

SANTOS

UM ESTUDO EM MEDIDAS REPETIDAS

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Sra. *Adriana Barbosa Santos* e aprovada pela Comissão Julgadora.

Campinas, 19 de Abril de 1991



Profa. Dra. Regina C. C. P. Moran
Orientadora

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

*Aos meus pais,
Josias e Tereza,
com carinho*

AGRADECIMENTOS

Reservo esta página de meu trabalho para agradecer às importantes contribuições que foram dadas durante o período em que estive cursando o mestrado.

Inicialmente, gostaria de agradecer o apoio, amizade e ajuda de vários colegas e amigos que contribuíram para que o mestrado se tornasse uma experiência marcante e agradável.

A Profa. Regina Moran, pelo incentivo e, principalmente, pelos conhecimentos que adquiri sob a sua orientação.

Aos professores do Departamento de Estatística, de modo geral, que contribuíram para minha melhor formação.

Aos colegas do IBILCE-UNESP, por colaborarem para que, através de afastamentos parciais, fosse possível chegar mais facilmente ao término deste trabalho.

Ao CNPq e CAPES, por fornecerem recursos financeiros.

Aos meus pais, pelo grande apoio e por sempre incentivarem o prosseguimento de meus estudos.

A "Deus", por me permitir este momento de satisfação.

O presente trabalho consiste do estudo de algumas técnicas de análise de curva de crescimento, visando fornecer resultados e interpretações decorrentes de adaptações e complementações adequadas destas técnicas.

Neste contexto, destaca-se:

(i) as interrelações que envolvem o Modelo de Curva de Crescimento de Potthoff & Roy(64) e o Ajuste por covariáveis de Rao(65,66).

(ii) complementações do método das Diferenças Sucessivas, cujas idéias foram introduzidas por Box(50) e posteriormente exploradas por Hills(68) e outros.

(iii) o estudo dos dados incompletos, ressaltando a generalização proposta por Kleinbaum(73) (Modelo de Curva de Crescimento Generalizado) e a adaptação do método das Diferenças Sucessivas para dados incompletos.

(iv) a ilustração de alguns resultados apresentados nos itens anteriores e comparação de técnicas de análise, através de um experimento real.

ÍNDICE

Capítulo 1 - Introdução	01
1.1 <i>Considerações Iniciais</i>	01
1.2 <i>Análise de Medidas Repetidas: Breve Revisão</i>	03
1.3 <i>Estrutura e Conteúdo Geral do Estudo</i>	09
Capítulo 2 - Alguns Métodos de Análise de Medidas Repetidas	13
2.1 <i>Introdução</i>	13
2.2 <i>Método de Análise Univariada</i>	14
2.3 <i>Métodos Multivariados; O Modelo de Curva de Crescimento e o Ajuste por Covariáveis</i>	23
Capítulo 3 - Uma Abordagem do Método das Diferenças Sucessivas	30
3.1 <i>Introdução</i>	29
3.2 <i>O Método DS no Caso de um Grupo</i>	31
3.3 <i>Extensão do Método DS para mais de um Grupo</i>	41
3.4 <i>Considerações sobre a Matriz de Variâncias e Covariâncias</i>	44
Capítulo 4 - Alguns Aspectos Sobre Dados Incompletos	47
4.1 <i>Introdução</i>	47
4.2 <i>O Modelo de Curva de Crescimento Generalizado</i>	48
4.3 <i>Utilização do Método das Diferenças</i>	

<i>Sucessivas em Dados Incompletos</i>	55
Capítulo 5 - Ilustração	72
5.1 <i>Introdução</i>	72
5.2 <i>O Experimento</i>	72
5.3 <i>Análise do Experimento</i>	73
5.4 <i>Comentários</i>	88
Considerações Finais	89
Apêndice .A <i>Observações sobre o Modelo Auto-Regressivo de Primeira Ordem</i>	91
Apêndice .B <i>Resultados numéricos da análise experimental</i>	93
Apêndice .C <i>Programas utilizados para obtenção dos resultados</i>	100
Apêndice .D <i>Dados utilizados na ilustração</i>	109
Referências Bibliográficas	115

CAPÍTULO 1

1.1 CONSIDERAÇÕES INICIAIS

Um experimento de medidas repetidas (MR) se distingue pelo fato de que uma sequência de medidas de uma mesma variável é tomada de cada unidade experimental, constituindo um perfil da unidade. Dados provenientes de estudos com esta característica podem ser representados através de vetores aleatórios p -variados $y_{ij} = (y_{ij1}, \dots, y_{ijp})'$, $i=1, \dots, a$ e $j=1, \dots, n_i$, onde y_{ijk} é a resposta observada da j -ésima unidade (indivíduo) do i -ésimo grupo na k -ésima ocasião (dose, tempo, subparcela, etc).

O fato de uma mesma variável ser observada ao longo de p ocasiões num mesmo indivíduo enseja aparecimento de correlação entre as respostas y_{ijk} , isto é, comumente nos estudos de MR, a matriz de variâncias e covariâncias associada a y_{ij} não é diagonal, sendo convenientemente estruturada em grande parte deles.

Considere que cada vetor aleatório multivariado tenha valor esperado θ_{-i} , $\theta_{-i} = (\theta_{i1}, \dots, \theta_{ip})'$, matriz de variâncias e covariâncias Σ e que os perfis sejam independentes uns dos outros,

cada um com distribuição normal p -variada. Portanto, os vetores θ_1 's constituem os perfis médios dos grupos. A estes perfis podem ser associadas curvas que, comparadas, estabelecem as diferenças e semelhanças entre os grupos de indivíduos. Estas comparações envolvem principalmente as questões do paralelismo via ajuste de polinômios.

Destacam-se como estruturas de Σ o padrão de uniformidade e o de correlação serial, pois, nestes casos, considerando algumas correções nos graus de liberdade, é possível proceder com uma análise de variância univariada (Capítulo 2).

Dentre os diversos problemas abordados nos estudos de MR, merece destaque a questão do ajuste de polinômios que descrevam os perfis médios associados à determinação dos coeficientes por métodos multivariados, com ou sem covariáveis (Rao(65,66,67); Leech e Healy(59); Grizzle e Allen(69); Potthoff e Roy(64); Lee(74); Kenward(85), etc).

Em grande parte dos experimentos é apropriada a metodologia multivariada, entretanto, há situações nas quais sua aplicação pode não ser conveniente. Quando o vetor de respostas y_{ij} não é inteiramente observado, surgem problemas de dimensão nos vetores y_{ij} que definem as linhas da matriz de dados Y , usualmente associada a um modelo multivariado, além de dificuldades de obtenção de estimativas de Σ , considerando as observações existentes. Uma das opções é recorrer ao método alternativo das Diferenças Sucessivas (DS) ou optar pela generalização multivariada de Kleinbaum(73) (Capítulos 3 e 4).

1.2 ANÁLISE DE MEDIDAS REPETIDAS: Breve Revisão

Nota-se na literatura que o precursor das discussões de Análise de Curva de Crescimento foi Wishart(38). Em seu artigo Wishart apresentou a análise para taxas de crescimento em um estudo de nutrição com porcos. Em seu estudo, utilizou polinômios ortogonais para transformar os dados originais e aplicou mínimos quadrados para estimar as taxas de crescimento e a mudança nas taxas de crescimento para cada animal. Comparou, ainda, grupos de tratamento via análise de variância univariada. Box(50) fez um estudo dos problemas envolvendo o padrão de uniformidade da matriz de variâncias e covariâncias através de um exemplo, enfocando o desgaste de equipamentos. Dando outras contribuições, Rao(58) discute algumas generalizações dos resultados de Wishart(38), utilizando polinômios ortogonais.

Na comparação de métodos univariados e multivariados, Elston e Grizzle(62) destacam-se pelo resultado de seu trabalho, onde mostram que as bandas de confiança calculadas pela suposição do modelo misto são compatíveis com as de Rao(59) ou possivelmente mais estreitas. Trabalham, no entanto, sob hipótese de matriz de variâncias e covariâncias com padrão de uniformidade, enfocando a utilização de métodos univariados.

Os efeitos de ignorar a dependência das observações, tomadas nos diferentes pontos do tempo, e a validade dos métodos estatísticos que não supõem esta correlação entre as medidas foram discutidos por Hoel(64) que, através de simulações, concluiu sobre o perigo da não consideração da dependência.

Com grande destaque nas questões abrangendo Análise de Curva de Crescimento, aparecem Potthoff e Roy(64) originando o Modelo de Curva de Crescimento (MCC). Este consiste em uma

generalização do usual Modelo de Análise de Variância Multivariada. Entretanto, embora bem geral, o método de estimação da matriz de parâmetros ξ , adotado pelos autores, gerou algumas discussões. Potthoff e Roy(64) utilizam-se da estimação por mínimos quadrados ponderados para obter $\hat{\xi}$, introduzindo uma matriz de ponderação G , a qual é arbitrária e positiva definida. O MCC é expresso como um modelo multivariado, $E(Y_0) = A\xi P$, no qual A é a matriz de especificação dos indivíduos dentro dos grupos e P , matriz relativa as ocasiões, relaciona o grau do polinômio às ocasiões. A escolha da matriz de ponderação é um dos pontos discutidos na literatura.

Criticas quanto a esta arbitrariedade foram lançadas nos artigos de Rao (65,66), nos quais optou-se pela utilização de um método baseado no ajuste por covariáveis. Reescreve o MCC como um modelo condicional, no qual as covariáveis (coeficientes polinomiais de maior grau) de esperança zero são introduzidas para melhorar a estimação dos parâmetros, isto é, a fim de reduzir a perda de informação.

Com a intenção de investigar a escolha de G , Khatri(66) mostrou que o estimador de máxima verossimilhança de ξ para o MCC é dado pela ponderação feita com a estimativa de máxima verossimilhança da matriz de variâncias e covariâncias S .

Este resultado é idêntico ao de Rao(65,66), não usando as covariáveis, assim como ao de Potthoff e Roy(64), tomando $G=S$ como matriz de ponderação.

O estimador dado acima, como mostram Grizzle e Allen(69), é também não viciado e foi obtido utilizando os resultados de Rao(67). O trabalho de Grizzle e Allen(69) apresenta uma discussão com respeito à escolha das covariáveis no modelo condicional de Rao(65,66), principalmente no sentido da redução do número de

covariáveis. A discussão é ilustrada através de um exemplo onde exploram a possibilidade de um subconjunto das covariáveis conter toda informação do conjunto de todas as covariáveis.

Vale ressaltar que no modelo de Análise de Variância Multivariada (MANOVA) pelo menos três diferentes estatísticas são utilizadas para testar hipóteses do tipo $H_0: C\xi V = 0$ (C e V matrizes conhecidas). Todas baseadas em soma de quadrados devido ao erro e à hipótese, SQ_E e SQ_H , respectivamente. As estatísticas mencionadas são expressas como segue:

- i. $|SQ_E|/|SQ_E + SQ_H|$: critério de Wilks
- ii. $\text{tr}(SQ_H SQ_E^{-1})$: estatística do traço de Lawley-Hotelling
- iii. o maior autovalor de $(SQ_H SQ_E^{-1})$: estatística de Pillai

De acordo com Khatri(66), estas estatísticas aplicam-se analogamente ao MCC.

Outro trabalho interessante envolvendo relações entre os resultados de Potthoff e Roy(64) e Rao(65,66,67) é a nota de Lee(74), que verifica a equivalência das estimativas dos parâmetros propostos pelos autores acima quando o ajuste de Rao é feito sem covariáveis.

Beauchamp e Hoel(73) utilizam-se da análise canônica, técnica multivariada, na diferenciação de grupos de observação multivariados com relação de dependência entre respostas de uma unidade experimental. Kabe(75) encontrou a distribuição de $\hat{\xi}$ e $\hat{\Sigma}$ (estimadores de máxima verossimilhança de ξ e Σ) em termos das variáveis originais e discute a aplicação dos resultados para o problema de predição de uma observação futura. Tubbs et al(75) relaxam a suposição de independência entre as linhas de Y_0 ,

considerando não mais $V(y_0) = I \otimes \Sigma$ mas $V(y_0) = \psi \otimes \Sigma$, onde y_0 é um vetor das linhas de Y_0 empilhadas e ψ é uma matriz $(n \times n)$ definida positiva. Neste trabalho, determinam o estimador de máxima verossimilhança dos parâmetros do MCC com esta estrutura de covariâncias geral. Zerbe e Walker(77) e Zerbe(79) introduziram um teste de aleatorização para comparar curvas médias de crescimento, relaxando as suposições de normalidade multivariada. Utilizam-se de tempos de observação idênticos para todos os indivíduos e mesmo grau de polinômio ajustado para as medidas de todos os indivíduos.

Para verificação da estrutura de correlação serial da matriz de variâncias e covariâncias, Hearne et al(83) derivam o teste da razão de verossimilhança para verificar a validade da hipótese da referida estrutura.

Num trabalho mais recente, Kenward(85,87) discute ganhos de precisão no ajuste por covariáveis sob certas estruturas de covariâncias, alertando para o fato de que em alguns casos estes ganhos podem ser ilusórios. Também discutindo o modelo condicional de Rao(65,66,67), Verbyla(86) sugere a redução a priori do número de covariáveis sob algumas estruturas de matrizes de variâncias e covariâncias, mostrando que a redução a priori depende da relação entre a estrutura da inversa da matriz de variâncias e covariâncias e a matriz de delineamento do perfil. Oza e Shukla(86) consideram a questão da predição sob a incerteza do grau do polinômio no modelo condicional de Rao(65) com covariáveis. Reconsiderando os resultados de Rao(66), quanto aos testes de hipóteses envolvendo as covariáveis, Saw(87) diverge do pensamento de Rao mostrando seus resultados em função do poder de teste.

A metodologia usada por Nel e Merwe(86) para testar a igualdade de vetores de médias sob a condição de Behrens-Fisher (matrizes de variâncias e covariâncias desiguais) foi estendida no

trabalho de Nel(87), que permitiu a análise de perfis sob desigualdade de matrizes de variâncias e covariâncias.

O tamanho da amostra na análise de medidas repetidas e a predição e estimação de curvas de crescimento com estruturas específicas da matriz de variâncias e covariâncias foram pesquisadas por Vonesh e Schork(86) e Lee(88) respectivamente.

Diggle(88) propõe um método de análise de dados de MR envolvendo modelagem, estruturas de variâncias e covariâncias e verificação de adequabilidade de modelo.

Stanek III e Diehl(88) ilustram como o MCC pode ser adaptado para a situação categorizada. Propõem um modelo de crescimento categórico, onde funções das respostas marginais são modeladas ao longo do tempo.

Métodos de análise de MR para dados categóricos foram discutidos por Conaway(89). Em seu artigo, explora relações entre vários métodos e fornece uma aproximação baseada no método da verossimilhança condicional.

A questão dos dados incompletos foi abordada por Kleinbaum(73), através de uma generalização do MCC que permite vetores de observação incompletos, isto é, de dimensão variando a cada indivíduo. O modelo proposto, $E(Y_{or}) = A_r \xi P B_r$, é uma adaptação que incorpora a matriz B_r (matriz de incidência) ao modelo de Potthoff e Roy(64) para cada subconjunto E_r , contendo os indivíduos observados nos mesmos tempos t_k 's. O autor utiliza a teoria de Wald(43) para obter o melhor estimador assintoticamente normal, a matriz de variância e covariância assintótica e estatísticas de testes. Além disso, Kleinbaum(73) faz uma modificação da estatística de Wald quando o ajuste é feito com covariáveis.

Trabalhos como os de Woolson et al(78) e Woolson e Leeper(80) referem-se aos métodos de análise de dados incompletos com enfoque na generalização de Kleinbaum(73), que é utilizada nos estudos longitudinais mistos.

Hartley e Hocking(71) derivam estimadores de máxima verossimilhança para θ e Σ num sentido mais geral, tal que permite que um vetor incompleto consista em uma combinação linear das componentes do vetor original e resolve a questão analiticamente.

Dempster et al(77) dão origem a um procedimento computacional, algoritmo EM, para estimação por máxima verossimilhança com aplicação às situações em que os dados podem ser vistos como incompletos.

Ainda envolvendo discussões sobre a generalização de Kleinbaum(73), Leeper e Woolson(82) aplicam a metodologia deste para produzir um estimador consistente de Σ , via aproximação por máxima verossimilhança, além de estudos de Monte Carlo para a estatística de Wald. Liski(85) propõe uma técnica computacional para calcular as estimativas dos parâmetros do MCC e usa o algoritmo EM no processo iterativo e no ajuste por covariáveis de Rao(65,66).

Visando estudar as propriedades dos estimadores e obter procedimentos de teste de máxima verossimilhança, Srivastava(85) explicitou a função de verossimilhança estimando valores perdidos e derivando testes nos problemas de uma e duas amostras.

Com o intuito de fornecer procedimentos alternativos no estudo de dados incompletos, mas não exclusivamente, Schwertman et al(85) e Schwertman e Heilbrun(86) recorrem ao método DS de Hills(68) e derivam o teste para paralelismo de perfis médios. No primeiro artigo, Schwertman investiga o poder de vários métodos

por estudos de Monte Carlo. No segundo, considera tempos de observação aleatórios e deriva o teste de paralelismo recorrendo ao fator de correção de Geisser e Greenhouse(58).

Outros resultados como os de Schwertman e Allen(79) e Schwertman(78) aparecem no contexto de análise de split-plot de dados incompletos. Ainda relacionado ao split-plot e às DS, Gill(88) adverte para a utilização das primeiras diferenças na análise de split-plot, devido à perda de graus de liberdade e a problemas relativos à homogeneidade da matriz de variâncias e covariâncias, assim como a redução da correlação na utilização de diferenças.

1.3 ESTRUTURA E CONTEÚDO GERAL DO ESTUDO

Essencialmente, o presente trabalho trata de um estudo de algumas técnicas de análise de dados com estrutura de MR. Este estudo visa, mais especificamente, detalhar aspectos metodológicos envolvidos em experimentos com MR nas situações em que os dados podem não ser completamente observados.

Através de uma breve revisão bibliográfica, foram destacados os trabalhos de Potthoff e Roy(64) e de Rao(65), de onde foi possível estabelecer inter-relações entre os métodos multivariados propostos por estes autores. Estas inter-relações dizem respeito, principalmente, à questão das escolhas da matriz de ponderação G do modelo de Potthoff e Roy(64), ao grau do polinômio e ao ajuste por covariáveis de Rao(65).

Do estudo dos trabalhos mencionados acima, notou-se limitações com respeito à possibilidade dos dados não serem

completamente observados (dados incompletos). Esta limitação dos métodos acima foi abordada através da metodologia derivada por Kleinbaum(73), a qual fornece uma generalização do Modelo de Curva de Crescimento. Tal generalização permite que se classifique o modelo de Potthoff e Roy(64) como um caso particular para situações em que os dados foram completamente observados.

O método das Diferenças Sucessivas é formalizado a fim de estudar a questão da determinação de grau de polinômio ajustado ao perfil médio de um grupo de indivíduos. Esta formalização foi feita de modo que permitisse sua utilização no contexto de dados incompletos, ou seja, visando a apresentação de uma alternativa simplificada de análise de dados, nas situações experimentais em que os dados não foram completamente observados para todos os indivíduos.

Assim, descrições de métodos usuais de análise de MR são apresentados no **Capítulo 2**: o modelo misto univariado, o Modelo de Curva de Crescimento e o ajuste por covariáveis de Rao(65). Aqui destaca-se as inter-relações dos métodos de Potthoff e Roy(64) e Rao(65) e considerações envolvendo estruturas da matriz de variâncias e covariâncias. De fato, o ajuste por covariáveis de Rao(65) nem sempre supera o modelo de Potthoff e Roy(64). Nota-se na **seção 2.3.2** que Rao(65) procurou incorporar as covariáveis a fim de evitar a perda de informação que elas possam conter, utilizando de um procedimento metodologicamente diferenciado do proposto por Potthoff e Roy(64). Do desenvolvimento deste capítulo, concluiu-se na **seção 2.3.3** que estas inter-relações estão associadas à estrutura de variâncias e covariâncias.

Acrescenta-se aqui, ainda, que a análise univariada não deve ser aplicada com o intuito de substituir a multivariada, mas sim de preceder a aplicação de outras técnicas multivariadas em certos contextos.

O método de Hills(68) é apresentado com uma formulação apropriada no **Capítulo 3**. Esta formulação foi introduzida visando esclarecer como as DS permitem a determinação de grau de polinômio. Nas seções 3.2.1 e 3.2.2 é abordada a questão da determinação do grau do polinômio via testes alternativos. Na seção 3.4, através de um tratamento algébrico foram iluminados aspectos interessantes com respeito às inter-relações entre as DS, através das matrizes de variâncias e covariâncias supondo estrutura uniforme.

Tratando de situações experimentais em que os dados são incompletos, uma das finalidades do **Capítulo 4** é ressaltar a generalização do Modelo de Curva de Crescimento proposta por Kleinbaum(73), destacando a forma estruturada da metodologia no tratamento de situações em que os dados são incompletos. Esta estrutura foi explorada na seção 4.3 para aplicação do método DS, visando teste de paralelismo de perfis médios. Diante da metodologia desenvolvida, principalmente nas seções 4.3.1 e 4.3.3, o que se nota é que a questão dos dados incompletos requer certas especificidades e detalhes a se considerar em cada situação experimental. Sob o aspecto teórico, o procedimento elaborado para derivação do teste de paralelismo, via diferenças de primeira ordem, fornece resultados interessantes. No entanto, estender o procedimento para determinação de grau do polinômio foi deixado para trabalho futuro.

O **Capítulo 5** consiste na ilustração de alguns resultados apresentados nos capítulos anteriores, através de um experimento real. O objetivo do experimento está no estudo comparativo de dois tipos de armazenamento de ovos, utilizando a acompanhamento do peso dos ovos ao longo de 7 ocasiões (120 dias). A particularidade no experimento está, principalmente, no fato de que há desinformação de algumas medidas de peso e esta se deve à causas

não aleatórias. Para analisar os dados dos ovos, foram aplicadas duas técnicas multivariadas (o MCCG e o MCC) e uma univariada (split-plot), com as devidas restrições. Na seção 5.3.4 são fornecidas algumas considerações relativas às metodologias aplicadas nas seções 5.3.1, 5.3.2 e 5.3.3. Em termos comparativos, os tipos de armazenamento são comentados na seção 5.4 .

CAPÍTULO 2

2.1 INTRODUÇÃO

A principal finalidade deste capítulo é colocar alguns dos importantes resultados metodológicos envolvendo estudos com medidas repetidas (MR), através de uma descrição de métodos usuais.

De modo geral, os resultados são colocados visando ressaltar o conteúdo metodológico no contexto de estudos de curva de crescimento.

Desde a década de 30, métodos univariados, multivariados e adaptações destes têm sido usados, muitos deles considerando os casos mais gerais, como técnica ideal de análise. Diante deste fato, cabe ressaltar os trabalhos de Wishart(38), Potthoff e Roy(64), Rao(65.66.67), Kleinbaum(73), entre outros.

