

CURVAS DE CRESCIMENTO: ANÁLISE DE DADOS LONGITUDINAIS  
INCOMPLETOS COM RESÍDUOS AUTOCORRELACIONADOS

LUIZ DE SOUZA

Licenciado em Matemática

Orientador: Prof. Dr. ANTONIO FRANCISCO IEMMA

Tese apresentada à Escola Superior  
de Agricultura "Luiz de Queiroz",  
da Universidade de São Paulo, para  
obtenção do título de Doutor em  
Agronomia. Área de concentração:  
Estatística e Experimentação  
Agronômica.

PIRACICABA

Estado de São Paulo - Brasil

Outubro, 1989

Ficha catalográfica preparada pela Seção de Livros da  
Divisão de Biblioteca e Documentação - PCAP/USP

---

S729c Souza, Luiz de  
Curvas de crescimento; análise de dados longitu-  
dinais incompletos com resíduos autocorrelacionados.  
Piracicaba, 1989.

84p.

Tese - ESALQ  
Bibliografia.

1. Análise estatística 2. Dado estatístico 3. Es-  
tatística matemática 4. Modelo matemático I. Escola  
Superior de Agricultura Luiz de Queiroz, Piracicaba

CDD 519.5

CURVAS DE CRESCIMENTO: ANÁLISE DE DADOS LONGITUDINAIS  
INCOMPLETOS COM RESÍDUOS AUTOCORRELACIONADOS

LUIZ DE SOUZA

Aprovado em: 24/11/89

Comissão Julgadora:

Prof. Dr. Humberto de Campos	ESALQ/USP
Prof. Dr. Antonio Francisco Iemma	ESALQ/USP
Profa. Dra. Clarice Garcia Borges Demétrio	ESALQ/USP
Prof. Dr. Carlos Roberto Padovani	IB/UNESP
Prof. Dr. Agostinho Roberto de Abreu	ESAL/UF

Orientador: Prof. Dr. Antonio Francisco Iemma



À memória de meus pais,

Geraldo e Lâmia;

Aos meus irmãos,

Dalva, Rubens e Celso;

À minha esposa e filhas,

Suely, Ivana e Andrea;

Dedico.



## AGRADECIMENTOS

O autor agradece

- \* ao Prof. Dr. Antonio Francisco Iemma, do Departamento de Matemática e Estatística da ESALQ-USP, pela orientação;
- \* à Pesquisadora Josepha Bandeira de Souza, da Faculdade de Medicina de Ribeirão Preto, USP, pela cessão dos dados;
- \* à Profa. Dra. Maria Aparecida de Paiva Franco e ao pós-graduando Mauro Andrea, da Faculdade de Medicina de Ribeirão Preto, USP, por proporcionarem discussões sobre o assunto;
- \* aos colegas do curso de Doutorado, João Riboldi, Luís Carlos Baida e Walter Veriano Valério Filho, por muitas colaborações;
- \* aos Professores do Departamento de Matemática e Estatística da ESALQ-USP, pelos ensinamentos;
- \* aos colegas da Escola de Enfermagem de Ribeirão Preto, USP, pelo apoio;
- \* aos amigos do Departamento de Bioestatística do Instituto de Biociências, UNESP - Botucatu, pelo incentivo e amizade.
- \* ao Francisco José Barroso Vessi, funcionário da Escola de Enfermagem de Ribeirão Preto, USP, pela confecção dos gráficos;
- \* à Profa. Maria Suely Crocci de Souza, pela revisão dos textos.

## S U M Á R I O

RESUMO .....	iv
ABSTRAT .....	v
1. INTRODUÇÃO .....	1
2. REVISÃO DA LITERATURA .....	7
2.1. Modelo Linear Multivariado (MLM) .....	10
2.2. Modelo Linear Multivariado de Crescimento (MLMC) ....	18
2.3. Modelo Linear Multivariado de Crescimento Gene- ralizado (MLMCG) .....	30
3. DESENVOLVIMENTO TEÓRICO: MODELO LINEAR MULTIVARI- ADO DE CRESCIMENTO GENERALIZADO (MLMCG) COM A RESTRIÇÃO $ARC(1)$ .....	38
3.1. Estrutura da Matriz de Covariâncias de Resíduos Autocorrelacionados .....	40
3.2. Estimação dos Parâmetros .....	47
3.3. Testes de Hipóteses .....	55

4. APLICAÇÕES .....	58
4.1. Análise do MLMC com todas as unidades experi- mentais .....	60
4.2. Análise do MLMC com parte das unidades experi- mentais .....	67
4.3. Análise do MLMCG com a restrição $ARC(1)$ no de- lineamento incompleto .....	69
4.4. Comparação dos Resultados .....	72
5. CONCLUSÕES .....	76
6. REFERÊNCIAS BIBLIOGRÁFICAS .....	78
APÊNDICE .....	83

CURVAS DE CRESCIMENTO: ANÁLISE DE DADOS LONGITUDINAIS  
INCOMPLETOS COM RESÍDUOS AUTOCORRELACIONADOS

Autor: Luiz de Souza

Orientador: Dr. Antonio Francisco Iemma

R E S U M O

O presente trabalho tem por objetivo estudar dados longitudinais incompletos usando modelos de curvas de crescimento. Supondo-se que os resíduos seguem um processo auto-regressivo de primeira ordem, obtem-se a matriz de covariâncias relativa a cada grupo de indivíduos de mesmo padrão de observações perdidas. Estimam-se os parâmetros do modelo polinomial proposto (matriz de parâmetros, variância e coeficiente de correlação de dois resíduos consecutivos) pelo método de máxima verossimilhança. Realizam-se testes de hipóteses da forma  $C\xi U = \phi$ , utilizando-se a estatística de Wald. Ilustra-se o método com parte dos dados de BANDEIRA (1989), relativos ao comprimento de crianças de baixo peso no nascimento, obtidos no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

GROWTH CURVES: ANALYSIS OF INCOMPLETE LONGITUDINAL DATA WITH  
AUTOCORRELATED ERRORS

Author: Luiz de Souza

Adviser: Prof. Dr. Antonio Francisco Iemma

ABSTRAT

The aim of this work is to study incomplete longitudinal data using growth curve models. Under the assumption of a first-order autoregressive process for the errors covariance matrices, each one relative to a set of individuals showing the same pattern of missing data, are obtained. Maximum likelihood estimators for the parameters of a proposed polynomial model, for the innovation variance and the one-lag correlation coefficient, are obtained. Test of hypothesis  $C\xi U = \phi$  are performed using Wald's statistics. One application is shown with part of data from SOUZA (1989), relative to the growth curve for length of low-birth-weight children born at the "Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo".

## 1. INTRODUÇÃO

Pretende-se, neste capítulo, introduzir o leitor em estudos longitudinais e apresentar os objetivos deste trabalho. Para tanto, são expostos alguns conceitos, mostrando as várias formas em que se pode apresentar um conjunto de dados longitudinais e são citados alguns métodos que podem ser aplicados na sua análise. Para um estudo introdutório mais completo o leitor deve consultar SINGER & ANDRADE (1986).

Entende-se por dados longitudinais aqueles obtidos de uma variável medida, para cada indivíduo, em um conjunto de pontos (por exemplo: tempos, doses de uma droga, etc...). Estudos longitudinais têm por finalidade avaliar o comportamento de uma variável ao longo da dimensão considerada.

Os experimentos planejados com dados longitudinais envolvem um conjunto de unidades experimentais (indivíduos), cada qual avaliada em várias condições (tempos, por exemplo) e classificadas em subpopulações (devido aos tratamentos).

A análise de dados longitudinais implica estudar o efeito dos tratamentos, das condições de avaliação, da interação entre os dois efeitos e ajustar modelos (no caso deste trabalho, modelos polinomiais) para explicar a variação da resposta média.

Considere-se  $r$  subpopulações  $i$  ( $i = 1, \dots, r$ ), com  $n_i$  indivíduos. Três situações podem se apresentar:

1) Todos os indivíduos são observados nas mesmas condições de avaliação.

A análise pode ser realizada através dos perfis de médias (como descrita em MORRISON, 1976) ou através de curvas de crescimento, cujo trabalho precursor é o de POTTHOFF & ROY (1964). A estrutura dos dados deste experimento é apresentada no Quadro 1.1.

Quadro 1.1: Apresentação dos Resultados de um Experimento com Dados Longitudinais

subpopulação	unidade experimental	condições de avaliação		
		1	...	p
1	1	$Y_{111}$	...	$Y_{11p}$
	...	.....		
	$n_1$	$Y_{1n_1 1}$	...	$Y_{1n_1 p}$
.....	.....	.....		
r	1	$Y_{r11}$	...	$Y_{r1p}$
	...	.....		
	$n_r$	$Y_{rn_r 1}$	...	$Y_{rn_r p}$

2) Existem falhas de observações.

Tais falhas podem ser atribuídas a perda casual ou mesmo a planejamento. KLEIMBAUM (1973) observa que às vezes é conveniente que o experimento seja delineado na forma incompleta, como, por exemplo, no caso em que a pesquisa tem um custo fixo. Um exemplo de experimento delineado na forma incompleta é visto em RAO & RAO (1966) ou (o mesmo) em WOOLSON et alii (1978).

Muitas vezes, o número de unidades experimentais com falhas de observação é tão grande que, se retiradas do estudo, comprometerá seriamente os resultados. Por esse motivo métodos multivariados para analisar dados incompletos têm sido objeto



de estudo de muitos pesquisadores. Uma revisão da literatura, nesse sentido, encontra-se em LAVANGE (1983).

O interesse deste trabalho é analisar dados longitudinais através de curvas de crescimento. A análise pode ser realizada pelo método proposto por KLEIMBAUM (1973) ou pelo modelo de efeitos aleatórios sugerido em LAIRD & WARE (1982). Em ambos os casos pode-se exigir uma estrutura de covariâncias para os resíduos.

3) As condições de avaliação são determinadas a partir de variáveis aleatórias.

É o caso mais geral. Provavelmente não haverá repetições das condições de avaliação nos  $n = \sum_i n_i$  indivíduos observados. A análise é realizada através do modelo de efeitos aleatórios e exigirá uma estrutura de covariâncias para os resíduos. Esse caso não será tratado aqui.

Vale a pena citar as observações de SINGER & ANDRADE (1986) sobre as vantagens de se realizar a análise através de curvas de crescimento:

- a) As comparações podem ser mais sensíveis devido ao número reduzido de parâmetros;
- b) Podem-se fazer previsões a respeito do comportamento médio das respostas em condições de avaliação distintas daquela em que o experimento foi planejado;

c) É possível adquirir um maior conhecimento sobre o processo de variação da resposta ao longo das condições de avaliação.

O objetivo deste trabalho é estudar dados longitudinais incompletos através de modelos lineares de curvas de crescimento. Tal estudo será uma extensão do trabalho de KLEIMBAUM (1973 ). Será adotado um modelo de matriz de covariâncias, mais especificamente, o modelo auto-regressivo de 1a. ordem, pois a forma de estimar a matriz de covariâncias proposta por Kleimbaum é falha quando os dados estão muito dispersos, isto é, a estimativa da matriz de covariâncias pode não ser positiva definida.

No capítulo 2, será apresentada uma revisão da literatura sobre análise de dados longitudinais completos com curvas de crescimento, e incompletos, no sentido da linha de KLEIMBAUM (1973 ).

No capítulo 3, será proposto um modelo utilizado em séries temporais, o auto-regressivo de 1a. ordem, AR(1), para a estrutura da matriz de covariâncias das variáveis respostas definidas nas  $p$  condições de avaliação. Estima-se a matriz de parâmetros do modelo polinomial proposto (Modelo Linear Multivariado de Crescimento Generalizado - MLMCG) e sua matriz de covariâncias, em experimentos onde há falhas de observação.

No capítulo 4, serão realizados três estudos: o primeiro,

com todos os dados do Apêndice (Quadro A); o segundo, só com as dezessete primeiras unidades experimentais; e, finalmente, o terceiro, com todas as unidades experimentais, mas eliminadas as observações com \*. Comparam-se os resultados dos três estudos.

No capítulo 5, será comentada a aplicabilidade da teoria desenvolvida no capítulo 3 e será sugerida a realização de alguns trabalhos.

## 2. REVISÃO DA LITERATURA

São apresentados, neste capítulo, alguns métodos utilizados para se analisarem dados longitudinais através de curvas de crescimento. Um trabalho que pode ser considerado um marco nessa área é o de POTTHOFF & ROY (1964), pois é uma generalização do Modelo Linear Multivariado (MLM) de Gauss-Markoff; ou seja, através de transformações convenientes no Modelo Linear Multivariado de Crescimento (MLMC) de Potthoff & Roy, obtém-se o MLM. Desse modo, esta análise apoia-se na teoria desse modelo. Trabalhos anteriores resolveram problemas mais particulares, dentre os quais serão citados, rapidamente, alguns.

Segundo KOWALSKI & GUIRE (1974), o primeiro estudo de crescimento foi realizado pelo Conde de Montbeillard durante os anos de 1759 a 1777 e foi publicado por Buffon em um suplemento para Histoire Naturelle. WISHART (1938), em um experimento realizado com porcos, comparou taxas de crescimento e aumentos nessas taxas, relativas a pesos, através de curvas polinomiais do 2o. grau. BOX (1950) apresenta diferentes campos de

aplicações onde aparece o problema de comparar curvas de crescimento. RAO (1958) comparou  $k$  grupos de indivíduos substituindo as observações originais pela observação inicial e diferenças das observações sucessivas. RAO (1959), utilizando o modelo polinomial, analisou dados longitudinais provenientes de apenas um grupo, aplicando métodos de análise multivariada. ELSTON & GRIZZLE (1962), considerando, também, apenas um grupo de indivíduos, compararam o método multivariado de RAO (1959) com dois métodos univariados: o primeiro, método da análise de regressão, supõe independência entre as observações de um mesmo indivíduo; o segundo, método do modelo misto, supõe um modelo de covariância uniforme. Outros estudiosos trabalharam nesse assunto, como LEECH & HEALY (1959) e HOEL (1964), mas foram POTTHOFF & ROY (1964) que apresentaram um modelo que permitiu analisar casos bastante gerais de um conjunto de dados longitudinais completos.

Não se pretende apresentar uma revisão bibliográfica completa, mas apenas aquela parte que forma uma base para a proposta deste trabalho. Estudos de revisão mais completos poderão ser vistos em SINGER (1977), WOOLSON & LEEPER (1980) e SINGER & ANDRADE (1986).

Em 2.1 se fará uma breve apresentação do MLM. Maiores detalhes poderão ser vistos em MORRISON (1976). Em 2.2

se verá, propriamente, a análise de dados longitudinais completos através do MLMC, proposto por POTTHOFF & ROY (1964). Finalmente, em 2.3 se apresentará a proposta de KLEIMBAUM (1973) para analisar dados longitudinais incompletos com curvas de crescimento.

