UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CÂMPUS DE BOTUCATU

PROGRAMA COMPUTACIONAL PARA AJUSTE DE CURVAS POLINOMIAIS EM EXPERIMENTOS ENVOLVENDO DADOS LONGITUDINAIS

MARIE OSHIIWA

Tese apresentada à Faculdade de Ciências Agronômicas da UNESP – Câmpus de Botucatu, para obtenção do título de Doutor em Agronomia (Energia na Agricultura).

BOTUCATU – SP
Fevereiro - 2005

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CÂMPUS DE BOTUCATU

PROGRAMA COMPUTACIONAL PARA AJUSTE DE CURVAS POLINOMIAIS EM EXPERIMENTOS ENVOLVENDO DADOS LONGITUDINAIS

MARIE OSHIIWA

Orientador: Prof. Dr. Carlos Roberto Padovani

Tese apresentada à Faculdade de Ciências Agronômicas da UNESP – Campus de Botucatu, para obtenção do título de Doutor em Agronomia (Energia na Agricultura).

BOTUCATU – SP
Fevereiro – 2005

OFERECIMENTO

Aos meus pais que, mais uma vez, me ajudaram a superar os meus limites.
“Não tenho palavras para traduzir o meu amor por vocês.
Obrigada pela vida, pelo exemplo e por estarem sempre
ao meu lado em todos momentos de minha vida.
Amo vocês”.

Aos meus irmãos de sangue e de coração, pela paciência, pelo incentivo e pelo carinho.
"Quem tem um amigo, mesmo que um só, não importa onde se encontre, jamais sofrerá de solidão; poderá morrer de saudades, mas não estará só."
Amir Klink

Ao Prof. Dr. Carlos Roberto Padovani, minha eterna gratidão pela orientação do trabalho, pela amizade e pelo exemplo de pessoa.
“*Feliz aquele que transfere o que sabe*
e aprende o que ensina”
(Cora Coralina)

Ao Luiz Gustavo Minardi, pelo orgulho de desenvolver mais um projeto juntos, pela competência e paciência.
Diz o ditado: “*O maior orgulho do mestre*
é quando o seu aprendiz o supera”

AGRADECIMENTOS

Ao Prof. Dr. Carlos Roberto Padovani pela orientação, paciência e exemplo profissional.

Ao *Webmaster* e professor Luiz Gustavo Minardi da Universidade de Marília pelo desenvolvimento do programa computacional, pela competência e paciência.

A Universidade Metodista de Piracicaba pelo apoio na confecção da tese.

Aos coordenadores e diretores da UNIMAR e da UNIMEP pelo apoio e estímulo ao processo de doutoramento.

Ao mestrando Fabiano Alan Martins pela colaboração nos coeficientes dos polinômios ortogonais.

Ao Dr. Luiz de Souza pela contribuição na metodologia do trabalho.

Ao Dr. Julio da Motta Singer, pela sugestão bibliográfica e esclarecimentos metodológicos.

Ao Dr. Guilherme J. M. Rosa pela contribuição nas referências bibliográficas.

Ao Dr. Luciano Soares de Souza pelos dados fornecidos para exemplos.

Ao Dr. Flávio Ferrari Aragon pela colaboração com o Maple.

Ao engenheiro agrônomo Isao Imaizumi pela colaboração do exemplo ilustrativo do trabalho.

Ao *webdesigner* da Universidade de Marília José Roberto Aleixo Júnior pela logomarca do programa computacional.

À professora Maria Cristina Maldonato pela revisão gramatical.

À bibliotecária Viviane Sousa pela revisão nas referências bibliográficas.

Aos funcionários do Departamento de Bioestatística, UNESP/Botucatu, pela atenção.

Aos funcionários da Biblioteca e da Seção de Pós-Graduação sempre atenciosos e prestativos.

À senhora Silvia Padovani pelo carinho que sempre me recebeu em sua casa.

E a todas as pessoas que, de forma direta ou indireta, colaboraram com o desenvolvimento deste trabalho.

SUMÁRIO

		Página
LISTA DE TABELAS		VI
LISTA DE QUADROS		VII
LISTA DE FIGURAS		VIII
LISTA DE APÊNDICES		IX
1	INTRODUÇÃO	01
2	SUMMARY	02
3	INTRODUÇÃO	03
4	REVISÃO DE LITERATURA	05
	4.1 Dados longitudinais	05
	4.2 Curvas de Crescimento	07
5	MATERIAL E MÉTODOS	16
	5.1 Dados longitudinais	16
	5.2 Análise de dados longitudinais	18
	5.3 Análise de Variância Multivariada (MANOVA)	20
	5.4 Modelo Linear Multivariado de Crescimento (MLMC)	26
	5.4.1 Proposta de Potthoff & Roy	30
	5.4.2 Proposta de ajuste por covariância	31
	5.4.3 Proposta de Rao	35
	5.5 Desenvolvimento proposto	40
	5.5.1 Modelo de curvas de crescimento polinomial	40
	5.5.2 Testes estatísticos	43
	5.5.2.1 Teste para o ajuste do MLMC	43
	5.5.2.2 Teste de coincidências das curvas	45
	5.5.2.3 Medida de Adequacidade do Modelo	46
	5.5.2.4 Intervalos de Confiança para contrastes de coeficientes polinomiais de grupos	47
	5.6 Programa Computacional CRESYS	47
	5.6.1 Descrição do programa	48
	5.6.2 Entrada de dados	49
	5.7 Exemplo ilustrativo	49
6	Resultados e Discussão	50
7	Conclusão	54
8	Referências Bibliográficas	55

LISTA DE TABELAS

		Página
1	Resultados do procedimento estatístico de “descarte” do polinômio (grau mais alto)	45
2	Resultados do Teste de Coincidências das Curvas Polinomiais.....	46
3	Médias e desvios-padrão da porcentagem de ganho de peso dos frutos de Araribá dos grupos escarificados com ácido sulfúrico por 10 minutos (1) e escarificados com ácido sulfúrico por 15 minutos (2), em quatro momentos.....	50
4	Estimativas de $\hat{\xi}$ e dos desvios-padrão de $Var\hat{(}\hat{\xi})$ do polinômio das curvas de crescimento que descrevem o comportamento médio da porcentagem de ganho de peso dos frutos de Araribá segundo o grupo.....	52

LISTA DE QUADROS

		Página
1	Estrutura ilustrativa de um experimento casualizado em três tratamentos com 20 plantas de alho, avaliados em quatro semanas consecutivas.....	17
2	Estrutura genérica de valores observados em um experimento com planejamento longitudinal.....	19
3	Resultados do Procedimento de descarte do grau do polinômio do ganho de peso dos frutos de Araribá.....	51

LISTA DE FIGURAS

		Página
1	Curva de crescimento médio da altura das plantas alho do tipo Caçador, dos grupos Testemunha e Tratamento, em centímetros, aos 20, 30, 40 e 50 dias após a plantação.....	51

LISTA DE APÊNDICES

		Página
1	Altura, em centímetros, dos grupos Testemunha e Tratamento de plantas de alho do tipo Caçador, avaliados em quatro momentos.....	61
2	Manual do usuário do CRESYS.....	62

1. RESUMO

O presente trabalho discutiu aspectos teóricos e práticos do comportamento da variável resposta nos diferentes grupos e condições de avaliação utilizando o Ajuste de Curvas de Crescimento, procedimento multivariado de análise de dados experimentais que possibilita fazer previsões sobre o comportamento médio da resposta para situações diferentes daquelas para as quais o estudo foi planejado, além de propiciar análise comparativa das curvas dos grupos de interesse. Considerando a dificuldade existente quanto a programas computacionais acessíveis a pesquisadores das áreas agronômicas, biológicas e da saúde, e a falta de entendimento da complexidade da estrutura de análise dos dados longitudinais, elaborou-se programa computacional em linguagem que permita ao usuário facilidade de manuseio, e torná-lo disponível, para pesquisadores das áreas aplicadas e, finalmente, discutir as vantagens do procedimento multivariado na preservação da estrutura de dependência dos dados em relação aos procedimentos convencionais utilizados na experimentação agronômica.

COMPUTER PROGRAM FOR POLINOMIAL GROWTH CURVE ADJUST IN EXPERIMENTAL WORK INVOLVING LONGITUDINAL DATA. Botucatu, 2005, 80p. Tese (Doutorado em Agronomia/Energia na Agricultura) – Faculdade de Ciências Agronômicas, Universidade Estadual Paulista.
Author: MARIE OSHIIWA
Adviser: CARLOS ROBERTO PADOVANI

2. SUMMARY

The purpose of the present paper is to discuss theoretical and practical aspects of the behavior of response variables in different groups and evaluation conditions by using Growth Curves methodology. This methodology refers to a multivariate procedure of experimental data analysis that makes forecasts about the average behavior of the response variable for different situations from that for which the study was planned. In additien, the methodology enables comparative analysis of the curves between each of the experimental groups. Considering the lack of easy-use computer programs for researchers in the agronomical, biological and health fields, and the difficulty to understand the complexity of the data structure in longitudinal studies, a computer program will be proposed and written using high level language. The software will be of simple handling, easy access to researchers of applied areas and available. This work will also discuss the advantages of using the multivariate procedure of analysis compared to the conventional procedures commonly used in agronomic experimentation, related to the preservation of the structure of data dependence.

Keywords: growth curve, program computer, longitudinal data.

3. INTRODUÇÃO

A utilização de estudos longitudinais é muito comum na experimentação agronômica, por exemplo, nas pesquisas prospectivas com o objetivo de observar a influência de determinado nutriente ao longo de um tempo específico, nos ensaios para investigar os efeitos da variação na dosagem de determinado fertilizante no experimento, na avaliação de produção nas mesmas parcelas em plantas perenes em anos seguidos, entre outros. De forma geral, neste tipo de estudo, é avaliado o comportamento de uma variável resposta ao longo de uma dimensão específica (por exemplo: tempo, dosagem) num conjunto de unidades de investigação classificada em diferentes grupos, segundo um ou mais fatores ou tratamentos.

A análise estatística dos dados obtidos em estudo longitudinal pode ser efetuada sob dois aspectos:

univariado: análise de variância para experimentos em parcelas subdivididas no tempo (*split plot in time*);

multivariado: análise de perfil (*profile analysis*) e análise de curvas de crescimento.

A grande vantagem do procedimento multivariado está na utilização de matrizes de variâncias e covariâncias amostrais que preservam a estrutura de associação das respostas, composta da variação entre e dentro das características observadas.

Deve ser destacado que, na análise multivariada, o nível de significância pré-estabelecido é assegurado em todas as conclusões (as variáveis são avaliadas

conjuntamente), enquanto que na univariada tal fato não ocorre. Basta lembrar que, num contexto univariado, para cada variável as conclusões são discutidas no nível fixado; porém, quando se faz a conclusão conjunta envolvendo todas elas, verifica-se que o valor de significância fica alterado para valor maior que o indicado inicialmente, implicando em falsas diferenças.

O ajuste de curvas de respostas em estrutura de dados longitudinais pode ser considerado como uma complementação da análise de perfil para grupos independentes. Segundo Singer *et a.l* (2004), dentre as vantagens que podem estar associadas a esse tipo de análise destacam-se: a possibilidade de comparações mais sensíveis em função do número reduzido de parâmetros quando comparado à análise de perfil; previsões sobre o comportamento médio das respostas para ocasiões de avaliação diferentes daquelas para as quais o estudo foi planejado e a aquisição de um maior conhecimento sobre o processo de variação da resposta ao longo das condições avaliadas.

Neste sentido, o presente trabalho tem por objetivo discutir aspectos teóricos e práticos da análise de curvas de crescimento, elaborar programa computacional de fácil acesso e simples manuseio, para pesquisadores das áreas aplicadas e apresentar exemplos ilustrativos de operacionalização do programa computacional na experimentação agronômica.