Nas seções seguintes, aspectos importantes de alguns métodos de análise univariados e multivariados são explorados e comparados.

2.2 MÉTODO DE ANÁLISE UNIVARIADA

Os métodos de análise em MR são baseados essencialmente em dois tipos de procedimentos: univariado e multivariado. No primeiro, o conjunto de dados é analisado considerando principalmente o número total de observações, já o último, baseia-se no número de unidades experimentais. Este fato faz com que algumas suposições envolvidas nos métodos univariados sejam desnecessárias nos multivariados, como a estrutura de dispersão dos dados, por exemplo.

2.2.1 Modelo Misto Univariado

Considere que observações de um experimento em MR são feitas conforme a estrutura dada na Tabela 2.2.1, onde y_{ijk} representa o valor da variável resposta da j -ésima unidade experimental (indivíduo) dentro do i -ésimo grupo na k -ésima ocasião (tempo), $i=1, \dots, a$; $j=1, \dots, n_i$ e $k=1, \dots, p$.

O modelo que descreve $y_{ij} = (y_{ij1}, y_{ij2}, \dots, y_{ijp})'$ pode ser escrito como:

$$y_{ijk} = \theta + \alpha_i + \beta_k + s_{j(i)} + \alpha\beta_{ik} + s\beta_{jk(i)} + \epsilon_{ijk} \quad (2.2.1)$$

sob as restrições $\sum \alpha_i = \sum \beta_k = \sum \alpha\beta_{ik} = \sum s\beta_{jk(i)} = 0$

Tabela 2.2.1 - Dados de um estudo em MR com a grupo
ocasiões

grupo	indivíduo	ocasiões			
		1	2	. . .	p
1	1	y_{111}	y_{112}	. . .	y_{11p}
	\vdots	\vdots	\vdots	\vdots	\vdots
	n_1	$y_{1n_1 1}$	$y_{1n_1 2}$. . .	$y_{1n_1 p}$
	média	$\bar{y}_{1.1}$	$\bar{y}_{1.2}$. . .	$\bar{y}_{1.p}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a	1	y_{a11}	y_{a12}	. . .	y_{a1p}
	\vdots	\vdots	\vdots	\vdots	\vdots
	n_a	$y_{an_a 1}$	$y_{an_a 2}$. . .	$y_{an_a p}$
	média	$\bar{y}_{a.1}$	$\bar{y}_{a.2}$. . .	$\bar{y}_{a.p}$
	média geral	$\bar{y}_{..1}$	$\bar{y}_{..2}$. . .	$\bar{y}_{..p}$

onde, θ , α_i e β_k são a média geral da população, o efeito do i-ésimo grupo de tratamento e o efeito da k-ésima ocasião respectivamente.

$s_{j(i)}$ é o efeito aleatório associado com o j-ésimo indivíduo no i-ésimo grupo.

$\alpha\beta_{ik}$ é o efeito da interação para o i-ésimo grupo de tratamento na k-ésima ocasião.

$s\beta_{jk(i)}$ é o efeito da interação do j-ésimo indivíduo e k-ésima ocasião dentro de i-ésimo grupo de tratamento

ϵ_{ijk} é o termo de erro aleatório.

Para a Tabela 2.2.1, sendo os efeitos aleatórios independentes e normalmente distribuídos,

$$s_{j(i)} \sim N(0, \sigma_s^2) ; s\beta_{jk(i)} \sim N(0, \sigma_{s\beta}^2) ; \epsilon_{ijk} \sim N(0, \sigma_e^2)$$

então, estas componentes são combinadas tal que:

$$\pi_{ijk} = s_{j(i)} + s\beta_{jk(i)} + \epsilon_{ijk} ,$$

pode ser escrito como π_{ij} , que tem uma distribuição normal p-variada com esperança zero e matriz de variâncias e covariâncias, por grupo, com padrão de uniformidade,

$$\Sigma = \sigma^2 [(1-\rho)I + \rho 11'] \quad (2.2.2)$$

onde 1 é um vetor de 1's p-variado, para todo i , $i=1, \dots, a$; $\sigma^2 = \sigma_{s\beta}^2 + \sigma_s^2 + \sigma_e^2$ e $\rho = \sigma_s^2 / \sigma^2$.

Este padrão de Σ afeta diretamente as estatísticas que envolvam comparações "dentro" da unidade experimental, mas não aquelas que estejam associadas a comparações "entre" unidades experimentais.

Dadas algumas facilidades de interpretação das estatísticas univariadas, alguns autores como Box(54), Geisser e Greenhouse(58) e Huynh e Feldt(70,76) preocuparam-se tanto em corrigir os graus de liberdade das estatísticas F da análise de variância univariada associada ao modelo (2.2.1), como em validar

o teste F exato através da estrutura tipo H da matriz de variâncias e covariâncias. Esta estrutura tipo H permite que a análise univariada seja feita, apenas introduzindo o fator de correção aos graus de liberdade. Os elementos de Σ , segundo tal estrutura são dados por:

$$\sigma_{kk'} = \alpha_k + \alpha_{k'} + \lambda \delta_{kk'}$$

onde, $\lambda > 0$, α_k e $\alpha_{k'}$ são constantes. Para o padrão de uniformidade $\alpha_k = \alpha_{k'} = \sigma^2 \rho / 2$ e $\lambda = \sigma^2 (1 - \rho)$.

Estes recursos foram utilizados procurando evitar que este método de análise se limitasse à estrutura (2.2.2), já que essa estrutura de Σ (padrão de uniformidade) é, na verdade, pouco realística em situações práticas.

Uma outra forma estruturada de Σ associada aos estudos de MR é quando Σ corresponde ao padrão de correlação serial obtido através de um processo auto-regressivo de primeira ordem (Apêndice A). A partir daí, Σ é dada pelo seguinte:

$$\Sigma = \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{p-1} \\ & 1 & \rho & \rho^2 & \dots & \rho^{p-2} \\ & & & & \dots & \\ & & & & & 1 \end{pmatrix} \quad (2.2.3)$$

dependente também dos parâmetros ρ e σ^2 . Mesmo para este padrão de Σ , a análise via estatísticas F-aproximadas é possível, já que Walleistein e Fleiss(79) expressaram o fator ϵ de Box(54) como função de ρ e p , ajustando os graus de liberdade.

Um fato importante na análise univariada é a suposição de normalidade multivariada dos π_{1j} , que pode ser preocupante em caso de testes de matrizes de variâncias e covariâncias.

2.2.2 O Padrão de Uniformidade da Matriz de Variâncias e Covariâncias

No contexto da análise univariada de MR, a estrutura uniforme de Σ aparece como consequência do modelo misto. Dependente dos parâmetros ρ e σ^2 , este tipo de estrutura de Σ pode ser verificado, via teste de hipóteses, pelo critério da razão de verossimilhança, como descrito por Box(50).

Assim, para testar

$$H_0: \Sigma_1 = \Sigma_u \quad \text{versus} \quad H_1: \Sigma_1 \neq \Sigma_u$$

onde Σ_u é dada por (2.2.2), $\Lambda = \Delta_1/\Delta_0$ estabelece o critério de teste.

As quantidades Δ_1 e Δ_0 correspondem aos determinantes das matrizes de variâncias e covariâncias amostrais, sob hipótese de uniformidade, elevados a $n/2$, ($n = \sum n_1$), respectivamente.

O denominador da razão Λ , Δ_0 , é estimado pela expressão:

$$\left\{ \left[\bar{s}_{kk} + (p-1)\bar{s}_{kk'} \right] \left[\bar{s}_{kk} - \bar{s}_{kk'} \right]^{p-1} \right\}^{n/2}$$

$$\text{com } \bar{s}_{kk'} = \sum_k s_{kk'}/p \quad \text{e} \quad \bar{s}_{kk'} = \left\{ \sum_{k=1}^p \sum_{k'=1+1}^{p-1} s_{kk'} \right\} / \{p(p-1)/2\}$$

sendo as médias das variâncias e covariâncias amostrais, respectivamente.

Pelo critério de Wilks(46), o teste de H_0 versus H_1 é, então, dado pela seguinte estatística:

$$\Lambda = \frac{|\hat{\Sigma}_1^*|^{n/2}}{|\hat{\Sigma}_u|^{n/2}} = \frac{|\hat{\Sigma}_1^*|^{n/2}}{\{[\bar{s}_{kk} + (p-1)\bar{s}_{kk'}][\bar{s}_{kk} - \bar{s}_{kk'}]^{p-1}\}^{n/2}}$$

onde, $\hat{\Sigma}_1^* = \sum_{i=1}^a (n_i/n)\hat{\Sigma}_i$ é o estimador de máxima verossimilhança da matriz de variâncias e covariâncias comum aos grupos.

Calculando as expressões abaixo,

$$f_1 = (p(p+1)/2) - 2 \quad \text{e} \quad A_1 = \frac{\{p(p+1)^2(2p-3)\}}{\{6n(p-1)(p^2+p-4)\}}$$

então, $(A_1 - 1)2 \ln \Lambda$ tem distribuição χ^2 com f_1 graus de liberdade.

Uma outra aproximação proposta por Box(50) utiliza os cálculos das seguintes expressões:

$$A_2 = \frac{(p-1)p(p+1)(p+2)}{6n^2(p^2+p-4)} \quad ; \quad f_2 = \frac{f_1 + 2}{A_2 - A_1^2} \quad ; \quad b = f_1 / (1 - A_1 - f_1/f_2)$$

onde, $-2\ln\Lambda/b$ tem distribuição aproximadamente F com f_1 e f_2 graus de liberdade.

Deste modo, o padrão de uniformidade pode ser verificado e, segundo Box(50), ambas aproximações podem ser empregadas adequadamente. Ressaltando, apenas, o caso em que p é grande e n é pequeno, pois, nesta situação, a aproximação χ^2 é menos precisa que a F.

2.2.3 O Padrão de Correlação Serial da Matriz de Variâncias e Covariâncias

Sendo a estrutura de Σ_1 , matriz de variâncias e covariâncias do grupo 1, de grande importância dentro do contexto de medidas repetidas, parece razoável questionar alguns pontos relacionados ao padrão de correlação serial de Σ_1 como em (2.2.3).

Portanto, considerando o teste da razão de verossimilhança generalizado, deduzido por Hearne et al(83) a fim de verificar a validade do padrão serial, pode-se descrever o mecanismo de estimação desta matriz que envolve apenas dois parâmetros principais, ρ e σ^2 .

A estatística de teste de razão de verossimilhança de modo geral é expressa por:

$$\Lambda = L(\hat{w}) / L(\hat{\Omega})$$

onde,

\hat{w} é o estimador de máxima verossimilhança do parâmetro sob o espaço da hipótese nula,

$\hat{\Omega}$ é o estimador de máxima verossimilhança do parâmetro sob o

espaço geral.

$L(\hat{w})$ e $L(\hat{\Omega})$ são as funções de verossimilhança calculadas em \hat{w} e $\hat{\Omega}$, respectivamente.

Os estimadores de máxima verossimilhança, sob hipótese que as observações são normalmente distribuídas, são encontrados abaixo.

Considere a transformação $x_{1jk} = y_{1jk} - \bar{y}_{1.k}$ para testar as seguintes hipóteses:

$$H_0: \Sigma_1 = \Sigma_p \quad \text{versus} \quad H_1: \Sigma_1 \neq \Sigma_p$$

onde Σ_p é como em (2.2.3).

Inicialmente, o objetivo é determinar \hat{w} que satisfaça $\partial L(w)/\partial \rho = 0$. Isto equivale a :

$$-N\rho^2 + (1-\rho^2) (A_1 - \rho A_2) = 0 \quad (2.2.4)$$

$$A_1 = \sum_{i=1}^a n_i / (n_i - 1) \sum_{j=1}^{n_i} \sum_{k=2}^p x_{1jk} x_{1j(k-1)}$$

e

$$A_2 = \sum_{i=1}^a n_i / (n_i - 1) \sum_{j=1}^{n_i} \sum_{k=2}^{p-1} x_{1jk}^2$$

Além disso, de $\partial L(\hat{w})/\partial \sigma^2 = 0$ segue a outra equação para determinação de \hat{w} .

$$-Np\sigma^2 + A_2 + A_3 - 2A_1\rho + A_2\rho^2 = 0 \quad (2.2.5)$$

onde,

$$A_3 = \sum_i n_i / (n_i - 1) \sum_j (x_{1ji}^2 + x_{1jp}^2)$$

Resolvendo (2.2.5) para σ^2 e substituindo o resultado em (2.2.4), obtem-se,

$$(1-1/p)A_2\rho^3 - (1-2/p)A_1\rho^2 - \{(1+1/p)A_2 + A_3/p\}\rho + A_1 = 0 \quad (2.2.6)$$

A solução de (2.2.6) é usada para resolver (2.2.5) e obter σ^2 . Koopmans(42) sugere a escolha de $\hat{\rho}$ mais próxima de $A_1/(A_2+A_3)$, pois do polinômio (2.2.6) são encontrados 3 valores de ρ .

Tomando agora Ω , $\hat{\Sigma}_1^* = \sum_{i=1}^a (n_i/n)\hat{\Sigma}_i$ é o estimador de máxima verossimilhança não viciado da matriz de variâncias e covariâncias comum aos grupos.

Segundo Hearne et al(83), o valor de Λ é encontrado pela expressão :

$$\Lambda = \frac{|\hat{\Sigma}_1^*|^{n/2}}{|\hat{\Sigma}_p|^{n/2}} = \frac{|\hat{\Sigma}_1^*|^{n/2}}{\{\sigma^{2p} / (1-\hat{\rho}^2)\}^{n/2}}$$

Então, a estatística $-2\ln\Lambda$ é aproximadamente

qui-quadrado com $[(p(p+1)/2) - 2]$ graus de liberdade.

Finalizando, pode-se notar que a estimação de Σ_{ρ} não é obtida de modo trivial, devido à estrutura seriada. Além disso, a preocupação com a forma da matriz de variâncias e covariâncias deve estar sempre presente em estudos de medidas repetidas, pois sempre que possível é recomendável testes de hipóteses envolvendo a forma de Σ_1 ou algum procedimento que dê maior indicação de como é a verdadeira estrutura de dispersão inerente aos dados.

O fator ϵ aproximado deduzido por Wallestein e Fleiss(79) é uma opção em caso de não rejeição de H_0 , permitindo a análise univariada. Entretanto, em caso contrário, os procedimentos multivariados que independem da forma de Σ_1 são mais adequados, ressaltando que aqueles procedimentos que dependem de uma estrutura de covariância específica devem ser preferidos, ou seja, métodos multivariados com a suposição de uma estrutura específica de matriz de variâncias e covariâncias.

2.3 MÉTODOS MULTIVARIADOS: O MODELO DE CURVA DE CRESCIMENTO E O AJUSTE POR COVARIÁVEIS

A análise de MR via métodos multivariados possui algumas vantagens sobre a dos usuais univariados brevemente apresentados.

Pensando em obter ferramentas mais gerais para análise de curvas de crescimento, Potthoff e Roy(64) introduziram um modelo de análise de variância multivariada generalizado. Por suas várias aplicações em problemas de curvas de crescimento, entre outras, tornou-se o Modelo de Curva de Crescimento (MCC). Entretanto, o

método desenvolvido por estes autores deu origem a muitas discussões sobre a maneira que aborda a estimação e a escolha da matriz de ponderação G .

Questionando a escolha de G arbitrária, Rao(65,66) utilizou uma metodologia baseado em ajuste de covariáveis, a fim de não perder informação. Reduziu o MCC a um modelo de Gauss-Markoff com covariáveis, de tal forma que as variáveis que não fizessem parte do polinômio poderiam ser utilizadas para melhorar a estimação dos parâmetros.

2.3.1 O Modelo de Curva de Crescimento - MCC

O MCC proposto por Potthoff e Roy(64) consiste basicamente de uma generalização do já conhecido modelo linear geral multivariado (modelo MANOVA), cuja finalidade em aplicações de problemas de curva de crescimento é: ajustar modelo, testar hipóteses gerais e obter bandas de confiança.

Estes autores propuseram o seguinte modelo:

$$E(Y_0(n \times p)) = A\xi P \quad (2.3.1)$$

onde,

p é o número de ocasiões

n é o número de indivíduos, $n = \sum_{i=1}^a n_i$

a é o número de grupos

q é o número de parâmetros

$A(n \times a)$ é a matriz que define os grupos,

$$A = \text{diag}(1_{n_1}, \dots, 1_{n_a})$$

$P(q \times p)$ é a matriz de especificação de posto completo $q \leq p$, cuja finalidade é associar os níveis do fator, relacionado às ocasiões, ao polinômio desejado.

O estimador de ξ em (2.3.1) pelo critério da máxima verossimilhança ou de mínimos quadrados ponderados é dado por:

$$\hat{\xi} = (A'A)^{-1} A' Y_0 G^{-1} P' (P G^{-1} P')^{-1} \quad (2.3.2)$$

com,

$$Y = Y_0 G^{-1} P' (P G^{-1} P')^{-1} \quad (2.3.3)$$

e G a matriz de ponderação sendo definida positiva garante que $(P G^{-1} P')$ também o seja.

Pela transformação (2.3.3), (2.3.1) é reduzido ao modelo de Gauss-Markoff,

$$E(Y) = A\xi \quad (2.3.4)$$

e

$$\Sigma = V(y_{ij}) = (P G^{-1} P')^{-1} P G^{-1} \Sigma_0 G^{-1} P' (P G^{-1} P')^{-1}$$

onde,

Σ_0 é a matriz de variâncias e covariâncias de cada linha de Y_0 e Σ , a correspondente para cada linha de Y (matriz de dados transformada).

A escolha de G é um ponto importante dentro do contexto do método de Potthoff e Roy(64) diante do fato de que $\hat{\xi}$ depende de G . No sentido da estimação não viciada de mínima variância, a escolha ideal de G corresponde a $G=\Sigma_0$, que na prática é desconhecida. Os autores propuseram a utilização de G^{-1} que se afastasse o mínimo possível de Σ^{-1} , pois distanciamento provocaria um aumento na variância de $\hat{\xi}$,

$$V(\hat{\xi}) = (A'A)^{-1} \{ (PG^{-1}P')^{-1} PG^{-1} \Sigma_0 G^{-1} P' (PG^{-1}P')^{-1} \}$$

Dentre as escolhas possíveis de G , incluem-se, $G=I$, $G=S$ ou G baseada em uma estimativa de Σ_0 obtida de estudos anteriores.

2.3.2 Ajustando Covariáveis

Dando um enfoque via ajustes por covariáveis à questão das curvas de crescimento, Rao(65) propõe um modelo no qual utiliza covariáveis.

Considere o seguinte modelo:

$$E(Y_0) = A\xi P_1 \quad (P_1 \text{ ortogonal})$$

Seja $Y = Y_0 P'_*$,

onde $P'_* = (P'_1 | Q')$, P'_* uma matriz ortogonal quadrada ($p \times p$) que assim particionada satisfaz $P'_1 Q' = 0$, ou seja, P'_* consiste em uma base de polinômios ortogonais de graus 0 a p e P_1 uma parte desta base, dos polinômios de graus inferiores a q ($q \leq p$).

Substituindo P_* ,

$$Y = Y_0(P_1' | Q')$$

como $P_1'Q=0$, então, se $Z=Y_0Q'$, $E(Z)=0$ e, portanto,

$$E(Y) = A\xi P_1' P_1'$$

mas

$$E(Y) = E(E(Y|Z))$$

logo, Rao(65) propôs o seguinte modelo condicional:

$$E(Y|Z) = A\xi_* + Z\gamma \quad (2.3.5)$$

onde, $\xi_* = \xi P_1' P_1'$, $Z(n \times (p-q))$ é a matriz que contém a informação dos $(p-q)$ parâmetros não incorporados no modelo (2.3.1) e $\gamma((p-q) \times q)$ é a matriz dos coeficientes destes parâmetros.

Escrevendo o modelo (2.3.5) como um modelo linear multivariado,

$$E(Y) = X\beta \quad (2.3.6)$$

onde,

$$X = (A|Z) \text{ e } \beta' = (\xi_*' | \gamma')$$

O estimador de mínimos quadrados de β é:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2.3.7)$$

de onde facilmente é possível encontrar que a estimativa de γ e ξ_* é dada pelas expressões seguintes:

$$\hat{\gamma} = (Z'Z - Z'A(A'A)^{-1}A'Z)^{-1}(Z'Y - Z'A(A'A)^{-1}A'Y) \quad (2.3.8)$$

$$e \quad \hat{\xi}_* = (A'A)^{-1}A'Y - (A'A)^{-1}A'Z\hat{\gamma} \quad (2.3.9)$$

Definindo $\hat{\xi}_Y = (A'A)^{-1}A'Y$ e $\hat{\xi}_Z = (A'A)^{-1}A'Z$,

$\hat{\gamma}$ e $\hat{\xi}_*$ podem ser escritos por:

$$\hat{\gamma} = (Z'Z - Z'A\hat{\xi}_Z)^{-1}(Z'Y - \hat{\xi}_Z'A'Y) \quad (2.3.10)$$

$$\hat{\xi}_* = \hat{\xi}_Y - \hat{\xi}_Z\hat{\gamma} \quad (2.3.11)$$

Sendo $Z=Y_0Q'$, substituindo em (2.3.8) e (2.3.9), a expressão da matriz $\hat{\xi}_*$ é reescrita como segue

$$\hat{\xi}_* = (A'A)^{-1}A'Y_0\{P' - Q'(QSQ')^{-1}QSP'\} \quad (2.3.12)$$

onde,

$$nS = Y_0'(I - A(A'A)^{-1}A')Y_0$$

2.3.3 Relações Envolvendo o Modelo de Curva de Crescimento

Quando o modelo saturado é ajustado em (2.3.1) (polinômio de grau $p-1$), para G arbitrária, não tendo sentido o ajuste por covariáveis em (2.3.5), verifica-se que as estimativas provenientes de ambos coincidem, quando P é ortogonal. Se P não é ortogonal também há coincidência, a menos de reparametrização.

Ajustando um modelo de grau inferior a $q-1$, portanto, não saturado e escolhendo G arbitrária, a estimativa (2.3.2) não coincide com (2.3.9), dada a perda de informação contida nos $(p-q)$ parâmetros restantes (covariáveis). Caso a matriz de ponderação seja S , para P ortogonal, então os modelos (2.3.1) e (2.3.5) são praticamente equivalentes, pois mínimos quadrados ponderados coincide com mínimos quadrados corrigidos (Rao(67)).

Segue ainda que, quando Σ é uniforme, há independência entre os coeficientes polinomiais. Isto significa que o ajuste de covariáveis é inadequado e, neste caso, não há relação entre os modelos referidos acima, independentemente da ortogonalidade ou não da matriz P .

Na presença do padrão de correlação serial para Σ , o ajuste por covariáveis de Rao(65) deve ser cuidadoso, dada a não correlação entre os coeficientes polinomiais relativos aos graus pares com ímpares, como mencionado por Kenward(85). O ajuste por covariáveis deve ser feito dentro dos subconjuntos de grau par e de grau ímpar, como aponta Moyano(90).