### 2.1. Modelo Linear Multivariado (MLMV)

O modelo linear multivariado que será considerado é expresso por

$$E(Y) = X \beta \quad (2.1.1)$$

$$\text{Var}(Y) = \Sigma \otimes I_n$$

onde

$$Y (n \times p) = \begin{bmatrix} Y_{111} & \dots & Y_{11p} \\ \dots & \dots & \dots \\ Y_{1n_1 1} & \dots & Y_{1n_1 p} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ Y_{r11} & \dots & Y_{r1p} \\ \dots & \dots & \dots \\ Y_{rn_r 1} & \dots & Y_{rn_r p} \end{bmatrix}; \quad X (n \times r) = \begin{bmatrix} 1 & \dots & 0 \\ \dots & \dots & \dots \\ 1 & \dots & 0 \\ \vdots & & \vdots \\ \vdots & & \vdots \\ 0 & \dots & 1 \\ \dots & \dots & \dots \\ 0 & \dots & 1 \end{bmatrix} \otimes I_n$$

$$\beta (r \times p) = \begin{bmatrix} \beta_{11} & \dots & \beta_{1p} \\ \dots & \dots & \dots \\ \beta_{r1} & \dots & \beta_{rp} \end{bmatrix}$$

Y : é a matriz de observações,

X : é a matriz de delineamento entre indivíduos,

$\beta$  : é a matriz de parâmetros desconhecidos.

$r$  : e' o número de grupos de tratamentos,

$n_1, \dots, n_r$  : são os números de unidades experimentais dos grupos 1, ..., r, e tais que  $\sum_{i=1}^r n_i = n$ ,

$p$  : e' o número de condições de avaliação em que cada unidade experimental é submetida.

Será suposto que as linhas de  $Y$  são distribuídas como normais multivariadas.

O estimador de mínimos quadrados (o qual coincide com o de máxima verossimilhança de  $\beta$ ) é expresso por

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{2.1.2}$$

$$\text{Var}(C\hat{\beta}U) = U'\Sigma U \otimes C(X'X)^{-1}C'$$

onde  $C$  ( $c \times r$ ) e  $U$  ( $p \times u_1$ ) são matrizes selecionadas de tal forma a obter as combinações lineares desejadas dos parâmetros.

As hipóteses a serem testadas são da forma  $H_0: C\beta U = \Gamma$ .

Tem-se que

$$S(p \times p) = Y'[I - X(X'X)^{-1}X']Y \tag{2.1.3}$$



$$E(u_1 \times u_1) = U'SU \sim W(n-r, U'\Sigma U, \phi) \quad (2.1.4)$$

e

$$H(u_1 \times u_1) = (C\hat{\beta}U - \Gamma)' [C(X'X)^{-1}C']^{-1} (C\hat{\beta}U - \Gamma) \sim W(c, U'\Sigma U, [C\beta U - \Gamma]' [C(X'X)^{-1}C']^{-1} [C\beta U - \Gamma]) \quad (2.1.5)$$

onde  $W$  representa a distribuição de Wishart e  $\phi$  a matriz nula de ordem  $u_1$ .

As matrizes  $E$  e  $H$  são ditas matrizes de quadrados e produtos cruzados relativas aos resíduos e hipótese, respectivamente.

Sob  $H_0$ ,  $H \sim W(c, U'\Sigma U, \phi)$ . Ainda mais,  $E$  e  $H$  são independentes. Os testes para essa hipótese são realizados com estatísticas que são funções das raízes características não nulas  $(\lambda_1, \dots, \lambda_s)$  de  $HE^{-1}$ . Defina-se

$$\theta_\nu = \frac{\lambda_\nu}{1 + \lambda_\nu}, \quad \nu = 1, \dots, s; \quad (2.1.6)$$

$$s = \min(c, u_1); \quad (2.1.7)$$

$$m_1 = \frac{|c - u_1| - 1}{2}; \quad (2.1.8)$$

$$m_2 = \frac{n - r - u_1 - 1}{2}; \quad (2.1.9)$$

A seguir, serão apresentados alguns testes estatísticos para a hipótese  $H_0: C\beta U = \Gamma$ , que também se encontram em MORRISON (1976) e GODOI (1985).

a) Teste da união e intersecção de Roy

A hipótese  $H_0: C\beta U = \Gamma$  é testada através da estatística

$$\theta_1 = \max_{1 \leq v \leq s} \theta_v \quad (2.1.10)$$

cuja distribuição encontra-se tabelada em HECK (1960) ou MORRISON (1976), dentre outros, e cujos parâmetros são  $s$ ,  $m$  e  $n^*$ .

Se  $s = 1$ , então

$$\frac{m_2 + 1}{+ 1} \frac{\theta_1}{1 - \theta_1} \sim F(2m_1 + 2, 2m_2 + 2) \quad (2.1.11)$$

b) Teste da Razão de Verossimilhança Generalizada de Wilks

A estatística de Wilks, estabelecida a partir do princípio da razão de verossimilhança generalizada, é expressa por

$$\Lambda = \frac{|E|}{|H + E|} = \prod_{j=1}^s (1 - \theta_j) = \prod_{j=1}^s (1 + \lambda_j)^{-1} \quad (2.1.12)$$

Se  $s = 1$ , então

$$\frac{m_2 + 1}{m_1 + 1} \frac{1 - \Lambda}{\Lambda} \sim F(2m_1 + 2, 2m_2 + 2) \quad (2.1.13)$$

Se  $s = 2$ , então

$$\frac{2m_2 + 2}{2m_1 + 3} \frac{1 - \Lambda^{1/2}}{\Lambda^{1/2}} \sim F[4m_1 + 6, 4(m_2 + 1)] \quad (2.1.14)$$

Para outros casos, mas quando o tamanho da amostra é

grande, utiliza-se a aproximação de Bartlett:

$$- [n - r - \frac{1}{2} (u_1 - c + 1)] \ln \Lambda \sim \chi^2 (c \cdot u_1) \quad (2.1.15)$$

### c) Teste de Lawley-Hotelling

A estatística de Lawley-Hotelling é expressa por

$$LH = \text{tr}(HE^{-1}) = \sum_{j=1}^s \lambda_j \quad (2.1.16)$$

A aproximação F para esta estatística é

$$\frac{2(sm_2 + 1)}{s^2(2m_1 + s + 1)} \cdot LH \sim F[s(2m_1 + s + 1), 2(sm_2 + 1)] \quad (2.1.17)$$

d) Teste de Pillai

Seja PI a estatística de Pillai, tal que

$$PI = \text{tr } HCH + E)^{-1} = \sum_{v=1}^s \theta_v \quad (2.1.18)$$

A aproximação F para a estatística de Pillai é

$$\frac{2m_2 + s + 1}{2m_1 + s + 1} \frac{PI}{s - PI} \sim F[s(2m_1 + s + 1), s(2m_2 + s + 1)] \quad (2.1.19)$$

As hipóteses, na forma  $H_0: C\beta U = \phi$ , que geralmente se testam são as seguintes:

i) Paralelismo dos perfis das r subpopulações:

$$C((r-1) \times r) = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

$$U(p \times (p-1)) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & -1 \end{bmatrix}$$

ii) Coincidência dos perfis das r subpopulações

Se o paralelismo for aceito, as matrizes C e U para o teste de coincidência são as seguintes:

$$C((r-1) \times r) = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix} \quad U(p \times 1) = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

Se se deseja testar a coincidência diretamente, sem

testar o paralelismo, então  $U = I_p$ .

iii) Paralelismo dos perfis ao eixo das abscissas

$$C(1xr) = [1 \quad 1 \quad 1 \quad \dots \quad 1]$$

$$U_{(p \times (p-1))} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & -1 \end{bmatrix}$$

## 2.2. Modelo Linear Multivariado de Crescimento (MLMC)

Proposto por POTTHOFF & ROY (1964), o MLMC é uma generalização do MLM. É expresso por

$$E(Y) = X\xi G \quad (2.2.1)$$

$$\text{Var}(Y) = \Sigma \otimes I_n$$

onde X e Y são os mesmos do modelo (2.1.1),

G é a matriz de delineamento dentro de indivíduos,

$\xi$  é a matriz de parâmetros, tais que

$$G(q \times p) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_p \\ \dots & \dots & \dots & \dots \\ t_1^{q-1} & t_2^{q-1} & \dots & t_p^{q-1} \end{bmatrix}$$

$$\xi(r \times q) = \begin{bmatrix} \xi_{01} & \xi_{11} & \dots & \xi_{q-1,1} \\ \xi_{02} & \xi_{12} & \dots & \xi_{q-1,2} \\ \dots & \dots & \dots & \dots \\ \xi_{0r} & \xi_{1r} & \dots & \xi_{q-1,r} \end{bmatrix}$$

POTTHOFF & ROY (1964) apresentam várias situações onde este modelo pode ser aplicado. Para estimar  $\xi$ , consideram a transformação

$$Y_* = Y\Delta^{-1}G'(G\Delta^{-1}G')^{-1} \quad (2.2.2)$$

onde  $\Delta$  é qualquer matriz simétrica e positiva definida ou não singular tal que  $G\Delta^{-1}G'$  tenha posto completo.

Obtém-se, então, o seguinte MLM:

$$E(Y_*) = X\xi \quad (2.2.3)$$

$$\text{Var}(Y_*) = (G\Delta^{-1}G')^{-1}G\Delta^{-1}\Sigma\Delta^{-1}G'(G\Delta^{-1}G')^{-1} \otimes I_n$$

Aplicando-se a teoria desenvolvida no item anterior, determina-se o estimador de mínimos quadrados (o qual coincide com o estimador de máxima verossimilhança), expresso por

$$\hat{\xi} = (X'X)^{-1}X'Y_* = (X'X)^{-1}X'Y\Delta^{-1}G'(G\Delta^{-1}G')^{-1} \quad (2.2.4)$$

$$\text{e } \text{Var}(\hat{\xi}) = (G\Delta^{-1}G')^{-1}G\Delta^{-1}\Sigma\Delta^{-1}G'(G\Delta^{-1}G')^{-1} \otimes (X'X)^{-1} \quad (2.2.5)$$

O estimador de  $\xi$  será sempre não viciado para qualquer  $\Delta$ , observadas as restrições citadas. No entanto, POTTHOFF & ROY (1964) demonstraram que  $\Delta = \Sigma$  constitui a escolha ótima, no sentido de obter a menor  $\text{Var}(\hat{\xi})$ . Nesse caso, tem-se o BLUE ("Best Linear Unbiased Estimator") de  $\xi$ , segundo LEE (1974). Como essa escolha é impraticável ( $\Sigma$  é sempre desconhecida),



POTTHOFF & ROY (1964) propõem utilizar uma estimativa de  $\Sigma$  baseada em informações anteriores. RAO (1965 e 1966) mostra que o estimador de POTTHOFF & ROY (1964) não aproveita as informações contidas nas  $p-q$  variáveis desprezadas e propõe o estimador de mínimos quadrados com ajuste por covariâncias. Também, utilizando o modelo condicional, KHATRI (1966) obteve o estimador de máxima verossimilhança de  $\xi$ , igual ao de POTTHOFF & ROY (1964), com  $\Delta = S$ . GRIZZLE & ALLEN (1969) mostram que os estimadores obtidos pela forma proposta por RAO (1965) com todas as covariáveis, por KHATRI (1966) e por POTTHOFF & ROY (1964), com  $\Delta=S$ , são idênticos. Esse estimador pode ser determinado da seguinte forma (veja GRIZZLE & ALLEN, 1969): considerem-se o modelo (2.2.1) e as matrizes  $G_1(p \times q)$  e  $G_2(p \times (p-q))$  tais que  $r(G_1) = q$ ,  $r(G_2) = p - q$ ,  $GG_1 = I_q$  e  $GG_2 = \phi$  ( $q \times (p-q)$ ). Estas condições são satisfeitas tomando-se  $G_1 = G'(GG')^{-1}$  e  $G_2$  como sendo uma base do espaço coluna de  $I - G'(GG')^{-1}G$ . Se as condições de avaliação forem equi-espaciaadas, toma-se, em geral,  $G$  como sendo a matriz de polinômios ortogonais.

Pós-multiplicando-se  $E(Y) = X\xi G$  por  $[G_1; G_2]$ , obtém-se:

$$E[Y_1; Y_2] = X [\xi; \phi] \quad (2.2.6)$$

$$\text{Var}[Y_1; Y_2] = \Sigma_1 \otimes I_n,$$

onde

$$\Sigma_1(p \times p) = \begin{bmatrix} G_1' \Sigma G_1 & G_1' \Sigma G_2 \\ G_2' \Sigma G_1 & G_2' \Sigma G_2 \end{bmatrix}$$

é a matriz de covariâncias de qualquer linha de  $[Y_1|Y_2]$ ,

$$Y_1 (n \times q) = YG_1, \quad E(Y_1) = X\xi,$$

$$Y_2 (n \times (p-q)) = YG_2, \quad E(Y_2) = \phi$$

Conseqüentemente, veja, por exemplo, ANDERSON, (1958: p.29), o modelo linear multivariado condicional de  $Y_1$  dado  $Y_2$  e' expresso por

$$E(Y_1/Y_2) = X\xi + Y_2(G_2'\Sigma G_2)^{-1}G_2'\Sigma G_1$$

isto é,

$$E(Y_1/Y_2) = [X \ Y_2] \begin{bmatrix} \xi \\ \eta \end{bmatrix}, \quad \eta = (G_2'\Sigma G_2)^{-1}G_2'\Sigma G_1$$

Dai,

$$E(Y_1/Y_2) = D \gamma \tag{2.2.7}$$

$$\text{Var}(Y_1/Y_2) = \Sigma_2 \otimes I_n$$

onde

$$D (n \times (r+p-q)) = [X \ Y_2]$$

$$\gamma ((r+p-q) \times q) = \begin{bmatrix} \xi \\ \eta \end{bmatrix}$$

e

$$\Sigma_2 = G_1'\Sigma G_1 - G_1'\Sigma G_2(G_2'\Sigma G_2)^{-1}G_2'\Sigma G_1 \tag{2.2.8}$$

Por KHATRI (1966, lema 1, p.76), tem-se:

$$G_2(G_2'\Sigma G_2)^{-1}G_2' = \Sigma^{-1} - \Sigma^{-1}G'(G\Sigma^{-1}G')^{-1}G\Sigma^{-1} \quad (2.2.9)$$

Substituindo (2.2.9) em (2.2.8), obtém-se:

$$\Sigma_2(qxq) = (G\Sigma^{-1}G')^{-1} \quad (2.2.10)$$

O estimador de mínimos quadrados (o qual coincide com o estimador de máxima verossimilhança) de  $\gamma$  será expresso por

$$\hat{\gamma} = (D'D)^{-1}D'Y_1 \quad (2.2.11)$$

Dai,  $\hat{\xi}$  é obtido considerando-se r primeiras linhas de  $\hat{\gamma}$ . A expressão algébrica de  $\hat{\xi}$  é dada por

$$\hat{\xi} = (X'X)^{-1}X'YS^{-1}G'(G\Sigma^{-1}G')^{-1} \quad (2.2.12)$$

A obtenção dessa expressão pode ser vista em SINGER (1977, p.36), com detalhes. GRIZZLE & ALLEN (1969) determinaram  $\text{Var}(\hat{\xi})$ , expressa por

$$\text{Var}(\hat{\xi}) = \frac{(n-r)-1}{(n-r)-(p-q)-1} (G\Sigma^{-1}G')^{-1} \otimes (X'X)^{-1} \quad (2.2.13)$$

cujo estimador é

$$\widehat{\text{Var}}(\hat{\xi}) = \frac{(n-r) - 1}{(n-r) - (p-q) - 1} \frac{1}{(n-r) - (p-q)} (GS^{-1}G')^{-1} \otimes (X'X)^{-1} \quad (2.2.14)$$

BAKSALARY et alii (1978) mostraram que a classe de estimadores PR (Potthoff & Roy) e de estimadores R (Rao) coincidem. Comentam que a escolha de um método ou de outro depende da conveniência. Em suas opiniões, o método de Rao deve ser escolhido sempre que não existir informações a respeito de  $\Sigma$ .

