis of the double and triple logistic growth curves. *Ann. Hum. Biol.* 5, 389-394.
- Fearn, R. (1975) A bayesian approach to growth curves. *Biometrika* 62, 89-100.
- Füller, W.A. (1976) *Introduction to statistical time series analysis* . Wiley: NY.
- Gallant, A.R. (1975) Seemingly unrelated nonlinear regressions. *Journal of Econometrics* 3, 35-50.
- Gallant, A.R. (1987) *Nonlinear statistical models*. Wiley:NY.
- Gasser, T., Müller, H.G., Köhler, W., Molinari, L. e Prader, A. (1984) Nonparametric regression analysis of growth curves. *The Annals of Statistics* 12, 210-229.
- Gennings, C., Chinchilli, V.M. e Carter, W.H. (1989) Response surface analysis with correlated data: A nonlinear model approach. *JASA* 84, 805-809.
- Glasbey, C.A. (1979) Correlated residuals in nonlinear regression applied to growth data. *Applied Statistics* 28, 251-259.
- Gosh, M., Grizzle, J.E. e Sen, P.K. (1973) Nonparametric methods in longitudinal studies. *JASA* 68, 29-36.
- Hartley, H.O. (1961) The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. *Technometrics* 3, 264-280.
- Hirst, K., Zerbe, G.O., Boyle, D.W. e Wilkening, R.B. (1991) On nonlinear random effects models for repeated measurements. *Commun. Statist.- Simula.* 20, 463-478.
- Jenrich, R.I. (1969) Asymptotic properties of nonlinear least squares estimation. *Ann. Math. Stat.* 40, 633-643.
- Jenrich, R.I. e Schluchter, M.D. (1986) Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 42, 805-820.
- Jones, R.H. (1993) *Longitudinal data with serial correlation: a state space approach*. Seminar on Biostatistics, Health and Environment. IASI - UNICAMP.

- Laird, N.M., Lange, N. e Stram, D. (1987) Maximum likelihood computations with repeated measures: Application of the EM algorithm. *JASA* 82, 97-105.
- Laird, N.M. e Ware, J.H. (1982) Random-effects models for longitudinal data. *Biometrics* 38, 963-974.
- Levenberg, K. (1944) A method for the solution of certain problems in least squares. *Quart. Appl. Math.* 2, 164-168.
- Lindstrom, M.J. e Bates, D.M. (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics* 46, 673-687.
- Malinvaud, E. (1966) *Statistical methods of econometrics*. Amsterdam: North-Holland.
- Malinvaud, E. (1970) The consistency of nonlinear regressions. *Ann. Math Stat.* 41, 856-969.
- Mardia, K.V., Kent, J.T. e Bibby, J.M. (1979) *Multivariate analysis*. Academic Press, London.
- Marquardt, D.W. (1963) An algorithm for least squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* 11, 431-441.
- Mazucheli, J. (1995) *Análise bayesiana e discriminação de modelos não-lineares*. Dissertação de Mestrado. ICMSC/USP.
- Morrison, D.F. (1976) *Multivariate statistical methods*. 2nd. Ed. Mc Graw-Hill, New York.
- Müller, H.G. (1989) *Nonparametric regression analysis of longitudinal data*. Lecture Notes in Statistics, 46. Springer-Verlag.
- Nelder, J.A. (1961) The fitting of a generalization of logistic curve. *Biometrics* 17, 89-110.
- Nelder, J.A. (1962) An alternative form of a generalized logistic equation. *Biometrics* 18, 614-616.
- Phillips, P.C.B. (1976) The iterated minimum distance estimator and the quasi-maximum likelihood estimator. *Econometrica* 44, 449-460.
- Pinheiro, J.C. e Bates, D.M. (1995) Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* 4, número 1, 12-35
- Pompeu, F.S. (1994) *Proposta de protocolo ergométrico para a determinação da curva de acúmulo do lactato sanguíneo em pista de atletismo*. Dissertação de Mestrado - EEFD-UFRJ.

- Potthoff, R.F. e Roy, S.N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51, 313-326.
- Preece, M.A. e Baines, M.J. (1978) A new family of mathematical models describing the human growth curve. *Ann. Humm. Biol.* 5, 1-24.
- Racine-Poon, A. (1985) Bayesian approach to nonlinear random-effects models *Biometrics* 41, 1015-1023.
- Ratkowsky, D.A. (1983) *Nonlinear regression modeling - An unified practical approach.* Dekker:NY.
- Reinsel, G. (1985) Mean squared error properties of empirical Bayes estimators in a multivariate random-effects general linear model. *JASA* 80, 642-650.
- Richards, F.J. (1959) A flexible growth function for empirical use. *J. Exp.Botany* 10, 290-300.
- Sandland, R.L. (1983) Mathematics and the growth of organisms - some historical impressions. *Math. Scientist.* 8, 11-30.
- Seber, G.A.F. e Wild, C.J. (1989) *Nonlinear regression.* Wiley. NY.
- Sheiner, L.B. e Beal, S.L. (1980) Evaluation of methods for estimating population pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics* 8, 553-571.
- Singer, J.M. (1987) *Análise de curvas de crescimento.* Dissertação de Mestrado. IME-USP.
- Singer, J.M. e Leite, J.G. (1990) *Métodos assintóticos em estatística.* IX SINAPE - ABE-IME-USP.
- Suyama, E. (1995) *Identificação de modelos de efeitos aleatórios.* Tese de Doutorado. IME-USP
- Von Bertalanffy, L. (1957) Quantitative laws in metabolism and growth. *The Quarterly Review of Biology* 32, no.3, 217-231.
- Vonesh, E.F. e Carter, R.L. (1987) Efficient inference for a random coefficient growth curve model with unbalanced data. *Biometrics* 43, 617-628.
- Vonesh, E.F. e Carter, R.L. (1992) Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics* 48, 1-17.
- Vonesh, E.F. (1992) Non-Linear models for the analysis of longitudinal data. *Statistics in Medicine* 11, 1929-1954.

- Vonesh, E.F. (1993) MIXNLIN: a SAS procedure for nonlinear mixed-effects models. Version 2.1.- Technical Report TR92M-0300A - Applied Statistics Center - Baxter Healthcare Corporation.
- Vonesh, E.F. e Chinchilli, V.M.(1997) Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker, Inc. New York.
- Ware, J.H. (1985) Linear models for the analysis of longitudinal studies. *The American Statistician* 39, 95-101.
- Wu, C.F. (1981) Asymptotic theory of nonlinear least squares estimation. *Ann. Stat.* 9, 501-503.