UMA ABORDAGEM DO MÉTODO DAS DIFERENÇAS SUCESSIVAS

CAPÍTULO 3

3.1 INTRODUÇÃO

O método das Diferenças Sucessivas (DS) é apresentado por Hills(68), abordando o estudo da determinação do grau do polinômio que descreve o perfil médio, na situação de um grupo de indivíduos. No contexto de Hills(68) há uma semelhança evidente entre as estatísticas de teste baseadas no método DS e no ajuste por polinômios ortogonais de Rao(59). Esta semelhança pode ser vista como uma relação entre os métodos.

As DS, podendo ser feitas de várias ordens, possuem a particularidade de a matriz de transformação conter informação sobre os ajustes polinomiais. Esta matriz de transformação depende da ordem da diferença estudada.

Se a resposta é explicada por um modelo polinomial de grau $q-1$, a DS de ordem q se anula, ou seja, há uma forte correspondência entre a ordem da diferença e o grau do polinômio que se ajusta às respostas médias.

Uma consequência da aplicação do método DS é o fato de que as diferenças estabelecem comparações entre médias, como contrastes sucessivos de médias. Isto pode ser notado pela construção da matriz de transformação U_m e sua aplicação sobre o vetor de respostas médias.

Na aplicação deste método há também uma estreita relação entre as matrizes de variâncias e covariâncias e as matrizes U_m , dependendo da ordem m da diferença que se está fazendo.

Neste capítulo, o método DS é formalizado, utilizando uma notação apropriada e visando algumas colocações em termos da estruturação da matriz de variâncias e covariâncias dos dados originais.

3.2 O MÉTODO DS NO CASO DE UM GRUPO

A exposição do método DS segundo a formalização e notação, convenientemente adotadas aqui, requer uma suposição quanto ao número de ocasiões. Considere que um indivíduo foi observado ao longo de $p+1$ períodos de tempo, com respostas $y_j = (y_{j0}, \dots, y_{jp})'$, de acordo com a Tabela 3.1.

Cada indivíduo j descrito pelo vetor y_j é considerado distribuído multinormalmente,

$$y_j \sim N_{p+1}(\theta, \Sigma) \quad , \quad j=1, \dots, n \quad (3.2.1)$$

onde $\underline{\theta} = (\theta_0, \dots, \theta_p)'$ é o vetor de respostas médias nos tempos t_0, \dots, t_p .

Tabela 3.1 - Estrutura de dados no caso de um grupo de indivíduos

indivíduo	tempos				
	t_0	t_1	t_2	\dots	t_p
1	y_{10}	y_{11}	y_{12}	\dots	y_{1p}
2	y_{20}	y_{21}	y_{22}	\dots	y_{2p}
.	.	.	.	\dots	.
n	y_{n0}	y_{n1}	y_{n2}	\dots	y_{np}
média	\bar{y}_0	\bar{y}_1	\bar{y}_2	\dots	\bar{y}_p

Das observações do conjunto descrito pela Tabela 3.1 segue que:

$\bar{\mathbf{y}} = (\bar{y}_0, \bar{y}_1, \dots, \bar{y}_p)'$ é uma estimativa de $\underline{\theta}$, portanto, também multinormalmente distribuído,

$$\bar{\mathbf{y}} \sim N_{p+1}(\underline{\theta}, n^{-1}\Sigma)$$

Suponha que as observações são feitas nos indivíduos sob intervalos de tempo igualmente espaçados, logo

$t_k = t_{k-1} + c$; $k=1, \dots, p$ e c é uma constante que define o espaçamento entre os tempos.

Tomando um indivíduo em particular, as diferenças de ordem m são definidas vetorialmente como segue:

$$\Delta^m y_j = U_m y_j = (u_{m(1)} \dots u_{m(p-m+1)})' y_j \quad ; \quad m=1, \dots, p \quad (3.2.2)$$

onde U_m é uma matriz de transformação de ordem m com dimensão $(p-m+1) \times (p+1)$.

Fica claro de (3.2.2) que as diferenças são combinações lineares das respostas originais, através da matriz U_m , que é obtida pela definição das diferenças de ordem m . Particularmente, sendo t_0 o tempo inicial e

y_{j0} o valor inicial para o elemento j ,

$\sum_{k=1}^p \Delta^1 y_{jk} / p$ é denominada a velocidade média de crescimento e

$\sum_{k=2}^p \Delta^2 y_{jk} / (p-1)$ é denominada a aceleração média de crescimento.

A diferença de ordem 1 que define o vetor $\Delta^1 y_j$ para o indivíduo j é calculada por:

$$\Delta^1 y_{jk} = y_{jk} - y_{j(k-1)} ; k=1, \dots, p \quad (3.2.3)$$

logo,

$\Delta^1 y_j = (\Delta^1 y_{j1}, \dots, \Delta^1 y_{jp})'$ é o vetor p-variado cuja matriz de transformação U_1 é a que segue,

$$U_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

As diferenças de ordem 2 são obtidas da diferença de ordem 1 sob o vetor $\Delta^1 y_j$, a partir daí, obtendo $\Delta^2 y_j$, isto é,

$$\Delta^2 y_j = (\Delta^2 y_{j2}, \dots, \Delta^2 y_{jp})'$$

com dimensão $(p-1) \times 1$ e cuja matriz U_2 possui seus elementos definidos a partir de :

$$\Delta^2 y_{jk} = \Delta^1 y_{jk} - \Delta^1 y_{j(k-1)}$$

resultando em U_2 dada a seguir.

$$U_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & -2 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Para interpretação das DS considere que o perfil individual pode ser descrito por um polinômio de grau $q-1$,

$$E(y_j) = P' \underline{\tau} \tag{3.2.4}$$

no qual, P é a mesma matriz considerada em (2.3.1) (dos tempos naturais), no entanto, com dimensões $((p+1) \times q)$ e $\underline{\tau}(q \times 1)$ é o vetor de parâmetros.

De acordo com a expressão (3.2.2),

$$E(\Delta^m y_j) = U_m P' \underline{\tau} \tag{3.2.5}$$

Assim, verificando algumas situações, para valores de q e m em (3.2.5), de cada resposta y_j , segue que:

(i) se $q=2$ e $m=1$

$$E(\Delta^1 y_{jk}) = c \tau_1, \text{ ou ainda, } E(\Delta^1 y_{jk}) = \tau_0^+$$

(ii) se $q=3$ e $m=1$

$$E(\Delta^1 y_{jk}) = (c\tau_1 + c^2\tau_2) + 2c\tau_2 t_k, \text{ ou ainda, } E(\Delta^1 y_{jk}) = \tau_0^+ + \tau_1^+ t_k$$

(iii) se $q=2$ e $m=2$

$$E(\Delta^2 y_{jk}) = \tau_0 + \tau_1 t_{k+1} - 2\tau_0 - 2\tau_1 t_k + \tau_0 + \tau_1 t_{k-1} = 0$$

Em termos gerais, para $m=1, \dots, p$

(i) se $q=m+1$

$$E(\Delta^m y_{jk}) = m!c^m \tau_m, \text{ ou ainda, } E(\Delta^m y_{jk}) = \tau_0^+ \quad (3.2.6)$$

(ii) se $q > m+1$

$$E(\Delta^m y_{jk}) = \tau_0^+ + \tau_1^+ t_k + \dots + \tau_{q-m-1}^+ t_k^{q-m-1} \quad (3.2.7)$$

(iii) se $q \leq m$

$$E(\Delta^m y_{jk}) = 0 \quad (3.2.8)$$

Os resultados de (i), (ii) e (iii) são provenientes do ajuste polinomial feito em (3.2.4) de onde, na presença de ortogonalidade da matriz P , é possível ressaltar alguns resultados,

(a) quando o grau do polinômio e a ordem da diferença são iguais, cada componente do vetor $\Delta^m y_j$ corresponde ao coeficiente do termo de grau m do polinômio ajustado entre os pontos de y_j que definem a componente de $\Delta^m y_j$, ou seja, $p-m+1$ ajustes de polinômio de grau $q-1$ são feitos sobre $m+1$ pontos consecutivos para compor o vetor $\Delta^m y$.

Observando $u_{m(k)}$ (k -ésima linha de U_m) nota-se, ainda, que seus elementos não nulos coincidem com os valores dos polinômios ortogonais saturados de grau m , isto é, o ajuste é feito via uma base de polinômios ortogonais sobre $m+1$ pontos.

(b) pelo ajuste polinomial (3.2.4) e pela aplicação das DS sobre as respostas y_j espera-se que, quando o grau do polinômio supera a ordem da diferença, os elementos de $\Delta^m y_j$ correspondam às respostas de um polinômio de grau $q-m-1$, de acordo com a expressão em (3.2.7).

(c) se o grau do polinômio for inferior à ordem da diferença, nota-se que o valor esperado da DS se anula e, portanto, qualquer DS de ordem superior também se anula. Isto pode ser verificado observando que

$$E(\Delta^m y_j) = 0, \text{ para qualquer } m > q-1, \text{ pois, } U'_m P = 0, \text{ para } m > q-1$$

ou seja, a medida que $\Delta^m y_j$ se anula, então, $\tau_m = \dots = \tau_p = 0$

Diante das observações (a), (b) e (c), as DS podem ser interpretadas em função de sua ordem e do grau do polinômio que modela os valores esperados das DS, ou seja, se uma diferença de ordem m for constante, espera-se que qualquer diferença de ordem superior se anule e que o grau do polinômio que descreve o perfil do indivíduo seja $q=m+1$, a ordem da DS mais um.

Por outro lado, se a DS reflete um comportamento polinomial, isto significa que o grau do polinômio é superior à ordem da diferença considerada, como observado em (3.2.7), ou seja, a DS de ordem m ainda não indica o grau do polinômio.

Nota-se, ainda, que o grau do polinômio pode ser testado

onde, $\Delta^q \underline{\theta} = U_q \underline{\theta}$ e $\Sigma_q = U_q \Sigma U_q'$.

Considere a seguinte hipótese nula:

$H_0: \Delta^q \underline{\theta} = \mathbf{0}$, ou seja, a curva média de crescimento é de grau $q-1$.
Consequentemente, de (3.2.10)

$$n^{1/2}(\Delta^q \bar{y}) \sim N_{p+1-q}(0, \Sigma_q) \quad \text{e} \quad nS_q \sim W_{p-q+1}(\Sigma_q, n-1)$$

de onde fazendo $S_q^* = (n/n-1) S_q$ implica que

$$(n-1) (\Delta^q \bar{y})' S_q^{-1} (\Delta^q \bar{y}) = n (\Delta^q \bar{y})' (S_q^*)^{-1} (\Delta^q \bar{y}) \tag{3.2.11}$$

tem distribuição T^2 de Hotelling com parâmetros $(p-q+1)$ e $(n-1)$ e, ainda, S_q^* é a estimativa não viciada da matriz de variâncias e covariâncias.

3.2.2 Relação com o Método dos Polinômios Ortogonais

Retomando o modelo polinomial em (3.2.4) e de acordo com (3.2.1),

$$\underline{\theta} = P' \underline{\tau} \tag{3.2.12}$$

ou seja, (3.2.12) são médias em cada ocasião descritas por um modelo polinomial de grau $q-1$, através da matriz conhecida P e dos q parâmetros em $\underline{\tau}$.

onde, $\Delta^q \underline{\theta} = U_q \underline{\theta}$ e $\Sigma_q = U_q \Sigma U_q'$.

Considere a seguinte hipótese nula:

$H_0: \Delta^q \underline{\theta} = 0$, ou seja, a curva média de crescimento é de grau $q-1$.
Consequentemente, de (3.2.10)

$$n^{1/2}(\Delta^q \bar{y}) \sim N_{p+1-q}(0, \Sigma_q) \quad e \quad nS_q \sim W_{p-q+1}(\Sigma_q, n-1)$$

de onde fazendo $S_q^* = (n/n-1) S_q$ implica que

$$(n-1) (\Delta^q \bar{y})' S_q^{-1} (\Delta^q \bar{y}) = n (\Delta^q \bar{y})' (S_q^*)^{-1} (\Delta^q \bar{y}) \quad (3.2.11)$$

tem distribuição T^2 de Hotelling com parâmetros $(p-q+1)$ e $(n-1)$ e, ainda, S_q^* é a estimativa não viciada da matriz de variâncias e covariâncias.

3.2.2 Relação com o Método dos Polinômios Ortogonais

Retomando o modelo polinomial em (3.2.4) e de acordo com (3.2.1),

$$\underline{\theta} = P' \underline{\tau} \quad (3.2.12)$$

ou seja, (3.2.12) são médias em cada ocasião descritas por um modelo polinomial de grau $q-1$, através da matriz conhecida P e dos q parâmetros em $\underline{\tau}$.

Considere a matriz B tal que $BP' = 0$.

$$By_j \sim N_{p-q+1}(0, B\Sigma B') \quad (3.2.13)$$

Sendo adequado um modelo polinomial de grau $q-1$, os parâmetros τ_q, \dots, τ_p são não significativos para explicar as respostas y_j .

Por (3.2.13),

$$B\bar{y} \sim N_{p-q+1}(0, n^{-1}B\Sigma B') \quad (3.2.14)$$

e para verificar a adequabilidade do modelo (3.2.12), ou melhor, testar a hipótese $H_0: \tau_q = \dots = \tau_p = 0$, Rao(59) sugere a estatística

$$n(B\bar{y})'(BSB')^{-1}(B\bar{y}) \quad (3.2.15)$$

que é distribuída de acordo com T^2 de Hotelling de parâmetros $(p-q+1)$ e $(n-1)$.

Retomando a expressão (3.2.9), substituindo s por q ,

$$\Delta^q \bar{y} = U_q \bar{y}$$

por equivalência em (3.2.11),

$$n(U_q \bar{y})'(U_q S U_q')^{-1}(U_q \bar{y}) \sim T^2(p-q+1, n-1) \quad (3.2.16)$$

Sob o aspecto do teste do grau do polinômio, e segundo Hills(68), os testes (3.2.15) e (3.2.16) são alternativos, já que as estatísticas possuem a mesma distribuição e ambos os testes fornecem informação sobre o grau do polinômio e adequabilidade do modelo ajustado.

3.3 EXTENSÃO DO MÉTODO DS PARA MAIS DE UM GRUPO

O caso de mais de um grupo apresentado a seguir consiste em uma extensão da abordagem dada anteriormente e, conseqüentemente, dos resultados de Hills(68).

A estrutura de dados agora se identifica com a apresentada na Tabela 2.2.1, entretanto, convenientemente adequada tomando t_0 como a primeira ocasião (tempo) observada.

Assim, y_{ijk} é a observação do indivíduo j do grupo i no tempo t_k e,

$y_{ij} = (y_{ij0}, \dots, y_{ijp})'$ é o vetor de respostas supostamente multinormal,

$$y_{ij} \sim N_{p+1}(\theta_i, \Sigma_i) \text{ para } i=1, \dots, a \text{ e } j=1, \dots, n_i. \quad (3.3.1)$$

De modo análogo ao caso de um grupo, as diferenças são determinadas pela transformação definida pela matriz U_m . Assim, $\Delta^m y_{ij} = U_m y_{ij}$ é a diferença de ordem m do indivíduo j do grupo i $m=1, \dots, p-1$.

Portanto, para um grupo i a diferença de ordem m sobre a média amostral consiste, basicamente, em um vetor $(p-m+1)$ -variado

determinado pela expressão

$$\Delta^m \bar{y}_i = U_m \bar{y}_i \quad (3.3.2)$$

onde,

$$\bar{y}_i = (\bar{y}_{i0}, \dots, \bar{y}_{ip})' ; i=1, \dots, a \text{ e } m=1, \dots, p$$

Sendo y_{ij} (p+1)-variado,

$$V(y_i) = I_{n_i} \otimes \Sigma_i \quad ; \quad y_i = (y'_{i1}, \dots, y'_{in_i})' \quad (3.3.3)$$

Escrevendo as diferenças e correspondentes matrizes de variâncias e covariâncias,

$$\Delta^m y_i = (I_{n_i} \otimes U_m) y_i$$

e

$$V(\Delta^m y_i) = I_{n_i} \otimes (U_m \Sigma_i U) = I_{n_i} \otimes \Sigma_{mi} \quad (3.3.4)$$

3.3.1 Testes de Hipóteses

Utilizando a formulação dada anteriormente é possível proceder com a idéia de determinação de grau de polinômio através de DS, para situações em que mais de um grupo de indivíduos é investigado.

De forma análoga ao caso de um grupo, para determinar que grau de polinômio pode descrever o perfil médio de um grupo i , segundo as expressões (3.3.2), (3.3.3) e (3.3.4), então:

$$E(\Delta^m \bar{y}_i) = \Delta^m \theta_{-1} \quad \text{e} \quad V(\Delta^m \bar{y}_i) = 1/n \sum_{mi} \quad (3.3.5)$$

onde a estimativa de \sum_{mi} é dada por $S_{mi} = U_m S_i U'_m$.

A hipótese de teste, assim como na seção 3.2.1 é $H_0: \Delta^q \theta_{-1} = 0$ que está relacionada à curva média de crescimento dos indivíduos do grupo i . Para testar H_0 , a estatística de teste é dada por (3.3.6), a seguir

$$(n_i - 1)(\Delta^q \bar{y}_i)' (S_{qi}^{-1})(\Delta^q \bar{y}_i) = n_i (\Delta^q \bar{y}_i)' (S_q^*)^{-1} (\Delta^q \bar{y}_i) \quad (3.3.6)$$

cuja distribuição é T^2 de Hotelling com parâmetros $(p-q+1)$ e $(n_i - 1)$.

Esta estatística (3.3.6) é consequência do fato que

$$n_i^{1/2} (\Delta^q \bar{y}_i) \sim N_{p-q+1}(0, \Sigma_{qi}) \quad \text{e} \quad n_i S_{qi} \sim W_{p-q+1}(\Sigma_{qi}, n_i - 1)$$

no qual S_{qi}^* é a estimativa não viciada da matriz de variâncias e covariâncias da diferença de ordem q associada aos indivíduos do grupo i .

Através do teste de H_0 proposto acima, os graus dos polinômios são determinados independentemente para cada grupo,

podendo mais de um grupo ser descrito por polinômios de mesmo grau. Este procedimento não descarta a possibilidade de que um mesmo polinômio descreva um conjunto de a 's grupos, desde que esses a ' grupos possuam as curvas médias de crescimento paralelas. Os testes de paralelismo, utilizando as DS, serão discutidos posteriormente num contexto mais amplo, isto é, abordando as situações em que os dados são incompletos.

3.4 CONSIDERAÇÕES SOBRE A MATRIZ DE VARIÂNCIAS E COVARIÂNCIAS

A determinação da matriz de variâncias e covariâncias das DS de ordem m provém do cálculo da expressão:

$$\Sigma_m = U_m \Sigma U_m'$$

Introduzindo o padrão de uniformidade a Σ , a matriz Σ_m independe da correlação entre as ocasiões, conseqüentemente, da variação intra-indivíduo, pois,

$$\Sigma_m = \sigma^2 U_m U_m' = (\sigma_{kk'})_m \quad (3.4.1)$$

$$(\sigma_{kk'})_m = \begin{cases} \sigma^2 (2m)! / (m!)^2, & k=k' \\ \sigma^2 (-1)^l (2m)! / ((m+1)!(m-l)!), & k \neq k' \text{ e } l=|k-k'|, l < m \\ 0, & k \neq k' \text{ e } l > m \end{cases}$$

Do cálculo de (3.4.1) é possível verificar que há um tipo de estrutura em Σ_m , relacionada, de certa forma, com U_{2m} .

Portanto, a matriz Σ_m , juntamente com U_m possui a característica particular de relacionar as diferenças de ordem m com as de ordem $2m$.

A relação mencionada no penúltimo parágrafo diz respeito aos elementos das linhas de Σ_m e U_{2m} que, por construção, coincidem a menos de sinal, da constante σ^2 e dimensão. Observa-se daí que os elementos que compõem $(1/\sigma^2)\Sigma_1$ são os elementos de uma linha de $(-1)U_2$ e que o mesmo acontece com $(1/\sigma^2)\Sigma_2$ e U_4 e assim por diante.

Dois fatos interessantes envolvendo a matriz Σ_m , podem ser notados :

(i) *sob hipótese de uniformidade ou independência entre as observações por indivíduo, nota-se um aumento das variâncias à medida que a ordem da diferença vai aumentando. O mesmo ocorre com as covariâncias, entretanto, estas diminuem à medida que o cálculo é feito entre diferenças envolvendo tempos distanciados, isto é, crescem com m e diminuem ao longo das linhas de Σ_m . Este fato torna inadequado o cálculo de diferenças de ordem elevada.*

(ii) *a mesma similaridade verificada entre os elementos de $u_{2m(i)}$ (uma linha de U_{2m}) com os valores dos polinômios ortogonais (vide Draper e Smith(77)) aparece em Σ_m , a menos de sinal.*

Segue ainda da expressão (3.4.1) que quando não há correlação entre as observações ou quando esta correlação é constante, a variabilidade devido ao indivíduo inexistente ou desaparece

com a aplicação de diferenças, bastando apenas uma diferença de ordem um para este fato ser verificado.

CAPÍTULO 4

4.1 INTRODUÇÃO

Em situações onde os dados são incompletos, nota-se uma necessidade de adaptação dos métodos multivariados devido à variação das dimensões dos vetores multivariados associados às unidades experimentais (indivíduos).

Uma das adaptações estudadas aqui é o Modelo de Curva de Crescimento Generalizado proposta por Kleinbaum(73), o qual consiste em uma adaptação do MCC de Potthoff e Roy(64). No presente trabalho, o método de Kleinbaum(73) é descrito com modificações convenientemente adotadas para maior clareza e complementação dos resultados.

Outra alternativa de análise de dados incompletos é estudada com base nos trabalhos de Schwertman et al(85,86), utilizando o método DS com o intuito de comparar grupos de indivíduos, via teste de paralelismo.

Além disso, alguns resultados adicionais são fornecidos nas seções 4.3.1 e 4.3.2 relativos à estrutura dos dados.

A finalidade deste capítulo é, portanto, descrever e formalizar as adaptações mencionadas acima, sem desconsiderar a possibilidade de que na presença de dados com observações perdidas existam procedimentos mais adequados de análise, como por exemplo o algoritmo EM. Assim, não é objetivo no capítulo comparar os métodos, mas sim explorá-los sobre alguns aspectos.

4.2 O MODELO DE CURVA DE CRESCIMENTO GENERALIZADO

Sabendo que o modelo (2.3.1), o MCC, não era adequado na presença de dados incompletos, Kleinbaum(73) propôs, então, uma generalização que acomodasse tal situação, utilizando-se de um modelo multivariado, o MCCG.

Nas seções 4.2.1 e 4.2.2 a proposta de Kleinbaum(73) é reescrita, sendo complementada na seção 4.2.3 com uma expressão para o intervalo de confiança simultâneo e faixas de confiança.