RAO (1965) mostrou que nem sempre o estimador de  $\xi$  com ajuste por covariância (onde utiliza todas as variáveis de  $Y_2$  como covariáveis) é o melhor, no sentido de obter a menor variância. Propôs um estimador que utilizasse um subconjunto de variáveis de  $Y_2$ , aquele que tem correlação com o conjunto de variáveis principais ( $Y_1$ ).

Para escolher as covariáveis que compõem o modelo, MORRISON (1976) descreve um teste múltiplo para "a" coeficientes de correlação, utilizando a desigualdade de Bonferroni e a transformação de Fisher. As hipóteses são estabelecidas como

$$H_{0_{kk'}} : \rho_{kk'} = 0 \quad \text{contra} \quad H_{A_{kk'}} : \rho_{kk'} \neq 0, \quad k \neq k'.$$

Rejeita-se  $H_{0_{kk'}}$ , se  $|r_{kk'}| > \tanh \frac{z}{\sqrt{n-a}}$ ,

onde

$r_{kk'}$  é a correlação estimada entre a  $k$ -ésima variável principal e a  $k'$ -ésima covariável,

$z$  é tal que  $P(Z > z) = \frac{\alpha}{2a}$ ,  $Z \sim N(0,1)$ ,

$a$  é o número de correlações e

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Aquelas variáveis para as quais se rejeitou  $H_0$  deverão ser consideradas como covariáveis.

Segundo RAO (1967), o estimador de  $\xi$  não é necessariamente o melhor quando, para a sua determinação, utiliza-se a matriz  $S$ ; o estimador pode ser melhorado com o conhecimento da estrutura da matriz  $\Sigma$ . A escolha das variáveis auxiliares depende desta estrutura. Veja também SINGER (1977, p.42).

GRIZZLE & ALLEN (1969) mostraram como testar hipóteses a respeito de funções lineares dos elementos de  $\xi$ . O primeiro teste que deve ser realizado é o do ajuste do modelo (2.2.1). Fazendo a transformação  $[Y_1; Y_2] = Y[G_1; G_2]$ , obtém-se:

$$E [Y_1; Y_2] = X [\xi; \psi]$$

onde  $\psi = \phi$  (matriz nula), se o modelo acima for adequado.

Esse teste é realizado considerando-se a hipótese  $H_0: E(Y_2) = X\xi G G_2 = \phi$  no modelo (2.2.1) ou, equivalentemente,  $H_0: C\xi_* U = \phi$  no modelo  $E(Y) = X\xi_*$ , com  $\xi_* = \xi G$ ,  $C = I$  ( $r \times r$ ) e  $U = G_2$  ( $p \times (p-q)$ ). Tem-se

$$E_A = G_2' S G_2 \sim W(n-r, G_2' \Sigma G_2, \phi) \quad (2.2.15)$$

$$H_A = G_2' Y' X (X' X)^{-1} X' Y G_2 \sim W(r, G_2' \Sigma G_2, G_2' \xi_*' X' X \xi_* G_2) \quad (2.2.16)$$

Sob  $H_0$ ,  $H_A \sim W(r, G_2' \Sigma G_2; \phi)$  e independente de  $E_A$ . Portanto, qualquer das estatísticas citadas em 2.2 servirá para testar  $H_0: E(Y_2) = \phi$ . Agora os parâmetros serão

$$s = \min(r, p-q) \quad (2.2.17)$$

$$m_1 = \frac{|r-(p-q)|-1}{2} \quad (2.2.18)$$

$$\ominus \quad m_2 = \frac{n-r-(p-q)-1}{2} \quad (2.2.19)$$

A estatística de Wilks tornar-se-á

$$-[n-r - \frac{1}{2}(p-q-r+1)] \ln \Lambda \sim \chi^2[r(p-q)] \quad (2.2.20)$$

Se o modelo (2.2.1) é adequado, podem-se testar hipóteses da forma  $H_0: C\xi U = \Gamma$ . Considerando-se o modelo condicional (2.2.7), a hipótese acima é equivalente a

$$H_0: [C \quad \phi] \begin{bmatrix} \xi \\ \eta \end{bmatrix} U = \Gamma$$

Dai, tem-se

$$E_1 = U' Y_1' [I - D(D'D)^{-1} D'] Y_1 U = U' S_1 U \sim W[(n-r)-(p-q), U' \Sigma_1 U, \phi] \quad (2.2.21)$$

onde  $S_1 = (G'S^{-1}G)^{-1}$ ,

$$H_1 = (C\hat{\xi}U - \Gamma)' (CRC')^{-1} (C\hat{\xi}U - \Gamma) \sim W[c, U' \Sigma_1 U, (C\hat{\xi}U - \Gamma)' (CRC')^{-1} (C\hat{\xi}U - \Gamma)] \quad (2.2.22)$$

onde

$$R_{(r \times r)} = (X'X)^{-1} + (X'X)^{-1} X' Y [S^{-1} - S^{-1} G' (GS^{-1}G')^{-1} GS^{-1}] Y' X (X'X)^{-1} \quad (2.2.23)$$

é o menor principal de ordem r de  $(D'D)^{-1}$ . Veja a determinação de R em KHATRI (1966).

Sob  $H_0$ ,  $H_1 \sim W(c, U' \Sigma_1 U, \phi)$  e independe de  $E_1$ . Então a hipótese  $H_0: C\xi U = \Gamma$  pode ser testada através de alguma das

estatísticas citadas no item 2.2. Os parâmetros serão:

$$s = \min (c, u_1); \quad (2.2.24)$$

$$m_1 = \frac{|c - u_1| - 1}{2} \quad (2.2.25)$$

$$m_2 = \frac{n - r - (p - q) - u_1 - 1}{2} \quad (2.2.26)$$

A estatística de Wilks tornar-se-á

$$-[n-r-(p-q) - \frac{1}{2}(u_1 - c + 1)] \ln \Lambda \sim \chi^2(cu_1) \quad (2.2.27)$$

No caso de se utilizar um subconjunto de covariáveis de  $Y_2$ , seja  $Y_3$ , deve-se considerar  $D = [X \ ; \ Y_3]$  e daí determinar  $\hat{\xi}$  como sendo as  $r$  primeiras linhas de  $\hat{\gamma}$ . Tem-se que

$$\hat{\gamma} = (D_1' D_1)^{-1} D_1' Y_1 \quad (2.2.28)$$

Daí

$$\hat{\xi} = (X'X)^{-1} X'Y [G_1 - G_3 (G_3' S G_3)^{-1} G_3' S G_1] \quad (2.2.29)$$

$$\text{Var}(\hat{\xi}) = \frac{n-r-1}{(n-r)-v-1} \frac{1}{(n-r)-v} [G_1' S G_1 - G_1' S G_3 (G_3' S G_3)^{-1} G_3' S G_1] (X'X)^{-1} \quad (2.2.30)$$

onde  $v$  é o número de covariáveis (= no. de colunas de  $Y_3$ ).



As hipóteses da forma  $H_0: C\xi U = \Gamma$  são testadas considerando-se

$$E_2 = U' Y_1' [I - D(D'D)^{-1} D'] Y_1 U \quad (2.2.31)$$

$$H_2 = (C\hat{\xi}U - \Gamma)' (C(RC')^{-1} (C\hat{\xi}U - \Gamma)) \quad (2.2.32)$$

onde  $R$  é o menor principal de  $(D'D)^{-1}$ , diferente, portanto, daquele da expressão (2.2.23). Os parâmetros  $s$  e  $m_1$  são os mesmos de (2.2.24) e (2.2.25) e  $m_2 = (n - r - v - u_1 - 1)/2$ .

Se nenhuma variável de  $Y_2$  for utilizada como covariável, aplica-se a técnica de POTTHOFF & ROY (1964), com  $\Delta = I$ . Nesse caso, de (2.2.4) e (2.2.5), tem-se:

$$\hat{\xi} = (X'X)^{-1} X' Y G' (GG')^{-1} \quad (2.2.33)$$

$$\widehat{\text{Var}}(\hat{\xi}) = \frac{1}{n - r} (GG')^{-1} G S G' (GG')^{-1} \otimes (X'X)^{-1} \quad (2.2.34)$$

Para os testes de hipóteses da forma  $H_0: C\xi U = \Gamma$ , determina-se

$$E_3 = U' G_1' S G_1 U \quad (2.2.35)$$

$$H_3 = (C\hat{\xi}U - \Gamma)' [C(X'X)^{-1} C']^{-1} (C\hat{\xi}U - \Gamma) \quad (2.2.36)$$

Os parâmetros  $s$  e  $m_1$  são os mesmos de (2.2.24) e (2.2.25) e  $m_2 = (n - r - u_1 - 1)/2$ .

GRIZZLE & ALLEN (1969) apresentam uma estratégia para se analisarem dados longitudinais completos através de programas de modelos lineares multivariados. Esta estratégia foi adotada por ANDRADE & SINGER (1988) na elaboração do aplicativo "Uso do SOC na análise de curvas de crescimento" implementado no sistema computacional "Software Científico" (SOC).

### 2.3. Modelo Linear Multivariado de Crescimento Generalizado

(MLMCG)

Sob a denominação de dados longitudinais incompletos considera-se aquele conjunto em que o vetor-resposta de algumas unidades experimentais está incompleto.

Na realização da análise de dados dessa natureza, os  $n$  indivíduos observados em um estudo longitudinal serão partidos em  $g$  subconjuntos disjuntos  $(S_1, S_2, \dots, S_g)$  de tal forma que os indivíduos de um mesmo subconjunto sejam observados nas mesmas condições de avaliação. Dessa forma, o grupo  $S_u$  será composto de  $n_u$  indivíduos  $(n_1 + n_2 + \dots + n_g = n)$ , todos observados nas  $p_u$  condições de avaliação.

Para a análise desses dados, KLEIMBAUM (1973) propôs o modelo linear multivariado de crescimento generalizado (MLMCG) que é uma generalização daquele proposto por POTTHOFF & ROY (1964), o qual permite a representação de um conjunto de dados com observações perdidas. É expresso por

$$\begin{aligned} E(Y_u) &= X_u \xi \otimes A_u & (2.3.1) \\ \text{Var}(Y_u) &= A_u' \Sigma A_u \otimes I_{n_u} \end{aligned}$$

onde

$Y_u (n_u \times p_u)$  é a  $u$ -ésima matriz de observações relativa ao padrão  $u$  de observações perdidas.

$X_u (n_u \times r)$  é a u-ésima matriz de planejamento entre indivíduos,

$\xi (r \times q)$  e  $G (q \times p)$  são as mesmas matrizes definidas no modelo (2.2.1),

$A_u (p \times p_u)$  é a u-ésima matriz de incidência.

Além disso, supõe-se que as linhas da matriz  $Y_u$  ( $u=1, \dots, g$ ) são distribuídas como multinormais.

A fim de se tornarem mais claras estas definições, considere-se o experimento cuja estrutura dos dados está apresentada no Quadro 2.3.1.

Quadro 2.3.1: Representação dos resultados de um experimento com dados longitudinais incompleto.

grupos de tratamentos	indivíduos	condições de avaliação				
		1	3	4	5	6
1	1	$Y_{111}$	$Y_{112}$			$Y_{115}$
	2		$Y_{122}$	$Y_{123}$		
	3	$Y_{131}$	$Y_{132}$			$Y_{135}$
	4	$Y_{141}$	$Y_{142}$	$Y_{143}$	$Y_{144}$	
2	1		$Y_{212}$	$Y_{213}$		
	2	$Y_{221}$	$Y_{222}$			$Y_{225}$
	3	$Y_{231}$	$Y_{232}$	$Y_{233}$	$Y_{234}$	
	4	$Y_{241}$	$Y_{242}$	$Y_{243}$	$Y_{244}$	
	5	$Y_{251}$	$Y_{252}$	$Y_{253}$	$Y_{254}$	

As matrizes relativas a este conjunto são apresentadas no Quadro 2.3.2.

Quadro 2.3.2: Matrizes  $Y_u, X_u$  e  $A_u$ ,  $u = 1, 2$  e  $3$ , relativas aos dados do quadro anterior.

$u$	$Y_u$				$X_u$		$A_u$			
1	$Y_{141}$	$Y_{142}$	$Y_{143}$	$Y_{144}$	1	0	1	0	0	0
	$Y_{231}$	$Y_{232}$	$Y_{233}$	$Y_{234}$	0	1	0	1	0	0
	$Y_{241}$	$Y_{242}$	$Y_{243}$	$Y_{244}$	0	1	0	0	1	0
	$Y_{251}$	$Y_{252}$	$Y_{253}$	$Y_{254}$	0	1	0	0	0	1
							0	0	0	0
2	$Y_{111}$	$Y_{112}$	$Y_{115}$		1	0	1	0	0	
	$Y_{131}$	$Y_{132}$	$Y_{135}$		1	0	0	1	0	
	$Y_{221}$	$Y_{222}$	$Y_{225}$		0	1	0	0	0	
							0	0	0	
							0	0	1	
3	$Y_{122}$	$Y_{123}$			1	0	0	0		
	$Y_{212}$	$Y_{213}$			0	1	1	0		
							0	1		
							0	0		
							0	0		

As matrizes  $Y_1, Y_2$  e  $Y_3$  foram formadas segundo o padrão de observações perdidas. A matriz  $Y_1$  não tem observações relativas

à 5a. condição de avaliação,  $Y_2$  não tem observações relativas à 3a. e 4a. condições de avaliação e  $Y_3$  não tem observações relativas à 1a., 4a. e 5a. condições de avaliação.

As matrizes  $X_1$ ,  $X_2$  e  $X_3$  possuem duas colunas porque existem dois tratamentos. O 1 da 1a. coluna indica que a unidade experimental correspondente se submeteu ao tratamento 1, e o 1 da 2a. coluna, ao tratamento 2.

As matrizes  $A_1$ ,  $A_2$  e  $A_3$  são indicadoras de presença ou ausência de observações nas condições de avaliação. A 5a. linha de  $A_1$  é nula então não existe observação na 5a. condição de avaliação do subconjunto 1, etc...