4. REVISÃO DE LITERATURA

4.1 Dados longitudinais

Na área agronômica é comum encontrarem-se experimentos com medidas repetidas, que se caracterizam pelo processo de obtenção de dados. Basicamente, duas formas de obtenção são consideradas: a primeira corresponde à atribuição aleatória da ordem com que as medidas sob diferentes condições de avaliação são realizadas na mesma unidade de investigação, como os dados obtidos de delineamentos com parcelas sub-divididas ("split-plot") ou delineamentos com intercâmbio ("crossover"). A segunda ocorre quando as medidas realizadas sob as diferentes condições de avaliação estão dispostas ao longo de uma escala ordenada, ou seja, são obtidas seqüencialmente e, neste caso, os estudos são denominados longitudinais (Singer *et al*, 2004).

Geralmente, a pesquisa longitudinal envolve observações de um conjunto de unidades de investigação (árvores, frutos, animais, indivíduos) classificados em diferentes subpopulações (variedade genética, raça, sexo) ou tratamentos (com e sem adubação, diferentes concentrações de determinado nutriente, entre outros) ao longo de diversas condições de avaliação (tempo, doses). As observações de cada unidade amostral são comumente denominadas perfis multivariados de resposta, refletindo a possível correlação entre essas medições.

Soler (1997) destaca duas classes de funções paramétricas de crescimento: (i) regressões polinomiais; sendo o uso de funções polinomiais recomendado

principalmente no ajuste de curvas de crescimento a séries curtas de dados; (ii) curvas com assíntotas; indicado para situações onde a resposta se aproxima de uma assíntota horizontal à medida que os valores na escala onde varia a condição de avaliação crescem, sugere-se o uso de curvas propostas por Gompertz ou von Bertalanffy. A autora enfatiza o uso de funções polinomiais como alternativa aos modelos assintóticos que apresentam problemas metodológicos relacionados com a estimação e testes de hipóteses sobre os parâmetros envolvidos.

Aubin (1984) descreve a técnica para dados obtidos de tratamentos avaliados em ocasiões sucessivas, em uma mesma unidade experimental, como análise de dados longitudinais, análise de curvas de crescimento ou análise de medidas repetidas, sendo as duas primeiras denominações mais freqüentemente usadas para experimentos cujo objetivo é monitorar a evolução de determinada variável no tempo. No último caso, é mais usual o estudo de efeito dos níveis dos fatores envolvidos no experimento. Dados oriundos de experimento com medidas repetidas podem ser analisados de duas formas: univariada e multivariada. A decisão sob qual modelo adotar depende da natureza das suposições que o pesquisador é capaz de fazer sobre os seus dados

A natureza de dados longitudinais é ilustrada por Suyama (1995) por meio de exemplos práticos e apresentação gráfica dos dados. O autor destaca que os modelos estatísticos para dados longitudinais devem, idealmente, (a) permitir a incorporação de uma estrutura de dependência entre as observações repetidas numa mesma unidade amostral; (b) não se restringir a dados balanceados ou regulares no tempo, e; (c) permitir a inclusão de covariáveis na análise.

As vantagens da análise multivariada aplicada em dados longitudinais sobre as técnicas convencionais de análise de variância para delineamento inteiramente casualizado e para parcelas subdivididas foram discutidas por Rosa & Padovani (1993). A importância do nível de significância pré-estabelecido na análise multivariada em relação à conclusão conjunta das análises individualizadas nos momentos de mensuração, foi enfatizada considerando dados relativos à densidade de estocagem de peixes.

O procedimento estatístico da análise de perfil para comparação dos vetores de médias de dados com planejamento longitudinal foi discorrido por Rosa (1994), onde foram destacadas as exigências estatísticas da metodologia: distribuição normal dos

dados, homocedasticia e que o número de unidades de investigação em cada grupo tem que ser maior do que o número de condições de avaliação. O trabalho apresentou, ainda, programa computacional denominado PERFIL de fácil manuseio, ilustrado com um ensaio para avaliação do desempenho produtivo de alevinos de carpa, distribuídos em três densidades de estocagem (2, 4 e 6 peixes/aquário), avaliados em quatro períodos de mensuração.

Os dados obtidos de estudos longitudinais em diversos tratamentos também podem ser ajustados por função logística (Carvalho & Pinho, 1996). Neste trabalho, as porcentagens do peso de frutos de araribá foram submetidas a quatro tratamentos e com avaliação de seis em seis horas, durante 84 horas. Os modelos das curvas logísticas foram comparadas por meio da estatística de t de Student.

Dois métodos de análise para dados longitudinais, freqüentemente utilizados na pesquisa biomédica, com enfoque para um público de não-estatísticos, foram discutidos por Ghost *et al.* (1999). O primeiro método examinava as variáveis pela análise de variância univariada (ANOVA) e o segundo por meio de modelo de efeitos mistos. Para a análise dos dados longitudinais foram utilizados vários *softwares* estatísticos, aos dados obtidos de um estudo pré-clínico conduzidos no ZymoGenetics.

Dados longitudinais podem ser estudados dentro de uma perspectiva bayesiana, sendo que a condução da análise é “a posteriori”. Rosa *et al*. (2004) discutem modelos e as distribuições normal contaminada, normal reduzida e t de Student sob este novo enfoque de análise estatística. O conjunto de dados longitudinais apresentado em Potthoff & Roy (1964) foi analisado sob os critérios dos modelos de efeitos mistos. Os resultados mostraram alternativas robustas, interessantes para modelos lineares mistos.

4.2 Curvas de Crescimento

Um estudo detalhado dos problemas que envolvem o procedimento de ajuste de curvas polinomiais a conjuntos de dados utilizando o modelo linear multivariado de crescimento foi realizado por Singer (1977). Uma discussão dos problemas que abrangem a estimação dos parâmetros em modelos lineares multivariados e a aplicação de testes de hipóteses para esses modelos foi desenvolvida com a utilização do estimador de máxima verossimilhança da matriz de parâmetros ξ e com a generalização da análise de variância

multivariada (MANOVA). Três exemplos práticos foram utilizados para discussão do procedimento em questão, um deles descrito por Potthoff e Roy (1964), Grizzle e Allen (1969) e um conjunto de dados submetidos ao Setor de Estatística Aplicada do Departamento de Estatística do IME – USP/SP.

Um modelo auto-regressivo simplificado no qual os erros apresentaram uma distribuição probabilística que incorporou a correlação existente entre dados de um mesmo animal , com menor número de parâmetros, foi apresentado por Crocci (1984). Os parâmetros do modelo matemático estudado foram estimados pelo método dos mínimos quadrados generalizados. Considerou-se, para exemplificação do procedimento, dados obtidos de experimento com vacas em lactação.

O ajuste de curvas polinomiais de crescimento a conjunto de dados longitudinais em avicultura, sob o enfoque multivariado, foi utilizado no estudo de Lima (1988), que discutiu a teoria do modelo linear de crescimento obtida da generalização do modelo usual da análise de variância multivariada. Os principais resultados foram apresentados envolvendo a estimação e a construção de testes de significância para os parâmetros do modelo, além de um procedimento básico de análise que utilizou um programa de MANOVA. O polinômio que melhor se ajustou aos perfis médios de pesos dos dois grupos de frangos de corte, machos e fêmeas, foi o de segundo grau, com duas variáveis auxiliares incluídas no ajuste. Para as duas linhagens de poedeiras – vermelhas e brancas – o polinômio de menor grau que melhor se ajustou aos perfis médios de pesos foi o do terceiro grau, com uma variável auxiliar incluída como covariável.

Para casos em que os dados longitudinais estão incompletos, o modelo de curva de crescimento também pode ser ajustado, como no trabalho de Souza (1989), onde a matriz de covariâncias relativa a cada grupo de indivíduos de mesmo padrão de observações perdidas foi obtida supondo-se que os resíduos seguem um processo auto-regressivo de primeira ordem. Os parâmetros do modelo polinomial proposto foram estimados pelo método da máxima verossimilhança. O método foi ilustrado com dados relativos ao comprimento de crianças de baixo peso no nascimento, obtidos no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

Partindo de um ponto de vista estatístico, o critério de ajuste de curvas individuais de crescimento para dados longitudinais envolve dois problemas discutidos por

Milani *et al.* (1989): (a) escolher a função matemática ideal, em que haja um número mínimo de parâmetros para descrever o processo individual de crescimento num determinado período de tempo; (b) estimar a matriz de covariâncias dos dados longitudinais de um grupo homogêneo de indivíduos e calcular intervalos de confiança para a construção de gráficos. Sob o ponto de vista prático, a falta de interpretação biológica dos coeficientes da curva de crescimento tem sido uma preocupação para os pesquisadores.

Kroll (1990) estuda o crescimento de seis vacas leiteiras, avaliado pelos pesos corporais, anotados durante 60 meses, em duas situações: a primeira, considerando a estrutura de independência dos erros, tomada como situação usual, e a segunda, com estrutura de erros autocorrelacionados (autoregressivo). Os modelos de crescimentos utilizados foram: Polinomial do 3º grau, de Mitscherlich, de Gompertz e o Logístico. Com os resultados obtidos na análise estatística, o autor concluiu que, embora o processo autorregressivo seja mais sensível a flutuações dos pesos das vacas, deve ser preferido em relação ao usual, frente à estrutura de autocorrelação dos erros.

A dinâmica do crescimento em peso de 95 recém-nascidos pré-termo adequados para a idade gestacional, do nascimento até o sexto mês pós-termo foi uma outra aplicação do modelo polinomial na área médica. Souza *et al.* (1992) ajustaram aos dados um polinômio de 2º grau onde a variável antropométrica foi considerada dependente e a idade atual da criança, em semanas, a variável independente. As crianças estudadas foram divididas em quatro grupos, segundo a idade gestacional, tendo sido pesadas ao nascer e depois a cada duas ou quatro semanas. A discussão dos resultados foi realizada comparando-se os picos das curvas, porém nenhum teste estatístico foi aplicado para comparação dos mesmos.

Um polinômio de segundo grau foi o modelo ajustado aos dados das alturas de 690 gêmeos, considerando o período do nascimento até os dois anos de idade, segundo o sexo. Baker *et al.* (1992) estimaram os parâmetros das curvas de crescimento que permitiram concluir que os meninos tiveram um crescimento significativamente superior.

Amorim & Bergamin Filho (1992) utilizaram a análise de variância multivariada para comparação das curvas de crescimento da evolução da doença do carvão em quatro variedades de cana-de-açúcar. As avaliações foram realizadas durante quatro anos e as curvas de progresso da doença foram construídas para cada variedade. As variáveis: número cumulativo de chicotes no primeiro patamar, tempo entre o plantio ou corte e o

estabelecimento do primeiro patamar, o número máximo de chicotes e a área sob a curva de progresso da doença foram os parâmetros para a determinação do modelo considerado.

Silveira Jr. *et al.* (1992) mostraram como a utilização de curvas de crescimento tem contribuído no estudo da sazonalidade de elementos climáticos, característicos de regiões de clima temperado, que influenciam a disponibilidade de pastagens, e, conseqüentemente, as alterações do desenvolvimento ponderal de bovinos. Um método para a obtenção de estimativas preliminares dos parâmetros em modelos de crescimento não-lineares foi apresentado e aplicado ao modelo raiz cúbica de von Bertalanffy com componente sazonal.

Silva (1993) enfatiza que ao se analisar uma variável é necessário estabelecer um modelo sobre o qual será construída uma hipótese, que será depois testada e discutida. Como esse modelo é uma representação da variável, sob as mensurações realizadas, com os possíveis fatores envolvidos na sua manifestação fenotípica e, por definição, jamais será perfeito. Portanto, esses fatores deverão ser incluídos no modelo de forma biologicamente aceitável.