4.2.1 A estimação no MCCG

A idéia central da generalização consiste basicamente em dividir os n indivíduos em u conjuntos disjuntos, E_r com n_r elementos cada. Cada conjunto E_r contém as observações daqueles indivíduos observados nos mesmos tempos (ocasiões). Portanto, cada conjunto E_r é visto como um subconjunto dos dados com estrutura completa, isto é, as observações dos n_r indivíduos constituem vetores de mesma dimensão correspondentes aos mesmos tempos. Além disso, os conjuntos E_r e $E_{r'}$ são tais que um indivíduo observado em p_r ocasiões pertence a um único E_r , ou seja, não possuem medidas nos mesmos p_r pontos de tempo, embora dois conjuntos possam

ter o mesmo número de medidas $p_r = p_r$.

A fim de indicar quais pontos do tempo foram observados pelos indivíduos em E_r , define-se uma matriz de incidência de zeros e uns, $B_r (p_r \times p_r)$, incorporada ao MCC como segue:

$$E(Y_{or}) = A_r \xi P B_r \quad e \quad V(Y_{or}) = I_{n_r} \otimes B_r' \Sigma B_r \quad (4.2.1)$$

$r=1, \dots, u$

com as correspondentes dimensões :

$Y_{or} (n_r \times p_r)$, $A_r (n_r \times a)$, $\xi (a \times q)$, $P (q \times p)$ e $\Sigma (p \times p)$, ou seja, a grupos de indivíduos observados em p tempos, divididos em u conjuntos, como descrito por polinômios de grau $q-1$, conforme a expressão (4.2.1).

Neste modelo, as linhas de Y_{or} constituem vetores p_r -variados com matriz de variâncias e covariâncias definida positiva $B_r' \Sigma B_r$, pois B_r é de posto coluna completo. A matriz A_r é a matriz de especificação dos indivíduos dentro dos grupos em cada conjunto E_r .

Escrevendo o modelo (4.2.1) na forma vetorial empilhando as linhas de Y_{or} e ξ como vetores y_r e ξ de dimensões $(n_r p_r \times 1)$ e $(a \times 1)$, respectivamente, este passa a ser:

$$E(y) = L \xi \quad e \quad V(y) = \Omega \quad (4.2.2)$$

onde, $y = (y_1', y_2', \dots, y_u)'$; $y_r (n_r p_r \times 1)$, $r=1, \dots, u$

$$L = \left[(A_1 \otimes B_1' P')' \mid \dots \mid (A_u \otimes B_u' P')' \right]' = \left[L_1' \mid \dots \mid L_u' \right]'$$

$$\Omega = \text{bloc diag} \left(I_{n_1} \otimes B_1' \Sigma B_1, \dots, I_{n_u} \otimes B_u' \Sigma B_u \right)$$

$$\xi = (\xi_1', \dots, \xi_a')'; \quad \xi_i = (\xi_{i0}, \dots, \xi_{iq-1})' \quad i=1, \dots, a$$

Assim como no MCC, no modelo (4.2.1) há interesse na estimação de conjuntos lineares estimáveis da forma $C\xi V$, no qual $C(g \times a)$ e $V(q \times v)$ são matrizes conhecidas. Entretanto, para o MCCG não é diretamente possível obter uma aproximação de ξ e Σ pelas equações de máxima verossimilhança. Deste modo, Kleinbaum(73) sugere um estimador com tão boas propriedades quanto o estimador de máxima verossimilhança na presença de grandes amostras. Esse consiste no "Melhor Estimador Assintoticamente Normal" (MEAN).

Escrevendo $C\xi V$ na forma de um vetor ($gv \times 1$), $g(\xi) = M\xi$, onde, $M=C \otimes V'$, então um estimador de um conjunto linear estimável $M\xi$ é dado como segue:

$$M \hat{\xi} = M (L'(\hat{\Omega})^{-1} L)^{-1} L'(\hat{\Omega})^{-1} y \tag{4.2.3}$$

onde $(L'(\hat{\Omega})^{-1} L)^{-1}$ é uma inversa generalizada de $L'(\hat{\Omega})^{-1} L$, $\hat{\Omega}$ é uma matriz desconhecida, cuja estimativa é determinada com base em uma estimativa consistente de Σ . Uma estimativa mais diretamente proposta por analogia algébrica é dada por:

$$\hat{\Sigma} = (\hat{\sigma}_{kk'}) = \frac{1}{n_{kk'} - p(L_{kk'})} x'_{kk'} \left[I - L_{kk'} (L'_{kk'} L_{kk'})^{-1} L'_{kk'} \right] x_{kk'} \tag{4.2.4}$$

k, k' = 1, ..., p

sendo,

$n_{kk'} (\geq 2)$ o número de unidades experimentais no qual t_k e $t_{k'}$ são ambos observados

$x_{kk'}(n_{kk'}, x1)$ o vetor de observação no qual os tempos t_k e $t_{k'}$ são ambos observados

$L_{kk'}(n_{kk'}, xa)$ a matriz de delineamento que consiste das linhas de A_r , correspondente aos elementos em $x_{kk'}$ e

$p(L_{kk'})$ o posto de $L_{kk'}$.

No entanto, esta estimativa como aponta Little e Rubin (87) não é necessariamente positiva definida, mas Schwertman e Allen(79) fornecem um procedimento para obtenção de uma versão alisada que torne a estimativa pelo menos semidefinida positiva.

A matriz de variâncias e covariâncias assintótica de $M_{\xi}^{\hat{}}$ é obtida diretamente do cálculo de:

$$V(M_{\xi}^{\hat{}}) = M(L'(\hat{\Omega})^{-1}L)^{-1}M' \quad (4.2.5)$$

mas,

$$\begin{aligned} L'(\hat{\Omega})^{-1}L &= L'_1 \left[I_{n_1} \otimes (B'_1 \hat{\Sigma} B_1)^{-1} \right] L_1 + \dots + L'_u \left[I_{n_u} \otimes (B'_u \hat{\Sigma} B_u)^{-1} \right] L_u \\ &= \sum_{r=1}^u A'_r A_r \otimes P B_r (B'_r \hat{\Sigma} B_r)^{-1} B'_r P' \end{aligned}$$

pois, $L_r = A_r \otimes B'_r P'$. Logo,

$$L'(\hat{\Omega})^{-1}L = \text{diag} \left(\sum_{r=1}^u n_{r1} P B_r (B'_r \hat{\Sigma} B_r)^{-1} B'_r P', \dots, \sum_{r=1}^u n_{ra} P B_r (B'_r \hat{\Sigma} B_r)^{-1} B'_r P' \right) \quad (4.2.6)$$

onde n_{ri} corresponde ao número de elementos do grupo i dentro do conjunto E_r .

$$A_r' A_r = \text{diag}(n_{r1}, \dots, n_{ra}), \text{ pois } A_r = (A_{r1}' | \dots | A_{ra}')'$$

sendo cada matriz A_{ri} a matriz de especificação dos indivíduos do grupo i dentro do conjunto E_r .

Da expressão (4.2.6) resulta que

$$V(\hat{M}\hat{\xi}) = M \{ \text{bloc diag } (V(\hat{\xi}_1), \dots, V(\hat{\xi}_a)) \} M' \quad (4.2.7)$$

4.2.2 Teste de Hipóteses

A estatística de teste proposta por Kleinbaum(73) para testar hipóteses lineares do tipo $H_0: M\xi=0$ baseia-se no conhecimento de um MEAN de $M\xi$ e de um estimador consistente de Σ . Esta estatística consiste em uma forma quadrática cuja distribuição assintótica é qui-quadrado central. Isso significa que à medida que o tamanho da amostra aumenta, a estatística de teste se aproxima cada vez mais dos valores da distribuição qui-quadrado central.

Pelo critério da razão da máxima verossimilhança, a estatística de teste de $H_0: M\xi=0$ é uma forma quadrática, no qual $M\hat{\xi}$ e $\hat{\Sigma}$ são estimativas de máxima verossimilhança de $M\xi$ e Σ , respectivamente. A estatística de Wald(43) para testar H_0 , dada acima, assintoticamente equivalente à baseada na máxima verossimilhança, é dada por:

$$W_n = (M\hat{\xi})' \left[M \left(\sum_{r=1}^u A_r' A_r \otimes P B_r (B_r' \hat{\Sigma} B_r)^{-1} B_r' P' \right) M' \right]^{-1} (M\hat{\xi}) \quad (4.2.8)$$

onde $M\hat{\xi}$ é o MEAN de $M\xi$ e $\hat{\Sigma}$ é um estimador consistente de Σ . Para M de posto completo, gv, W_n é assintoticamente qui-quadrado com gv graus de liberdade.

Utilizando propriedades assintóticas, (4.2.8) consiste na estatística de teste para verificar hipóteses lineares como $M\xi$ formuladas no MCCG.

4.2.3 Faixas de Confiança

Nesta seção são fornecidas algumas sugestões para determinação das faixas de confiança utilizando algumas informações contidas na seção anterior.

Inicialmente, considerando que a matriz de delineamento L , dada em (4.2.2), tenha posto completo e que $\hat{\Sigma}$, dada por (4.2.4), seja definida positiva, então, é possível estimar $\hat{\xi}$ unicamente por (4.2.3), não sendo necessária a utilização da inversa generalizada na expressão, e arranjar-lo matricialmente como $\hat{\xi}^+(axq)$. Tomando a parte completa de Y_0 , Y_{0+} e escrevendo adequadamente a matriz de especificação dos indivíduos dentro dos grupos A_+ é possível fornecer a partir daí, com os devidos cálculos, um intervalo de confiança simultâneo. A determinação deste intervalo pode ser feita aplicando a descrição de Morrison(76).

Retomando a hipótese multivariada $H_0: C\xi V = 0$ e o Modelo de Curva de Crescimento, para testar a hipótese H_0 faz-se necessário o cálculo da estatística de Pillai

$$\theta^+ = c_h / (1 + c_h)$$

onde c_α é a maior raiz característica de $SQ_H SQ_E^{-1}$. As somas de quadrados SQ_H e SQ_E são determinadas a partir das expressões:

$$SQ_H = (C\hat{\xi}^+V)'(CRC')^{-1}C\hat{\xi}^+V \quad e \quad SQ_E = V'(P\hat{\Sigma}^{-1}P')^{-1}V$$

onde,

$$R = (A_+^+A_+)^{-1} + (A_+^+A_+)^{-1}A_+^+ Y_{0+} \hat{\Sigma}^{-1} Y_{0+}' A_+ (A_+^+A_+)^{-1} - \hat{\xi}^+ (P\hat{\Sigma}^{-1}P')^{-1} \hat{\xi}^+ \quad (4.2.9)$$

A estatística θ^+ é referida à tabela de Pillai com parâmetros

$$h = \min(v, g) \quad l = (|v - g| - 1) / 2 \quad e \quad f = (n - a - p + q - v - 1) / 2$$

Sabendo que para $\forall b, d, H_0$ é verdadeira se, e só se, $b' C \xi V d = 0$ (Mardia et al(79)), então

$$P(b' C \xi V d \in b' C \hat{\xi}^+ V d \pm \{ \theta_\alpha^+ (d' S Q_E d) (b' C R C' b) \}^{1/2}, \forall d, b) = 1 - \alpha \quad (4.2.10)$$

onde $c_\alpha \equiv c_{\alpha; h, l, f}$ é o valor crítico da tabela de Pillai. No caso de $h=1$, $\theta_\alpha^+ = [(1+l)/(f+1)] F_{\alpha; 2l+2, 2f+2}$

Portanto, a expressão em (4.2.10) permite a determinação do intervalo de confiança simultâneo. Conseqüentemente, as faixas

de confiança são calculadas de (4.2.10), especificando o vetor d adequadamente para o grau do polinômio envolvido na comparação definida por C e V , considerando agora que ao invés da estimativa consistente de Σ dada por (4.2.4), fosse utilizada uma estimativa baseada somente no conjunto E_r , no qual todas as ocasiões foram observadas, conjunto completo (vide 4.3.7). Rearranjando o vetor $\hat{\xi}_+$, obtido na forma matricial $\hat{\xi}_+ (axq)$, e tomando A_+ e Y_{0+} , descritos anteriormente, substituindo $\hat{\xi}_+$ em (4.2.9) e nas somas de quadrados, assim como a $\hat{\Sigma}$ sugerida, determina-se uma nova estimativa do intervalo de confiança simultâneo para o polinômio ajustado.

4.3 UTILIZAÇÃO DO MÉTODO DAS DIFERENÇAS SUCESSIVAS EM DADOS INCOMPLETOS

De acordo com a abordagem dada no Capítulo 3, através do método DS, é possível indicar que grau de polinômio descreve a curva média de crescimento.

No caso de uma estrutura de dados incompletos a formulação do método DS precisa ser adaptada apropriadamente, visando a determinação do grau do polinômio que descreve o perfil médio de um grupo de unidades experimentais. A adaptação é feita apropriadamente a seguir e a determinação do grau do polinômio é uma extensão para diferenças de maior ordem do tratamento detalhado para diferenças de ordem um, no contexto de paralelismo. No entanto, esta extensão não será tratada aqui.

As seções 4.3.1 e 4.3.2, a seguir, são colocadas com o intuito de expor algumas expressões utilizadas posteriormente, isto é, a fim de permitir a adaptação do método DS para dados incompletos e de fornecer a estatística de teste de paralelismo de

perfis médios, na existência de mais de um grupo de indivíduos, também via o método DS. Isto feito recorrendo essencialmente à diferença de ordem 1. Embora o tratamento seja análogo para diferenças de maior ordem, não será documentado aqui.

4.3.1 Dados Preliminares

Seja $\hat{y}_{ij}(p \times 1)$ o vetor de respostas do indivíduo j do grupo i não completamente observado em p tempos, $i=1, \dots, a$; $j=1, \dots, n_i$.

O perfil de um indivíduo, neste contexto, pode ser escrito em função do vetor p -variado completo através de uma transformação linear.

Desta forma, considere a matriz C_{ij} , correspondente ao indivíduo j do grupo i , como segue:

$$C_{ij} = \text{diag} (c_{ij1}, \dots, c_{ijp})$$

$$\text{onde } c_{ijk} = \begin{cases} 1 & \text{se o indivíduo } j \text{ do grupo } i \text{ é observado na ocasião } k \\ & \text{ou no tempo } t_k. \\ 0 & \text{caso contrário.} \end{cases}$$

Então o perfil individual é dado pela expressão:

$$\hat{y}_{ij} = C_{ij} y_{ij}, \quad i=1, \dots, a; \quad j=1, \dots, n_i$$

Para um grupo em particular, o vetor que descreve o perfil médio amostral é dado por:

$$\dot{\bar{y}}_1 = (\dot{\bar{y}}_{11}, \dots, \dot{\bar{y}}_{1p})'$$

onde cada componente de $\dot{\bar{y}}_1$ pode ser expressa como:

$$\dot{\bar{y}}_{1k} = \sum_{j=1}^{n_1} y_{1jk} c_{1jk} / n_{1(k)} \quad ; \quad n_{1(k)} = \sum_j c_{1jk} \quad \text{é o número de}$$

respostas observadas no tempo t_k , $k=1, \dots, p$.

A estimativa da matriz de variâncias e covariâncias $\Sigma_1(p \times p)$, associada a cada grupo, na sua forma mais simples, pode ser expressa por:

$$S_1 = (s_{1(k'k)}), \text{ com } s_{1(k'k)} = \sum_j \frac{c_{1jk}(y_{1jk} - \dot{\bar{y}}_{1k}^*)(y_{1jk} - \dot{\bar{y}}_{1k}^*)c_{1jk}}{n_{1(k'k)} - 1} \quad (4.3.1)$$

onde $\dot{\bar{y}}_{1k}^*$ é a média amostral calculada no tempo t_k baseando-se nos indivíduos que possuíam as observações em t_k e $t_{k'}$, de um total de $n_{1(k'k)} = \sum_j c_{1jk} c_{1jk'}$ indivíduos, nos quais t_k e $t_{k'}$ foram observados. Esta estimativa S_1 não é, necessariamente, definida positiva, mas procedimentos para alcançar esta importante propriedade encontram-se citados no artigo de Ware(85).

Para efeito de analogia, considere o caso completo, tomando o empilhamento por linha.

$$\bar{y}_1 \sim N_p(\theta_1, \Sigma_1^{-1}) \quad \theta_1 = (\theta_{11}, \dots, \theta_{1p})'$$

e

$$E(y_1) = 1_{n_1} \otimes \theta_1 \quad e \quad V(y_1) = I_{n_1} \otimes \Sigma_1$$

$y_1 = (y'_{11}, \dots, y'_{1n_1})'$ é o vetor que corresponde ao empilhamento das linhas da matriz de dados do grupo 1.

No caso da estrutura de dados incompletos, os estimadores, respectivas distribuições e modelo podem ser expressos como se segue, levando-se em conta que a notação introduzida foi modificada para evitar degeneração na formulação matricial.

Seja,

$$D_1 = \text{diag} \left(1/n_{1(1)}, \dots, 1/n_{1(p)} \right)_{p \times p}$$

O vetor que contém os dados do grupo i empilhando as linhas, com zeros nas ocasiões em que os dados são faltantes, é dado por:

$$\dot{y}_1 = C_1 y_1 \quad ; \quad i=1, \dots, a$$

com $C_1 = \text{bloc diag} \left(C_{11}, \dots, C_{1n_1} \right)$

Note que a matriz $C_1(n_1 p \times n_1 p)$ é indexada pelo grupo e cada bloco C_{1j} de C_1 é indexado pelo indivíduo e grupo. De fato, o produto de cada bloco de p linhas de C_1 pelo vetor $y_1(n_1 p \times 1)$ reproduz, para o indivíduo do bloco correspondente, o vetor de observações nos p tempos. Naturalmente, por construção, com as respostas nos tempos observados e zeros nos tempos nos quais faltaram as respostas. Para eliminar as linhas de zero, considere

as modificações em C_{ij} e C_i para C_{ij}^* e C_i^* , respectivamente. Na verdade, essas últimas serão responsáveis por uma nova combinação linear em y_i , que reproduz as ocasiões em que efetivamente houve resposta, ou seja,

$$y_i^* = C_i^* y_i \quad (4.3.3)$$

Aplicando agora a matriz modificada $C_{ij}^*(p_{ij} \times p_{ij})$, na qual p_{ij} é o número de respostas observadas no indivíduo j do grupo i . A matriz C_i^* é constituída de blocos semelhantes aos de C_i , no entanto, extraindo desta última as linhas de zeros. Desta forma, sendo $\sum p_{ij}$ o número de respostas observadas no grupo i , a dimensão de C_i^* é $(\sum p_{ij} \times n_i \times p)$. As respostas esperadas segundo tal formulação são dadas por:

$$E(y_i^*) = C_i^* (I_{n_i} \otimes \theta_i) \quad (4.3.4)$$

$$V(y_i^*) = C_i^* (I_{n_i} \otimes \Sigma_i) C_i^* \quad (4.3.5)$$

O vetor de médias amostral \bar{y}_i^* é calculado baseando-se nas respostas observadas e no número de respostas observadas por ocasião. O mesmo ocorre com a matriz de variâncias e covariâncias.

Para muitas comparações onde cabe uma estimativa comum da matriz de variâncias e covariâncias, ou seja, quando há igualdade entre as Σ_i 's, uma forma de usar (4.3.1) para produzir esta estimativa comum Σ_c é dada abaixo:

$$S_c = (s_{c(k'k)})_i s_{c(k'k)} = \frac{s_{1(k'k)} n_{1(k'k)} + \dots + s_{a(k'k)} n_{a(k'k)}}{n_{1(k'k)} + \dots + n_{a(k'k)}} \quad (4.3.6)$$

de tal forma que $S_c = \sum_c (p \times p)$. (vide Schwertman e Allen(79) caso S_c não seja definida positiva)

4.3.2 Desprezando os Dados Incompletos

Aqui são considerados os dados completamente observados, isto é, aqueles nos quais o indivíduo possui respostas ao longo dos k tempos de observação.

Como mencionado anteriormente, se somente os dados completamente observados são considerados na análise, ignorando qualquer natureza da perda, é natural de se esperar que hajam vantagens e desvantagens nesse procedimento. Uma das vantagens é a *simplicidade*, já que poderão ser analisados segundo as técnicas estatísticas para dados completos, outra é a *comparabilidade*, pois um mesmo número de respostas é comum a todos os indivíduos, facilitando até a aplicação dos métodos multivariados convencionais. Como desvantagens, aparece a perda de informação no descarte dos casos incompletos e, ainda, a possibilidade dos casos completos selecionados constituírem uma sub amostra viciada.

Supondo que o número de unidades experimentais é suficientemente grande para que o número de observações completas seja superior ao número de tempos de observação em estudo, então:

$\bar{y}_{1+} = 1'_{n_{1(.,.)}} Y_{01+} 1/n_{1(.,.)}$ é o vetor de média amostral dos dados completos do grupo i , contidos em Y_{01+} (matriz de dados completos do grupo i), ignorando os $n_i - n_{1(.,.)}$ incompletos.

A matriz de variâncias e covariâncias amostral dos dados completos do grupo i S_{1+} possui os elementos calculados pela expressão:

$$S_{1+} = (Y'_{01+} H Y_{01+}) / n_{1(\dots)} \quad (4.3.7)$$

onde, $H = I - 11' / n_{1(\dots)}$,

ou, ainda,

$S_{1+}^* = \left(n_{1(\dots)} / (n_{1(\dots)} - 1) \right) S_{1+}$ é uma estimativa de Σ_1 definida positiva, para $n_{1(\dots)} > p$

Sob suposição de multinormalidade dos y_{1j+} desta sub-amostra considerada, então:

$$n_{1(\dots)}^{-1/2} (\bar{y}_{1+} - \theta_1) \sim N_p(0, \Sigma_1)$$

e, como H é uma matriz idempotente de posto $n_{1(\dots)} - 1$ e os $(y_{1j+} - \theta_1)$ multinormais independentes $(0, \Sigma_1)$, então,

$$n_{1(\dots)} S_{1+} \sim W_p(\Sigma_1, n_{1(\dots)} - 1)$$

isto é, tem distribuição de Wishart central com $n_{1(\dots)} - 1$ graus de liberdade.

A matriz de variâncias e covariâncias amostral comum, para os a' grupos ($a' \leq a$) com dados completos S_{c+} , é obtida a partir de:

$$S_{c+} = (n_{1(\dots)} S_{1+} + \dots + n_{a'(\dots)} S_{a'+}) / n_{\dots} ; \quad (4.3.7)$$

$n_{\dots} = \sum_{i=1}^{a'} n_{i(\dots)}$: número de indivíduos completamente observados nos a' grupos.

4.3.3 Adaptando o Método DS para Teste de Paralelismo

O propósito desta seção é expor um método de comparação de perfis médios. A abordagem se restringe a constatar o paralelismo (ou ausência de) entre os perfis médios de a' grupos de indivíduos ($a' \leq a$), baseando-se num modelo linear geral ajustado para as **Diferenças Sucessivas de ordem 1**. A comparação de perfis médios, naturalmente, vai além da checagem de paralelismo.