Se, por exemplo, se deseja ajustar uma curva do 1o. grau, ter-se-á

$$\xi = \begin{bmatrix} \xi_{01} & \xi_{11} \\ \xi_{02} & \xi_{12} \end{bmatrix} \quad \text{e} \quad G = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 5 & 6 \end{bmatrix}$$

KLEIMBAUM (1973.) transformou o modelo (2.3.1) em um modelo de regressão linear univariado, desenvolvendo cada matriz  $Y_u$  e  $\xi$  por colunas. Obteve-se

$$E(y) = F\varphi \tag{2.3.2}$$

onde

$$y \text{ (N x 1)} = [y'_1 \quad \dots \quad y'_g]';$$

$y_u$  ( $n_u p_u \times 1$ ) é o vetor obtido das colunas de  $Y_u$ ;

$$F (N \times qr) = [F'_1 \quad \dots \quad F'_g]';$$

$$F_u (n_u p_u \times qr) = A'_u G' \otimes X_u ;$$

$\phi$  ( $qr \times 1$ ) é o vetor obtido das colunas de  $\xi$ ;

$$N = \sum_{u=1}^g n_u p_u$$

O modelo (2.3.2) é constituído de modelos da forma

$$E(y_u) = F_u \phi \tag{2.3.3}$$

$$\text{Var}(y_u) = A'_u \Sigma A_u \otimes I_{n_u}$$

Portanto, (2.3.2) torna-se

$$E \begin{bmatrix} y_1 \\ \vdots \\ y_g \end{bmatrix} = \begin{bmatrix} A'_1 G' \otimes X_1 \\ \dots \\ A'_g G' \otimes X_g \end{bmatrix} \phi \rightarrow E(y) = F \phi \tag{2.3.4}$$

$$\Omega(N \times N) = \text{Var}(y) = \begin{bmatrix} A'_1 \Sigma A_1 \otimes I_{n_1} & & \phi \\ & \dots & \\ \phi & & A'_g \Sigma A_g \otimes I_{n_g} \end{bmatrix}$$

Daí, o estimador de máxima verossimilhança de  $\varphi$  será expresso por

$$\hat{\varphi} = (F' \Omega^{-1} F)^{-1} F' \Omega^{-1} y$$

e, desenvolvendo-se:

$$\hat{\varphi} = \left( \sum_{u=1}^g [G A_u (A_u' \Sigma_u A_u)^{-1} A_u' G' \otimes X_u' X_u] \right)^{-1} \sum_{u=1}^g [G A_u (A_u' \Sigma_u A_u)^{-1} \otimes X_u'] y_u \quad (2.3.5)$$

Como geralmente  $\Sigma$  é desconhecida, KLEIMBAUM (1973) propõe a utilização de um estimador consistente  $\hat{\Sigma}$ , o que implicará em um estimador BAN ("Best Asymptotically Normal") de  $\varphi$ . Para amostras de tamanho grande, esse estimador possui as mesmas propriedades que o estimador de máxima verossimilhança. O estimador de sua variância é expresso por:

$$\hat{\text{Var}}(\hat{\varphi}) = (F' \hat{\Omega}^{-1} F)^{-1}$$

Isto é:

$$\hat{\text{Var}}(\hat{\varphi}) = \left[ \sum_{u=1}^g G A_u (A_u' \hat{\Sigma}_u A_u)^{-1} A_u' G' \otimes X_u' X_u \right]^{-1} \quad (2.3.6)$$

onde  $\hat{\Sigma}$  é um estimador consistente de  $\Sigma$ .

Um estimador consistente de  $\Sigma = ((\sigma_{kk}))$  é obtido por



$$\hat{\sigma}_{kk'} = \frac{1}{n_{kk'} - r(B_{kk'})} x'_{kk'} [I - B_{kk'} (B'_{kk'} B_{kk'})^{-1} B'_{kk'}] x_{k',k} \quad (2.3.7)$$

onde

$n_{kk'}$  ( $> 1$ ) é o número de unidades experimentais observadas em ambas as condições de avaliação ( $k$  e  $k'$ ),

$x_{kk'}(n_{kk'} \times 1)$  é o vetor observado na  $k$ -ésima condição de avaliação nas unidades experimentais onde a  $k'$ -ésima condição de avaliação é observada,

$x_{k'k}(n_{k'k} \times 1)$  é o vetor observado na  $k'$ -ésima condição de avaliação nas unidades experimentais onde a  $k$ -ésima condição de avaliação é observada,

$B_{kk'}(n_{kk'} \times r)$  é a matriz de delineamento correspondente às  $n_{kk'}$  unidades experimentais.

KLEIMBAUM (1973) propõe um refinamento para este estimador de  $\Sigma$ . Por (2.3.5) e (2.3.7) determina  $\hat{\rho}$ . Daí, calcula os resíduos, que fornecerão um novo  $\hat{\Sigma}$ , e assim por diante.

O teste da hipótese  $H_0: C \xi U = \phi$  ou, equivalentemente,  $H_0: L' \rho = \phi$ , onde  $L' = U' \otimes C$ , é realizado com base na estatística de Wald (1943) - uma forma quadrática que utiliza o estimador BAN (Best Asymptotically Normal) de  $\rho$  e um estimador consistente de  $\Sigma$ , ou seja:

$$\begin{aligned} W^* &= (L' \hat{\rho})' [\text{Var}(L' \hat{\rho})]^{-1} (L' \hat{\rho}) \\ &= (L' \hat{\rho})' [L' (\sum_{u=1}^g G A_u (A_u' \hat{\Sigma} A_u)^{-1} A_u' G' \otimes X_u' X_u)^{-1} L]^{-1} (L' \hat{\rho}) \quad (2.3.8) \end{aligned}$$

Esta estatística é assintoticamente distribuída como  $\chi^2(cu_1)$ .

Alternativamente, KLEIMBAUM (1973) apresenta um teste onde se utiliza o modelo ajustado por covariância, como visto em 2.3 para dados completos. A estatística de Wald, se aplicada a dados completos, produzirá a estatística de Lawley-Hotelling.

### 3. DESENVOLVIMENTO TEÓRICO: MODELO LINEAR MULTIVARIADO DE CRESCIMENTO GENERALIZADO (MLMCG) COM A RESTRIÇÃO $ARC(1)$

A análise de dados longitudinais incompletos, quase sempre, exige a estimativa da matriz de covariâncias  $\Sigma$ , relativa ao vetor de observações.

A estimativa da matriz  $\Sigma$  proposta por KLEIMBAUM (1973), apesar de consistente, possui uma característica indesejável: não é necessariamente positiva definida. Um método para resolver este problema é o do "alisamento", proposto por Schwertman & Allen<sup>4</sup>, citado por WOOLSON & LEEPER (1980), que consiste em encontrar uma matriz semi-positiva definida mais próxima da  $\Sigma$  estimada. A desvantagem desse procedimento está relacionada com o desconhecimento da distribuição da estatística de Wald em pequenas amostras, quando se utiliza essa estimativa de  $\Sigma$ .

---

<sup>4</sup>SCHWERTMAN, N.C. & ALLEN, D.M. The Smoothing of an indefinite matrix with applications to growth curve analysis with missing observations. Dept. Statist. Tech. Report. N.56, University of Kentucky, 1973.

Uma outra proposta, trabalhada por LAVANGE (1983) no GIM ("General Incomplete Models"), é a de adotar modelos teóricos para a estrutura de  $\Sigma$ ; mais especificamente, supor que os resíduos são correlacionados segundo o modelo AR(1). O que se pretende aqui, com a adoção deste modelo, é estimar os parâmetros e realizar testes de hipóteses lineares da forma  $C\xi U = \phi$ , no MLMCG.

Como já foi dito no capítulo 1, o caso mais geral de dados longitudinais, não necessariamente o mais freqüente, é aquele em que as condições de avaliação são determinadas a partir de variáveis aleatórias. As condições de avaliação, provavelmente, não se repetirão e deve-se admitir uma matriz de covariâncias para cada indivíduo. Porém, o caso que será tratado neste trabalho, é aquele onde há perda de algumas observações e os dados podem ser agrupados conforme o padrão definido por essas falhas. Nessas condições, em 3.1, será determinada a estrutura da matriz de covariâncias supondo que os resíduos estejam autocorrelacionados segundo o modelo AR(1). Em 3.2, incorporar-se-á ao MLMCG de KLEIMBAUM (1973), a restrição AR(1), o que implicará no estudo do modelo admitindo a estrutura de covariâncias determinada em 3.1. Serão determinados um estimador consistente de  $\Sigma$  e um estimador BAN ("Best Asymptotically Normal") de  $\xi$ . E então, em 3.3, será estabelecida a estatística de Wald para o teste da hipótese  $H_0: C\xi U = \phi$ .

3.1. Estrutura da Matriz de Covariâncias de Resíduos Autocorrelacionados

Em um conjunto de dados longitudinais incompletos, as condições de avaliação podem ser representadas pelo quadro abaixo:

Quadro 3.1.: Condições de avaliação para um conjunto de dados longitudinais incompletos

subconjunto	condições de avaliação				
1	$t_{11}$	...	$t_{1k}$	...	$t_{1p_1}$
.....					
u	$t_{u1}$	...	$t_{uk}$	...	$t_{up_u}$
.....					
g	$t_{g1}$	...	$t_{gk}$	...	$t_{gp_g}$

onde  $u = 1, \dots, g; k = 1, \dots, p_u$ .

Cada linha  $(t_{u1} \dots t_{uk} \dots t_{up_u})$  é um subconjunto das

$p(t_1 \dots t_l \dots t_p)$  condições de avaliação total. Então  $p_u \leq p$ , para qualquer  $u$ .

O que se quer é determinar a matriz de covariâncias relativa a cada subconjunto  $u$ , impondo o modelo AR(1) aos resíduos.

Seja  $\delta$  um divisor comum (de preferência o máximo) de  $(t_l - t_{l-1}, l=2, \dots, p)$ . Divida o intervalo  $[t_1, t_p]$  em  $m$  sub-intervalos de amplitude  $\delta$ , de modo que os  $m+1$  pontos sejam numerados como  $\{1, \dots, m+1\}$  e marcados para todos os indivíduos. Pode-se pensar em  $m+1$  variáveis definidas nestes pontos, ou seja, cada ponto representará uma suposta condição de avaliação. Dividindo estas condições de avaliação por  $\delta$ , o novo conjunto terá o divisor comum igual a 1. Neste delineamento, o novo conjunto teria a forma de dados longitudinais completos com  $m+1$  condições de avaliação. O modelo empregado para a análise destes dados é o de Potthoff & Roy (2.2.1), ou seja,  $Y = X\xi G + E$ , onde  $E$  é a matriz de resíduos.

Suponha-se que os resíduos de cada indivíduo estejam relacionados por uma função auto-regressiva de primeira ordem, isto é, considere-se uma linha qualquer de  $E$ , representada por  $[e_1 \dots e_{m+1}]$ . Então,

$$e_{\tau} = \rho \cdot e_{\tau-1} + \varepsilon_{\tau} \quad (3.1.1)$$

com  $e_0 = \text{constante};$   
 $\tau = 1, \dots, m+1$  e  
 $-1 \leq \rho \leq 1.$

Considere-se, também, que  $\varepsilon_{\tau} \sim \text{IIDN}(0, \sigma^2).$

Por substituições sucessivas em (3.1.1), obtém-se:

$$e_{\tau} = \sum_{v=0}^{\infty} \rho^v \varepsilon_{\tau-v} \quad (3.1.2)$$

Portanto,

$$E(e_{\tau}) = 0,$$

$$\text{Var}(e_{\tau}) = \frac{\sigma^2}{1 - \rho^2}$$

$$\text{Cov}(e_{\tau}, e_{\tau-h}) = E(e_{\tau} \cdot e_{\tau-h}) = \frac{\sigma^2}{1 - \rho^2} \rho^h \quad (3.1.3)$$

Observe-se que  $\rho$  é a correlação entre duas variáveis respostas definidas em duas condições de avaliação sucessivas separadas pela distância 1.

A covariância entre dois resíduos relativos às condições

de avaliação originais,  $t_{uk}$  e  $t_{uk'}$  ( $t_{uk} > t_{uk'}$ ) é expressa por

$$\text{cov}(e_{uk}, e_{uk'}) = \frac{\sigma^2}{1 - \rho^2} \rho^{t_{uk} - t_{uk'}} \quad (3.1.4)$$

Para o conjunto onde existem observações em todas as condições de avaliação ( $t_1, \dots, t_p$ ), a matriz de covariâncias é expressa por

$$\Sigma = \sigma^2 V \quad (3.1.5)$$

onde

$$V(p \times p) = \frac{1}{1 - \rho^2} \begin{bmatrix} 1 & \rho^{t_2 - t_1} & \dots & \rho^{t_p - t_1} \\ \rho^{t_2 - t_1} & 1 & \dots & \rho^{t_p - t_2} \\ \dots & \dots & \dots & \dots \\ \rho^{t_p - t_1} & \rho^{t_p - t_2} & \dots & 1 \end{bmatrix}$$

Considerando-se o subconjunto  $u$ , de uma maneira geral, por (3.1.4), tem-se:



$$\Sigma_u = \sigma^2 V_u \quad (3.1.6)$$

onde

$$V_u(p_u \times p_u) = \frac{1}{1 - \rho^2} \begin{bmatrix} 1 & \rho^{t_{u2} - t_{u1}} & \dots & \rho^{t_{up_u} - t_{u1}} \\ \dots & \dots & \dots & \dots \\ \rho^{t_{up_u} - t_{u1}} & \rho^{t_{up_u} - t_{u2}} & \dots & 1 \end{bmatrix}$$

Determinaram-se, algebricamente, as matrizes  $V_u^{-1}$ ,  $\Lambda_u$  tal que  $\Lambda_u' \Lambda_u = V_u^{-1}$  e  $\Lambda_u^{-1}$ , cujas formas são descritas a seguir:

$$V_u^{-1} = (1 - \rho^2) (v_{kk'}) \quad (3.1.7)$$

onde

$$v_{kk'} = \begin{cases} \frac{1}{1 - \rho^{2(t_{u2} - t_{u1})}} & \text{se } k = k' = 1; \\ \frac{1}{1 - \rho^{2(t_{up_u} - t_{up_u-1})}} & \text{se } k = k' = p_u; \\ \frac{1 - \rho^{2(t_{uk'} - t_{uk'})}}{[1 - \rho^{2(t_{uk} - t_{uk'})}][1 - \rho^{2(t_{uk'} - t_{uk})}]}, & \text{se } k = k' = 2, \dots, p_u-1; \\ -\frac{\rho^{t_{uk'} - t_{uk}}}{1 - \rho^{2(t_{uk'} - t_{uk})}} & \text{se } k = k+1 \text{ e } \\ & k = 1, \dots, p_u-1 \end{cases}$$

$$\left\{ \begin{array}{ll} \frac{\rho(t_{uk} - t_{uk'})}{1 - \rho^2(t_{uk} - t_{uk'})}, & \text{se } k = k+1 \text{ e} \\ & k = 1, \dots, p_{u-1} \\ \\ 0, \text{ em outros casos.} \end{array} \right.$$

A matriz  $V_u^{-1}$  é simétrica e definida positiva. Então, existe  $\Lambda_u$  tal que  $\Lambda_u' \Lambda_u = V_u^{-1}$ , sendo

$$\Lambda_u = \sqrt{1 - \rho^2} ((\lambda_{kk'})), \quad (3.1.8)$$

onde

$$\lambda_{kk'} = \left\{ \begin{array}{ll} 1, & \text{se } k = k' = 1; \\ \\ \frac{1}{\sqrt{1 - \rho^2(t_{uk} - t_{uk'})}}, & \text{se } k = k' = 2, \dots, p_u; \\ \\ -\lambda_{kk} \rho(t_{uk} - t_{uk'}), & \text{se } k = 2, \dots, p_u \\ & \text{e } k' = k-1; \\ \\ 0, \text{ em outros casos.} \end{array} \right.$$

A inversa de  $\Lambda_u$  é expressa por

$$\Lambda_u^{-1} = \frac{1}{\sqrt{1 - \rho^2}} ((w_{kk'})); \quad (3.1.9)$$

onde

$$w_{kk'} = \begin{cases} 1, & \text{se } k = k' = 1; \\ \rho^{t_{uk} - t_{u1}}, & \text{se } k=2, \dots, p_u \text{ e } k'=1; \\ \sqrt{1 - \rho^2(t_{uk} - t_{uk'})}, & \text{se } k = k' = 2, \dots, p_u; \\ \rho^{t_{uk} - t_{uk'}} \sqrt{1 - \rho^2(t_{uk'} - t_{uk'-1})}, & \text{se } k' < k \leq p_u; \\ & k'=2, \dots, p_u - 1; \\ 0, & \text{se } k' > k. \end{cases}$$

### 3.2. Estimação dos parâmetros

Considere-se o modelo (2.3.1) e a estrutura de covariâncias para  $\Sigma_u$  determinada em (3.1.6). Na forma (2.3.4), este modelo torna-se:

$$E(y) = F \phi, \quad (3.2.1)$$

$$\text{Var}(y) = \sigma^2 \Omega,$$

$$\text{onde } \Omega = \begin{bmatrix} V_1 \otimes I_{n_1} & & \phi \\ & \dots & \\ \phi & & V_g \otimes I_{n_g} \end{bmatrix}$$

o  $V_u$  é a matriz definida em (3.1.6).