A aplicabilidade de técnicas multivariadas de análises de perfis e de curvas de crescimento usando modelos mistos para a análise de conjuntos de dados longitudinais foi discutida por Lima (1996). Para o estudo, foram utilizados dados experimentais relativos aos níveis de adubação, espaçamento e consorciação no crescimento de árvores das espécies *Pinus Caribaea var. Hondodorensis bar. et golf* e *Liquidambar styraciflua l.*. As variáveis respostas foram: altura, diâmetro e volume em idades de 4, 10 e 28 anos. Para a identificação de observações e perfis individuais discrepantes foram utilizados dois diagramas paralelos de dispersão individuais.

Uma das principais finalidades do estudo de curvas de crescimento na Zootecnia é a estimativa da exigência nutricional, a qual pode ser fornecida pela simulação da composição corporal durante as diversas fases de crescimento do animal. Essa previsão é realizada por meio do conhecimento prévio dos parâmetros de crescimento de um grupo de animais, tanto do peso adulto como das respectivas taxas de crescimento relativas. Sob estas considerações, Tedeschi (1996) estudou o comportamento dos parâmetros de curvas de crescimento de 151 machos e 147 fêmeas de sete grupos raciais submetidos a três tipos de suplementação (sem, com suplementação durante o período da seca e com suplementação

durante o ano todo), segundo as funções: Brody, Gompertz, Logística, von Bertalanffy, Richards e algumas variações de Brody e Gompertz. Procedimentos estatísticos univariados e multivariados foram utilizados na análise dos resultados.

Utilizando as mesmas funções que Tedeschi (1996), Braccini Neto *et al.* (1996) verificaram a qualidade de ajuste da curva de crescimento a dados de peso- idade de galinhas de postura usando o coeficiente de determinação, quadrado médio do resíduo e o erro de predição médio. Os indicativos avaliados para a escolha da melhor curva de crescimento foram as dificuldades computacionais, a qualidade de ajuste e a interpretação biológica dos parâmetros.

A análise de crescimento da *Solanum crinitum*, uma espécie arbórea pioneira de porte médio, comum como planta invasora de pastagens, áreas agrícolas abandonadas, e clareiras em áreas de floresta na Amazônia brasileira, foi realizada por Dias Filho (1997). A fisiologia, a morfologia e o crescimento deste arbusto foram observados em quatro momentos durante 66 dias, em câmaras de crescimento, sob alto e baixo regimes de luz. A análise estatística foi considerada sob os enfoques da da Análise de Variância Univariada e Multivariada. Os programas computacionais MINITAB, SYSTAT e STATISTICA foram necessários para a análise dos dados.

Elias (1998) considerou os mesmos modelos não-lineares estudados por Tedeschi (1996), para pesos de 629 fêmeas da raça Nelore, 358 da raça Guzerá e 176 da raça Gir, nascidas no período de 1978 a 1991. A estimação dos parâmetros da curva de crescimento, considerando as formas não-ponderadas e ponderadas pelo inverso das variâncias dos pesos, com o objetivo de levar em consideração a heterocedasticidade dos dados, foi realizada pelo programa computacional SAS ("Statistical Analysis System"), supondo os erros não correlacionados. Quatro métodos de estimação de parâmetros dos modelos foram considerados: Gauss-Newton, Marquardt, , método da secante (ou DUD – "doesn't use derivatives") e Newton, sendo o primeiro destes o escolhido para o ajuste de dados com base na porcentagem de convergência e tempo de processamento. De maneira geral, considerando-se os procedimentos dos quadrados mínimos, todos os modelos se ajustaram adequadamente aos dados, porém, o modelo Brody ponderado foi o escolhido por descrever melhor o crescimento das vacas.

Littell *et al.* (1998) utilizaram o programa computacional SAS na análise dos dados de medidas repetidas observados em um experimento para a avaliação de potencial genético de touros, considerando-se um modelo linear misto. Na construção do modelo da estrutura de covariância dos dados foram utilizados os procedimentos GLM e MIXED do programa computacional SAS.

Peressin *et al.* (1998) avaliaram a produção e o acúmulo de matéria seca das plantas de mandioca em diferentes períodos a partir do plantio, na presença ou ausência de plantas infestantes, considerando-se a interferência do mato no desenvolvimento das mesmas. As análises de crescimento da cultura evidenciaram que as parcelas mantidas por períodos sob a interferência do mato, apresentam drástica redução no acúmulo de matéria seca pelas plantas, estando as perdas de raízes próximas de 90%. As curvas de acúmulo de matéria seca nas raízes foram explicadas pelo modelo sigmoidal de Boltzman, embora, para os períodos crescentes na presença de plantas infestantes, para ambos os experimentos, os coeficientes de determinação não tenham sido representativos. As maiores produções de matéria seca nas raízes foram obtidas aos 360 dias do plantio.

Os testes estatísticos para a hipótese do modelo linear multivariado de crescimento podem ser avaliados de forma semelhante aos procedimentos aplicados na MANOVA, que utilizam quatro critérios considerados principais para se testar a hipótese de igualdade dos efeitos dos vetores médios dos tratamentos. Cecchetti (1999) avalia, por meio de simulação, o poder dos testes e as taxas de erro tipo I desses critérios: Wilks, Lawley-Hotelling, Roy e Pillai. Por meio de experimentos simulados, pode-se concluir que a estrutura de correlação entre as variáveis não afetou as taxas de erro tipo I, porém tiveram influências no poder dos quatro critérios.

Um estudo da precisão da análise de dados longitudinais, com diferentes estruturas para a matriz de variâncias e covariâncias, quando se emprega o esquema em parcelas subdivididas foi realizado por Malheiros (1999), com utilização de dados simulados em computador. Os resultados mostraram que: a) existem situações em que Σ não satisfaz as condições de circularidade e a precisão da análise é satisfatória, mas, nesses casos, a indicação de significância é superestimada; b) a precisão da análise independe do fato dos experimentos serem balanceados ou não, com ou sem caselas vazias.

Rosa *et al*.(2000) avaliaram o modelo linear que melhor se ajustava aos pesos de 32 cordeiros aos 24, 35, 46 e 57 dias de idade, segundo o sexo. O método univariado não foi apropriado por causa da diferença entre as estruturas de variâncias e covariâncias dos dois grupos. Na análise multivariada foram utilizados quatro critérios: Wilks, Pillai, Hotteling-Lawley e Roy para verificar o efeito do tempo (dia). O melhor modelo de ajuste apresentado para a matriz Σ foi o autorregressivo de 1ª ordem, com heterogeneidade das variâncias e o critério utilizado foi o de máxima verossimilhança.

Uma aplicação do estudo de dados longitudinais na área de Obstetrícia são as curvas percentuais empregadas no acompanhamento da evolução temporal de diferentes fenômenos. Shinzato *et al*. (2000) estudaram a variação do peso fetal em gestações gemelares normais duplas, em 40 gestações em dois ou mais instantes ao longo de 12 a 39 semanas, com uso do ultra-som. Levando em conta fatores como a possível correlação entre as medidas realizadas no mesmo feto e em fetos irmãos, heterocedasticidade e distribuição do peso fetal estimado para cada idade gestacional, foram considerados modelos não polinomiais ou polinomiais de 2º ou 3º graus para representar as curvas de interesse.

Equações de primeiro e segundo graus descrevem o comportamento de dados obtidos em dois experimentos conduzidos para comparar metodologia de determinação do coeficiente de cobertura do solo pelo melão, quando submetido a condições diferenciadas de lâminas e nível de salinidade na água em Mossoró (RN). Dutra *et al.* (2000) analisaram esses dados considerando a ANOVA e a MANOVA sob os critérios de Wilks, Pillai e Hotelling-Lawley.

Uma discussão da escala do tempo no estudo do ajuste de curvas de crescimento linear foi apresentada por Stoel & Wittenboer (2000) considerando-se o procedimento multivariado. Os autores enfatizam que a quantidade de trabalhos que discutem a relação entre a condição inicial e a taxa de crescimento segundo a escala de tempo envolvida no modelo é pequena, e também que medidas de tempo diferentes conduzem a estimativas diferentes destes dois parâmetros do crescimento, tanto quanto suas variâncias e covariâncias. Através de procedimentos matemáticos, a natureza dessa relação é ilustrada por meio de modelos com diferentes escalas de tempo para os mesmos dados empíricos.

A estrutura de dependência entre as medidas na mesma gestante foi considerada por Martinelli *et al*. (2001) no estudo realizado com 100 gestantes para a

construção de uma curva de altura uterina em função da idade gestacional confiável, sem intercorrências clínicas e/ou obstétricas, de modo prospectivo e com mensurações efetuadas por um único observador. O modelo ajustado para o conjunto de dados foi uma função polinomial de segundo grau por meio do programa computacional SAS. O trabalho também analisou as diferenças com as curvas já existentes e a influência da cor, paridade e peso materno na variação da altura uterina.

Três funções matemáticas (Gama Incompleta, Polinomial Inversa e Quadrática Logarítmica) para ajustes da curva de lactação de vacas Caracu foram comparadas por Faro & Albuquerque (2002). Os dados foram ajustados através de processos iterativos do SAS e as estatísticas usadas para avaliação da qualidade do ajuste foram o coeficiente de determinação ajustado ao número de parâmetros de cada modelo e a análise dos resíduos, cujo resultado acusou autocorrelação residual positiva. Apesar da dificuldade na interpretação biológica dos parâmetros dessas funções matemáticas, os autores concluíram que a maior importância no uso dessas funções está na estimação das produções totais, a partir das produções parciais. Os resultados indicaram que a lactação média, normalmente usada como critério para escolha da função de melhor ajuste, pode ser um indicativo pobre do ajuste das lactações individuais.

Um estudo comparativo de três sistemas de alimentação de frangos de corte coloniais, envolvendo peso corporal, consumo de ração, conversão alimentar e custo da ração para produzir um quilo de frango vivo, foi realizado por meio de ajuste de modelos polinomiais mistos. Coldebella *et al.* (2003) utilizaram o SAS para a análise dos dados ajustados ao polinômio de segundo grau. As estimativas dos parâmetros das matrizes de covariâncias permitiram evidenciar associações entre as diversas fases da criação.

Uma aplicação do modelo Multivariado de Curvas de Crescimento proposto por Potthoff & Roy (1964) é ilustrado por Lima (2003) a um conjunto de dados completos e balanceados em relação ao tempo. Medidas de pesos de frangos de corte, com heterocedasticidade de variâncias ao longo das sete semanas de experimentação foram explicados por polinômios do segundo grau e incluindo como covariáveis, as colunas referentes aos polinômios de graus três e quatro. As análises foram realizadas utilizando-se o programa computacional SAS.

A importância de ajustar modelos adequados a dados observacionais foi discutida considerando de resultados de análises granulométricas de 130 perfis de solos descritos no Projeto RadamBrasil. Esses dados foram utilizados por Silva *et al.* (2004) no ajuste de 14 modelos de curvas de crescimento, selecionados em diversas fontes, com potencial para descrição do traçado da curva granulométrica. A comparação entre os modelos foi baseada na soma de quadrados dos erros verificados entre os valores observados e ajustados. O número máximo de parâmetros no ajuste dos modelos foi limitado a três. Tratando-se de uma otimização de modelos não-lineares, a escolha dos valores iniciais dos parâmetros a serem ajustados no Excel foi fundamental na convergência do processo de otimização do mínimo global.