Considere as diferenças sucessivas de primeira ordem aplicada sobre o vetor de médias y_1^* , como expresso em (3.2.2)

$$\Delta^1 y_1^* = U_{11} y_1^* = U_{11} C_1^* y_1 \quad (4.3.9)$$

onde $U_{11} ((\sum p_{1j}) \times \sum p_{1j})$ é como ilustrado na seção 3.2.

Particionando U_{11} convenientemente em blocos, de modo que cada bloco tenha a matriz de transformação U_{11j} de cada indivíduo, tem-se:

$$U_{11} = \text{bloc diag} \left(U_{111}, \dots, U_{11n_1} \right)$$

cada U_{11j} é idêntica a matriz U_1 descrita na seção 3.2, entretanto, com dimensão $(p_{1j}-1) \times p_{1j}$.

O valor esperado da expressão em (4.3.9) é, portanto,

$$E(\Delta^1 y_1^*) = U_{11} C_1^* (1_{n_1} \otimes \theta_1) \quad (4.3.10)$$

ou ainda, de forma mais simplificada, tomando as diferenças para cada indivíduo,

$$E(\Delta^1 y_{1j}^*) = U_{11j} C_{1j}^* \theta_{-1} \quad (4.3.11)$$

Nota-se pela expressão (4.3.10) e, mais detalhadamente, por (4.3.11) que as DS de ordem 1 fornecem a mudança de reposta de um tempo t_k observado para um tempo $t_{k'}$, também observado.

Considere que a' grupos estejam envolvidos no teste de paralelismo objetivado. Tomando os y_i^* 's destes a' grupos e, escrevendo compactadamente as observações de todos os grupos em um só vetor y^* , pode-se obter as DS de ordem 1 para este novo vetor da seguinte forma

$$\Delta^1 y^* = \text{bloc diag} \left\{ U_{11}, \dots, U_{1a'} \right\} \left\{ y_1^*, \dots, y_{a'}^* \right\} \quad (4.3.12)$$

cujo valor esperado e matriz de variâncias e covariâncias são dados abaixo

$$E(\Delta^1 y^*) = \text{bloc diag} \left\{ U_{11}, \dots, U_{1a'} \right\} \left\{ \left(C_1^* (1_{n_1} \otimes \theta_{-1}) \right)', \dots, \left(C_{a'}^* (1_{n_{a'}} \otimes \theta_{-1}) \right)' \right\}$$

e

$$(4.3.13)$$

$$V(\Delta^1 y^*) = \text{bloc diag} \left\{ U_{11} C_1^* C_1 (1_{n_1} \otimes \Sigma_1) C_1' C_1' U_{11}', \dots, U_{1a'} C_{a'}^* C_{a'} (1_{n_{a'}} \otimes \Sigma_{a'}) C_{a'}' C_{a'}' U_{1a'} \right\}$$

Retomando à questão do teste de paralelismo colocada no início desta seção, para viabilizar a formulação da hipótese de paralelismo, é necessário recorrer à formulação do MCGG. De maneira análoga, considere que n indivíduos estejam divididos em u conjuntos disjuntos E_r , com m_r elementos cada. As observações

nesses conjuntos correspondem às respostas dos indivíduos observados nos mesmos tempos (ou ocasiões), independentemente do grupo a que ele pertence.

Para incorporar as DS neste contexto, torna-se necessário um rearranjo de y^* baseando-se nos conjuntos E_r a na forma que os indivíduos estão dispostos nesses conjuntos E_r . Uma reordenação dos indivíduos implica numa reformulação que permita expressar adequadamente as DS dentro de cada conjunto E_r .

Notando por y^0 o novo vetor decorrente do empilhamento das respostas por conjunto, isto é,

$$y^0 = (y^1, \dots, y^u),$$

onde y^r contém as respostas dos indivíduos que compõem o conjunto E_r , ou seja, aqueles que foram observados nos mesmos p_r tempos.

O valor esperado da DS de ordem 1 aplicada sobre y^0 é dada pela expressão (4.3.14).

$$E(\Delta^1 y^0) = \text{bloc diag}(I_{m_1} \otimes U_1^1, \dots, I_{m_u} \otimes U_1^u) E(y^0) \quad (4.3.14)$$

onde U_1^r é a matriz que estabelece a diferença para os m_r indivíduos do conjunto E_r e que é comum a todos, independentemente de grupo.

A esperança do vetor reordenado, para cada conjunto E_r , pode ser expresso por:

$$E(y^r) = (I_{m_r} \otimes C^{*r}) \bar{e}^r \quad (4.3.15)$$

com $\underline{\theta}^r$ correspondendo ao vetor de médias associado aos grupos a que pertencem os indivíduos do conjunto E_r e a matriz C^{*r} é responsável por selecionar os tempos observados para os m_r indivíduos do conjunto E_r , assim como anteriormente foi usada C_{ij} . Como todos os indivíduos de um mesmo conjunto foram observados nos mesmos tempos não há necessidade de indexar o indivíduo e grupo, somente o conjunto E_r .

Como consequência natural de (4.3.15), a esperança do vetor reordenado inteiro é dada por:

$$E(y^0) = (((I_{m_1} \otimes C^{*1}) \underline{\theta}^1)', \dots, ((I_{m_u} \otimes C^{*u}) \underline{\theta}^u)')' \quad (4.3.16)$$

Com isto, as diferenças aplicadas sobre os vetores individuais dentro de cada conjunto E_r correspondem aos mesmos contrastes de médias no que diz respeito às ocasiões, como exposto nas expressões (4.3.9) a (4.3.13).

A hipótese de paralelismo pode ser tratada tomando as correspondentes DS aplicadas sobre o vetor reordenado, para cada conjunto E_r , da seguinte forma:

$$E(\Delta^1 y^r) = (I_{m_r} \otimes U_1) (I_{m_r} \otimes C^{*r}) \underline{\theta}^r \quad (4.3.17)$$

Para explicitar a hipótese de paralelismo, torna-se necessário descrever o vetor $\underline{\theta}^r$ como segue:

$$\underline{\theta}^r = (\underline{\theta}_{(1)}^r, \dots, \underline{\theta}_{(s)}^r)'$$

ou seja, $\underline{\theta}_{(i)}^r$ é o vetor de médias do i -ésimo grupo que forneceu elementos para E_r , portanto contém apenas as médias nas ocasiões

definem E_r

Sob hipótese de paralelismo entre os perfis médios dos s grupos dentro do conjunto E_r , então:

$$U_1 C^{*r} \theta_{\sim(1)}^r = U_1 C^{*r} \theta_{\sim(2)}^r = \dots = U_1 C^{*r} \theta_{\sim(s)}^r$$

De fato, em termos de diferenças, o problema consiste em testar se os contrastes de médias estabelecidos pelas DS tem esperança constante, entre indivíduos de grupos distintos dentro de um mesmo conjunto. Como em um mesmo conjunto E_r deve ser verificada a presença de pelo menos dois grupos, a hipótese de igualdade das matrizes de variâncias e covariâncias através dos grupos é fator relevante no desenvolvimento do teste de paralelismo. De tal forma que uma estimativa comum pode ser empregada, sob hipótese de igualdade, como apresentado abaixo:

$$V(y_j^r) = (C^{*r}) \Sigma (C^{*r})' \quad \text{e} \quad V(\Delta^1 y_j^r) = U_1 C^{*r} \Sigma (C^{*r})' U_1' \quad (4.3.18)$$

A hipótese H_0 , de paralelismo dentro do conjunto E_r , pode ser formulada como em (4.3.19)

$$H_0: U_1 C^{*r} \theta_{\sim(1)}^r = \dots = U_1 C^{*r} \theta_{\sim(s)}^r \quad (4.3.19)$$

Definindo,

$$z_j^r = U_1 C^{*r} y_j^r \quad , \quad j=1, \dots, m_r$$

é possível escrever a variação dentro dos grupos que compreendem o conjunto E_r por D_r ,

$$D_r = \sum_{i=1}^s \sum_{j=1}^{m_{ri}} (z_{(i)j}^r - \bar{z}_{(i)}^r)(z_{(i)j}^r - \bar{z}_{(i)}^r)' \quad (4.3.20)$$

$m_{ri} \geq 2$

onde m_{ri} é o número de indivíduos do grupo i dentro do conjunto E_r , $\bar{z}_{(i)}^r$ é o vetor de médias amostral das DS aplicadas nos indivíduos do grupo i dentro de E_r e $z_{(i)j}^r$ é o vetor de diferenças de respostas do indivíduo j do grupo i dentro de E_r . Os vetores $\bar{z}_{(i)}^r$ e $z_{(i)j}^r$ não possuem dimensões necessariamente distintas de um conjunto para outro.

A variação total dentro do conjunto sem discriminar grupo é descrita pela soma de quadrados T_r

$$T_r = \sum_{i=1}^s \sum_{j=1}^{m_{ri}} (z_{(i)j}^r - \bar{z}_{(i)}^r)(z_{(i)j}^r - \bar{z}_{(i)}^r)' \quad (4.3.21)$$

onde $m_r = \sum_{i=1}^s m_{ri} \geq 2s$

Com base em (4.3.20) e (4.3.21) a estatística de teste de paralelismo é dada pela razão dos determinantes de D_r por T_r , isto é,

$$\Lambda_r = \frac{|D_r|}{|T_r|} \sim \Lambda(p_r, m_r - s, s-1) \quad (4.3.22)$$

tem distribuição de Wilks com os parâmetros discriminados acima.

Como os conjuntos E_r são disjuntos, este procedimento pode ser empregado para cada um independentemente. No entanto, uma estatística de teste razoável, que incorpore testes para todos os conjuntos simultaneamente, isto é, $\Lambda_1, \dots, \Lambda_u$ é de grande interesse. Em primeiro lugar, como os E_r são disjuntos, há

independência entre os Λ_r , $r=1, \dots, u$. Além disso a estatística de Wilks pode ser escrita como produto de variáveis do tipo beta independentes, com distribuição exata conhecida conforme Rathie e Mathai(71). Nesse trabalho, os autores esclarecem a utilização de produtos de variáveis do tipo beta para derivação da distribuição exata da estatística de Wilks. Para que seja mantida a estrutura de momentos, a estatística de teste conjunto, levando em conta todos os E_r , deve ser proposta como $\Lambda = \prod_{r=1}^u \Lambda_r$. A distribuição exata de Λ através da transformada de Melin poderá então ser caracterizada e seus percentis obtidos.

A validade das estatísticas derivadas dentro de cada E_r está condicionada à suposição de que as observações sejam provenientes de populações multinormais com médias possivelmente distintas e mesma matriz de variâncias e covariâncias. A presença necessária de indivíduos de grupos distintos em cada E_r , aliada à observação do período anterior, deixam claro que a validade do procedimento depende da hipótese de igualdade das matrizes de variâncias e covariâncias entre os grupos considerados no conjunto. Em caso contrário, o problema é bastante complexo pois a desigualdade das matrizes de variâncias e covariâncias leva ao problema de Behrens-Fisher multivariado que foi tratado por Nel(87), porém no caso de dois grupos. Neste contexto, a adaptação do teste dentro dos E_r só poderia ser abordada para dois grupos. Para mais de dois grupos, que é o caso considerado, é necessário estender os resultados de Nel(87) e depois adaptá-los para dados incompletos. Naturalmente, além do escopo deste trabalho.

O teste derivado em (4.3.22) informa limitações mínimas sobre indivíduos e grupos, pois como pode ser verificado, cada Λ_r pode evidenciar ausência de paralelismo apenas entre os grupos que figuram em E_r , levando a necessidade de em cada E_r haver indivíduos de pelo menos dois grupos. Por outro lado, o mínimo

para obter informação da variação dentro (4.3.20) é que em cada grupo se tenha pelo menos dois indivíduos. Os parâmetros da estatística de Wilks juntamente com a restrição associada a variação dentro sugerem, portanto, que o número de indivíduos em cada E_r supere 2s.

A constatação de paralelismo, através dos testes mencionados depende do exame das ocasiões que aparecem em cada E_r . Por exemplo, se em um grupo as observações concentram-se nas ocasiões iniciais e finais, com o teste não rejeitando H_0 , pode ser que não ocorra o paralelismo de fato. A diferença de padrão do perfil médio pode ser devida ao comportamento diferenciado na região sobre a qual há falta de informação. Não há mágica no procedimento de montagem do teste e menos ainda no resultado. Em cada situação, a configuração dos dados incompletos, com respeito às ocasiões e grupos, requer uma investigação detalhada dos conjuntos E_r . Ainda, esta visão estruturada deve preceder o uso de métodos iterativos, pois informa melhor as possíveis limitações dos resultados.

4.3.4 O Método DS no Ajuste Polinomial

Retomando a idéia já explorada no capítulo anterior, relativa à aplicação do método DS para um ou mais grupos, o intuito agora é expor a aplicação das DS, levando em conta a informação do conjunto completo, a fim de fornecer um indicador do grau do polinômio que descreve a curva média do grupo.

Considere que y_{ij+} seja o perfil de um indivíduo completamente observado, ou seja, pertencente ao conjunto E_r no qual existem respostas nos p -tempos.

Sendo \bar{y}_{i+} o vetor p -variado das médias amostrais, calcu-

lado sobre y_{1j^*} , como considerado na seção 4.3.2, então:

$$\Delta^s \bar{y}_{1+} = U_s \bar{y}_{1+}$$

logo,
$$E(\Delta^s \bar{y}_{1+}) = U_s \theta_{-1}$$

e
$$\Sigma_{s1} = U_s V(\bar{y}_{1+}) U_s' = U_s \Sigma_1 U_s' / n_{1(\dots)} \quad (4.3.23)$$

A estimativa de Σ_{s1} é obtida simplesmente da substituição de Σ_1 por $\hat{\Sigma}_1 = S_{1+}$, ou seja,

$$S_{s1+} = U_s S_{1+} U_s' / n_{1(\dots)} \quad (4.3.24)$$

De modo análogo ao exposto no capítulo anterior, o procedimento de determinação do grau do polinômio para a curva média segue o resultado para as $n_{1(\dots)}$ observações, ou seja, se q for o grau do polinômio que descreve a curva média do grupo i e $n_{1(\dots)} > p-1$, então

$$H_0: E(\Delta^s \bar{y}_{1+}) = 0 ; \text{ para } s=q+1, \dots, p-1$$

é a hipótese de interesse que pode ser testada pela estatística T^2 de Hotelling, da mesma forma que na seção 3.3.1

$$n_{1(\dots)}^{-1/2} (\Delta^s \bar{y}_{1+}) \sim N_{p-s}(0, \Sigma_{s1})$$

e

$$n_{1(\dots)} S_{s1+} \sim W_{p-s}(\Sigma_{s1}, n_{1(\dots)} - 1)$$

onde S_{s1+} é dada pela forma quadrática abaixo

$$S_{s1+} = (Y_{01+} \ U'_s)' (I - 1/n_{1(..)} 11') (Y_{01+} \ U'_s) = (Y_{01+} \ U'_s)' H (Y_{01+} \ U'_s)$$

4.3.4.1 Algumas Considerações

Diante da metodologia apresentada na seção anterior pode-se notar a importância da estrutura dos dados faltantes, mesmo utilizando a separação por conjuntos. Neste contexto, desprezar os dados incompletos significa eliminar algumas observações que poderiam influenciar no cálculo das variâncias e covariâncias. Mas não só devido a essa questão, pois, o tamanho de $n_{1(..)}$ é de fundamental importância, a não tendenciosidade dos dados contidos em Y_{01+} , e, ainda, a singularidade de S_{1+} se $n_{1(..)}$ for inferior a $p-1$.

Diante da determinação de q , uma alternativa pode ser um ajuste polinomial a fim de estimar os valores perdidos e, tendo um novo conjunto completo, utilizar algum dos métodos multivariados convencionais para determinação da curva média estimada, faixas de confiança e, possivelmente, teste de outras hipóteses, que não a de paralelismo.

Vale ressaltar que os resultados provenientes da seção 4.3.3 são relevantes diante do fato de que a estrutura de dados incompletos se mantém e, ainda assim, são notadas evidências sobre o paralelismo dos perfis. Na presença de dados incompletos, este teste pode preceder a determinação do indicador $q-1$, dependendo do interesse no problema. O teste específico do grau do polinômio incorporando a estrutura de dados incompletos é derivado de forma análoga ao desenvolvido para o teste de paralelismo no qual se utiliza as DS de ordem 1. Na verdade, o procedimento baseia-se em considerar as diferenças de ordens superiores nas expressões desenvolvidas na seção 4.3.2.

CAPÍTULO 5

5.1 INTRODUÇÃO

O presente capítulo tem a finalidade de ilustrar alguns dos aspectos abordados nos capítulos anteriores.

A ilustração é feita através de um experimento real que caracteriza a situação em que os dados são incompletos e, nesse caso, as causas são determinísticas.

A análise dos dados é feita utilizando não só a metodologia apresentada, mas, aplicando técnicas alternativas não formalizadas neste texto, entretanto, difundidas nos estudos de medidas repetidas.

5.2 O EXPERIMENTO

Trata-se de um experimento que foi conduzido na FEAGRI/UNICAMP em 1983 na área de processamento agrícola pela docente Marlene Rita de Queiroz com o objetivo de investigar

condições de armazenamento do ovo de galinha, por seu alto consumo pela população e grande parte das indústrias alimentícias.

A finalidade do estudo está na análise da variação de peso dos ovos ao longo do tempo, sob diferentes condições de armazenamento, visando a preservação da qualidade do produto.

Na parte aqui considerada e adaptada deste experimento, foram utilizados dois lotes de 90 ovos cada. Os ovos destes lotes eram ovos brancos de tamanho extra (peso variando de 60 a 65 gramas), os quais foram armazenados à temperatura de 8 °C e oleados por duas vezes para impermeabilização da casca. Em um lote foi aplicado o óleo tipo 1 (W.O.USP-85) e no outro o óleo tipo 2 (USP-180/190), possibilitando a comparação de tipos dos armazenamento.

Sob as condições descritas acima, os ovos foram acompanhados por 120 dias e pesados periodicamente a cada 20 dias. A cada período, 12 ovos selecionados ao acaso, foram quebrados para cálculo das dimensões da gema, de modo que ao final dos 120 dias somente 18 ovos foram pesados em 7 ocasiões.

5.3 ANÁLISE DO EXPERIMENTO

Conforme mencionado na seção 5.2, o estudo está, essencialmente, associado a uma comparação de dois grupos de tratamento, definidos pelo tipo de óleo aplicado nos ovos. O experimento ilustra claramente um estudo em medidas repetidas com dados incompletos. Neste caso, a falta de observações é devida a uma causa determinística (quebra intencional dos ovos), para atender ao interesse, na observação da variável índice de gema,

relativos à avaliação dos efeitos de grupo (tipo de armazenamento), tempo (ocasião) e da interação entre esses fatores. Nesta variável a estrutura dos dados é, portanto, de doze observações independentes em cada tempo.

5.3.1 Análise sob Estrutura de Dados Incompletos

Na situação experimental, a variável resposta de interesse y_{ijk} representa o peso do j -ésimo ovo o qual foi aplicado o óleo tipo i na ocasião k ; $j=1,\dots,90$, $i=1,2$ e $k=1,\dots,7$

A princípio são estimados as médias de peso e respectivos desvios padrões, por ocasião, de cada tipo de óleo. Estes constam na Tabela 5.3.1 e em formas de perfis nas figuras 5.3.1 e 5.3.2 .

Tabela 5.3.1 - Médias de pesos dos ovos e respectivos desvios padrões estimados, por ocasião, para cada tipo de óleo

ocasião	óleo tipo 1		óleo tipo 2	
	resposta média (g)	desvio padrao(g)	resposta média(g)	desvio padrao(g)
1 (90)*	62.15	1.30	62.25	1.35
2 (78)	62.16	1.33	62.16	1.42
3 (66)	62.08	1.36	62.04	1.45
4 (54)	61.89	1.43	61.89	1.46
5 (42)	61.71	1.51	62.00	1.11
6 (30)	61.70	1.64	62.16	1.15
7 (18)	61.75	1.76	62.05	1.21

* (n_k) : número de ovos pesados na ocasião k

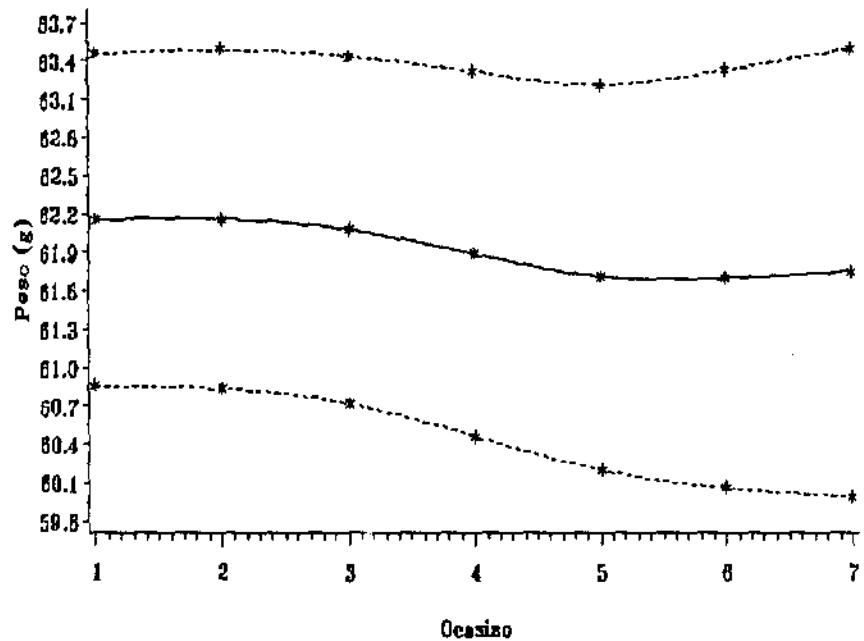


Figura 5.3.1 - Perfil Medio de Peso de Ovos do Grupo 1 por Ocasiao

—:Oleo Tipo 1

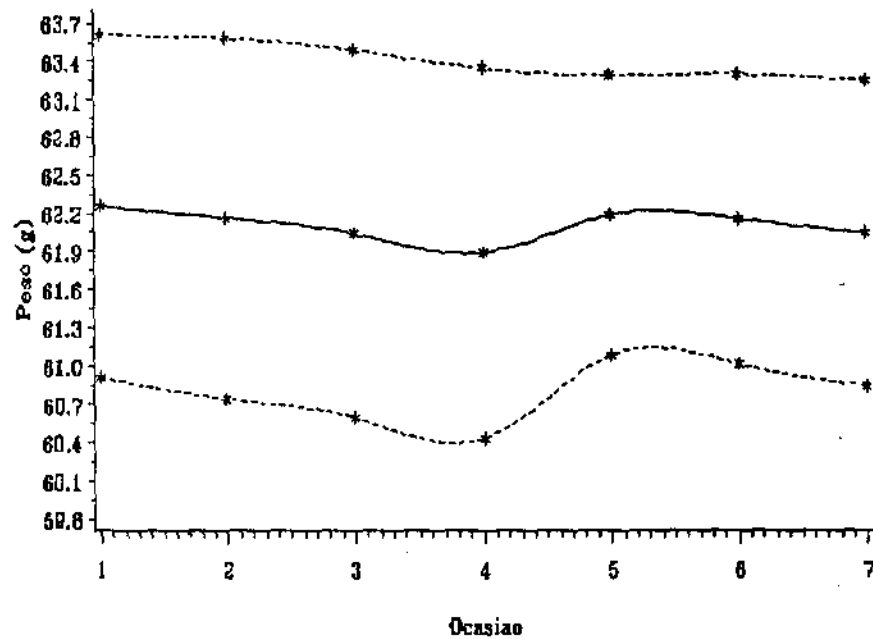


Figura 5.3.2 - Perfil Medio de Peso de Ovos do Grupo 2 por Ocasiao

--:Oleo Tipo 2

Como foi mencionado anteriormente, o processo de quebra dos ovos introduz a falta de informação e, dessa forma, os conjuntos disjuntos descritos na seção 4.2.1 ficam definidos. A cada ocasião em que ovos são quebrados, surge um novo conjunto. Assim, em cada conjunto E_r , estão os pesos daqueles ovos que foram quebrados na mesma ocasião, sendo que para um dos conjuntos os ovos foram pesados em todo o período de observação do experimento, isto é, as 7 ocasiões.