Se  $\rho$  é conhecido, então (2.3.5) resultano BLUE de  $\phi$ . No entanto, em quase todos os casos de aplicação, o pesquisador não conhece  $\rho$ , e este deve ser estimado a partir da amostra. Se em seu lugar for usado um estimador consistente, então o estimador de  $\phi$  será chamado BAN ("Best Asymptotically Normal") e tem as mesmas propriedades que o BLUE, para amostras de tamanho grande, ou seja, consistência, eficiência e normalidade. Testes de hipóteses a respeito de funções lineares de  $\phi$  poderão ser realizados desde que se utilize um estimador BAN de  $\phi$  e um estimador consistente de  $\rho$ .

A suposição de que os resíduos, de cada indivíduo, estão relacionados conforme o modelo AR(1), torna a estimação da matriz de covariâncias mais simples, já que envolve apenas dois parâmetros:  $\rho$  e  $\sigma$ .

Os estimadores de máxima verossimilhança de  $\varphi$ ,  $\sigma^2$  e  $\rho$ , serão obtidos considerando-se a função de verossimilhança

$$L = \left( (2\pi)^{N/2} |\sigma^2 \Omega|^{1/2} \right)^{-1} \exp \left( -\frac{1}{2\sigma^2} (y - F \varphi)' \Omega^{-1} (y - F \varphi) \right)$$

onde  $N = \sum n_u p_u$ . Mas,  $|\sigma^2 \Omega| = (\sigma^2)^N |\Omega|$ , com  $n = \sum n_u$ . Então, a expressão para  $\ln L$  será

$$\ln L = C - \frac{N}{2} \ln \sigma^2 - \frac{1}{2} \ln |\Omega| - \frac{1}{2\sigma^2} (y - F \varphi)' \Omega^{-1} (y - F \varphi),$$

onde  $C = \ln (2\pi)^{-N/2}$  é constante. Ou ainda,

$$\ln L = C - \frac{N}{2} \ln \sigma^2 - \frac{1}{2} \ln |\Omega| - \frac{1}{2\sigma^2} \sum_{u=1}^g [y_u - (A_u' G' \otimes X_u) \varphi]' (V_u^{-1} \otimes I_{n_u}) [y_u - (A_u' G' \otimes X_u) \varphi] \quad (3.2.2)$$

Derivando-se (3.2.2) em relação a  $\varphi$  e igualando-se o resultado ao vetor nulo, obtém-se

$$\hat{\varphi} = \left( \sum_{u=1}^g G A_u V_u^{-1} A_u' G' \otimes X_u' X_u \right)^{-1} \sum_{u=1}^g (G A_u V_u^{-1} \otimes X_u') y_u \quad (3.2.3)$$

$$\text{Var}(\hat{\varphi}) = \left\langle \sum_{u=1}^g GA_u V_u^{-1} A_u' G' \otimes X_u' X_u \right\rangle^{-1} \sigma^2 \quad (3.2.4)$$

Estes resultados, como era de se esperar, coincidem com aqueles obtidos por mínimos quadrados generalizados em (2.3.5), considerada a restrição AR(1) para os resíduos.

É fácil ver que  $\hat{\varphi}$  é um estimador não viciado de  $\varphi$ , pois

$$\begin{aligned} E(\hat{\varphi}) &= \left\langle \sum_{u=1}^g GA_u V_u^{-1} A_u' G' \otimes X_u' X_u \right\rangle^{-1} \sum_{u=1}^g (GA_u V_u^{-1} \otimes X_u') E(y_u) \\ &= \left\langle \sum_{u=1}^g GA_u V_u^{-1} A_u' G' \otimes X_u' X_u \right\rangle^{-1} \sum_{u=1}^g (GA_u V_u^{-1} \otimes X_u') (A_u' G' \otimes X_u) \varphi \\ &= \varphi \end{aligned}$$

A derivada da função (3.2.2), em relação ao parâmetro  $\sigma^2$ , igualada a zero, fornecerá o estimador de  $\sigma^2$ , ou seja:

$$\hat{\sigma}^2 = \frac{1}{N-qr-1} \sum_{u=1}^g [y_u - (A_u' G' \otimes X_u) \hat{\varphi}]' (V_u^{-1} \otimes I_{n_u}) [y_u - (A_u' G' \otimes X_u) \hat{\varphi}] \quad (3.2.5)$$

Na realidade, o quociente deste estimador é  $N$  e não  $N-qr-1$ . Tal correção foi realizada com o objetivo de tornar  $\hat{\sigma}^2$  um estimador não viciado de  $\sigma^2$ .

Para efeito de cálculo é conveniente que (3.2.5) seja apresentada na forma

$$\hat{\sigma}^2 = \frac{1}{N-qr-1} \sum_{u=1}^g \sum_{j=1}^{n_u} e'_{uj} V_u^{-1} e_{uj}$$

onde  $e'_{uj}$  é a  $j$ -ésima linha da matriz de erros  $E_u$ .

Para a estimação de  $\rho$  serão utilizadas apenas as observações consecutivas separadas pela mesma distância  $d$  nas condições de avaliação. Esta restrição deve ser considerada devido ao fato de as derivadas dos elementos de  $V_u^{-1}$  com relação a  $\rho$  (veja (3.1.7)) serem de difícil obtenção.

Considere-se o novo conjunto de observações e a função (3.2.2) redefinida adequadamente para este conjunto, ou seja:

$$\ln L^* = C - \frac{N^*}{2} \ln \sigma^2 - \frac{1}{2} \ln |\Omega^*| - \frac{1}{2\sigma^2} \sum_{u=1}^g (y_u^* - E(y_u^*))' (V_u^{*-1} \otimes I_{n_u^*}) (y_u^* - E(y_u^*)) \quad (3.2.7)$$

onde os elementos sobrepostos com \* indicam os mesmos elementos anteriores, mas considerando-se o novo conjunto de dados.

A função (3.2.7) pode ser escrita como

$$\ln L^* = C - \frac{N^*}{2} \ln \sigma^2 - \frac{1}{2} - \ln |\Omega^*| - \frac{1}{2\sigma^2} \sum_{u=1}^g \sum_{j=1}^{n_u^*} \epsilon_{uj}^* V_u^{*-1} \epsilon_{uj}^* \quad (3.2.8)$$

onde  $\epsilon_{uj}^*$  é a j-ésima linha da matriz de erros  $E_u$ .

$$\Omega^* = \begin{bmatrix} V_1^* & \otimes I_{n_1^*} & & \\ & & \dots & \\ & & & V_g^* & \otimes I_{n_g^*} \end{bmatrix} \quad (3.2.9)$$

Então

$$\begin{aligned} |\Omega^*| &= |V_1^* \otimes I_{n_1^*}| \dots |V_g^* \otimes I_{n_g^*}| \\ &= |V_1^*|^{n_1^*} \dots |V_g^*|^{n_g^*} \\ &= [1/(1-\rho^2)]^{n_1^*} \dots [1/(1-\rho^2)]^{n_g^*} \\ &= (1 - \rho^2)^{-n^*} \end{aligned} \quad (3.2.10)$$

pois  $|V_u^*| = 1/(1 - \rho^2)$ , para todo  $u = 1, \dots, g$ .

Substituindo-se (3.2.10) em (3.2.8), obtém-se:

$$\begin{aligned} \ln L^* &= C - \frac{N^*}{2} \ln \sigma^2 + \frac{n^*}{2} \ln (1 - \rho^2) + \\ &\quad - \frac{1}{2\sigma^2} \sum_{u=1}^g \sum_{j=1}^{n_u^*} \epsilon_{uj}^* V_u^{-1} \epsilon_{uj}^* \end{aligned} \quad (3.2.11)$$



Derivando-se (3.2.11) com relação a  $\rho$  e igualando-se o resultado a zero, determina-se a seguinte equação de verossimilhança para  $\hat{\rho}$ :

$$\frac{n \hat{\rho}^*}{1 - \hat{\rho}^2} + \frac{1}{2 \hat{\sigma}^2} \sum_{u=1}^g \sum_{j=1}^{n_u^*} e_{uj}^* \left( \frac{\partial}{\partial \hat{\rho}} \hat{V}_u^{*-1} \right) e_{uk}^* = 0 \quad (3.2.12)$$

Mas,

$$\frac{\partial}{\partial \hat{\rho}} \hat{V}_u^{*-1} (p_u^* \times p_u^*) = \begin{bmatrix} 0 & -1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2\hat{\rho} & -1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 2\hat{\rho} & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 2\hat{\rho} & -1 \\ 0 & 0 & 0 & 0 & \dots & -1 & 0 \end{bmatrix}$$

Dai, (3.2.12) torna-se:

$$r_1 \hat{\rho}^3 - r_2 \hat{\rho}^2 - (r_1 + n \hat{\sigma}^2) \hat{\rho} + r_2 = 0 \quad (3.2.13)$$

onde

$$r_1 = \sum_{u=1}^g \sum_{j=1}^{n_u^*} \sum_{k=2}^{p_u^*} e_{ujk}^* = \sum_{u=1}^g \sum_{k=2}^{p_u^*-1} e_{uk}^* e_{uk}^* \quad (3.2.14)$$

$$r_2 = \sum_{u=1}^g \sum_{j=1}^{n_u^*} \sum_{k=2}^{p_u^*} e_{ujk-1}^* e_{ujk}^* = \sum_{u=1}^g \sum_{k=2}^{p_u^*} e_{uk-1}^* e_{uk}^* \quad (3.2.15)$$

O subconjunto de condições de avaliação deve ser escolhido de tal forma a aproveitar o maior número possível de observações. Como  $d$  é a distância entre estas condições de avaliação, então o estimador que se obtém da resolução da

equação (3.2.13) será  $\hat{\rho}^d$ .

As estimativas dos parâmetros serão determinadas simultaneamente. O procedimento é supor, em uma primeira etapa, que os resíduos são totalmente independentes. Isso significa que cada  $V_u = I_{p_u}$ , ou seja, estimar os parâmetros pelo método dos mínimos quadrados ordinários. Dessa forma, determina-se  $\hat{\phi}$  por (3.2.3) e  $\hat{\sigma}^2$  por (3.2.5). Com estes resultados, calcula-se o valor de  $\hat{\rho}$  resolvendo a equação (3.2.13), onde se considera cada  $e_{uj}^*$  como sendo vetor de resíduos observados. Retorne-se à equação (3.2.3) para determinar um novo valor de  $\hat{\phi}$ , considerando-se o valor de  $\hat{\rho}$  encontrado anteriormente. Determina-se um novo  $\hat{\sigma}^2$  e daí, um novo  $\hat{\rho}$ . Este processo iterativo deve ser realizado até se obter uma estimativa de  $\rho$  com erro pré-fixado. Por exemplo, suponha-se que se fixe este erro em  $10^{-4}$ . Então o processo de estimação cessará quando  $|\hat{\rho}_i - \hat{\rho}_{i-1}| < 10^{-4}$ , ou seja, o valor absoluto da diferença das duas últimas iterações for menor do que  $10^{-4}$ . Com o valor de  $\hat{\rho}$  final, determina-se  $\hat{\phi}$  (ou  $\hat{\xi}$ ) e  $\hat{\sigma}^2$ .

Embora declare que necessita ser demonstrado, KLEIMBAUM (1973), baseado em estudo de simulação realizado por Kmenta & Gilbert<sup>1</sup> para estimadores de regressões aparentemente não

---

<sup>1</sup> KMENTA, J. & GILBERT, R.F. Small sample properties of estimator of seemingly unrelated regressions. J. Amer. Statist. Assoc., 62: 1180-1200, 1967.

correlacionadas. conclui que o procedimento de estimar parâmetros de forma iterativa como a descrita acima apresenta vantagens no que diz respeito a uma maior aproximação do parâmetro populacional, principalmente quando se trata de amostras pequenas.

Os estimadores de máxima verossimilhança são, em geral, consistentes. Se  $\hat{\rho}$  e  $\hat{\sigma}^2$  são estimadores consistentes de  $\rho$  e  $\sigma^2$ , então  $\hat{\Sigma}_u = \hat{\sigma}^2 \hat{V}_u$  será um estimador consistente de  $\Sigma_u = \sigma^2 V_u$  (veja prova em LAVANGE, 1983, p. 28). Daí,  $\hat{\xi}$  será um estimador BAN de  $\xi$ . Este fato possibilita a construção de intervalos de confiança e a realização de testes de hipóteses sobre funções lineares dos elementos de  $\xi$ , através da estatística de Wald.