Uma aplicação de ajuste de dados longitudinais em Saúde Pública foi realizada por Giolo *et al.* (2004) com o objetivo de reduzir os diâmetros de gotas do veículo aquoso utilizado na diluição de produtos químicos no controle do mosquito *Aedes Aegypti*. Três caldas foram preparadas e a capacidade de aumentar o tempo de evaporação foi avaliada por meio do ajuste de curvas de crescimento. Um estudo detalhado do modelo adequado ao banco de dados e da estrutura de covariâncias do modelo adotado foi descrito no estudo. Considerando-se o software SAS, um modelo polinomial de segunda ordem foi ajustado aos dados observacionais.

5. MATERIAL E MÉTODOS

5.1 Dados longitudinais

Para ilustrar uma situação genérica de aplicação de dados longitudinais na área de ciências agronômicas, considere que um pesquisador deseja estudar a influência de diferentes concentrações de um determinado adubo na altura média de uma certa espécie de alho durante quatro semanas consecutivas, a partir do sétimo dia de plantio. Neste sentido, 20 plantas são casualizadas em três tratamentos: um grupo controle (sem adubação) e dois grupos com concentrações diferentes do adubo. A altura de cada planta é medida semanalmente a partir do sétimo dia de plantio conforme o quadro a seguir:

Quadro1. Estrutura ilustrativa de um experimento casualizado em três tratamentos com 20 plantas de alho, avaliados em quatro semanas consecutivas.

Grupo experimental	**Unidade experimental**	**Tempo (semana)**			
		1	**2**	**3**	**4**
Tratamento 1 (controle)	1	y_{111}	y_{112}	y_{113}	y_{114}
	2	y_{121}	y_{122}	y_{123}	y_{124}
	⋮	⋮	⋮	⋮	⋮
	6	y_{161}	y_{162}	y_{163}	y_{164}
Tratamento 2 (concentração1)	1	y_{211}	y_{212}	y_{213}	y_{214}
	2	y_{221}	y_{222}	y_{223}	y_{224}
	⋮	⋮	⋮	⋮	⋮
	7	y_{271}	y_{272}	y_{273}	y_{274}
Tratamento 3 (concentração2)	1	y_{311}	y_{312}	y_{313}	y_{314}
	2	y_{321}	y_{322}	y_{323}	y_{324}
	⋮	⋮	⋮	⋮	⋮
	7	y_{371}	y_{372}	y_{373}	y_{374}

Como pode ser esperado, a altura de cada planta avaliada semanalmente, terá na medida da segunda semana, resposta igual ou maior que a da primeira; a da terceira semana será igual ou superior a da segunda e, conseqüentemente, a altura da quarta semana será igual ou superior a da terceira. Verifica-se, desta forma, a existência de uma estrutura de dependência entre as medidas de cada planta, fato este que deve ser relevado na decisão de não utilização dos modelos clássicos de Análise de Variância e de Regressão na análise estatística dos dados.

De uma forma geral, dados longitudinais caracterizam-se como aqueles obtidos de mensurações de uma variável resposta em condições de avaliação distintas, para cada unidade experimental. Essas condições podem ser, por exemplo, tempo, doses de adubos e outros. Os experimentos planejados com dados longitudinais envolvem um conjunto de unidades experimentais, cada qual avaliada em várias condições e classificadas em subpopulações (tratamentos ou grupos).

5.2 Análise de dados longitudinais

Para ilustrar um estudo longitudinal, considere o exemplo do alho, onde três grupos são avaliados em quatro momentos distintos. Se para cada grupo for ajustada uma reta relativa ao perfil médio de resposta, as comparações entre os grupos serão realizadas em termos de seis parâmetros (três coeficientes lineares e três coeficientes angulares) ao invés de 12 médias referentes aos quatro momentos de cada um dos três grupos.

Quando a estrutura de correlação entre as observações dos dados longitudinais é preservada e o efeito da subpopulação ou tratamento é estabelecido em função das condições de avaliação, o modelo estatístico apropriado para explicar o fenômeno em estudo é conhecido por “Curvas de Crescimento”.

Genericamente, as respostas observadas em um experimento com planejamento longitudinal, sem falhas de observação, considerando-se g grupos e p momentos de avaliação, podem estar organizadas conforme quadro apresentado a seguir:

Quadro2. Estrutura genérica de valores observados em um experimento com planejamento longitudinal.

Grupo Experimental	Unidade Experimental (Parcela)	Momento de Avaliação			
		t_1	t_2	...	t_p
1	1	y_{111}	y_{112}	...	y_{11p}
1	2	y_{121}	y_{122}	...	y_{12p}
M	M	M	M	M	M
1	n_1	$y_{n_1 1}$	$y_{n_1 2}$	...	$y_{n_1 p}$
2	1	y_{211}	y_{212}	...	y_{21p}
2	2	y_{221}	y_{222}	...	y_{22p}
M	M	M	M	M	M
2	n_2	$y_{1\,n_2 1}$	$y_{1\,n_2 2}$	...	$y_{1 n_2 p}$
M	M	M	M	M	M
g	1	y_{g11}	y_{g12}	...	y_{g1p}
g	2	y_{g21}	y_{g22}	...	y_{g2p}
M	M	M	M	M	M
g	n_g	$y_{g\,n_g 1}$	$y_{g\,n_g 2}$	...	$y_{g\,n_g p}$

A observação y_{ijk} é expressa pelos índices:

i = 1, ..., g (grupo),

j = 1, ..., n_i (parcela) e

k = 1, ..., p (momentos).

Portanto, $\mathbf{y}'_{ij} = (y_{ij1}\ ,\ y_{ij2}\ ,\ \ldots\ ,\ y_{ijp}\ )$ representa o vetor resposta correspondente às observações dos p momentos de avaliação para a j-ésima unidade experimental do i-ésimo grupo.

Como no presente estudo os modelos discutidos serão reduzidos ao modelo da Análise de Variância Multivariada (MANOVA), será realizada a seguir uma breve revisão deste procedimento analítico. Um estudo mais detalhado do procedimento metodológico pode ser encontrado em Morrison (1976).

5.3 Análise de Variância Multivariada (MANOVA)

O modelo clássico de Análise de Variância Multivariada (MANOVA) pode ser expresso pelo modelo linear usual de Gauss-Markov:

$$\mathbf{Y_{nxp}} = \mathbf{X_{nxg}}\ \beta_{\mathbf{gxp}} + \varepsilon_{\mathbf{nxp}},$$

onde,

$\mathbf{Y_{nxp}}$ é a matriz de observações (p medidas para cada uma das n unidades experimentais), com $n = \sum_{i=1}^{g} n_i$;

$$Y_{nxp} = \begin{bmatrix} y_{111} & \Lambda & y_{11p} \\ \mathrm{M} & \mathrm{M} & \mathrm{M} \\ y_{gn_g 1} & \Lambda & y_{gn_g p} \end{bmatrix};$$

$\mathbf{X_{nxg}}$ é a matriz de planejamento, constituída por 0´s e 1´s, de forma a associar cada unidade experimental ao respectivo grupo (g),

$$X_{nxg} = \begin{bmatrix} 1_{n_1} & 0 & \Lambda & 0 \\ 0 & 1_{n_2} & \Lambda & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 0 & 0 & \Lambda & 1_{n_g} \end{bmatrix},$$

onde 1_{n_i} é o vetor unidade com n_i componentes;

$\beta_{\mathbf{gxp}}$ é a matriz de parâmetros desconhecidos (cada linha é constituída por p elementos que representam os valores esperados da variável observada, em cada um dos instantes considerados para um determinado grupo),

$$\beta_{gxp} = \begin{bmatrix} \beta_{11} & \Lambda & \beta_{1p} \\ \Lambda & \Lambda & \Lambda \\ \beta_{g1} & \Lambda & \beta_{gp} \end{bmatrix};$$

ε_{nxp} é a matriz dos componentes aleatórios das observações (cada elemento representa o desvio entre o valor observado e o respectivo valor esperado). Ademais, β e ε, são matrizes não-observáveis.

Uma pressuposição subjacente do modelo considerado é que as linhas da matriz ε sejam não correlacionadas e possuam uma distribuição normal multivariada com vetor de valores esperados $\mathbf{0}_{1xp}$ e matriz de covariância não singular Σ_{pxp}, comum a todas as linhas. Ou seja:

$$E(\mathbf{Y}) = \mathbf{X}\beta$$

$$\text{Var}(\mathbf{Y}) = \Sigma \otimes \mathbf{I}_n.$$

O modelo descrito dessa forma é denominado "Modelo Linear Multivariado" (MLM) e indica-se a seguinte notação MLM ($\mathbf{Y}$, $\mathbf{X}\beta$, Σ).

O símbolo $\otimes$ indica produto tensorial ou de Kronecker (Gantmacher, 1989), cuja operação consiste em:

$$\mathbf{A}_{nxp} \otimes \mathbf{B}_{pxp} = \begin{bmatrix} a_{11} & \Lambda & a_{1p} \\ M & M & M \\ a_{n1} & \Lambda & a_{np} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & \Lambda & b_{1p} \\ M & M & M \\ b_{p1} & \Lambda & b_{pp} \end{bmatrix} = \begin{bmatrix} a_{11}B & \Lambda & a_{1p}B \\ M & M & M \\ a_{n1}B & \Lambda & a_{np}B \end{bmatrix},$$

resultando em uma matriz de ordem np x p^2.

As seguintes propriedades são básicas para determinação das estatísticas multivariadas:

i) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$;

ii) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$;

iii) $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$;

iv) $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}$.

A partir destas propriedades, fica fácil verificar que, se:

Var (Y) = $\mathbf{A} \otimes \mathbf{B}$, então:

i) Var $(\mathbf{Y}') = \mathbf{B} \otimes \mathbf{A}$;

ii) Var $(\mathbf{E}'\mathbf{Y}) = \mathbf{A} \otimes (\mathbf{E}'\mathbf{BE})$;

iii) Var $(\mathbf{YF}) = (\mathbf{F}'\mathbf{AF}) \otimes \mathbf{B}$;

iv) Var $(\mathbf{E}'\mathbf{YF}) = (\mathbf{F}'\mathbf{AF}) \otimes (\mathbf{E}'\mathbf{BE})$.

Sob as considerações descritas, o estimador de mínimos quadrados de β, que coincide com o de máxima verossimilhança, expresso por:

$$\hat{\beta} = (X'X)^{-1}X'Y \text{ e}$$

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

onde, $\mathbf{C_{cxg}}$ e $\mathbf{U_{pxu1}}$ são matrizes selecionadas de tal forma a obter as combinações lineares desejadas dos parâmetros.

As hipóteses testadas sob as considerações do modelo são escritas na forma geral:

$$H_0: \mathbf{C\beta U} = \Gamma\mathbf{.}$$

Para o teste estatístico de H_0, são consideradas as seguintes matrizes:

$\mathbf{S}_{pxp} = \mathbf{Y'[I - X(X'X)^{-1}X']Y}$ (matriz da soma de quadrados e produtos cruzados do resíduo);

$\mathbf{E}_{u1\ x\ u1} = \mathbf{U'SU} \sim \mathbf{W}$ (n-g, $\mathbf{U'\Sigma U}$, ∅)

e

$\mathbf{H}_{u1x\ u1} = (\mathbf{C}\hat{\beta}\mathbf{U}-\Gamma)'[\mathbf{C(X'X)^{-1}C'}]^{-1}(\mathbf{C}\hat{\beta}\mathbf{U}-\Gamma) \sim W(c,\mathbf{U'\Sigma U},(\mathbf{C}\hat{\beta}\mathbf{U}-\Gamma)'[\mathbf{C(X'X)^{-1}C'}]^{-1}(\mathbf{C}\hat{\beta}\mathbf{U}-\Gamma)$

onde W representa a distribuição de Wishard e ∅ a matriz nula de ordem u_1.

As matrizes **E** e **H** são formadas por soma de quadrados e produtos cruzados relativos aos resíduos e à hipótese do teste, respectivamente. A matriz E tem distribuição central, enquanto H tem não central.