Neste caso, $\theta_i = (\theta_{i1}, \dots, \theta_{i7})'$ é o vetor de pesos médios do grupo i e Σ_i a matriz de variâncias e covariâncias associada aos indivíduos do grupo i .

Sabendo que para a utilização de uma estimativa comum da matriz de variâncias e covariâncias a hipótese de igualdade entre Σ_1 e Σ_2 deve estar satisfeita, antes de proceder com qualquer análise, é conveniente o teste das seguintes hipóteses:

$$H_0: \Sigma_1 = \Sigma_2 \quad \text{versus} \quad H_1: \Sigma_1 \neq \Sigma_2$$

O teste da hipótese H_0 segue do cálculo das seguintes expressões:

$$\Lambda = \frac{\prod |\Sigma_1|^{n_1/2}}{|\Sigma|^{n/2}} \quad B = 1 - \left\{ \sum_1 \frac{1}{n_1 - 1} - \frac{1}{n - a} \right\} \frac{2p^2 + 3p - 1}{6(p+1)(k-1)}$$

onde, n_1 é o número de observações por grupo; p é o número de ocasiões; a é o número de grupos e $n = n_1 + \dots + n_a$ ($n_1 = n_2 = 18$ ovos; $p = 7$ ocasiões; $a = 2$ grupos e $n = 36$ ovos).

sob H_0 ,

$$-2B \ln \left\{ \frac{n^{np/2} \Lambda}{\prod (n_1)^{n_1 p/2}} \right\} \sim \chi^2_{(a-1)p(p+1)/2}$$

Para o ajuste do modelo (4.2.1), nos dados deste experimento, 7 conjuntos disjuntos são obtidos com $p_r=r$ ocasiões observadas, $n_r=12$ ovos em cada grupo para $r=1, \dots, 6$ e $n_7=18$ ovos, dando um total de 90 ovos por lote (180 ovos no total).

No experimento estudado aqui são utilizadas as estimativas das matrizes de variâncias e covariâncias baseadas na parte completa dos dados, isto é, nas medidas dos ovos que foram observados nas 7 ocasiões. Essas estimativas constam dos resultados apresentados no Apêndice B e foram obtidas, assim como os outros resultados, utilizando o pacote estatístico SAS. O valor do teste qui-quadrado foi 126.70 que com 28 graus de liberdade fornece uma probabilidade de significância de 0.001. Diante deste resultado, há evidências para rejeitar a homogeneidade entre as matrizes consideradas, isto é, a análise segue separadamente para os dois grupos.

Escrevendo as expressões em (4.2.1),

$$A_{1r} = \mathbf{1}_{12} \quad r=1, \dots, 6 \quad \text{e} \quad A_{17} = \mathbf{1}_{18} ; \quad i=1,2$$

onde $\mathbf{1}_{n_{1r}}$ é um vetor coluna de tantos uns quantos elementos tiverem do grupo i no conjunto r . Aqui os grupos são idênticos do ponto de vista estrutural.

Levando em conta que os ovos, com o passar do tempo, possuem a característica de perder peso, pensando no ajuste polinomial, parece razoável o ajuste de um polinômio de primeiro grau.

$$P = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{pmatrix} = [P_1] \dots [P_7]$$

$$B_r = \begin{pmatrix} I_r \\ O_{7-r} \end{pmatrix}, \quad r=1, \dots, 7$$

onde O_{7-r} é uma matriz de zeros de dimensão $((7-r) \times r)$.

$$\xi = \begin{pmatrix} \xi_{10} & \xi_{11} \\ \xi_{20} & \xi_{21} \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \quad (5.3.1)$$

e ξ_1 é a matriz de parâmetros, na qual ξ_{1q} é o coeficiente polinomial relativo ao grau q do ajuste feito ao grupo i .

De acordo com a notação pré estabelecida $\xi_1 = (\xi_{01}, \xi_{11})'$ é a matriz ξ_1 escrita vetorialmente, por linha.

A expressão de $V(Y_{0r})$ dada em (4.2.1) é portanto:

$$V(Y_{0r}) = I_n \otimes (B_r' \Sigma_1 B_r) \quad (5.3.2)$$

para $r=k=1, \dots, 6$, $V(Y_{0k}) = I_{12} \otimes \Sigma_{kk}$

e $V(Y_{07}) = I_{18} \otimes \Sigma_{77} = I_{18} \otimes \Sigma$, de modo que,

Σ_{kk} é o bloco de Σ que fornece as variâncias e covariâncias relativas as ocasiões de 1 a k , portanto de dimensão $(k \times k)$.

Nota: (i) Devem ficar claras, que as particularidades do experimento com respeito a forma estruturada de perda de dados facilitaram a explicitação das matrizes expressas por (5.3.2). Aqui, há o mesmo número de conjuntos e ocasiões, assim como

mesmo número de ovos por conjunto nos dois grupos, o que pode não ocorrer em outros experimentos.

(ii) A não homogeneidade entre as matrizes Σ_1 's pode ser problemática, a medida que não exista conjuntos nos quais o número de indivíduos supere o número de ocasiões observadas nesses conjuntos, isto é, quando $n_r \leq p_r$, $\forall r$.

Escrevendo agora as matrizes que definem o modelo em (4.2.2)

$$L_1 = A_1 \otimes P_1', \quad L_2 = A_2 \otimes [P_1 | P_2]', \quad \dots, \quad L_7 = A_7 \otimes P'$$

e a matriz de delineamento $L = [L_1' | \dots | L_7']'$ tem dimensão (378x4)

A matriz de variâncias e covariâncias também fica explícita a medida que Σ_{1r} é calculada para cada r, logo,

$$\Omega = \text{bloc diag} \left[I_{12} \otimes \Sigma_{11}, \dots, I_{18} \otimes \Sigma \right]$$

que, neste caso, tem dimensão (378x378). Cada bloco r corresponde a matriz de variâncias e covariâncias dos indivíduos que foram observados na ocasião $k=r$.

Sendo L uma matriz de posto completo, é possível estimar ξ_{s_1} por (4.2.3) sem necessidade da inversa generalizada. Assim,

$$\hat{\xi}_{s_1} = (62.2937 \quad -0.0942)' \quad \text{e} \quad \hat{\xi}_{s_2} = (62.2325 \quad -0.0438)'$$

são obtidos utilizando $\hat{\Sigma}_1$ e $\hat{\Sigma}_2$ calculadas por (4.3.6).

Para verificar a significância dos parâmetros ξ_{11} e ξ_{21} , foram testadas as seguintes hipóteses:

$H_{01}: \xi_{11}=0$ e $H_{02}: \xi_{21}=0$, cujo resultado obtido do cálculo da estatística (4.2.8) (vide B.2 no Apêndice B) revelou, a um nível de significância de 5%, que H_{01} pode ser rejeitada com fortes evidências contra H_{01} . No entanto, para H_{02} , verifica-se um resultado oposto, isto é, não se rejeita H_{02} e, conseqüentemente, uma constante passando pela média torna-se o ajuste mais apropriado na descrição do perfil médio dos pesos dos ovos armazenados com óleo tipo 2.

As faixas de confiança e os intervalos de confiança para os parâmetros da curva ajustada aos pesos dos ovos armazenados com óleo tipo 1, a um nível de significância de 5%, são:

$$(62.2937 - 0.0942) \pm 0.2(1.7501 + 0.0614t + 0.0039t^2)^{1/2}$$

$$IC_{\xi_{10}}(95\%)=(62.0291 ; 62.5583) \text{ e } IC_{\xi_{11}}(95\%)=(-0.1067 ; -0.0817)$$

Para os ovos armazenados com óleo tipo 2 ajustou-se uma constante. Portanto, a faixa e intervalo de confiança coincidem,

$$IC_{\xi_{20}}(95\%)=(61.9236 ; 62.2363)$$

Nas figuras 5.3.3 e 5.3.4 estão dispostos os perfis médios com as respectivas curvas ajustadas e faixas de confiança, para os dois tipos de armazenamento.

Dessas figuras, nota-se que os ovos armazenados com o óleo tipo 2 variam menos de peso que os armazenados com óleo tipo 1.

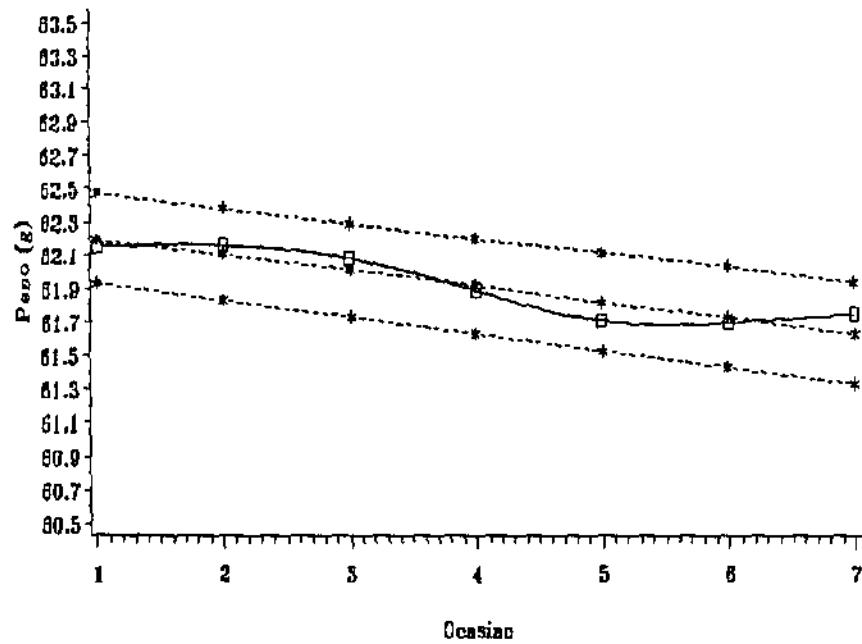


Figura 5.3.3 - Perfil do Peso de Ovos para Oleo Tipo 1
por Ocasiao - Com Dados Incompletos

--:Curva ajustada e faixa de confianca

—:Resposta media

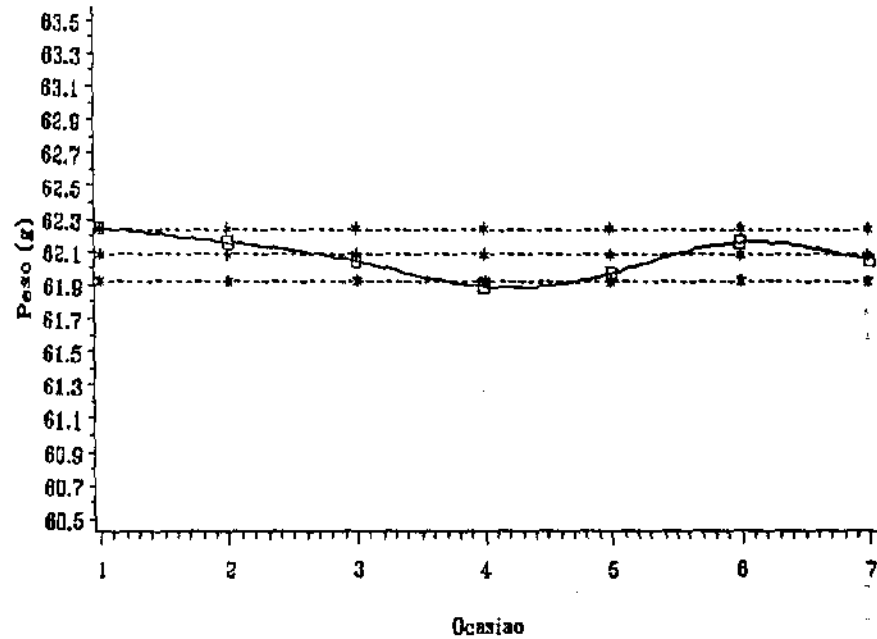


Figura 5.3.4 - Perfil do Peso de Ovos para Oleo Tipo 2
por Ocasiao - Com Dados Incompletos

--:Curva ajustada e faixa de confianca

—:Resposta Media

Não tendo evidências para crer na igualdade de Σ_1 e Σ_2 , há dificuldades em proceder com uma comparação entre os grupos, já que se caracteriza o problema do Behrens-Fisher multivariado. Como tal abordagem não encontra disponível no pacote estatístico utilizado, no presente trabalho não será aplicado de Nel e Merwe(86). Assim, a análise se restringe a uma comparação em função das curvas ajustadas.

Como pôde ser verificado, os polinômios ajustados aos perfis médios possuem graus distintos. Esse fato indica uma diferenciação de comportamento dos pesos dos ovos, entre os dois tipos de armazenamento, ou seja, um efeito significativo de interação grupo*ocasião (ausência de paralelismo). Para os ovos oleados com óleo tipo 2, não se nota influência do tempo, devido ao ajuste de uma constante. O mesmo não ocorre com os oleados com óleo tipo 1, pois, esses tendem a perder peso com o passar do tempo.

5.3.2 Análise Desprezando os Dados Incompletos

Desprezar os dados incompletos nesta situação experimental equivale a eliminar de 252 medições em cada lote de 90 ovos, ou seja, utilizar as observações que constam do conjunto E_7 descrito na seção anterior.

Dispondo agora de 18 ovos por grupo, observados em 7 ocasiões a aplicação do modelo (2.3.1) parece imediata, entretanto, devendo a análise ser feita separadamente para cada grupo, devido a não homogeneidade de Σ_1 e Σ_2 .

Por (2.3.2) as estimativas obtidas para ξ_1 e ξ_2 foram:

$$\hat{\xi}_1 = (62.7053 \quad -0.1340) \quad \text{e} \quad \hat{\xi}_2 = (62.5170 \quad -0.0672)$$

Utilizando os intervalos de confiança para os parâmetros ξ_{11} e ξ_{21} é possível verificar que são significativamente distintos de zero, rejeitando as hipóteses H_{01} e H_{02} descritas na seção anterior.

$$IC_{\xi_{11}}(95\%)=(-0.1465;-0.1215) \text{ e } IC_{\xi_{21}}(95\%)=(-0.0732;-0.0612)$$

Para os coeficientes lineares, os intervalos de confiança obtidos foram:

$$IC_{\xi_{10}}(95\%)=(62.4407;62.9698) \text{ e } IC_{\xi_{20}}(95\%)=(62.4764;62.5575)$$

As faixas de confiança calculadas pela expressão (4.2.10) para as duas retas ajustadas são as que seguem:

$$(62.7053 - 0.1340t) \pm 0.2(1.7501 + 0.0614t + 0.0039t^2)^{1/2}$$

para o óleo tipo 1

$$\text{e } (62.5170 - 0.0672t) \pm 0.2(1.0274 - 0.0248t + 0.0009t^2)^{1/2}$$

para o óleo tipo 2.

O ajuste das curvas pode ser visto graficamente através das figuras 5.3.5 e 5.3.6, nas quais constam as respostas médias observadas para os 18 ovos, as retas estimadas e respectivas faixas de confiança.

A fim de comparar os dois tipos de armazenamento, desprezando os dados incompletos, foi feita uma análise discriminante. O objetivo está em verificar se é possível, através de uma função discriminante, classificar os ovos como pertencentes a um grupo ou outro, levando em conta os seus pesos.

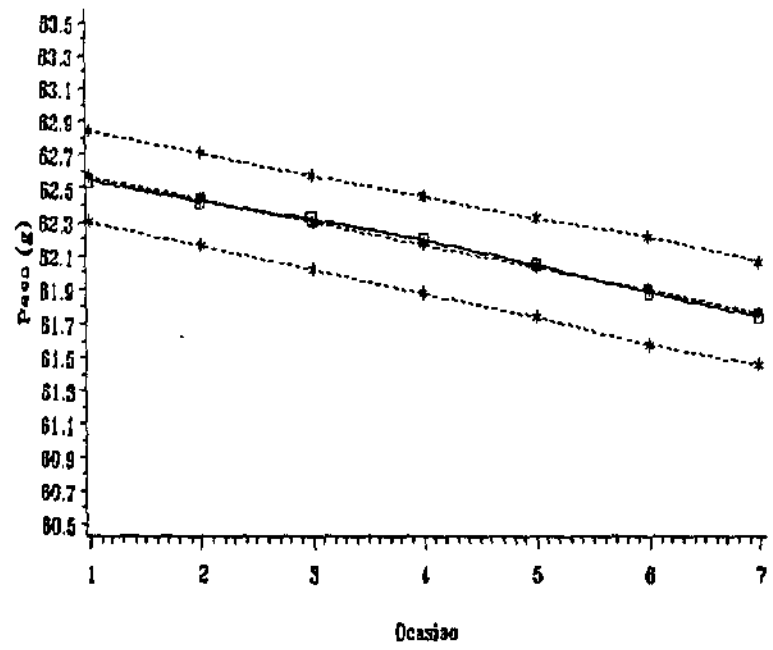


Figura 5.3.5 - Perfil do Peso de Ovos para Oleo Tipo 1 por Ocasiao - Sem Dados Incompletos
 ---:Curva ajustada e faixa de confianca ___:Resposta media

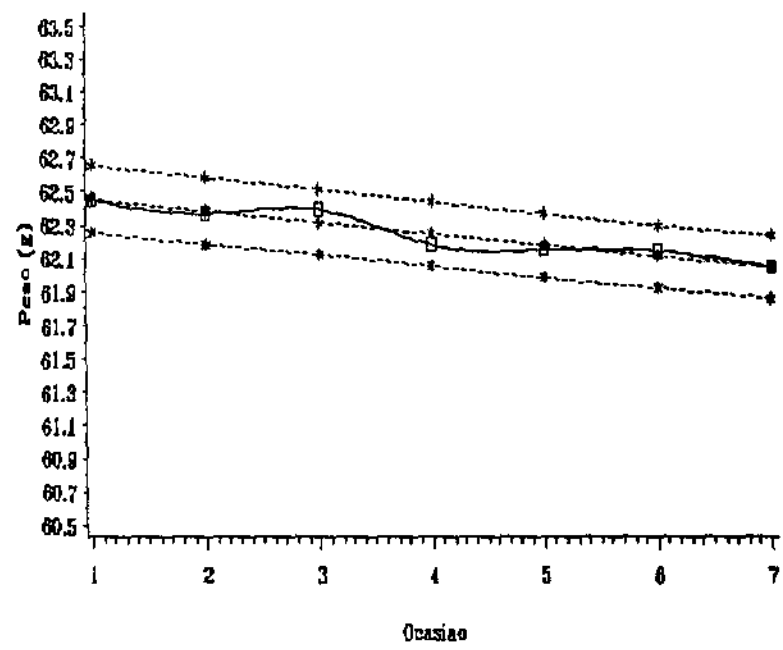


Figura 5.3.6 - Perfil do Peso de Ovos para Oleo Tipo 2 por Ocasiao - Sem Dados Incompletos
 ---:Curva ajustada e faixa de confianca ___:Resposta Media

Observando os resultados apresentados em B.5 no Apêndice B, considerando a não igualdade de Σ_1 e Σ_2 , verifica-se que todas as classificações foram corretas. Esse fato indica uma diferenciação entre as populações. Para esta situação os intervalos de confiança calculados para os coeficientes angulares não se sobrepõem. Isso permite, então, constatar a ausência de paralelismo entre os perfis médios dos pesos dos ovos nos diferentes tipos de armazenamento.

O resultado do ajuste desprezando os dados incompletos pode, em algumas situações, não ser adequado para descrever o comportamento do perfil médio. Nesta situação experimental, nota-se um nítida superestimação através das figuras 5.3.7 e 5.3.8. Nelas encontram-se expostos os perfis das respostas médias observadas com todas as observações disponíveis (peso dos 90 ovos) e as outras ajustadas desprezando os dados incompletos.

5.3.3 Análise alternativa

Pensando em diminuir o esforço de coleta de dados e mantendo o objetivo do experimento em verificar o efeito dos tipos de armazenamento, da ocasião (tempo) e da interação entre esses fatores, o split-plot parece uma alternativa simplificada de análise e, conseqüentemente, de interpretação de resultados.

Tomando 12 ovos a cada ocasião, a estrutura do split-plot é definida, associando às parcelas os tipos de armazenamento e às subparcelas as ocasiões pré-fixadas. O fator tipo de armazenamento possui 2 níveis e o fator ocasião possui 7 níveis, o que permite um total de 12 replicações do experimento (168 ovos).