### 3.3. Testes de Hipóteses

Como no modelo completo (capítulo 2), existe interesse em testar a hipótese  $H_0: C\xi U = \phi$ . Equivalentemente, esta hipótese é apresentada na forma  $H_0: L'\phi = \phi$ , onde  $L'(cu_1 \times qr) = U' \otimes C$  e  $\phi$  ( $qr \times 1$ ) é o vetor composto pelas colunas de  $\xi$ . A estatística de Wald é uma forma quadrática construída de um estimador BAN de  $L'\phi$  e um estimador consistente de  $\Sigma$ . Sob a hipótese  $H_0$ , é distribuída, assintoticamente, como uma variável quiquadrado central. Daí, se  $\hat{\Sigma} = \hat{\sigma}^2 \hat{V}$  é um estimador consistente de  $\Sigma$  e, conseqüentemente,  $L'\hat{\phi}$  é um estimador BAN de  $\phi$ , então a estatística de Wald será expressa por

$$W = (L'\hat{\phi})' [\text{Var} (L'\hat{\phi})]^{-1} (L'\hat{\phi}) \sim \chi^2 (c.u_1) \quad (3.3.1)$$

onde

$$\text{Var} (L'\hat{\phi}) = L' \left[ \sum_{u=1}^g (G A_u \hat{V}_u^{-1} A_u' G' \hat{\sigma}^2 \otimes X_u' X_u) \right]^{-1} L \quad (3.3.2)$$

Nas hipóteses sobre o paralelismo e a coincidência das curvas, as matrizes  $C$  e  $U$  são as mesmas do caso completo. A de coincidência será ilustrada no capítulo 4.

Um teste para o ajustamento de uma função polinomial é mencionado em WOOLSON et alii (1978). É semelhante àquele visto em 2.2 para dados completos. Considere-se o modelo (2.3.1) na forma

$$E(Y_u) = X_u \xi_* A_u \quad (3.3.3)$$

onde  $\xi_* = \xi G$ . Desenvolvendo-se a matriz  $Y_u$  ( $n_u \times p_u$ ) por colunas obtém-se o vetor  $y_u$  ( $n_u p_u \times 1$ ) e o modelo (3.3.3) será expresso por

$$E(y_u) = (A_u' \otimes X_u) \varphi_* \quad (3.3.4)$$

$$\text{Var}(y_u) = A_u' V A_u \sigma^2 \otimes I_{n_u}$$

onde  $\varphi_*$  é o vetor composto das colunas de  $\xi_*$ .

Pelo procedimento idêntico àquele de (2.3.2) a (2.3.4), obtém-se:

$$E(y) = F_* \varphi_* \quad (3.3.5)$$

$$\text{Var}(y) = \begin{bmatrix} A_1' V A_1 \otimes I_{n_1} & \phi & & \\ & \dots & \dots & \\ & & & A_g' V A_g \otimes I_{n_g} \end{bmatrix} \sigma^2$$

onde  $F_* = [A_1 \otimes X_1' \dots A_g \otimes X_g']'$ . Daí, um estimador BAN de  $\varphi_*$  é dado por

$$\begin{aligned} \hat{\varphi}_* &= \left( \sum_{u=1}^g (A_u \hat{V}_u^{-1} A_u' \otimes X_u' X_u) \right)^{-1} \sum_{u=1}^g (A_u \hat{V}_u^{-1} \otimes X_u') y_u \\ &= \sum_{u=1}^g [(A_u \hat{V}_u^{-1} \otimes X_u') (A_u' \otimes X_u)]^{-1} \sum_{u=1}^g (A_u \hat{V}_u^{-1} \otimes X_u') y_u \quad (3.3.6) \end{aligned}$$

e um estimador consistente de  $\text{Var}(\hat{\varphi}_*)$  é expresso por

$$\begin{aligned} \widehat{\text{Var}}(\hat{\varphi}_*) &= \left\langle \sum_{u=1}^g (A_u \hat{V}_u^{-1} A_u' \otimes X_u' X_u) \right\rangle^{-1} \hat{\sigma}^2 \\ &= \left\langle \sum_{u=1}^g [(A_u \hat{V}_u^{-1} \otimes X_u')(A_u' \otimes X_u)] \right\rangle^{-1} \hat{\sigma}^2 \end{aligned} \quad (3.3.7)$$

A hipótese que se quer testar (ajustamento) é expressa por

$$H_0 : G \xi_* U = \phi \quad \text{ou} \quad L' \varphi_* = \phi$$

onde  $L' [(p-q)r \times rp] = G' \otimes I_r$ .

A estatística de Wald será

$$W^* = (L' \hat{\varphi}^*)' [\widehat{\text{Var}}(L' \hat{\varphi}^*)]^{-1} (L' \hat{\varphi}^*) \sim \chi^2 [(p-q)r] \quad (3.3.8)$$

#### 4. APLICAÇÕES

Com a finalidade principal de ilustrar a teoria desenvolvida no capítulo 3, considerou-se parte do conjunto de dados obtido por BANDEIRA (1989). O experimento consistiu em observar medidas antropométricas (peso, comprimento e perímetro cefálico) em crianças de baixo peso, nascidas no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto. Para as aplicações apresentadas neste capítulo considerou-se a variável comprimento relativa a 57 crianças, nascidas aos 280 dias de gestação e com peso inferior a 2.500g, nas idades de 2,3,4,5 e 6 meses. As crianças foram divididas em 2 grupos: feminino.(1) e masculino (2), com 29 e 28 crianças, respectivamente. Os dados constam no Apêndice (Quadro A).

Neste capítulo serão aplicadas as teorias dos capítulos 2 e 3 aos dados de SOUZA (1989). Um delineamento incompleto foi escolhido ao acaso conforme o Quadro A do Apêndice, onde se excluem as observações com \*. As observações alocadas nesse delineamento correspondem a 77% do total. Realizam-se três análises de curvas de crescimento: em 4.1, com todas as

unidades experimentais (incluindo-se as unidades com \*); em 4.2, com as dezessete unidades experimentais que permaneceram completas no delineamento incompleto (isto é, excluindo-se as unidades com \*) e, em 4.3, com todas as unidades experimentais do delineamento incompleto (excluindo-se só as observações com \*). Comparações sobre os resultados das três análises são realizadas em 4.4.

Para as análises estatísticas dos três casos, utilizou-se o sistema computacional SOC (Software Científico), mais especificamente, as subrotinas MLM, MLMC e CM.



#### 4.1. Análise do MLMC com Todas as Unidades Experimentais

Relativamente aos dados do Quadro A, é apresentado abaixo, um resumo das estatísticas descritivas.

Quadro 4.1.1: Resumo das Estatísticas Descritivas da Variável Comprimento de Crianças (BANDEIRA, 1989).

SEXO	tempo	no. obs.	média	erro padrão	min.	max.
FEMININO	2	29	52,6552	0,2445	49,0	55,0
	3	29	55,9483	0,2536	52,5	58,5
	4	29	58,2241	0,2913	53,5	61,0
	5	29	60,4828	0,2714	56,5	63,0
	6	29	62,1552	0,2590	58,5	65,0
MASCULINO	2	28	52,9286	0,3743	48,5	56,0
	3	28	56,3036	0,3604	52,0	58,5
	4	28	58,9571	0,3883	54,0	62,0
	5	28	61,4464	0,3884	57,5	65,5
	6	28	63,5803	0,3818	59,5	66,5

O Gráfico a seguir representa os perfis de médias do comprimento de crianças no período de 2 a 6 meses.

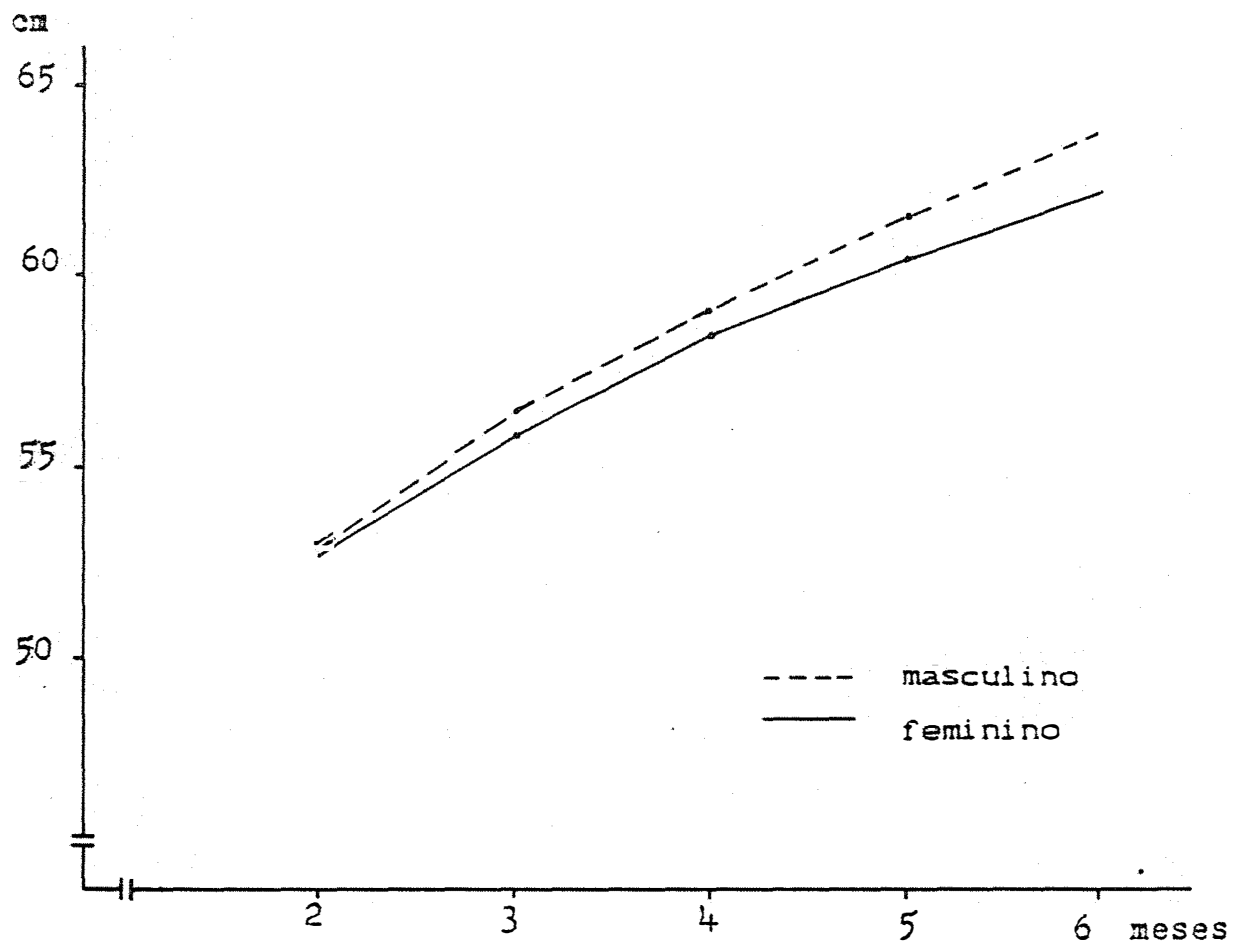


Gráfico 4.1.1: Perfis de Médias do Comprimento de Crianças ao Longo do Tempo.

O Gráfico 4.1.1 sugere que se devam ajustar polinômios de, no máximo, segundo grau, aos perfis médios de respostas. Considerando-se (2.2.1), isto é,  $E(Y) = X\xi G$ , tem-se:

$Y$  (57x5), a matriz de escores formada pelas 5 últimas colunas do Quadro A;

$X$  (57x2), a matriz de delineamento relativa à primeira coluna do Quadro A, isto é:

$$X = \begin{bmatrix} 1 & 0 \\ \dots \\ 1 & 0 \\ \hline 0 & 1 \\ \dots \\ 0 & 1 \end{bmatrix} \begin{matrix} 29 \text{ linhas} \\ \\ 28 \text{ linhas} \end{matrix}$$

$$\xi (2 \times 3) = \begin{bmatrix} \xi_{01} & \xi_{11} & \xi_{21} \\ \xi_{02} & \xi_{12} & \xi_{22} \end{bmatrix}$$

$$G (3 \times 5) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 & 6 \\ 4 & 9 & 16 & 25 & 36 \end{bmatrix}$$

A matriz de covariâncias estimada ( $\hat{\Sigma}$ ) relativa a cada indivíduo é expressa por

$$\hat{\Sigma} = \frac{S}{n - r},$$

onde, de (2.1.3),  $S = Y'[I - X(X'X)^{-1}X']Y$ ;

n = 57 crianças e

r = 2 sexos

A matriz abaixo expressa  $\hat{\Sigma}$  (parte superior) e as correlações entre as variáveis definidas nas condições de avaliação (parte inferior):

2,7438	2,3062	2,0619	1,6358	1,4178
0,8379	2,7608	2,5155	2,0124	1,9948
0,6906	0,8399	3,2488	2,7309	2,5136
0,5669	0,6952	0,8697	3,0347	2,5720
0,5002	0,7016	0,8149	0,8627	2,9287

O ajustamento do modelo é testado pela hipótese

$H_0: C\xi_* U = \phi$ , onde:

$$C = I_2,$$

$$U = G_2 = \begin{bmatrix} -1 & 1 \\ 2 & -4 \\ 0 & 6 \\ -2 & -4 \\ 1 & 1 \end{bmatrix}$$

considerando o modelo  $E(Y) = X\xi_*$ .

De (2.2.15) a (2.2.19), calcularam-se  $E_{\Delta}$ ,  $H_{\Delta}$ ,  $m_1$  e  $m_2$ . De (2.1.10), (2.1.14), (2.1.17) e (2.1.19) calcularam-se as estatísticas do Quadro 4.1.2.

Quadro 4.1.2: Resultados das Estatísticas Usadas no Teste de Ajustamento.

Parâmetros: $s = 3$ , $m_1 = 0$ , $m_2 = 18,5$			
Estatística	Resultado da Estatística	Distribuição da Estatística	Nível Descritivo
Roy	0,1106	Heck	> 0,05
Wilks	1,6554	FC(4,108)	0,1650
Lawley-Hotteling	1,6722	FC(4,106)	0,1610
Pillai	1,6375	FC(4,110)	0,1690

Portanto, não se rejeita o ajuste de uma função polinomial do segundo grau.

Na tentativa de ajuste de uma função de 1o. grau, os níveis descritivos encontrados para as estatísticas de Wilks, Lawley-Hotteling e Pillai foram todos menores do que  $10^{-4}$ . Portanto, será adotada a função do 2o. grau para explicar a variável comprimento de crianças nas idades de 2 a 6 meses.

A escolha das variáveis auxiliares (covariáveis) é realizada considerando-se a matriz de correlações entre as variáveis do espaço de estimação ( $Y_1 = YG_1$ ) e o espaço de erros ( $Y_2 = YG_2$ ), ou seja:

Y <sub>1</sub>			Y <sub>2</sub>	
1	0,0899	-0,3191	0,0277	0,1785
0,0899	1	-0,2811	-0,1704	0,0500
-0,3191	-0,2811	1	0,1048	-0,2685
0,0277	-0,1704	0,1048	1	0,0203
0,1785	0,0500	-0,2685	0,0203	1

$$e \quad G_1 = \begin{bmatrix} 1 & -2 & 2 \\ 1 & -1 & -1 \\ 1 & 0 & -2 \\ 1 & 1 & -1 \\ 1 & 2 & 2 \end{bmatrix}$$

O teste múltiplo consiste em calcular  $\tanh(z/\sqrt{n-3})$ , onde  $z$  é tal que  $P(Z > z) = \alpha/2a$ . Para  $\alpha = 0,05$  e  $a = 6$  correlações,  $z = 2,6380$ . Daí,  $\tanh z/\sqrt{n-3} = 0,3443$ . Como nenhuma das correlações entre as variáveis de  $Y_1$  e as de  $Y_2$  é maior do que 0,3443, então o modelo considerado será sem covariável.