Sob a veracidade de H_0, **H** tem distribuição central de Wishart W(c, $\mathbf{U'\Sigma U}$, ∅) e, também, as matrizes **E** e **H** são independentes. Os testes estatísticos para a hipótese geral são realizados com estatísticas que são funções das raízes características não nulas (λ_1, ..., λ_s) da equação $|\mathbf{H} - \lambda\mathbf{E}| = 0$, ou seja, $|\mathbf{HE}^{-1} - \lambda\mathbf{I}| = \mathbf{0}$.

Para o estabelecimento dos testes estatísticos para a hipótese H_0, considere as seguintes definições:

$\theta_\nu = \dfrac{\lambda_\nu}{1+\lambda_\nu}$, $\nu = 1, ..., s$;

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

$m_1 = \dfrac{|c-u_1|-1}{2}$;

$m_2 = \frac{n-g-u_1-1}{2}$.

Equivalentemente, $\theta_\nu = \frac{\lambda_\nu}{1+\lambda_\nu}$ para $\nu = 1,..., s$, são as raízes características de $|\mathbf{H} - \theta(\mathbf{H} + \mathbf{E})| = \mathbf{0}$, ou ainda $|\mathbf{H}(\mathbf{H} + \mathbf{E})^{-1} - \theta\mathbf{I}| = \mathbf{0}$.

Os testes para a hipótese geral são baseados em quatro critérios:

a) União e Intersecção de Roy

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

$\theta_1 = \max \theta_\upsilon$, maior raiz característica da equação determinante $|\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1} - \theta\mathbf{I}| = \mathbf{0}$, $1 \leq \nu \leq s$, cuja distribuição é aproximada (ábacos de Heck, 1960) e os parâmetros são s, m_1 e m_2.

Se s = 1, então $F = \frac{1-\theta}{\theta} * \frac{m_2+1}{m_1+1}$, com graus de liberdade $(2m_1 +2)$ e $(2m_2 +2)$.

b) Razão de Verossimilhança Generalizada de Wilks

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

$$\Lambda = \frac{|E|}{|H+E|} = \prod_{\upsilon=1}^{s}(1-\theta_\upsilon) = \prod_{\upsilon=1}^{s}(1+\lambda_\upsilon)^{-1}$$

- Se s = 1, então,

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

- Se s = 2, então,

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

Para outros casos, mas quando o tamanho da amostra é grande, utiliza-se a aproximação de Bartlett representada por RV (Razão de Verossimilhança):

$$RV = -[n - g - \frac{1}{2}(u_1 - c + 1)] \ln \Lambda \sim \chi^2 (c*u_1)$$

c) Lawley-Hotelling

A estatística de Lawley-Hotelling é expressa por

$$LH = \text{tr}(\mathbf{HE}^{-1}) = \sum_{\upsilon=1}^{s} \lambda_{\upsilon}$$

A aproximação F para esta estatística é

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

d) Pillai

A estatística de Pillai, é expressa por:

$$PI = \text{tr}\,\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1} = \sum_{\upsilon=1}^{s} \theta_{\upsilon} = \sum_{\upsilon=1}^{s} \frac{\lambda_{\upsilon}}{1+\lambda_{\upsilon}}.$$

A aproximação F para a estatística de Pillai é:

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

Em estudos longitudinais, os objetivos mais comuns são:

i) identificar a existência de interação entre o fator que define as subpopulações, ou grupos, e o fator que define os momentos de avaliação, ou seja, se os g grupos apresentam o mesmo comportamento ao longo dos p momentos de avaliação;

ii) verificar se os g grupos apresentam os perfis médios coincidentes;

iii) analisar a evolução dos momentos nos perfis médios de respostas para os g grupos;

iv) ajustar modelos, geralmente na forma polinomial, para explicar a variação das respostas médias como função dos momentos de avaliação.

Para atingir tais objetivos, as hipóteses testadas em relação aos perfis médios de resposta com medidas repetidas são basicamente três, sob as considerações da hipótese geral H_0: $\mathbf{C\beta U} = \mathbf{0}$:

i) Paralelismo: os perfis médios de resposta correspondentes às diversas subpopulações são paralelos, isto é, não existe interação entre o fator que define as subpopulações e o fator que define os momentos de avaliação. Tem-se:

$$H_{01}: \begin{bmatrix} \beta_{11} - \beta_{12} \\ \beta_{12} - \beta_{13} \\ M \\ \beta_{1(p-1)} - \beta_{1p} \end{bmatrix} = \begin{bmatrix} \beta_{21} - \beta_{22} \\ \beta_{22} - \beta_{23} \\ M \\ \beta_{2(p-1)} - \beta_{2p} \end{bmatrix} = ... = \begin{bmatrix} \beta_{g1} - \beta_{g2} \\ \beta_{g2} - \beta_{g3} \\ M \\ \beta_{g(p-1)} - \beta_{gp} \end{bmatrix}$$

Sendo as matrizes **C** e **U**:

$$C_{(g-1)xg} = \begin{bmatrix} 1 & -1 & 0 & \Lambda & 0 & 0 \\ 0 & 1 & -1 & \Lambda & 0 & 0 \\ M & M & M & M & M & M \\ 0 & 0 & 0 & \Lambda & 1 & -1 \end{bmatrix} \text{ e } U_{px(p-1)} = \begin{bmatrix} 1 & 0 & 0 & \Lambda & 0 \\ -1 & 1 & 0 & \Lambda & 0 \\ 0 & -1 & 1 & \Lambda & 0 \\ M & M & M & M & M \\ 0 & 0 & 0 & M & 1 \\ 0 & 0 & 0 & M & -1 \end{bmatrix}$$

ii) Coincidência: os perfis médios de resposta correspondentes às diversas subpopulações são coincidentes, isto é, não existe efeito desse fator. Portanto, tem-se:

$$H_{02}: \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ M \\ \beta_{1p} \end{bmatrix} = \begin{bmatrix} \beta_{21} \\ \beta_{22} \\ M \\ \beta_{2p} \end{bmatrix} = ... = \begin{bmatrix} \beta_{g1} \\ \beta_{g2} \\ M \\ \beta_{gp} \end{bmatrix}.$$

A matriz **C** é a mesma da hipótese anterior e a **U** será: $U_{px1} = \begin{bmatrix} 1 \\ 1 \\ M \\ 1 \end{bmatrix}$

iii) Paralelismo ao eixo das abscissas: os perfis médios de respostas correspondentes às diversas subpopulações são paralelos ao eixo das abscissas, isto é, não existe efeito do fator que define os momentos de avaliação.

$$H_{03}: \sum_{i=1}^{q} \beta_{i1} = \sum_{i=1}^{q} \beta_{i2} = ... = \sum_{i=1}^{q} \beta_{ip}$$

Sendo, $C_{1xg} = \begin{bmatrix} 1 & 1 & \Lambda & 1 \end{bmatrix}$ e a matriz **U** idêntica à descrita na hipótese de paralelismo.

Deve ser observado que os resultados dos testes de hipóteses são invariantes às escolhas de **C** e **U**.

5.4 Modelo Linear Multivariado de Crescimento (MLMC)

Na investigação científica, os dados obtidos de experimentos são transformados em informações, por meio de análise estatística. Especificamente em estudos longitudinais, uma das dificuldades encontradas pelos pesquisadores no processamento analítico dos dados é a redução do número de parâmetros envolvidos por meio da imposição de modelos estruturais que relacionam os efeitos dos grupos com os momentos de avaliação.

Os modelos estruturais consistem na modelagem matemática que abrange um conjunto de equações que refletem as relações e interações entre as variáveis envolvidas num dado fenômeno. Essas expressões algébricas são apresentadas a partir de indicadores matemáticos, que são considerados como uma ferramenta utilizada para reduzir a dimensão do problema com algum significado prático e podem ser produzidos pelo agrupamento e combinação de variáveis resultando, a partir delas, numa nova variável que possui interpretação de interesse.

A partir de um estudo exploratório, dados obtidos em estudo longitudinal serão estruturados em modelos univariados e multivariados. Os critérios utilizados para o modelo adequado às observações longitudinais são, basicamente, o número de unidades amostrais e a estrutura de covariâncias.

O primeiro modelo é sugerido quando o padrão de variação da resposta ao longo das condições de avaliação apresenta estruturas mais restritivas para a matriz de covariância entre as observações, ou ainda, nas situações em que o número de unidades amostrais é pequeno e existirem muitas ocasiões de avaliação. A análise estatística utilizada neste caso é análise de variância para experimentos em parcelas subdivididas no tempo (“split plot in time”), por exemplo. Mais informações sobre esta estrutura podem ser encontradas em Singer & Andrade (1986).

Soler (1997) o enfatiza que modelo multivariado é adequado quando não existe uma base teórica ou empírica para assumir uma estrutura de covariância especial e, portanto, é considerada uma matriz de covariância não estruturada para as medidas observadas na mesma unidade amostral. A estrutura de covariância requer um estudo detalhado e complexo para que possa apresentar informações que reflitam o comportamento das observações.

Duas técnicas multivariadas são utilizadas na análise de dados longitudinais: Análise de Perfis de Médias e Ajuste de Curvas de Crescimento.

Um estudo que discorre sobre o procedimento estatístico metodológico da Análise de Perfil com ilustração na experimentação zootécnica é encontrado em Rosa (1994), que apresenta também um programa computacional de fácil manuseio.

O Ajuste de Curvas aos perfis médios de respostas a dados longitudinais pode ser encarado como uma alternativa de Análise de Perfis. Esta técnica é considerada quando o objetivo do estudo abrange, pelo menos, um dos seguintes objetivos:

i) estabelecer um relacionamento entre o efeito dos tratamentos e o tempo;
ii) fazer comparações mais sensíveis em função do número reduzido de parâmetros;
iii) fazer previsões sobre o comportamento médio da resposta para ocasiões de avaliação diferentes daquelas para as quais o estudo foi planejado;
iv) possibilitar a aquisição de um maior conhecimento sobre o processo de variação da resposta ao longo das condições de avaliação.

Existem algumas situações específicas em que o comportamento da variável é facilmente identificado pela observação das médias, durante os momentos planejados, no diagrama de dispersão. Outras vezes, o fenômeno comportamental da variável é previsível e existe um modelo matemático pré-estabelecido. Nesses casos, não há necessidade do estudo de adequação do modelo aos dados observados.

Uma alternativa interessante para estudos experimentais e de observação é considerar como modelo matemático, para descrever o comportamento da variável resposta em função do tempo, uma curva de crescimento na forma polinomial, pois apresenta uma expressão funcional simples, o que facilita a interpretação dos resultados.

As curvas de crescimento podem ser estudadas sob dois aspectos: individual e média do grupo. No primeiro caso, como cada indivíduo, ou unidade experimental, associa-se à sua própria curva, a estrutura de covariância torna-se mais restritiva. Basicamente, nesta classe de procedimentos, tem-se a especificação de um modelo para as médias e de outro para a estrutura de covariâncias.

Para o presente estudo, será considerado o Ajuste de Curvas de Crescimento para as médias. Os métodos envolvidos nessa abordagem são apropriados para dados balanceados e completos.

Geralmente, o estudo das curvas de crescimento envolve procedimentos preliminares com objetivos exploratórios que auxiliam o pesquisador na decisão do modelo a ser considerado para descrever o comportamento dos dados longitudinais.