Assim, através de um modelo misto univariado é possível montar uma tabela de análise de variância, na qual os efeitos dos

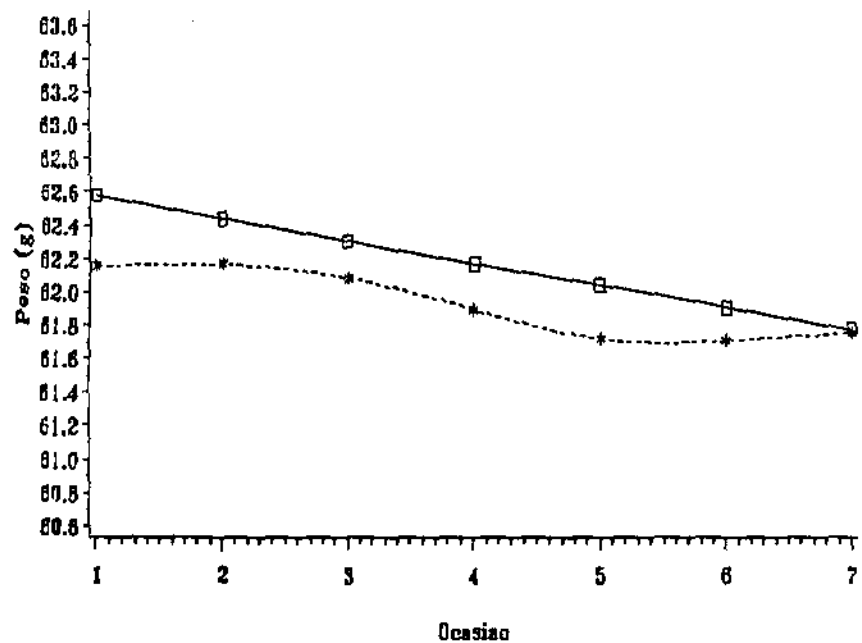


Figura 5.3.7 - Perfil do Peso de Ovos para Oleo Tipo 1 por Ocasiao com Ajuste Desconsiderando Dados Incompletos

---:Curva ajustada ---:Resposta media para dados incompletos

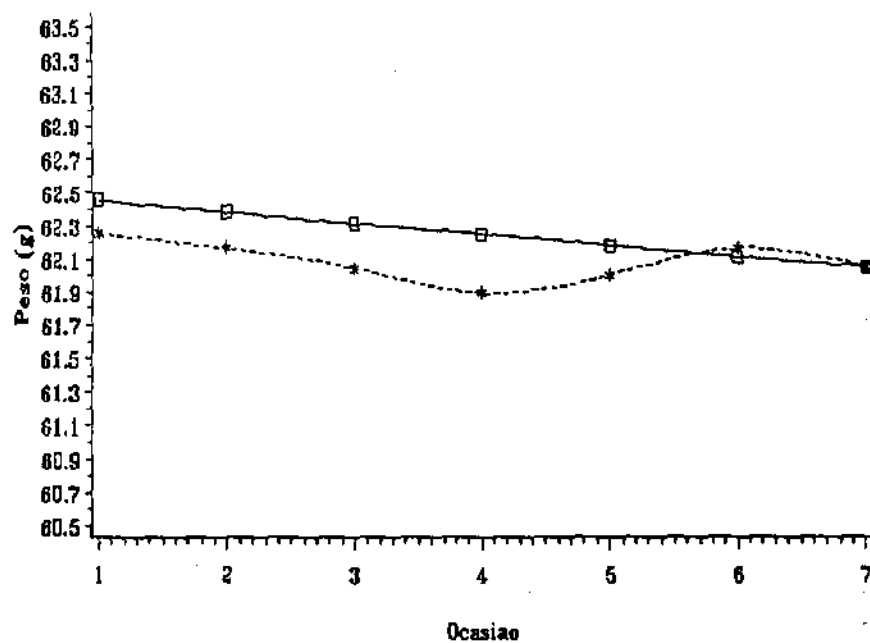


Figura 5.3.8 - Perfil do Peso de Ovos para Oleo Tipo 2 por Ocasiao com Ajuste Desconsiderando Dados Incompletos

---:Curva ajustada ---:Resposta media para dados incompletos

fatores e da interação entre eles sejam testados.

A tabela de análise de variância que se encontra em B.6 no Apêndice B traz os resultados provenientes deste rearranjo dos dados.

Como pode ser observado, não há efeito significativo de interação entre grupo e ocasião (p -value = 0.1854), o que permite concluir que, sob o aspecto do experimento, o comportamento dos perfis médios nos dois tipos de armazenamento apresentam mesmas tendências.

Como o efeito de grupo assim como o de ocasião foram não significativos (p -value = 0.4702 e 0.1479, respectivamente), conclui-se, portanto, que não houve diferença significativa entre os dois tipos de armazenamento, isto é, caracteriza-se a coincidência dos perfis médios.

5.3.4 Observações sobre a Metodologia Aplicada

Nota-se que para a análise do experimento em questão neste capítulo, foram ilustradas três técnicas de análise que poderiam fornecer as informações necessárias para atender às expectativas relacionadas ao experimento. Destas três técnicas, duas não contavam com os dados na sua totalidade, devido a eliminação das observações perdidas e tentando simplificar a interpretação de resultados.

Assim, as divergências dos resultados revelam que, embora mais complexa, a análise feita pela metodologia aplicada na seção 5.3.1 fornece resultados mais satisfatórios que as demais, nesta situação experimental. Embora mais simples sob o aspecto de interpretação e coleta de dados, a análise pelo split-plot seria

equivalente se os efeitos significativos notados para as outras técnicas fossem verificados. No entanto, contando com as observações utilizadas, não se revelou propriamente redundâncias, indicando que não foi desnecessário o esforço de amostragem para atender às expectativas envolvidas na experimentação.

Com relação ao descarte das observações incompletas, ocorre que a subamostra tomada contendo os 18 ovos pesados nas 7 ocasiões, pode não estar representando o verdadeiro comportamento médio dos ovos. Nota-se que essa subamostra parece superestimar os pesos dos ovos e, portanto, os ajustes não se tornam adequados.

5.4 COMENTÁRIOS

Diante dos resultados apresentados e descritos na seção 5.3, pode-se concluir com respeito aos propósitos do experimento que os dois tipos de armazenamento produzem um efeito distinto na preservação do peso dos ovos. Esta diferença de efeito, descrita através do comportamento dos perfis médios, é devida a não influência do tempo na perda de peso dos ovos armazenados com óleo USP-180/190, já que, em média, o peso dos ovos oleados com este tipo de óleo mantêm-se ao longo das 7 ocasiões estudadas. A maior viscosidade do óleo mencionado explica sua melhor efetividade na impermeabilização da casca e, portanto, menor perda de líquido da parte interna dos ovos.

CONSIDERAÇÕES FINAIS

O texto aqui apresentado procurou fornecer resultados provenientes de algumas adaptações e complementações decorrentes do estudo de medidas repetidas. Este estudo permitiu destacar as semelhanças, diferenças e equivalências entre os modelos de Potthoff & Roy(64) e Rao(65,66), principalmente sob o aspecto da escolha das matrizes de ponderação e dos tempos G e P, respectivamente. Daí se observa que, para cada situação, pode ou não haver vantagens na aplicação de um dos métodos, ou melhor, do ajuste de um dos modelos, pois, o que se nota é a existência de mais semelhanças do que diferenças entre eles.

Com relação ao método DS, o que se observa, quando abordado no contexto de dados incompletos, é que sofre influências tanto do número de observações perdidas quanto da estrutura de perda. Essa estrutura, olhadas ocasião e grupo como componentes dela, simultaneamente, é essencial para prever as limitações dos resultados. Para dados incompletos foi desenvolvido, sob a estruturação via modelos, o procedimento para teste de paralelismo. Ficou evidente a partir daí que a extensão para determinação de grau e outros testes é possível com esta abordagem, sendo objeto de trabalhos posteriores.

Neste sentido, a adaptação do método DS para dados incompletos baseou-se, essencialmente, na estruturação proposta por Kleinbaum(73).

A ilustração discutida no último capítulo contribuiu para destacar alguns pontos importantes na questão da falta de observações, ressaltando que o descarte de observações eventualmente pode não enfraquecer o diagnóstico via resultados mais simples em algumas situações, mas pode significar perda em outras.

Parte da proposta de esclarecer, complementar e adaptar resultados sobre as metodologias mencionadas anteriormente foi desenvolvida. No entanto, sem deixar de destacar alguns pontos interessantes para estudos posteriores. Por exemplo, a influência do número e estrutura das observações perdidas, o tratamento de matrizes de variâncias e covariâncias particulares na adaptação do método DS, a determinação de uma estimativa definida positiva da matriz de variâncias e covariâncias, na situação de dados incompletos, a extensão dos resultados da seção 4.3.3 para diferenças de maior ordem e, finalmente, a distribuição da estatística do teste proposto a partir de (4.3.22).

Observações sobre o Modelo Auto-Regressivo de Primeira Ordem

É interessante notar como o padrão de correlação serial é associado à matriz de variâncias e covariâncias em Medidas Repetidas. Assim, supondo que o tempo seja o fator responsável pela estrutura de dependência das respostas e, associando a esta estrutura um modelo estacionário,

$$y_{ijk} = \rho \cdot y_{ij(k-1)} + \gamma_{ijk} \quad , \quad k=1, \dots, p \quad (A.1)$$

$$, \quad |\rho| < 1$$

onde, ρ é o coeficiente auto-regressivo de primeira ordem

γ_{ijk} são variáveis aleatórias independentes com variância constante

y_{ijk} é a variável resposta observada no tempo t_k para o j -ésimo indivíduo do grupo i .

Do modelo (A.1) segue que:

$$V(y_{ijk}) = \rho^2 V(y_{ij(k-1)}) + V(\gamma_{ijk})$$

mas, supondo $\gamma_{ijk} \sim N(0, \sigma^2)$ e como o modelo é estacionário,

$$V(y_{ijk}) = V(y_{ij(k-1)})$$

logo,

$$V(y_{1jk}) = \sigma^2/(1-\rho^2)$$

Analogamente, verifica-se que

$$\text{cov}(y_{1jk}, y_{1j(k-1)}) = \rho\sigma^2/(1-\rho^2)$$

e de modo geral,

$$\text{cov}(y_{1jk}, y_{1j(k-t)}) = \rho^t\sigma^2/(1-\rho^2) \quad , k-t < p$$

de onde segue que

$$\text{corr}(y_{1jk}, y_{1j(k-t)}) = \rho^t$$

Dos resultados colocados acima, a matriz de variâncias e covariâncias de y_{1j} é

$$\Sigma = \frac{\sigma^2}{(1-\rho^2)} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{p-1} \\ & 1 & \rho & \dots & \rho^{p-2} \\ & & & \dots & \\ \text{(slm)} & & & & \\ & & & & 1 \end{pmatrix}$$

B.2 Valores das estatísticas de teste para o MCGG para $\alpha=5\%$ (PROC IML/SAS)

	Grupo 1	Grupo 2
W_n	11.09	2.39

B.3 Matrizes de somas de quadrados do erro (PROC IML/SAS):

$$SQ_E = \begin{pmatrix} 1.7501 & 0.0307 \\ 0.0307 & 0.0039 \end{pmatrix} \text{ para o grupo 1}$$

$$SQ_E = \begin{pmatrix} 1.0274 & -0.0124 \\ -0.0124 & 0.0009 \end{pmatrix} \text{ para o grupo 2}$$

B.4 Matrizes de variâncias e covariâncias dos parâmetros estimados (PROC IML/SAS):

$$V(\hat{\xi}_1) = \begin{pmatrix} 0.1343 & 0.0014 \\ 0.0014 & 0.0004 \end{pmatrix} \text{ e } V(\hat{\xi}_2) = \begin{pmatrix} 0.0954 & -0.00104 \\ -0.00104 & 0.00006 \end{pmatrix}$$

B.5 Análise Discriminante dos dados do conjunto com 18 ovos completamente observados (PROC DISCRIM/SAS)

Discriminant Analysis

36 Observations	35 DF Total
7 Variables	34 DF Within Classes
2 Classes	1 DF Between Classes

Simple Statistics

Total-Sample

Variable	N	Sum	Mean	Variance	Std Dev
T1	36	2250	62.49750	1.97821	1.40649
T2	36	2246	62.39889	2.04043	1.42844
T3	36	2245	62.35750	2.07976	1.44214
T4	36	2239	62.19000	2.20139	1.48371
T5	36	2236	62.10167	2.14455	1.46443
T6	36	2233	62.01750	2.19735	1.48235
T7	36	2228	61.89806	2.23887	1.49629

GRUPO = 1

Variable	N	Sum	Mean	Variance	Std Dev
T1	18	1126	62.54889	2.42340	1.55673
T2	18	1124	62.42889	2.57263	1.60394
T3	18	1122	62.32111	2.67490	1.63551
T4	18	1120	62.19944	2.77449	1.66568
T5	18	1117	62.05000	2.85678	1.69020
T6	18	1114	61.88889	2.97266	1.72414
T7	18	1111	61.74889	3.08646	1.75683

GRUPO = 2					
Variable	N	Sum	Mean	Variance	Std Dev
T1	18	1124	62.44611	1.64380	1.28211
T2	18	1123	62.36889	1.62634	1.27528
T3	18	1123	62.39389	1.60415	1.26655
T4	18	1119	62.18056	1.75759	1.32574
T5	18	1119	62.15333	1.55281	1.24612
T6	18	1119	62.14611	1.51627	1.23137
T7	18	1117	62.04722	1.47587	1.21485

Within Covariance Matrix Information

GRUPO	Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
1	7	-39.38475
2	7	-37.55287
Pooled	7	-33.70992

Test of Homogeneity of Within Covariance Matrices

Test Chi-Square Value = 126.705320
 with 28 DF Prob > Chi-Sq = 0.0001

Since the chi-square value is significant at the 0.1000 level, the within covariance matrices will be used in the discriminant function.

Pairwise Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \text{COV}_j^{-1} (\bar{X}_i - \bar{X}_j)$$

Squared Distance to GRUPO

From GRUPO	1	2
1	0	60.78993
2	232.42803	0

Pairwise Generalized Squared Distances Between Groups

$$D^2(i;j) = (\bar{X}_i - \bar{X}_j)' \text{COV}_j^{-1} (\bar{X}_i - \bar{X}_j) + \ln |\text{COV}_j|$$

Generalized Squared Distance to GRUPO

From GRUPO	1	2
1	-39.38475	23.23706
2	193.04329	-37.55287

Multivariate Statistics and Exact F Statistics

S=1 M=2.5 N=13

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.05048	75.2393	7	28	0.0001
Pillai's Trace	0.94952	75.2393	7	28	0.0001
Hotelling-Lawley Trace	18.8098241	75.2393	7	28	0.0001
Roy's Greatest Root	18.8098241	75.2393	7	28	0.0001

Resubstitution Summary using Quadratic Discriminant Function

Generalized Squared Distance Function:

$$D_j^2(X) = (\bar{X} - \bar{X}_j)' \text{COV}_j^{-1} (\bar{X} - \bar{X}_j) + \ln |\text{COV}_j|$$

Posterior Probability of Membership in each GRUPO:

$$\Pr(j|X) = \exp(-.5 D_j^2(X)) / \sum_k \exp(-.5 D_k^2(X))$$

Number of Observations and Percent Classified into GRUPO:

From GRUPO	1	2	Total
1	18	0	18
	100.00	0.00	100.00
2	0	18	18
	0.00	100.00	100.00
Total	18	18	36
Percent	50.00	50.00	100.00
Priors	0.5000	0.5000	

Error Count Estimates for GRUPO:

	1	2	Total
Rate	0.0000	0.0000	0.0000
Priors	0.5000	0.5000	

B.6 Análise de Variância para análise alternativa - Split-Plot
(PROC GLM/SAS)

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS LEVELS	VALUES
GRUPO 2	1 2
INDIV 12	1 2 3 4 5 6 7 8 9 10 11 12
OCASIAO 7	1 2 3 4 5 6 7

NUMBER OF OBSERVATIONS IN DATA SET = 168

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PESO

SOURCE	DF	TYPE I SS	F VALUE	PR > F
GRUPO	1	0.98900060	0.54	0.4702
ERRO	22	40.29639881		
OCASIAO	6	17.06706429	1.61	0.1479
GRUPO*OCASIAO	6	15.77934524	1.49	0.1854
ERRO	132	236.56947619		

Programas utilizados para obtenção dos resultados numéricos

C.1 Programa para obtenção de resultados envolvidos na análise de dados incompletos aplicando as adaptações da metodologia derivada por Kleinbaum(73).

```

                /* Leitura dos dados */
data conj1; infile arq1; input t1 ;obs=_n_;
data conj2; infile arq2; input t1-t2;obs=_n_;
data conj3; infile arq3; input t1-t3;obs=_n_;
data conj4; infile arq4; input t1-t4;obs=_n_;
data conj5; infile arq5; input t1-t5;obs=_n_;
data conj6; infile arq6; input t1-t6;obs=_n_;
data conj7; infile arq7; input t1-t7;obs=_n_;
PROC IML ;
                /* Matrizes de dados separadas por conjuntos */
use conj1;
read all var{t1} into t1;
Y1= t1;
use conj2;
read all var{t1} into t1;
read all var{t2} into t2;
Y2=t1||t2;
use conj3;

```

```

read all var{t1} into t1;
read all var{t2} into t2;
read all var{t3} into t3;
Y3=t1!!t2!!t3;
use conj4;
read all var{t1} into t1;
read all var{t2} into t2;
read all var{t3} into t3;
read all var{t4} into t4;
Y4=t1!!t2!!t3!!t4;
use conj5;
read all var{t1} into t1;
read all var{t2} into t2;
read all var{t3} into t3;
read all var{t4} into t4;
read all var{t5} into t5;
Y5=t1!!t2!!t3!!t4!!t5;
use conj6;
read all var{t1} into t1;
read all var{t2} into t2;
read all var{t3} into t3;
read all var{t4} into t4;
read all var{t5} into t5;
read all var{t6} into t6;
Y6=t1!!t2!!t3!!t4!!t5!!t6;
use conj7;
read all var{t1} into t1;
read all var{t2} into t2;
read all var{t3} into t3;
read all var{t4} into t4;
read all var{t5} into t5;
read all var{t6} into t6;
read all var{t7} into t7;
Y7=t1!!t2!!t3!!t4!!t5!!t6!!t7;

```

```

/* Matriz dos Tempos para Polinômios de Grau 3 */
P={1 1 1 1 1 1 1, 1 2 3 4 5 6 7};
PP=P(:,1);
/* Matrizes de Selecao dos Tempos Observados */
B1=I(1)//repeat({0},6,1);
B2=I(2)//repeat({0 0},5,1);
B3=I(3)//repeat({0 0 0},4,1);
B4=I(4)//repeat({0 0 0 0},3,1);
B5=I(5)//repeat({0 0 0 0 0},2,1);
B6=I(6)//repeat({0 0 0 0 0 0},1,1);
/* Rotina para calculo da estimativa da matriz de covariancias */
/* utilizando a parte completa dos dados (18 ovos)*/
start grcompl;
k=1:12; k1=1:18;
Y11=Y1(:,k); Y12=Y2(:,k); Y13=Y3(:,k); Y14=Y4(:,k);
Y15=Y5(:,k); Y16=Y6(:,k); Y17=Y7(:,k1);
mcl1=Y11(:,1)+Y12(:,1)+Y13(:,1)+Y14(:,1)+Y15(:,1)+
Y16(:,1)+Y17(:,1); MC11=MC11/90;
mcl2=Y12(:,2)+Y13(:,2)+Y14(:,2)+Y15(:,2)+Y16(:,2)+
Y17(:,2); MC12=MC12/78;
mcl3=Y13(:,3)+Y14(:,3)+Y15(:,3)+Y16(:,3)+Y17(:,3);
MC13=MC13/66;
mcl4=Y14(:,4)+Y15(:,4)+Y16(:,4)+Y17(:,4); MC14=MC14/54;
mcl5=Y15(:,5)+Y16(:,5)+Y17(:,5); MC15=MC15/42;
mcl6=Y16(:,6)+Y17(:,6); MC16=MC16/30;
mcl7=Y17(:,7)/18;
MC1=Y17(:,1)/18;
MC=MC11::MC12::MC13::MC14::MC15::MC16::MC17;
free mcl1 mcl2 mcl3 mcl4 mcl5 mcl6 mcl7;
S=(Y17'*Y17 - 18*mcl'*mcl)/17;
Y=(shape(Y11,1)::shape(Y12,1)::shape(Y13,1)::shape(Y14,1)::
shape(Y15,1)::shape(Y16,1)::shape(Y17,1))';
free Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y1 Y2 Y3 Y4 Y5 Y6 Y7 mcl;
finish;

```

```

start grcomp2;
  k=13:24;  k1=19:36;
  Y21=Y1(:,k,:);  Y22=Y2(:,k,:);  Y23=Y3(:,k,:);  Y24=Y4(:,k,:);
  Y25=Y5(:,k,:);  Y26=Y6(:,k,:);  Y27=Y7(:,k1,:);
  mc21=Y21(:,1:)+Y22(:,1:)+Y23(:,1:)+Y24(:,1:)+Y25(:,1:)+
Y26(:,1:)+Y27(:,1:);  MC21=MC21/90;
mc22=Y22(:,2:)+Y23(:,2:)+Y24(:,2:)+Y25(:,2:)+Y26(:,2:)+
  Y27(:,2:);  MC22=MC22/78;
mc23=Y23(:,3:)+Y24(:,3:)+Y25(:,3:)+Y26(:,3:)+Y27(:,3:);
MC23=MC23/66;
mc24=Y24(:,4:)+Y25(:,4:)+Y26(:,4:)+Y27(:,4:);  MC24=MC24/54;
mc25=Y25(:,5:)+Y26(:,5:)+Y27(:,5:);  MC25=MC25/42;
mc26=Y26(:,6:)+Y27(:,6:);  MC26=MC26/30;
mc27=Y27(:,7:)/18;
MC2=Y27(:,:)/18;
MC=MC21+MC22+MC23+MC24+MC25+MC26+MC27;
free mc21 mc22 mc23 mc24 mc25 mc26 MC27;
S=(Y27'*Y27 - 18*mc2'*mc2)/17;
Y=(shape(Y21,1)+shape(Y22,1)+shape(Y23,1)+shape(Y24,1)+
  shape(Y25,1)+shape(Y26,1)+shape(Y27,1))';
free Y1 Y2 Y3 Y4 Y5 Y6 Y7 mc2;
finish;

start mccg;
P1=P*B1; P2=P*B2; P3=P*B3; P4=P*B4; P5=P*B5; P6=P*B6;
IV1=INV(B1'*B1); IV2=INV(B2'*B2); IV3=INV(B3'*B3); IV4=INV(B4'*B4);
IV5=INV(B5'*B5); IV6=INV(B6'*B6); IV7=I(7);
free B1 B2 B3 B4 B5 B6;
L=REPEAT(P1',12,1)//REPEAT(P2',12,1)//REPEAT(P3',12,1)//
  REPEAT(P4',12,1)//REPEAT(P5',12,1)//REPEAT(P6',12,1)//
  REPEAT(P',18,1);
free P1 P2 P3 P4 P5 P6;
IV=BLOCK(I(12)@IV1, I(12)@IV2, I(12)@IV3, I(12)@IV4,
I(12)@IV5,I(12)@IV6,I(18)@IV7); free IV1 IV2 IV3 IV4 IV5 IV6 IV7 ;
  AUX=L'*IV*L;