O modelo adotado será o de POTTHOFF & ROY (1964). Por (2.2.33) e (2.2.34) obtiveram-se as seguintes estimativas:

$$\hat{\xi} = \begin{bmatrix} 45,2207 & 4,2155 & -0,2328 \\ 45,4307 & 4,1485 & -0,1877 \end{bmatrix}$$

$$DP(\hat{\xi}) = \begin{bmatrix} 0,6408 & 0,3160 & 0,0356 \\ 0,6521 & 0,3216 & 0,0362 \end{bmatrix}$$

A coincidência das 2 curvas é testada considerando-se a hipótese  $H_0: C \xi U = \phi$ , onde

$$C = \begin{bmatrix} 1 & -1 \end{bmatrix} \text{ e } U = I_3$$

De (2.2.35) e (2.2.36) obtiveram-se, respectivamente,  $E_3$  e  $H_3$ . Os parâmetros serão

$$s = \min(c, u_1) = \min(1, 3) = 1,$$

$$m_1 = \frac{|c - u_1| - 1}{2} = 0,5$$

$$m_2 = \frac{n - r - u_1 - 1}{2} = 25,5$$

De (2.1.10), (2.1.14), (2.1.17) e (2.1.19) calcularam-se as estatísticas do Quadro 4.1.3.

Quadro 4.1.3: Resultados do Teste de Coincidências das 2 Polinômiais (57 crianças, MLMC)

Parâmetros: $s = 1$ , $m_1 = 0,5$ e $m_2 = 25,5$			
Estatística	Resultado da Estatística	Distribuição da Estatística	Nível Descritivo
Roy	4,7217	F (3,53)	0,0054
Wilks	4,7217	F (3,53)	0,0054
Lawley-Hotteling	4,1923	F (3,53)	0,0098
Pillai	4,7217	F (3,53)	0,0054

Portanto, rejeita-se a hipótese de coincidência das duas curvas, ou seja, existe o efeito de sexo na variável comprimento de crianças nas idades de 2 a 6 meses.

#### 4.2. Análise do MLMC com Parte das Unidades Experimentais

Como já foi dito anteriormente, após o estabelecimento do delineamento incompleto, dezessete crianças continuaram com observações completas. Neste item será realizada a análise do MLMC considerando apenas estas crianças, cujas observações correspondem às dezessete primeiras linhas do Quadro A do Apêndice.

A matriz estimada dos parâmetros foi calculada como

$$\hat{\xi} = \begin{bmatrix} 44,6667 & 4,3008 & -0,2341 \\ 47,4250 & 3,3196 & -0,1071 \end{bmatrix}$$

e a estimativa da matriz dos desvios padrão de  $\hat{\xi}$ ,

$$\hat{DPC}(\hat{\xi}) = \begin{bmatrix} 1,1976 & 0,6457 & 0,0705 \\ 1,2702 & 0,6849 & 0,0748 \end{bmatrix}$$

A coincidência das 2 curvas foi testada considerando as mesmas matrizes da hipótese da análise anterior. Os resultados são apresentados a seguir.



Quadro 4.2.1: Resultados do Teste de Coincidências das 2 Polinômiais (MLMC, 17 crianças)

Parâmetros: $s = 1$ , $m_1 = 0,5$ e $m_2 = 5,5$			
Estadística	Resultado da Estadística	Distribuição da Estadística	Nível Descritivo
Roy	2,7477	F (3,13)	0,0850
Wilks	2,7477	F (3,13)	0,0850
Lawley-Hotteling	2,7477	F (3,13)	0,0850
Pillai	2,7477	F (3,13)	0,0850

Os resultados desta análise não permitem fazer distinção dos dois grupos (feminino e masculino). Um novo modelo deverá ser ajustado incorporando a nova informação. Este modelo será:

$$E(Y) = X\xi G$$

onde Y e G são as mesmas matrizes do modelo anterior,

$X(17 \times 1)$  é o vetor de 1's e

$$\xi(1 \times 3) = [\xi_0 \quad \xi_1 \quad \xi_2]$$

As estimativas das matrizes  $\xi$  e  $DPC(\xi)$  são, respectivamente, expressas por

$$\hat{\xi} = [ 45,9647 \quad 3,8391 \quad -0,1744 ]$$

$$\hat{DPC}(\hat{\xi}) = [ 0,9112 \quad 0,4711 \quad 0,0521 ]$$

#### 4.3. Análise do MLMCG com a Restrição ARC(1), no Delineamento Incompleto

Agrupando-se os dados de BANDEIRA (1989), conforme o delineamento incompleto citado, obtém-se o Quadro A, sem as observações com \*.

De maneira análoga a 4.1 e 4.2, propõe-se que o comprimento das crianças na faixa de 2 a 8 meses seja explicado por polinomiais do 2o. grau, ou seja:

$$E(Y_u) = X_u \xi + A_u$$

onde  $Y_u$  é a matriz definida pelo u-ésimo padrão de observações perdidas,  $u = 1, \dots, 9$ . Por exemplo,  $Y_1$  (17 x 5) é a matriz de observações relativa ao grupo 1 do Quadro A do Apêndice;  $Y_2$  (5 x 4) é a matriz relativa ao grupo 2, onde se desprezam as observações com \*, etc...

$X_u$  é a matriz de delineamento relativa a  $Y_u$ ; Por exemplo,

$$X_1 = \begin{bmatrix} 1 & 0 \\ \dots \\ 1 & 0 \\ \hline 0 & 1 \\ \dots \\ 0 & 1 \end{bmatrix} \begin{array}{l} 9 \text{ linhas} \\ \\ \\ 8 \text{ linhas} \end{array}; \quad X_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}; \text{ etc...}$$

$\xi$  e  $G$  são as mesmas matrizes das análises anteriores;

$A_u$  é a matriz de incidência relativa a  $Y_u$ . Por exemplo,

$$A_1 = I_5; \quad A_2 = \begin{bmatrix} I_4 & \\ 0 & 0 & 0 & 0 \end{bmatrix}; \text{ etc...}$$

Supondo que os resíduos são autocorrelacionados segundo o modelo AR(1), conforme visto no capítulo 3, estimaram-se os parâmetros,  $\xi$ ,  $\sigma^2$  e  $\rho$ . Utilizando-se (3.2.3) e (3.2.5), com  $V_u = I_{p_u}$ , calculou-se, em um primeiro passo,  $\hat{\rho}$  (ou  $\hat{\xi}$ ) e depois  $\hat{\sigma}^2$ . Com estes valores e considerando-se apenas as observações consecutivas separadas pela distância 1, calculou-se  $\hat{\rho}$  por (3.2.13). Retornando a (3.2.3) e (3.3.5), com  $V_u$  calculado de (3.1.6), determinou-se um novo  $\hat{\rho}$  (ou  $\hat{\xi}$ ) e um novo  $\hat{\sigma}^2$ . Esse processo foi realizado até obter uma diferença menor do que  $10^{-3}$  em duas estimativas consecutivas de  $\rho$ . Encontrou-se

$$\hat{\xi} = \begin{bmatrix} 44,9421 & 4,3903 & -0,2609 \\ 45,4136 & 4,3250 & -0,2220 \end{bmatrix}$$

$$\widehat{DP}(\xi) = \begin{bmatrix} 0,8243 & 0,4114 & 0,0494 \\ 0,8708 & 0,4302 & 0,0531 \end{bmatrix}$$

$$\hat{\sigma}^2 = 0,9248;$$

$$\hat{\rho} = 0,8136$$

A hipótese de ajustamento do modelo é formulada como nas análises anteriores, isto é,  $H_0: C\xi_*U = \phi$ , onde  $C = I_2$ ,  $U = G_2$  e  $\xi_* = \xi G$ . Esta formulação é equivalente a  $H_0: L'\phi_* = \phi$ , onde  $L' = G'_2 \otimes I_2$ . Determinou-se  $\hat{\phi}_*$  (ou  $\hat{\xi}_*$ ) por (3.3.6),  $\widehat{\text{Var}}(\hat{\phi}_*)$  por (3.3.7) e a estatística de Wald por (3.3.8). Para esta obteve-se o resultado  $W^* = 6,0653$ . O nível descritivo relativo a este valor, considerando a distribuição  $\chi^2(4)$ , foi de 0,1943, o que implica na não rejeição do modelo.

Outro teste realizado foi o da coincidência das 2 curvas. As matrizes C e U são as mesmas das análises anteriores (do teste de coincidência), ou seja, considera-se a hipótese  $H_0: C\xi U = \phi$ , ou  $H_0: L'\phi = \phi$  no modelo (3.2.1), onde  $C = [1 \ -1]$ ,  $U = I_3$  e  $L' = U' \otimes C$ . Utilizando-se  $\hat{\phi}$  (ou  $\hat{\xi}$ ),  $\hat{\sigma}^2$  e  $\hat{\rho}$  determinados acima, calcula-se  $\widehat{\text{Var}}(L'\hat{\phi})$  por (3.3.2) e daí, a estatística de Wald  $W^* = 9,9453$ , por (3.3.1). Considerando-se a distribuição  $\chi^2(3)$ , o nível descritivo é de 0,0190. Conclui-se que as curvas ajustadas para os sexos feminino e masculino são distintas. Portanto, a estimativa da matriz dos parâmetros do modelo ajustado e respectiva matriz de desvios padrão são aquelas já determinadas.

#### 4.4. Comparação dos Resultados

A comparação dos resultados de 4.1, 4.2 e 4.3 é importante no sentido de situar o estimador obtido de 4.3 com relação àqueles de 4.1 e 4.2. Por se basear em uma quantidade maior de dados, espera-se que o estimador de 4.1 seja mais preciso. Os resultados da análise confirmam esse fato. Espera-se, também, que o estimador de 4.3 esteja entre o de 4.1 e o de 4.2, em termos de precisão, se a suposição adicional relativa aos resíduos for correta. O trabalho consiste em melhorar a precisão do estimador de  $\xi$ , e isto deve ser feito utilizando-se o máximo de dados disponíveis, ou seja, não se desprezando as observações individuais incompletas. O que se nota pelo Quadro 4.4.1 é que o método proposto melhorou a precisão da estimativa de  $\xi$ , uma vez que a matriz dos desvios padrão diminuiu em relação àquela onde se utilizam apenas os 17 indivíduos de observações completas.

Quadro 4.4.1. Apresentação das Estimativas dos Desvios Padrão dos Parâmetros das Polinomiais de Segundo Grau Encontradas nas Três Análises.

Modelo	Feminino			Masculino		
MLMC (57 ind.)	0,6408	0,3160	0,0356	0,6521	0,3216	0,0362
MLMC (17 ind.)	0,9112	0,4711	0,0521	0,9112	0,4711	0,0521
MLMCG com ARC(1)	0,8243	0,4114	0,0494	0,8708	0,4392	0,0531

Na mesma disposição do Quadro 4.4.1, o Quadro 4.4.2 apresenta as estimativas dos parâmetros. O que se nota é que os parâmetros estimados na análise 4.3 estão mais próximos dos parâmetros estimados da análise 4.1 do que aqueles da análise 4.2.

Quadro 4.4.2. Apresentação das Estimativas dos Parâmetros das Polinomiais de Segundo Grau encontradas nas Três Análises.

Modelo	Feminino			Masculino		
MLMC (45 ind.)	45,2207	4,2155	-0,2328	45,4307	4,1485	-0,1877
MLMC (17 ind.)	45,9647	3,8391	-0,1744	45,9647	3,8391	-0,1744
MLMCG com ARC(1)	44,9421	4,3903	-0,2609	45,4136	4,3250	-0,2220

Relativamente a cada sexo, construíram-se os gráficos 4.1.2 e 4.1.3. Observa-se que, em ambos, a curva estimada no MLMCG está mais próxima da estimada no MLMC (57 indivíduos) do que no MLMC (17 indivíduos).

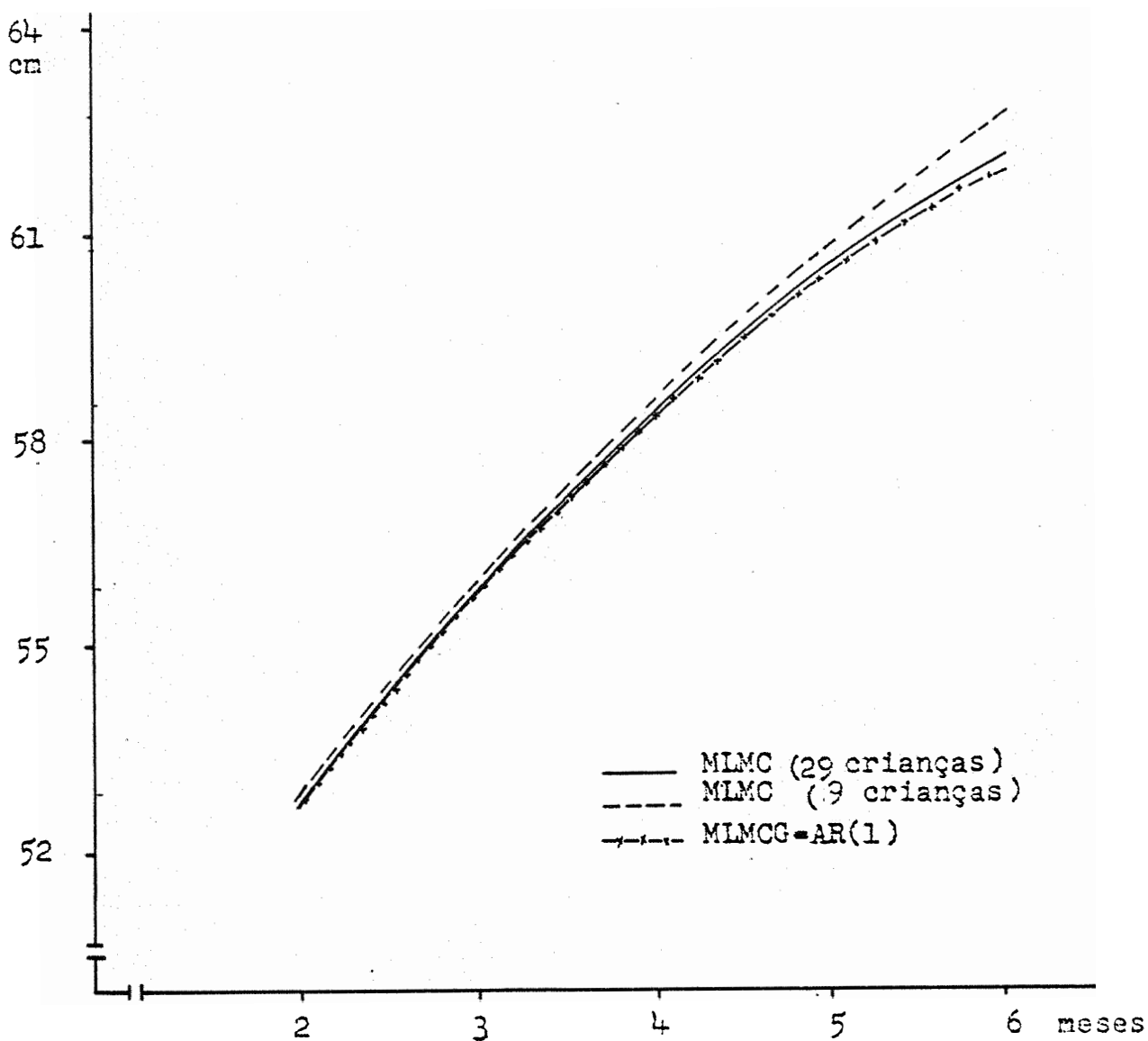


Gráfico 4.1.2: Polinômios do 2o. grau ajustados ao comprimento de crianças do sexo feminino.