Os primeiros estudos de crescimento foram realizados pelo Conde de Montbeillard (1759 a 1777) e o primeiro trabalho publicado foi em um suplemento para Histoire Naturelle. O estudo de curvas de crescimento foi intenso na primeira metade do século XX destacando-se as contribuições de Wishart (1938), Box (1950) e Rao (1958-1959), mas o grande marco nesta área aconteceu em 1964 quando Potthoff & Roy apresentaram o Modelo Linear Multivariado de Crescimento (MLMC) escrito na forma:

$\mathbf{Y_{nxp}} = \mathbf{X_{nxg}}\ \xi_{\mathbf{gxq}}\ \mathbf{G_{qxp}} + \varepsilon_{\mathbf{nxp}}$, com

$$E(\mathbf{Y}) = \mathbf{X}\ \xi\ \mathbf{G};$$

$$\text{Var}\ (\mathbf{Y}) = \Sigma \otimes \mathbf{I}_n\ ,$$

sendo,

$\mathbf{Y_{nxp}}$, $\mathbf{X_{nxg}}$ e $\varepsilon_{\mathbf{nxp}}$ idêntico ao MLM e,

$$\xi_{gxq} = \begin{bmatrix} \xi_{10} & \xi_{11} & \Lambda & \xi_{1q-1} \\ \xi_{20} & \xi_{21} & \Lambda & \xi_{2q-1} \\ \text{M} & \text{M} & \text{O} & \text{M} \\ \xi_{g0} & \xi_{g1} & \Lambda & \xi_{gq-1} \end{bmatrix},$$

$\xi_{\mathbf{gxq}}$ a matriz dos parâmetros (desconhecidos) das curvas de crescimento.

Cada linha desta matriz representa os coeficientes do polinômio de grau λ ($\lambda \leq$ p-1) relativo ao i-ésimo grupo do experimento.

$\mathbf{G_{qxp}}$ a matriz de delineamento dentro de indivíduos, de dimensão q x p, de posto completo q $\leq$ p, cuja finalidade é associar as respostas das unidades experimentais sob as condições de avaliação ao polinômio desejado. Ou seja, as colunas correspondendo aos momentos de avaliação (t_1, t_2, ..., t_p) e as linhas aos possíveis graus do polinômio.

Isto é,

$$\mathbf{G}_{qxp} = \begin{bmatrix} t_1^0 & t_2^0 & \Lambda & t_p^0 \\ t_1^1 & t_2^1 & \Lambda & t_p^1 \\ t_1^2 & t_2^2 & \Lambda & t_p^2 \\ \text{M} & \text{M} & \text{O} & \text{M} \\ t_1^{q-1} & t_2^{q-1} & \Lambda & t_p^{q-1} \end{bmatrix}.$$

A matriz de delineamento **G** pode ser estruturada considerando-se três situações mais comuns:

1ª) usar os próprios valores de t (momentos do planejamento experimental);

2ª) utilizar os valores dos afastamentos em relação à média dos momentos, ou seja, os valores codificados em $(t - \bar{t})$;

3ª) considerar os coeficientes dos polinômios ortogonais.

Em relação à estimação dos parâmetros, a escolha de G como matriz de planejamento dentro de indivíduos pode ser considerada como fator secundário, porém para os testes estatísticos o uso de G construída pelos coeficientes dos polinômios ortogonais facilitará muito a obtenção das estatísticas. As propostas 1 e 2 são irrestritas quando ao uso, mas a terceira é utilizada para condições igualmente espaçadas (equiespaçadas).

A estrutura matricial de $\xi\mathbf{G}$, descrita como β, pode ser representada por:

$$\beta = \xi G = \begin{bmatrix} \xi_{10} + \xi_{11}t_1 + \ldots + \xi_{1,q-1}t_1^{q-1} & \xi_{10} + \xi_{11}t_2 + \ldots + \xi_{1,q-1}t_2^{q-1} & \ldots & \xi_{10} + \xi_{11}t_p + \ldots + \xi_{1,q-1}t_p^{q-1} \\ \text{M} & \text{M} & \ldots & \text{M} \\ \xi_{g0} + \xi_{g1}t_1 + \ldots + \xi_{g,q-1}t_1^{q-1} & \xi_{g0} + \xi_{g1}t_2 + \ldots + \xi_{g,q-1}t_2^{q-1} & \ldots & \xi_{g0} + \xi_{g1}t_p + \ldots + \xi_{g,q-1}t_p^{q-1} \end{bmatrix}$$

+

Sob as considerações da estrutura da matriz **G** têm-se, respectivamente, os seguintes modelos polinomiais de crescimento:

1ª) $\beta_{ik} = \xi_{i0} + \xi_{i1}\, t_k + \xi_{i2}\, t_k^{\,2} + \ldots + \xi_{iq\text{-}1}\, t_k^{\,q\text{-}1} = \sum_{l=0}^{q-1} \xi_{il} t_k^l$;

2ª) $\beta_{ik} = \xi^*_{i0} + \xi^*_{i1}\left(t_k - \bar{t}\right) + \xi^*_{i2}\left(t_k - \bar{t}\right)^2 + \ldots + \xi^*_{iq\text{-}1}\left(t_k - \bar{t}\right)^{q\text{-}1} = \sum_{l=0}^{w} \xi^*_{il}(t_k - \bar{t})^l$, com $\bar{t} = \sum_{k=1}^{p} \frac{t_k}{p}$;

3ª) $\beta_{ik} = \xi_{i0}^{**} P_0(t_k) + \xi_{i1}^{**} P_1(t_k) + \xi_{i2}^{**} P_2(t_k) + ... + \xi_{iq-1}^{**} P_{q-1}(t_k) = \beta_{ik} = \sum_{l=1}^{q-1} \xi_{il}^{**} P_l(t_k)$, onde $P_l(t_k)$ é termo polinomial no grau λ na construção de polinômios ortogonais (Steel & Torrie, 1980).

Nos modelos apresentados verifica-se que, em cada grupo experimental, a resposta média de cada momento pode ser descrita por um polinômio de grau λ. É importante observar que a formulação das curvas descritas em β tenham o mesmo grau.

Considerando a formulação $\beta = \xi\mathbf{G}$, o Modelo Linear Multivariado de Crescimento pode ser entendido como uma generalização do Modelo Linear Multivariado, fato que propicia uma abordagem que encaminha às curvas de crescimento a possibilidade de utilizar para as análises os procedimentos da MANOVA.

A seguir serão apresentadas três propostas para estimar a matriz de parâmetros ξ: a de Potthoff & Roy, a que ajusta por covariâncias e a de Rao.

5.4.1 Proposta de Potthoff & Roy

Potthoff & Roy (1964) para estimar ξ, sugerem um método de transformação de variável, que possibilita reduzir o Modelo Linear Multivariado de Crescimento no Modelo Linear Multivariado.

A sugestão desses autores foi introduzir uma matriz de ponderação arbitrária Δ_{pxp} a transformação de variável expressa por:

$$\mathbf{Y}^* = \mathbf{Y}\Delta^{-1}\mathbf{G}'(\mathbf{G}\Delta^{-1}\mathbf{G}')^{-1}$$

onde, Δ é uma matriz simétrica e positiva definida ou não-singular, tal que $(\mathbf{G}\Delta^{-1}\mathbf{G}')^{-1}$, de ordem (q x q), tenha posto completo. Sob esta consideração, tem-se:

$$E(\mathbf{Y}^*) = \mathbf{X}\xi$$

e a matriz de variância-covariância Σ^*, não singular, dada por:

$\Sigma^* = (\mathbf{G}\Delta'^{-1}\mathbf{G}')^{-1}\ \mathbf{G}\Delta'^{-1}\Sigma\Delta^{-1}\mathbf{G}'(\mathbf{G}\Delta'^{-1}\mathbf{G}')^{-1}$, gerando um novo modelo para as observações transformadas conforme MLM ($\mathbf{Y}^*$, $\mathbf{X}\xi$, Σ^*).

Como $\mathbf{Y}$ tem distribuição normal multivariada com linhas independentes, então as linhas da matriz $\mathbf{Y}^*$ serão normais multivariadas independentes.

Portanto, um estimador não viciado de ξ pelo procedimento dos mínimos quadrados (que coincide com o estimador de máxima verossimilhança) é determinado por:

$$\hat{\xi} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}^{*} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\,\mathbf{Y}\Delta^{-1}\mathbf{G}'(\mathbf{G}\Delta^{-1}\mathbf{G}')^{-1}.$$

Os autores demonstraram que $\Delta = \Sigma$ é a melhor escolha (não-viciada de variância mínima) para estimar a matriz de parâmetros, se Σ for conhecida. Apesar de o estimador $\hat{\xi}$ obtido não ser viciado, independentemente da escolha da matriz Δ, sua variância aumenta conforme Δ^{-1} se distancia de Σ^{-1}, que é desconhecida. Sendo Σ desconhecida e não estruturada, uma opção é usar $\Delta = \mathbf{S}$, onde $\mathbf{S} = \mathbf{Y}'(\mathbf{I} - \mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X})\mathbf{Y}$. Uma outra opção apreciada foi $\Delta = \mathbf{I}$, pois esta matriz simplifica consideravelmente os cálculos. O uso da identidade como matriz arbitrária é uma condição interessante no caso da matriz **G** ser formada pelos coeficientes de polinômios ortogonais, pois $\mathbf{Y}^{*}=\mathbf{Y}\mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}$.

Singer (1977) destaca que se p = q então **G** é inversível e toma-se $\mathbf{Y}^{*} = \mathbf{Y}\mathbf{G}^{-1}$, não havendo necessidade da introdução de Δ. Ou seja, se o modelo ajustado às respostas observadas é um polinômio de grau máximo, então a solução será única. Tem-se que:

$$E(\mathbf{Y}^{*}) = \mathbf{X}\xi$$

$$\Sigma^{*} = (\mathbf{G}\Sigma^{-1}\mathbf{G}')^{-1}$$

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

5.4.2 Proposta de ajuste por covariância

A escolha arbitrária de Δ pode ser substituída por uma outra metodologia para estimar os parâmetros da matriz ξ, a qual consiste em considerar a informação adicional contida na estrutura de covariância.

Como os dados longitudinais apresentam dependência dentro dos vetores de observações, os modelos podem apresentar estruturas de covariâncias e este fato pode alterar a eficiência dos estimadores.

Considere, primeiramente, um problema geral de estimação de um vetor de parâmetros a partir de um conjunto de observações. Sejam $\mathbf{t}_{1\ (q\ x\ 1)}$ e $\mathbf{t}_{2\ (p\text{-}q\ x\ 1)}$ duas estatísticas vetoriais tais que:

$$E(\mathbf{t}_1) = \tau \qquad \text{e} \qquad E(\mathbf{t}_2) = \mathbf{0},$$

onde τ é um vetor de q parâmetros desconhecidos, que se deseja estimar.

Considere ainda que a matriz de variância-covariância Σ do vetor $\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$, (q+(p-q) x 1), seja não singular e conhecida e que seja particionada da seguinte forma:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

onde,

Var $(\mathbf{t}_1) = \Sigma_{11}$ (não singular), Var $(\mathbf{t}_2) = \Sigma_{22}$ (não singular) e Cov $(\mathbf{t}_1,\mathbf{t}_2) = \Sigma_{12}$.

Tem-se que a estatística $\mathbf{t}_1$ é o estimador não viciado de τ e tem matriz de variância-covariância não singular Σ_{11}.

Se Cov $(\mathbf{t}_1,\mathbf{t}_2) \neq 0$ podemos determinar τ^* como estimador não viciado de τ e mais eficientes que $\mathbf{t}_1$ (no sentido de variância mínima). Considere o seguinte estimador de τ:

$$\tau^* = \mathbf{t}_1 - \Sigma_{12}\, \Sigma_{22}^{-1}\, \mathbf{t}_2,$$

sendo,

$$E(\tau^*) = \tau \text{ (não viciado)}$$

$$\text{Var } (\tau^*) = \Sigma_{11} - \Sigma_{12}\, \Sigma_{22}^{-1}\, \Sigma_{21}$$

Comparando-se as Variâncias de τ^* e $\mathbf{t}_1$, tem-se:

Var $(\mathbf{t}_1)$ - Var $(\tau^*) = \Sigma_{12}\, \Sigma_{22}^{-1}\, \Sigma_{21}$, uma matriz não negativa definida. Portanto, tem-se a seguinte relação nas formas quadráticas:

$$\mathbf{a}'\text{Var } (\mathbf{t}_1)\, \mathbf{a} \geq \mathbf{a}'\text{Var } (\tau^*)\, \mathbf{a},$$

para qualquer **a**. Ou seja, o estimador τ^* é mais eficiente que $\mathbf{t}_1$.