```

```

C2=IV*Y;free IV Y;
IAUX=INV(AUX); FREE AUX;
CSI=IAUX*L'*C2; free L C2;
ESTIMAT=CSI*P;
VEST=P'*IAUX*P;

M=(0 1);
Wn=(M*CSI)'*INV(M*IAUX*M')*(M*CSI);
CSI=CSI';
finish;
start faixa;
SQE=INV(P*INV(S)*P');
A=REPEAT(1,18,1);
IA=INV(A'*A);          FREE A;
D1=P(1,1!);D2=P(1,2!);D3=P(1,3!);D4=P(1,4!);D5=P(1,5!);D6=P(1,6!);
D7=P(1,7!);          FREE P;
FX51=0.2*((D1'*SQE*D1)**0.5);
FXI51=CSI*D1-FX51; FXS51=CSI*D1+FX51;
FX52=0.2*((D2'*SQE*D2)**0.5);
FXI52=CSI*D2-FX52; FXS52=CSI*D2+FX52;
FX53=0.2*((D3'*SQE*D3)**0.5);
FXI53=CSI*D3-FX53; FXS53=CSI*D3+FX53;
FX54=0.2*((D4'*SQE*D4)**0.5);
FXI54=CSI*D4-FX54; FXS54=CSI*D4+FX54;
FX55=0.2*((D5'*SQE*D5)**0.5);
FXI55=CSI*D5-FX55; FXS55=CSI*D5+FX55;
FX56=0.2*((D6'*SQE*D6)**0.5);
FXI56=CSI*D6-FX56; FXS56=CSI*D6+FX56;
FX57=0.2*((D7'*SQE*D7)**0.5);
FXI57=CSI*D7-FX57; FXS57=CSI*D7+FX57;
LINF5=FXI51!!FXI52!!FXI53!!FXI54!!FXI55!!FXI56!!FXI57;
LSUP5=FXS51!!FXS52!!FXS53!!FXS54!!FXS55!!FXS56!!FXS57;
finish;

```

```

start saidal;
run grcomp1; run mccg; run faixa;
print "****          ANALISE DE DADOS INCOMPLETOS          ****";
print "*****          RESULTADOS DO GRUPO 1          *****";
print "Estimativa da matriz de covariâncias ", S;
print "Estimativa dos parametros ",csi;
print "Estimativa da matriz de covariâncias dos parametros ",Iaux;
print "Valor da estatística de teste",Wn;
print "Respostas médias observadas",mc;
print "Estimativa das respostas médias",estimat;
print "Soma de quadrados do ERRO",SQE;
print "Faixa de confiança para alfa=5%",linf5,lsup5 ;
finish;
start saida2;
run grcomp2; run mccg; F=PP; run faixa;
print "****          ANALISE DE DADOS INCOMPLETOS          ****";
print "*****          RESULTADOS DO GRUPO 2          *****";
print "Estimativa da matriz de variâncias e covariâncias ",S;
print "Estimativa dos parametros ",csi;
print "Estimativa da matriz de covariâncias dos parametros ",Iaux;
print "Valor da estatística de teste",Wn;
print "Respostas médias observadas",mc;
print "Estimativa das respostas médias",estimat;
print "Soma de quadrados do erro",SQE;
print "Faixa de confiança para alfa=5%",linf5,lsup5;
finish;
run saidal; run saida2;

```

C.2 Programa para obtenção de resultados, aplicando a metodologia desenvolvida por Potthoff e Roy(64), para a subamostra completamente observada - 18 ovos de cada grupo.

```

/* Leitura dos dados */
data conj7; infile arq7; input t1-t7;obs=_n_;
PROC IML ;
use conj7;
read all var{t1} into t1;
read all var{t2} into t2;
read all var{t3} into t3;
read all var{t4} into t4;
read all var{t5} into t5;
read all var{t6} into t6;
read all var{t7} into t7;
Y7=t1||t2||t3||t4||t5||t6||t7;
/* Matriz de definicao dos individuos dentro do grupo */
A=repeat(1,18,1);
/* Matriz dos Tempos para Polinomios de Grau 2 */
P={1 1 1 1 1 1 1 , 1 2 3 4 5 6 7 };
/* Rotina para calculo da estimativa da matriz de covariancias */
start grcomp1;
k1=1:18;
Yo=Y7(:,k1,:);
mc=Yo(:,:)/18;
S=(Yo'*Yo - 18*mc'*mc)/17; free Y7;
finish;
start grcomp2;
k1=19:36;
Yo=Y7(:,k1,:);
mc=Yo(:,:)/18;
S=(Yo'*Yo - 18*mc'*mc)/17;free Y7;

```



```

finish;
start mcc;
      /* Estimativa dos parametros */
AA=A'*A;
IA=INV(AA);   free AA;
IAA=IA*A';AUX=INV(P*P');AUX=P'*AUX;Y=Yo*AUX;
CSI=IAA*Y;   free Y;
      /* Estimativa da matriz de covariancias dos parametros */
C2=AUX'*S;C2=C2*AUX;VAR=IA@C2;FREE C2;
/* Estimativas das respostas/ Curva ajustada e de suas variancias */
ESTIMAT=CSI*P;VEST=P'*VAR*P;
finish;
      /* Rotina para calculo da faixa de confianca */
start faixa;
D1=P(:,1);D2=P(:,2);D3=P(:,3);D4=P(:,4);D5=P(:,5);
D6=P(:,6);D7=P(:,7);
SQE=inv(P*INV(S)*P');
FX51=0.2*((D1'*SQE*D1)**0.5);
FXI51=(CSI*D1)-FX51;   FXS51=(CSI*D1)+FX51;
FX52=0.2*((D2'*SQE*D2)**0.5);
FXI52=(CSI*D2)-FX52;   FXS52=(CSI*D2)+FX52;
FX53=0.2*((D3'*SQE*D3)**0.5);
FXI53=(CSI*D3)-FX53;   FXS53=(CSI*D3)+FX53;
FX54=0.2*((D4'*SQE*D4)**0.5);
FXI54=(CSI*D4)-FX54;   FXS54=(CSI*D4)+FX54;
FX55=0.2*((D5'*SQE*D5)**0.5);
FXI55=(CSI*D5)-FX55;   FXS55=(CSI*D5)+FX55;
FX56=0.2*((D6'*SQE*D6)**0.5);
FXI56=(CSI*D6)-FX56;   FXS56=(CSI*D6)+FX56;
FX57=0.2*((D7'*SQE*D7)**0.5);
FXI57=(CSI*D7)-FX57;   FXS57=(CSI*D7)+FX57;
LINF5=FXI51::FXI52::FXI53::FXI54::FXI55::FXI56::FXI57;
LSUP5=FXS51::FXS52::FXS53::FXS54::FXS55::FXS56::FXS57;
finish;

```

```
start saicompl;
RUN GRCOMP1;RUN MCC;RUN FAIXA;
print "***          ANALISE DE DADOS COMPLETOS -          ***";
print "*****          RESULTADOS DO GRUPO 1          *****";
print "Estimativa dos parametros ",csi;
print "Estimativa da Matriz de covariâncias dos parametros ",VAR;
print "Respostas Medias Observadas",mc;
print "Estimativa das respostas medias",estimat;
print "Estimativa da matriz de covariâncias das respostas
      estimadas",vest;
print "Soma de quadrados do ERRO",SQE;
print "Faixa de confiança para alfa=5% ",LINF5, LSUP5;
finish;
start saicomp2;
RUN GRCOMP2;RUN MCC;RUN FAIXA;
print "***          ANALISE DE DADOS COMPLETOS          ***";
print "*****          RESULTADOS DO GRUPO 2          *****";
print "Estimativa dos parametros ",csi;
print "Estimativa da Matriz de covariâncias dos parametros ",VAR;
print "Respostas Medias Observadas",mc;
print "Estimativa das respostas medias",estimat;
print "Estimativa da matriz de covariâncias das respostas
      estimadas",vest;
print "Soma de quadrados do ERRO",SQE;
print "Faixa de confiança para alfa=5% ",LINF5, LSUP5;
finish;
RUN SAICOMPI; RUN SAICOMP2;
```

Dados utilizados na ilustração

G: Grupo Oci: Ocasão i na qual o peso do ovo foi observado

G	OC1	OC2	OC3	OC4	OC5	OC6	OC7
1	62.62						
1	61.97						
1	61.78						
1	63.76						
1	60.45						
1	62.47						
1	60.74						
1	60.58						
1	61.08						
1	61.08						
1	60.19						
1	61.13						
2	63.36						
2	61.68						
2	62.46						
2	63.28						
2	63.15						
2	63.16						
2	61.83						
2	61.28						
2	62.00						
2	62.28						
2	61.36						
2	61.54						

1 61.85 61.77
1 62.25 62.15
1 60.06 60.04
1 64.38 64.36
1 61.16 61.11
1 62.34 62.30
1 62.14 62.03
1 64.51 64.48
1 62.53 62.48
1 60.58 60.54
1 61.09 61.01
1 62.68 62.68
2 63.27 63.36
2 63.23 63.28
2 62.00 62.10
2 62.00 62.10
2 61.82 61.93
2 63.41 63.46
2 62.74 62.58
2 63.35 63.22
2 62.02 61.93
2 64.87 64.79
2 61.95 61.88
2 64.08 63.85
1 62.11 61.99 61.94
1 62.53 62.44 62.41
1 63.17 63.16 63.21
1 63.53 63.53 63.63
1 63.92 63.81 63.84
1 61.66 61.49 61.38
1 63.13 63.11 63.03
1 61.79 61.75 61.67
1 63.14 62.94 62.64
1 63.63 63.60 63.53

1 62.77 62.69 62.59
 1 60.71 60.67 60.60
 2 64.45 64.27 64.28
 2 62.21 62.11 62.20
 2 60.59 60.48 60.60
 2 60.33 60.23 60.34
 2 62.55 62.40 62.53
 2 61.08 60.93 61.06
 2 60.36 60.17 60.18
 2 60.49 60.13 60.00
 2 62.80 62.47 62.29
 2 63.80 63.71 63.69
 2 63.34 63.28 63.34
 2 62.70 62.51 62.40
 1 62.12 62.08 62.02 62.08
 1 61.73 61.62 61.53 61.54
 1 63.44 63.41 63.37 63.38
 1 60.66 60.49 60.33 60.23
 1 62.44 62.41 62.36 62.42
 1 61.51 61.38 61.09 60.98
 1 63.12 63.09 63.04 62.94
 1 61.90 61.86 61.79 61.69
 1 63.89 63.86 63.79 63.65
 1 63.34 63.27 63.17 63.04
 1 63.35 63.25 63.13 62.96
 1 59.99 59.94 60.00 59.96
 2 61.82 61.76 61.76 61.70
 2 60.67 60.60 60.66 60.57
 2 60.29 60.21 60.22 60.23
 2 60.14 60.03 60.08 60.04
 2 64.19 64.03 64.11 64.06
 2 60.58 60.55 60.47 60.35
 2 60.13 60.04 60.06 59.91
 2 66.19 66.04 66.08 65.83

2 60.92 60.86 60.85 60.71
 2 63.21 63.03 63.00 62.81
 2 60.19 60.09 60.15 59.99
 2 60.22 60.06 60.08 59.89
 1 62.77 62.72 62.67 62.54 62.53
 1 61.79 61.79 61.76 61.68 61.68
 1 61.54 61.48 61.35 61.28 60.96
 1 60.96 60.95 60.90 60.81 60.78
 1 62.51 62.47 62.42 62.34 62.29
 1 62.29 62.27 62.19 62.07 62.03
 1 60.58 60.51 60.46 60.34 60.26
 1 61.28 60.79 60.30 59.78 59.24
 1 59.91 59.87 59.82 59.72 59.66
 1 63.08 63.06 62.99 62.88 62.76
 1 62.90 61.71 62.57 62.38 62.19
 1 62.91 62.88 62.79 62.68 62.44
 2 64.06 63.92 63.93 63.72 63.60
 2 60.20 60.08 60.15 60.03 60.07
 2 60.24 60.13 60.22 60.13 60.08
 2 60.80 60.79 60.77 60.64 60.57
 2 63.09 62.87 62.78 62.53 62.41
 2 62.34 62.20 62.20 62.00 61.89
 2 61.15 61.08 61.11 60.90 60.76
 2 63.53 63.44 63.45 63.27 63.16
 2 61.66 61.56 61.56 61.37 61.28
 2 59.99 59.99 60.00 59.83 59.77
 2 63.60 63.48 63.46 63.27 63.15
 2 61.41 61.31 61.27 61.00 60.67
 1 62.78 62.69 62.63 62.52 62.40 62.28
 1 61.79 61.74 61.70 61.61 61.54 61.48
 1 59.79 59.72 59.67 59.37 59.47 59.36
 1 60.14 60.10 60.07 60.02 59.95 59.90
 1 60.14 60.09 60.03 59.97 59.88 59.83
 1 64.13 64.08 64.03 63.97 63.89 63.80

1 61.88 61.79 61.74 61.64 61.56 61.46
 1 61.47 61.37 61.29 61.18 61.06 60.93
 1 64.43 64.29 64.21 64.12 64.03 63.94
 1 61.02 60.98 60.94 60.89 60.85 60.77
 1 63.39 63.28 63.21 63.15 63.06 62.98
 1 61.22 61.08 60.23 60.75 60.53 60.34
 2 63.57 63.48 63.50 63.38 63.23 63.26
 2 62.27 62.18 62.24 62.13 62.02 62.12
 2 61.60 61.53 61.59 61.49 61.42 61.35
 2 63.22 63.07 62.96 62.59 62.17 61.92
 2 62.94 62.87 62.85 62.68 62.51 62.38
 2 60.74 60.62 60.60 60.41 60.26 60.20
 2 64.93 64.87 64.88 64.65 64.53 64.50
 2 63.17 63.08 63.14 62.97 62.87 62.90
 2 62.38 62.27 62.33 62.19 62.05 62.09
 2 62.70 62.67 62.66 62.51 62.39 62.40
 2 61.89 61.71 61.78 61.66 61.56 61.57
 2 62.15 62.03 61.99 61.83 61.72 61.66
 1 60.41 60.31 60.26 60.16 60.06 59.93 59.82
 1 63.77 63.68 63.54 63.44 63.35 63.24 63.14
 1 64.96 64.84 64.76 64.68 64.58 64.48 64.38
 1 63.08 62.95 62.80 62.63 62.40 62.21 62.02
 1 61.81 61.69 61.59 61.51 61.40 61.30 61.20
 1 62.97 62.79 62.66 62.49 62.29 62.09 61.88
 1 62.61 62.46 62.39 62.29 62.19 62.06 61.97
 1 62.29 62.18 62.04 61.84 61.62 61.37 61.14
 1 64.96 64.85 64.75 64.65 64.49 64.34 64.21
 1 64.20 64.07 63.97 63.84 63.72 63.57 63.44
 1 60.45 60.37 60.32 60.28 60.24 60.16 60.10
 1 60.20 59.64 59.16 58.74 58.31 57.89 57.52
 1 61.81 61.66 61.60 61.48 61.34 61.17 61.03
 1 60.82 60.73 60.69 60.64 60.56 60.46 60.38
 1 64.75 64.60 64.58 64.52 64.36 64.23 64.16
 1 61.71 61.61 61.54 61.43 61.31 61.15 61.02

1 63.13 63.46 63.34 63.19 63.02 62.81 62.61
1 61.95 61.83 61.79 61.78 61.66 61.54 61.46
2 60.36 60.30 60.38 60.25 60.20 60.21 60.17
2 61.40 61.33 61.38 60.26 61.20 61.21 61.14
2 61.60 61.56 61.54 61.39 61.33 61.35 61.22
2 63.30 63.30 63.38 63.25 63.22 63.25 63.15
2 62.49 62.34 62.33 62.09 62.00 61.94 61.79
2 62.78 62.70 62.80 62.65 62.59 62.62 62.53
2 62.81 62.75 62.84 62.66 62.59 62.60 62.59
2 63.81 63.79 63.82 63.67 63.65 63.67 63.62
2 61.22 61.12 61.12 60.89 60.80 60.78 60.70
2 61.35 61.30 61.34 61.19 61.15 61.17 61.10
2 64.74 64.62 64.64 64.49 64.40 64.36 64.26
2 64.43 64.30 64.27 64.07 63.91 63.84 63.64
2 62.22 62.15 62.22 62.03 61.94 61.95 61.86
2 61.79 61.73 61.76 61.62 61.53 61.50 61.45
2 61.12 61.00 60.95 60.85 60.73 60.72 60.58
2 64.35 64.23 64.16 63.99 63.80 63.71 63.52
2 61.43 61.35 61.39 61.25 61.19 61.23 61.15
2 62.83 62.77 62.77 62.65 62.53 62.52 62.38

REFERÊNCIAS BIBLIOGRÁFICAS

Beauchamp, J. J. and Hoel, D. G. (1973) - Some Investigation and Simulation Studies of Canonical Analysis - *J. Stat. Comp. Simul.*, 2, 197-209.

Box, G. E. P.(1950) - Problems in the Analysis of Growth and Wear Curves - *Biometrics*, 6, 362-389.

Box, G. E. P.(1954) - Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, II. Effects of Inequality of Variance and Correlation Between Errors in the Two-Way Classification - *Ann. Math. Stat.*, 25, 484-498.

Conaway, M. R.(1989) - Analysis of Repeated Categorical Measurements with Conditional Likelihood Methods - *JASA*, 84, 405, 53-62.

Dempster, A. P., Laird, N. M. and Rubin, D. B.(1977) - Maximum Likelihood from Incomplete Data Via the EM Algorithm - *JRSS B*, 39, 29-38.

Diggle, P. J.(1988) - An Approach to the Analysis of Repeated Measurements - *Biometrics*, 44, 959-971.

Draper, N. and Smith, H.(1981) - *Applied Regression Analysis*, New York, Wiley.

Elston, R. C. and Grizzle, J. E.(1962) - Estimation of Time Response Curves and Their Confidence Bands - *Biometrics*, 18, 148-159.

Geisser, S. and Greenhouse, S. W.(1958) - An Extension of Box's Results on the Use of the F Distribution in Multivariate Analysis - *Ann. Math. Stat.*, 29, 885-891.

Gill, J. L.(1988) - Repeated Measurements: Split Plot Trend Analysis Versus Analysis of First Differences - *Biometrics*, 44, 289-297.

Grizzle, J. E., Allen, D. M.(1969) - Analysis of Growth and Dose Response Curves - *Biometrics*, 25, 375-381.

Hartley, H. O. and Hocking, R. R.(1971) - The Analysis of Incomplete Data - *Biometrics*, 27, 783-823.

Hearne, E. M., Clark, G. M. and Hatch, J. P.(1983) - A Test for Serial Correlation in Univariate Repeated Measures Analysis - *Biometrics*, 39, 237-243.

Hoel, P. G.(1964) - Methods for Comparing Growth Type Curves - *Biometrics*, 20, 859-872.

Hills, M.(1968) - A Note on the Analysis of Growth Curves - *Biometrics*, 192-196.

Huynh, H. and Feldt, L. S.(1970) - Conditions under which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distribution - *JASA*, 65, 332, 1582-1589.

_____ (1976) - Estimations of the Box Corrections for Degrees of Freedom from Sample Data in the

Randomized Block and Split-Plot Designs - *J. Educational Statistics*, 1, 69-82.

Kabe, D. G.(1975) - Some Results for the GMANOVA Model - *Comm. Stat.*, 4(9), 813-820.

Kenward, M. G.(1985) - The Use of Fitted Higher - Order Polynomial Coefficients as Covariates in the Analysis of Growth Curves - *Biometrics*, 41, 19-28.

_____ (1987) - A Method for Comparing Profiles of Repeated Measurements - *Appl. Stat.*, 36, 3, 296-308.

Kleinbaum, D. G.(1973) - A Generalization of the Growth Curve Model which Allows Missing Data - *J.Mult. Anal.*, 3, 117-124.

Khatri, C. G.(1966) - An note on a MANOVA model applied to problems in growth curve - *Ann. Inst. Statist. Math.*, 18, 75-88.

Koopmans, T.(1942) - Serial Correlation and Quadratic Forms in Normal Variables - *Ann. Math. Stat.*, 13, 14-33.

Lee, Y. K.(1974) - A Note on Rao's Reduction of Potthoff and Roy's Generalized Linear Model - *Biometrika*, 61, 2, 349-351.

_____ (1988) - Prediction and Estimation of Growth Curves with Special Covariance Structures - *JASA*, 83, 402, 432-440.

Leech, F. B. and Healy, M. J. R.(1959) - The Analysis of Experiments on Growth Rate - *Biometrics*, 98-106.

Leeper, J. D. and Woolson, R. F.(1982) - Testing Hypothesis for the Growth Curve Model when the Data are Incomplete - *J. Stat. Comp. Simul.*, 15, 97-107.

_____, Huseby, J. R. and Allen, D. M.(1980) -
Computation of the Estimated Parameters and Wald Statistic for the
Generalized Growth Curve Model - *Comm. Stat.*, B9(6), 675-693.

_____ and Allen, D. M.(1979) - Smoothing an Indefinite
Variance-Covariance Matrix - *J.Stat. Comp. Simul.*, 9 , 183-194.

_____, Flynn, W., Stein, S. and Schernk, K. L.(1985) - A
Monte Carlo Study of Alternative Procedures for Testing the
Hypothesis of Parallelism for Complete and Incomplete Growth Curve
Data - *J. Stat. Comp. Simul.*, 21, 1-37.

_____ and Heilbrun, L. K.(1986) - A Successive Differences
Method for Growth Curves with Missing Data and Random Observation
Times - *JASA*, 81, 396, 912-916.

Srivastava, M. S.(1985) - Multivariate Data with Missing
Observations - *Comm. Stat.* , 14(4), 775-792.

Stanek III, E. J. and Diehl, S. R.(1988) - Growth Curve Models of
Repeated Binary Response - *Biometrics*, 44, 973-983.

Tubbs, J. D., Lewis, T. O. and Duran, B. S.(1975) - A note on the
analysis of the MANOVA model and its applications to growth curves
- *Comm. Stat.*, 4(7), 643-653.

Vonesh, E. F. and Schork, M. A.(1986) - Sample Sizes in the
Multivariate Analysis of Repeat Measurements - *Biometrics*, 42,
601-610.

Verbyla, A. P.(1986) - Conditioning in the Growth Curve Model -
Biometrika, 73, 2, 475-483.

Wald, A.(1943) - Tests of Statistical Hypothesis Concerning Several Parameters when the Number of Observations is Large - *Trans. Amer. Math. Soc.*, 54, 426-482.

Wallestein, S. and Fleiss, J. L.(1979) - Repeated Measurements Analysis of Variance when the Correlations Have a Certain Pattern - *Psychometrika*, 44(2), 229-233.

Ware, J. H. (1985) - Linear Models for the Analysis of Longitudinal Studies - *JASA*, 39, 2, 95-101.

Wilks, S. S.(1946) - *Ann. Math. Statist.*, 17, 257.

Wishart, J.(1938) - Growth-Rate Determinations in Nutrition Studies with Bacon Pig and their Analysis - *Biometrika*, 30, 16-28.

Woolson, R. F., Leeper, J. D. and Clarke, W. R.(1978) - Analysis of Incomplete Data from Longitudinal and Mixed Longitudinal Studies - *JRSS A*, 141, part 2, 242-252.

_____ and Leeper, J. D.(1980) - Growth Curve Analysis of Complete and Incomplete Longitudinal Data - *Comm. Stat.*, A9(4), 1491-1513.

Zerbe, G. O.(1979) - Randomization Analysis of the Completely Randomized Design Extended to Growth and Response Curves - *JASA*, 74, p.365.

_____ and Walker, S. H.(1977) - A Randomization Test for Comparison of Groups of Growth Curves with Different Polynomial Design Matrices - *Biometrics*, 33, 653-657.