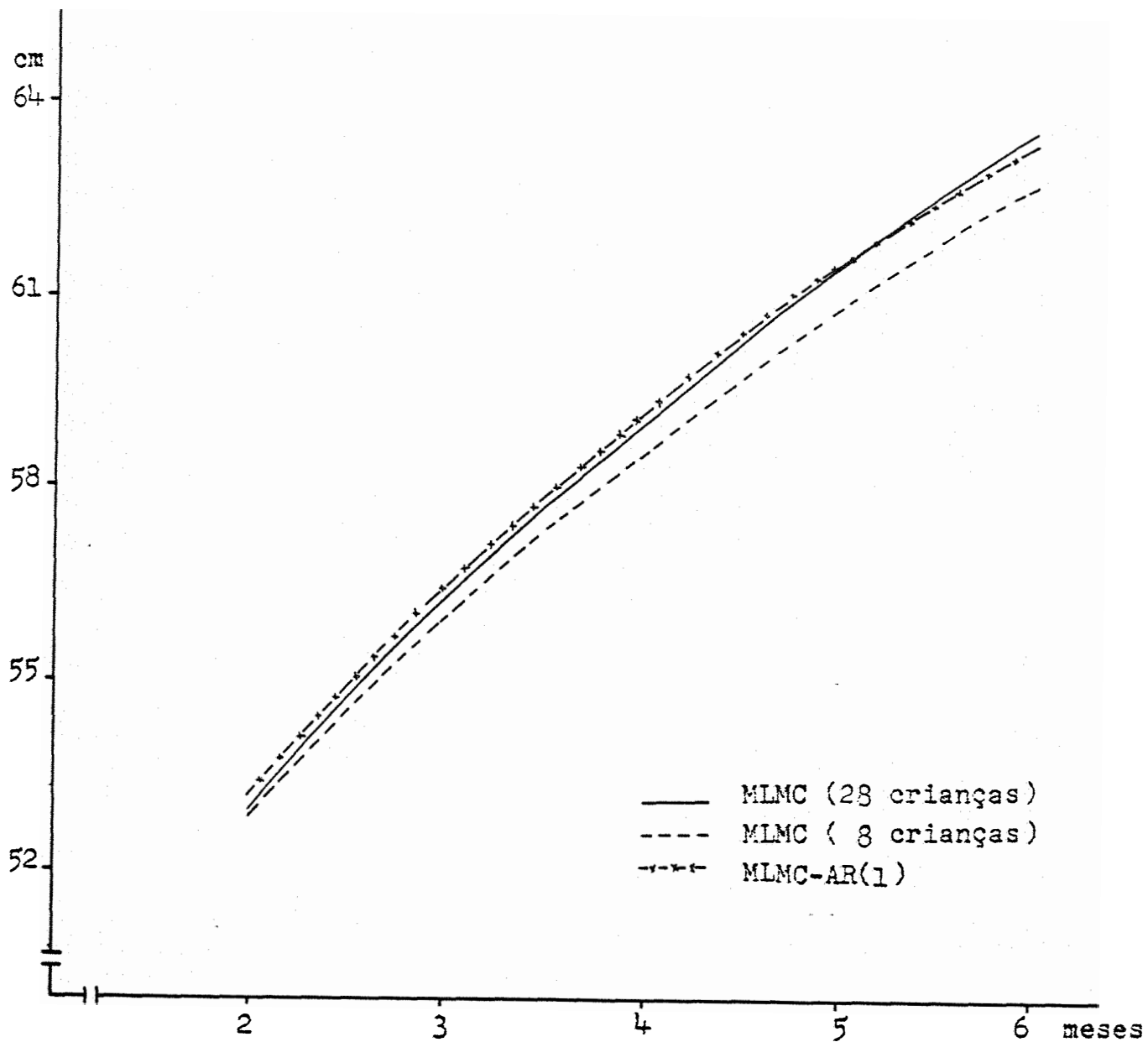


Gráfico 4.1.3: Polinômios do 2o. grau ajustados ao comprimento de crianças do sexo masculino.



## 5. CONCLUSÕES

Se houver perdas de observações ou se o experimento for planejado com perdas, o método proposto por KLEIMBAUM (1973) é indicado para a análise de curvas de crescimento. No entanto, pode acontecer da matriz de covariâncias estimada ser negativa definida. Nesse caso, uma solução é aplicar a metodologia desenvolvida no capítulo 3, desde que a restrição feita aos resíduos (modelo auto-regressivo de 1a. ordem) seja observada. Essa metodologia pode também ser aplicada em experimentos completos com muitas condições de avaliação, pois os métodos tradicionais (capítulo 2) implicam na estimação da matriz de covariâncias, a qual envolve  $p(p+1)/2$  parâmetros. É de se supor que muitos desses parâmetros serão mal estimados.

Uma outra aplicação da metodologia do capítulo 3 é analisar experimentos completos quando o número de indivíduos é menor do que o número de condições de avaliação. Nesse caso, pelo procedimento tradicional, a matriz de covariâncias estimada não será positiva definida.

Testes para auto-correlação existem para uma série longa

de observações, como, por exemplo, o teste de DURBIN & WATSON (1950). Para um futuro trabalho sugere-se que seja desenvolvido esse teste para o problema em questão, ou seja, quando se tem várias séries curtas. Outra sugestão é verificar, por simulação, o comportamento da estimativa de  $\rho$  da forma que foi proposta no capítulo 3, quanto à sua consistência.

Atualmente, o autor deste trabalho está desenvolvendo um programa computacional para a análise proposta, utilizando o sistema "Software Científico" (SOC).

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, T.W. . An Introduction to Multivariate Analysis, New York, John Wiley & Sons, Inc. , 1958, 374 p.
- ANDRADE, D.F. & SINGER, J.M. Uso do SOC na análise de curvas de crescimento. EMBRAPA, Campinas, S.P., 1988.
- BAKSALARY, J.K.; CORNSTEN, L.C. & KALA, R.. Reconciliation of two different views on estimation of growth curve parameters. Biometrika, London, 65 (3): 662-665, 1978.
- BOX, G.E.P. Problems in the analysis of growth and wear curves. Biometrics, Raleigh, 6: 362-371, 1950.
- DURBIN, J. & WATSON, G.S. Testing for serial correlation in least squares regressions. I. Biometrika, London, 409-428, 1950.
- ELSTON, R.C. & GRIZZLE, J.E. Estimation of time-response curves and their confidence bands. Biometrics, Raleigh, 18: 148-159, 1962.

- GODOI, C.R.M. Análise Estatística Multidimensional. In: I SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA e XXX REUNIÃO DA BIOMETRIA (RBRAS). Piracicaba, 1985, 187 p.
- GRIZZLE, J.E. & ALLEN, D.M. Analysis of growth and dose response curves. Biometrics, Raleigh, 25 (2): 357-381, 1969.
- HECK, D.L. Charts of some upper percentage points of distribution of the largest characteristic root. Annals of Mathematical Statistics, Baltimore, 31: 625 - 642, 1960.
- HOEL, P.G. Methods for comparing growth type curves. Biometrics, Raleigh, 20: 859-872, 1964.
- KHATRI, C.G. A note on a MANOVA model applied to problems in growth curve. Ann. Inst. Statist. Math., 18: 75-86, 1966.
- KLEIMBAUM, D.G. A generalization of the growth curve model which allow missing data. Journal of the Multivariate Analysis, 3: 117-124, 1973.
- KOWALSKI, J.C. & GUIRE, K.E. Longitudinal data analysis. Growth, 38: 131-169, 1974.
- LAIRD, N.M. & WARE, J.H. Random effects models for longitudinal data. Biometrics, Raleigh, 38: 963-974, 1982.

LAVANGE, L.M. Analysis of incomplete longitudinal data with constrained covariance structures. Chapel Hill, 1983, 134 p. (Ph.D. dissertation/ University of North Carolina)

LEE, Y.H.K. A note on Rao's reduction of Potthoff & Roy's generalized linear model. Biometrika, London, 61 (2): 349-351, 1974.

LEECH, F.B. & HEALY, M.J.R. The analysis of experiments on growth rate. Biometrics, Raleigh, 15: 98-106, 1959.

POTTHOFF, R.F. & ROY, S.N. A generalized multivariate of variance model useful especially for growth curve problems. Biometrika, London, 51 (3 & 4), 313- 326, 1964.

MORRISON, F. D. Multivariate Statistical Methods. Second edition. New York, McGraw-Hill Inc., 1976, 415 p.

RAO, C.R. Some statistical methods for comparison of growth curves. Biometrics, Raleigh, 14: 1-17, 1958.

RAO, C.R. Some problems involving linear hypotheses in multivariate analysis. Biometrika, London, 46: 49-58, 1959.

RAO, C.R. The theory of least squares when the parameters are stochastic and its application to the analysis of growth

curves. Biometrika, London, 52 (3 e 4): 447-458, 1965.

RAO, C.R. Covariance adjustment and related problems in multivariate analysis. In: KRISHNAIAH, P. R., ed. Multivariate Analysis II., New York, Academic Press, 1966, p. 87-103.

RAO, C.R. Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Proceedings of the 5<sup>th</sup> BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS, Berkeley, 1: 355-372, 1967.

RAO, M.N. & RAO, C.R. Linked cross-sectional study for determining norms and growth rates - a pilot survey on Indian school-going boys. Sankhya, series B, Calcutta, 28: 237-258, 1967.

SINGER, J.M. Análise de curvas de crescimento. São Paulo, 1977, 113 p. (Dissertação de Mestrado/ IME-USP).

SINGER, J.M. & ANDRADE, D.F. Análise de dados longitudinais. In: VII SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, Campinas, 1986, 106 p.

SOUZA, J.B. Peso, comprimento e perímetro cefálico de crianças de baixo peso no nascimento, até o 6o. mês pós-termo.

Ribeirão Preto, 1989, (Dissertação de Mestrado/ FMRP-USP)

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc., 54: 426-482, 1943.

WISHART, J. Growth-rate determination in nutrition studies with the bacon pig, and their analysis. Biometrika, London, 30: 16-28, 1938.

WOOLSON, R.F; LEEPER, J.D. & CLARKE, W.R. Analysis of incomplete data from longitudinal and mixed longitudinal studies. J. R. Statist. Soc., serie A, London, 141, (2): 242-252, 1978.

WOOLSON, R.F. & LEEPER, J. D. Growth curve analysis of complete and incomplete longitudinal data. Commun. Statist. - Theor. Meth., serie A, 9 (14): 1491-1513, 1980.

## APÊNDICE

Quadro A: Valores Observados de Comprimento de 57 Crianças com Baixo Peso no Nascimento, no HC da Faculdade de Medicina de Ribeirão Preto, USP (Dados de SOUZA, 1989)

1: Feminino; 2: Masculino

sexo	tempo após o nascimento (meses)				
	2	3	4	5	6
1	52,0	55,5	60,0	63,0	64,5
1	51,0	55,0	56,0	59,0	59,5
1	52,0	56,5	57,5	59,0	62,5
1	53,5	56,5	58,0	61,5	62,0
1	51,0	52,5	53,5	56,5	58,5
1	52,0	55,0	58,0	61,0	63,5
1	54,0	58,5	59,5	61,0	62,5
1	52,0	55,0	59,0	61,5	63,0
1	53,0	56,5	59,0	62,0	62,0
2	56,0	57,0	58,5	60,0	62,0
2	50,5	53,5	56,5	59,0	61,0
2	52,0	55,0	57,0	60,0	63,5
2	54,5	57,5	59,0	62,0	64,0
2	53,5	58,0	60,5	62,5	64,5
2	53,0	57,0	60,5	63,5	65,0
2	54,0	58,0	60,0	62,0	65,5
2	54,5	58,0	58,5	61,0	63,0
1	52,0	56,5	57,5	59,0	62,5*
1	55,0	57,5	60,0	62,0	63,0*
1	54,5	56,5	58,5	60,5	62,0*
2	55,5	58,0	62,0	63,5	65,5*
2	54,5	57,5	60,0	61,0	62,5*
1	54,0	56,5	59,0*	61,0	62,0
1	52,0	55,5	59,0*	61,0	62,5
1	53,0	55,5	56,5*	59,0	61,5
1	52,0	55,0	58,0*	61,0	64,0
2	53,0	58,0	60,5*	62,5	64,5
2	52,0*	55,0	57,0	59,0	61,0
2	54,5*	58,5	61,0	63,5	66,0
2	53,0*	55,5	59,0	61,0	63,0
2	48,5*	52,0	56,0	59,0	60,5
2	53,0*	56,5	57,5	59,5	61,5

continua



continuação do Quadro A

sexo	tempo após o nascimento (meses)				
	2	3	4	5	6
1	52,0	55,5*	59,0	60,0	61,5
1	53,0	56,5*	59,5	62,5	62,5
2	51,0	55,0*	59,0	63,5	65,0
2	51,5	53,5*	56,0	57,5	59,5
2	55,5	58,5*	61,5	64,0	66,5
1	53,0	55,5	58,0	59,5*	60,0*
1	52,0	54,0	58,0	59,5*	60,0*
2	54,0	58,0	60,5	62,0*	65,0*
2	49,5	55,0	58,0	60,0*	63,5*
2	54,5	57,0	59,5	62,5*	64,5*
1	51,5*	56,0	57,0	60,0	62,5*
1	53,5*	56,5	59,0	61,5	63,0*
1	55,0*	58,0	61,0	62,0	65,0*
2	51,0*	55,0	57,0	61,0	64,0*
2	54,0*	57,5	61,0	62,0	65,0*
1	52,0*	56,0*	57,5	61,0	62,0
2	53,0*	57,0*	61,0	64,0	66,0
2	49,5*	52,0*	54,0	58,0	60,0
2	52,0*	55,0*	58,0	61,5	62,5
2	54,5*	58,0*	61,5	65,5	66,0
1	54,0*	57,0*	58,0	60,0*	62,0
1	52,0*	54,0*	56,0	58,5*	61,5
1	54,0*	57,0*	58,5	62,0*	63,0
1	53,0*	58,5*	61,0	61,0*	63,0
1	49,0*	54,0*	57,0	58,5*	61,5