Considerando $\hat{\Sigma}$ a estimativa da matriz de variância-covariância (Σ), dada por:

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{bmatrix}, \text{ tem-se}$$

$$\hat{\tau} = t_1 - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}t_2 ,$$

que é o estimador ajustado por covariância.

Se $\hat{\Sigma}_{ij}$ for distribuída independentemente de $\mathbf{t}_i$ (i, j = 1, 2), tem-se:

$$E(\hat{\tau}) = E(\mathbf{t}_1) - E(\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1})\, E(\mathbf{t}_2) = E(\mathbf{t}_1) = \boldsymbol{\tau},$$

$$\text{Var}(\hat{\tau}) = \Sigma_{11} + E[\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\Sigma_{22}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}] - E[\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\Sigma_{21}] - E[\Sigma_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}],$$

onde as esperanças são consideradas em relação a $\hat{\Sigma}_{ij}$.

Sob estas circunstâncias, considere as observações:

- $\Sigma_{21} = \mathbf{0} \Leftrightarrow \mathbf{t}_1$ é mais eficiente que $\boldsymbol{\tau}^*$ (espera-se que Var($\mathbf{t}_1$) – Var ($\hat{\tau}$) seja negativa definida);
- $\Sigma_{21} \neq \mathbf{0} \Leftrightarrow \boldsymbol{\tau}^*$ é mais eficiente que $\mathbf{t}_1$ (espera-se Var($\mathbf{t}_1$) – Var ($\hat{\tau}$) seja não negativa definida).

A decisão da aplicação da técnica depende da covariância entre as variáveis auxiliares e as variáveis principais, pois há possibilidade de encontrar um estimador que resulte num decréscimo de eficiência. Mesmo que $\mathbf{t}_2$ não se constitua na escolha ótima das variáveis auxiliares para o ajuste por covariância, é possível que uma seleção conveniente de elementos de $\mathbf{t}_2$ como variáveis auxiliares possa fornecer estimadores ajustados mais eficientes.

O conhecimento da estrutura de Σ ou uma estimativa da mesma podem fornecer uma aproximação acurada da escolha ótima.

No estudo das distribuições dos estimadores ajustados por covariância devem ser consideradas a distribuição normal de:

$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \sim N_{q+(p-q)} \left[\begin{pmatrix} \tau \\ 0 \end{pmatrix}, \Sigma \right]$$ e a distribuição de Wishart para $f\hat{\Sigma} \sim W_{q+(p-q)}$ (f, Σ, .) central.

Demonstra-se que as distribuições condicionais de $\hat{\tau} = \mathbf{t}_1 - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{t}_2$ dado t_2 e $\hat{\Sigma}_{11}^{*} = \hat{\Sigma}_{11} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$ dada $\hat{\Sigma}_{22}$ são independentes e são dadas por:

$$\hat{\tau} \sim \mathrm{N_q}\,[\tau,\ \Sigma_{11}^{*}(1+T_{p-q}^{2}\ )]$$

$$\hat{\Sigma}_{11}^{*} \sim W_q(f-(p-q),\Sigma_{11}^{*},.)\ \text{central}$$

onde

$$\Sigma_{11}^{*}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\ \text{e}\ T_r^2=\frac{1}{f}t_2\hat{\Sigma}_{22}^{-1}t_2\,.$$

Além disso, pode-se demonstrar(Rao, 1973) que:

$$\frac{f-(p-q)+1}{2}T_{p-q}^{2}\ \sim F(p-q,\,f-(p-q)+1).$$

Para a estimação com ajuste por covariâncias em modelos lineares multivariados, considere $\mathbf{Y}_{1(nxq)}$ como MLM ($\mathbf{Y}$, $\mathbf{X}\beta$, Σ_{11}) e uma matriz de observações auxiliares $\mathbf{Y}_2$, de ordem (n x p-q) para a qual admitem-se válidas as suposições do MLM ($\mathbf{Y}_2$, $\mathbf{0}$, Σ_{22}).

A matriz geral de observações para contemplar as variáveis auxiliares pode ser representada por $[\mathbf{Y}_1\ \ \mathbf{Y}_2]_{nxp}$ para a qual tem-se que MLM ($[\mathbf{Y}_1\ \ \mathbf{Y}_2]$, $\mathbf{X}\,[\beta\ \ \mathbf{0}]$, Σ). A distribuição condicional de $\mathbf{Y}_1$ dadas as observações da variável auxiliar$\mathbf{Y}_2$, estabelece:

$\mathrm{E}\,(\mathbf{Y}_1 \mid \mathbf{Y}_2) = \mathbf{X}\beta + \mathbf{Y_2}\Sigma_{22}^{-1}\Sigma_{21}$ e

$\mathrm{Var}(\mathbf{Y}_1 \mid \mathbf{Y}_2) = \Sigma_{11}^{*} = \Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21}$.

Fazendo $\mathbf{D} = [\mathbf{X}\ \ \mathbf{Y}_2]$, $\eta = \mathbf{Y}_2\Sigma_{22}^{-1}\ \Sigma_{21}$ e $\gamma = \begin{bmatrix}\beta\\ \eta\end{bmatrix}$, o modelo condicional acima é equivalente ao MLM ($\mathbf{Y}$, $\mathbf{D}\gamma$, Σ_{11}^{*}), e um estimador não viciado de γ é dado por:

$$\hat{\gamma} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{Y}_1.$$

Singer (1977) demonstra detalhadamente o procedimento de estimação de parâmetros por covariância e obtém o estimador de $\mathbf{C}\beta$, $\mathbf{C}\beta\mathbf{U}$ e suas respectivas variâncias dadas por:

$$\mathbf{C}\,\hat{\beta} = \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y_1} - \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\,\mathbf{Y}_2\mathbf{S}_{22}^{-1}\mathbf{S}_{21};$$

$$\mathrm{Var}\,(\mathbf{C}\,\hat{\beta}\,) = \frac{(n-g)-1}{(n-g)-(p-q)-1}\Sigma_{11}^{*}\otimes[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'];$$

$$\mathbf{C}\,\hat{\beta}\,\mathbf{U} = \mathbf{C}[(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y_1} - (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}_2\mathbf{S}_{22}^{-1}\mathbf{S}_{21}]\mathbf{U};$$

$$\text{Var}\,(\mathbf{C}\,\hat{\beta}\,\mathbf{U}) = \frac{(n-g)-1}{(n-g)-(p-q)-1}[\mathbf{U}'\boldsymbol{\Sigma}_{11}{}^{*}\mathbf{U}] \otimes [\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'].$$

5.4.3 Proposta de Rao

Rao (1965 e 1966) considera as sugestões descritas por Potthoff & Roy (1964) frágeis, pois na prática, Σ é desconhecida. Não obstante, demonstra que o estimador apresentado por estes autores não aproveita as informações contidas nas (p-q) variáveis desprezadas e propõe o estimador de mínimos quadrados com ajuste por covariâncias para contornar o problema da arbitrariedade na escolha de Δ para transformar o Modelo Linear Multivariado de Crescimento no modelo usual de Análise de Covariância Multivariada, ou seja, inclusão de variáveis concomitantes.

Seja o MLMC ($\mathbf{Y}$, $\mathbf{X}\xi\mathbf{G}$, Σ), os estimadores de ξ, $\mathbf{C}\xi\mathbf{U}$ e Σ serão obtidos considerando-se $\mathbf{G}_{1(pxq)}$ e posto ($\mathbf{G}_1$) = q, $\mathbf{G}_{2\ (p\ x\ p\text{-}q)}$ e posto ($\mathbf{G}_2$) = p-q, tais que: $\mathbf{G}\mathbf{G}_1 = \mathbf{I}_{qxq}$ e $\mathbf{G}\mathbf{G}_2 = \mathbf{0}_{(q\ x\ (\ p\text{-}q))}$.

Considere a transformação,

$$\mathbf{Y}[\mathbf{G}_1 \ \ \mathbf{G}_2] = [\mathbf{Y}\mathbf{G}_1 \ \ \mathbf{Y}\mathbf{G}_2] = [\mathbf{Y}_{1(nxq)} \ \ \mathbf{Y}_{2(n\ x\ p\text{-}q)}].$$

Tomando-se $\mathbf{G}_1 = \mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}$ e $\mathbf{G}_2$ com as linhas sendo uma base gerada pelas colunas de $[\mathbf{I} - \mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}\mathbf{G}]$ as condições iniciais estarão satisfeitas e dois conjuntos independentes de variáveis ($\mathbf{Y}_1$ e $\mathbf{Y}_2$) são obtidos. Em síntese, $\mathbf{Y}_2$ não fornece qualquer informação sobre $\mathbf{Y}_1$ e pode ser desprezada para finalidade de inferências sobre ξ.

A transformação permite concluir que:

$$\text{E}\,(\mathbf{Y}_1) = \mathbf{X}\xi\mathbf{G}\mathbf{G}_1 = \mathbf{X}\xi \ \text{ e } \ \text{E}\,(\mathbf{Y}_2) = \mathbf{X}\xi\mathbf{G}\mathbf{G}_2 = \mathbf{0},$$

ou seja, $\mathbf{Y}_1$ é formada por q vetores pertencentes ao espaço de estimação, e $\mathbf{Y}_2$ por (p-q) vetores pertencentes ao espaço de erros. Sob essas considerações, é possível obter um estimador de ξ mais eficiente do que aquele obtido diretamente do modelo E ($\mathbf{Y}$) = $\mathbf{X}\xi$ (Potthoff & Roy, 1964) com $\Delta = \mathbf{I}$. Na transformação imposta aos dados, as colunas de $\mathbf{Y}_1$ são referidas como variáveis principais e as de $\mathbf{Y}_2$ como variáveis concomitantes. Portanto, se as covariâncias em ($\mathbf{G}_1'\Sigma\mathbf{G}_2$) forem diferentes de zero, inferências sobre ξ realizadas com base no modelo condicional de $\mathbf{Y}_1$ dado $\mathbf{Y}_2$ podem conduzir a estimativas mais precisas.

Dessa forma, o modelo linear que pode ser definido em função de $\mathbf{Y}_1$ e $\mathbf{Y}_2$ é expresso por:

$$[\mathbf{Y}_1 \quad \mathbf{Y}_2] = \mathbf{X}\,[\xi \;\; \Delta] + \varepsilon_1.$$

O MLMC ($\mathbf{Y}$, $\mathbf{X}\xi\mathbf{G}$, Σ) será adequado, se para um teste estatístico de ajuste do modelo inicial configurar $\Delta = \mathbf{0}$.

Sob esta hipótese, pode-se obter o MLM ($[\mathbf{Y}_1\ \mathbf{Y}_2]$, $\mathbf{X}[\xi\ \mathbf{0}]$, Σ_1), onde:

$$\text{E}\,[\mathbf{Y}_1 \quad \mathbf{Y}_2] = \mathbf{X}\,[\xi \;\; \mathbf{0}] \text{ e Var } [\mathbf{Y}_1 \quad \mathbf{Y}_2] = \Sigma_1 \otimes \mathbf{I}_n,$$

com $\Sigma_1 = \begin{bmatrix} G_1' \\ G_2' \end{bmatrix} \Sigma\,[\mathbf{G}_1 \quad \mathbf{G}_2] = \begin{bmatrix} G_1'\Sigma G_1 & G_1'\Sigma G_2 \\ G_2'\Sigma G_1 & G_2'\Sigma G_2 \end{bmatrix}$.

Considere a distribuição condicional de $\mathbf{Y}_1$ dado $\mathbf{Y}_2$. Supondo a distribuição $\mathbf{Y}_1|\mathbf{Y}_2$ como normal multivariada e o modelo condicional de $\mathbf{Y}_1$ dado $\mathbf{Y}_2$ é representado por:

$$\text{MLM(cond)}\ (\mathbf{Y}_1|\mathbf{Y}_2;\ [\mathbf{X} \quad \mathbf{Y}_2] \begin{bmatrix} \xi \\ \eta \end{bmatrix}, \Sigma_2) = \text{MLM}\ (\mathbf{Y}_1|\mathbf{Y}_2;\ \mathbf{X}\xi + \mathbf{Y}_2\eta;\ \Sigma_2)$$

onde,

$\eta = (\mathbf{G}_2'\Sigma\mathbf{G}_2)^{-1}\,(\mathbf{G}_2'\Sigma\,\mathbf{G}_1)$ é a matriz de parâmetros das variáveis concomitantes, e

$\Sigma_2 = \mathbf{G}_1'\Sigma\mathbf{G}_1 - \mathbf{G}_1'\Sigma\,\mathbf{G}_2\,(\mathbf{G}_2'\Sigma\mathbf{G}_2)^{-1}\,\mathbf{G}_2'\Sigma\,\mathbf{G}_1$.

Ou seja,

$$\text{E}\,[\mathbf{Y}_1|\mathbf{Y}_2] = \mathbf{X}\xi + \mathbf{Y}_2\,(\mathbf{G}_2'\Sigma\mathbf{G}_2)^{-1}(\mathbf{G}_2'\Sigma\mathbf{G}_1) \text{ e}$$

$$\text{Var}\,[\mathbf{Y}_1|\mathbf{Y}_2] = \Sigma_2 \otimes \mathbf{I}_n.$$

Retomando, agora, $\mathbf{D} = [\mathbf{X} \quad \mathbf{Y}_2]$ e $\gamma = \begin{bmatrix} \xi \\ \eta \end{bmatrix}$, tem-se $\text{E}\,[\mathbf{Y}_1|\mathbf{Y}_2] = \mathbf{D}\gamma$, e assim é possível representar o modelo MLM(cond) $(\mathbf{Y}_1|\mathbf{Y}_2;\ [\mathbf{X} \quad \mathbf{Y}_2] \begin{bmatrix} \xi \\ \eta \end{bmatrix}, \Sigma_2)$ através do MLM $(\mathbf{Y}_1, \mathbf{D}\gamma, \Sigma_2)$.

Portanto, o estimador de mínimos quadrados de γ será dado por:

$$\hat{\gamma} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{Y}_1.$$

Utilizando-se a técnica da obtenção da inversa de uma matriz particionada (Godoi, 1985), tem-se o estimador não viciado de ξ expressado por:

$$\hat{\xi} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}\mathbf{G}_1 - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{G}_2(\mathbf{G}_2'\mathbf{S}\mathbf{G}_2)^{-1}\mathbf{G}_2'\mathbf{S}\mathbf{G}_1,$$

sendo o primeiro fator corresponde às variáveis principais e o segundo às concomitantes. A variância do estimador pode ser escrita como $\mathrm{Var}(\hat{\xi}) = \dfrac{(n-g)-1}{(n-g)-(p-q)-1}\Sigma_2 \otimes (\mathbf{X}'\mathbf{X})^{-1}$.

Considerando-se o lema de Khatri, o qual descreve que:

"Sendo as matrizes $\mathbf{G}_{qxp}$ e $\mathbf{G}_{2(px\ p-q)}$, de postos q e p – q, respectivamente, tais que $\mathbf{G}\mathbf{G}_2 = \mathbf{0}$, então se S_{pxp} for uma matriz simétrica e positiva definida, tem-se:

$$\mathbf{G}_2\,(\mathbf{G}_2'\mathbf{S}\mathbf{G}_2)^{-1}\mathbf{G}_2' = \mathbf{S}^{-1} - \mathbf{S}^{-1}\,\mathbf{G}'(\mathbf{G}\mathbf{S}^{-1}\mathbf{G}')^{-1}\mathbf{G}\mathbf{S}^{-1}",$$

obtém-se as seguintes expressões para $\hat{\xi}$, Σ_2 e Var ($\hat{\xi}$):

$$\hat{\xi} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}\,\mathbf{S}^{-1}\,\mathbf{G}'(\mathbf{G}\mathbf{S}^{-1}\mathbf{G}')^{-1};$$

$$\Sigma_2 = (\mathbf{G}\Sigma^{-1}\mathbf{G}')^{-1} \text{ e}$$

$$\hat{Var}(\hat{\xi}) = \frac{(n-g)-1}{(n-g)-(p-q)-1}\ (\mathbf{G}\Sigma^{-1}\mathbf{G}')^{-1} \otimes (\mathbf{X}'\mathbf{X})^{-1},$$

independentes de $\mathbf{G}_1$ e $\mathbf{G}_2$.

Neste sentido, um estimador não viciado de $\mathbf{C}\xi\mathbf{U}$ é dado por:

$$\mathbf{C}\,\hat{\xi}\,\mathbf{U} = \mathbf{C}[(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}\,\mathbf{S}^{-1}\,\mathbf{G}'(\mathbf{G}\mathbf{S}^{-1}\mathbf{G}')]\mathbf{U},$$

e sua matriz de variância-covariância pode ser representada por:

$$\mathrm{Var}\,(\mathbf{C}\,\hat{\xi}\,\mathbf{U}) = \frac{(n-g)-1}{(n-g)-(p-q)-1}\ [\mathbf{U}'(\mathbf{G}\Sigma^{-1}\mathbf{G}')^{-1}\,\mathbf{U}] \otimes [\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'].$$

À semelhança do MLM (Y, Xβ, Σ), onde o estimador não-viciado de Σ é dado por:

$$\hat{\Sigma} = \frac{1}{n-g}\ \mathbf{Y}'[\mathbf{I} - \mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y} = \frac{1}{n-g}\mathbf{S},$$

sendo sua distribuição proporcional a W_p (n – g, Σ, **.**) central, tem-se para o modelo condicional MLM(cond) ($\mathbf{Y}_1$, $\mathbf{D}\gamma$, Σ_2) que:

$$\hat{\Sigma}_2 = \frac{1}{(n-g)-(p--q)}S_1, \text{ com } \mathbf{S}_1 = \mathbf{Y}_1'[\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{Y}_1 \text{ e } \mathbf{S}_1 \text{ com}$$

distribuição W_q [(n-g) – (p-q), Σ_2,**.**] central.

Neste sentido, estabelece-se um estimador não viciado da matriz de variância-covariância do estimador de $\hat{\xi}$ dado por:

$$\hat{V}ar(\hat{\xi}) = \frac{(n-g)-1}{(n-g)-(p-q)-1}\ \hat{\Sigma}_2 \otimes (\mathbf{X}'\mathbf{X})^{-1},$$

ou, equivalentemente,

$$\hat{V}ar(\hat{\xi}) = \frac{(n-g)-1}{(n-g)-(p-q)-1}\ \frac{1}{(n-g)-(p-q)} (\mathbf{G}\mathbf{S}^{-1}\mathbf{G}')^{-1} \otimes (\mathbf{X}'\mathbf{X})^{-1}.$$

Soler (1997) destaca que vários autores que têm estabelecido equivalências entre a classe de estimadores de ξ resultantes do procedimento de Rao e da proposta de Potthoff & Roy. Descreve que há possibilidade de reduzir um estimador ao outro com uma escolha adequada da matriz de ponderação Δ. A proposta de Rao é mais abrangente no que se refere ao procedimento de escolha dessas matrizes de ponderação, além de permitir o julgamento do quanto a inclusão de tais matrizes pode melhorar o ajuste do modelo.

Após um estudo detalhado dessa medida de eficiência do estimador ajustado por covariáveis relativamente ao estimador não ponderado, Soler (1997) conclui que essa relação depende do número de variáveis concomitantes, que por sua vez, está relacionada com o coeficiente de correlação destas variáveis com as principais. Ou seja, um ganho na eficiência pode ser obtido quando o número de variáveis concomitantes puder ser diminuído, sem prejuízo da quantidade de informação contida nas covariáveis. Na prática, a eficiência do estimador ajustado é função da covariância entre as variáveis auxiliares ($\mathbf{Y}_2$) e as variáveis principais ($\mathbf{Y}_1$), ou seja, expressa em $\mathbf{G}_1'\Sigma\mathbf{G}_2$.

O procedimento de escolha de variáveis concomitantes ou auxiliares ou covariáveis é realizado considerando-se a matriz de correlações entre as variáveis do espaço de estimação e o espaço de erros (Grizzle & Allen, 1969). Ou seja, a partir da matriz $\mathbf{G}_{pxp}$, duas matrizes serão consideradas $\mathbf{G}_{1\ (pxq)}$ e $\mathbf{G}_{2\ (px(p-q))}$, sendo que na primeira estarão contidas as linhas correspondentes ao grau do polinômio ajustado (graus de baixa ordem) e na segunda matriz, as linhas dos graus descartados (graus de ordem mais alta). Considere ainda que $r(\mathbf{G}_1) = q$, $r(\mathbf{G}_2) = p-q$, $\mathbf{G}\mathbf{G}_1 = I_q$ e $\mathbf{G}\mathbf{G}_2 = \mathbf{0}_{(qx(p-q))}$. Estas condições estarão satisfeitas tomando-se $\mathbf{G}_1 = \mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}$ e $\mathbf{G}_2$ como sendo uma base do espaço coluna de $\mathbf{I} - \mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}\mathbf{G}$.

Souza (1989) descreve um teste estatístico para o estudo conjunto de m coeficientes de correlação baseando-se na desigualdade de Bonferroni e a transformação de Fisher.

Ou seja,

H_0: $\rho_{kk'} = 0$ "versus" H_1: $\rho_{kk'} \neq 0$ para $k \neq k'$,

com a seguinte regra de decisão: se $|r_{kk'}| > \frac{z_{\frac{\alpha}{2m}}}{\sqrt{n-3}}$, rejeita-se H_0, onde $r_{kk'}$ é a correlação linear de Pearson estimado entre a k-ésima variável principal e k´-ésima variável concomitante.

O teste múltiplo utilizado verifica a inclusão ou não de variáveis concomitantes a partir de tanh ($\frac{z}{\sqrt{n-3}}$), onde z é tal que P(Z > z) = $\frac{\alpha}{2m}$ sendo m o número de correlações entre $\mathbf{Y}_1$ e $\mathbf{Y}_2$. Se nenhuma das correlações entre as variáveis for maior que o valor obtido, então o modelo considerado pode ser construído sem covariável, ou seja, o modelo proposto por Potthoff & Roy (1964). Porém, quando as variáveis apresentarem correlação significativa, estas devem ser consideradas como variáveis concomitantes, pois tornam o estimador mais eficiente.

Sob as considerações de Rao (1965 e 1966) ao MLM condicional, considere $\mathbf{Y}_1$ a matriz das q variáveis principais, $\mathbf{Y}_2$ como o espaço das ν concomitantes e $\mathbf{Y}_3$ como as (p – q – ν) variáveis serão desprezadas do modelo. Considere $\mathbf{D} = [\mathbf{X} : \mathbf{Y}_2]$ e para determinar $\hat{\xi}$ como sendo as g primeiras linhas de $\hat{\gamma}$, sendo:

$$\hat{\gamma} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{Y}_1;$$

$$\hat{\xi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}[\mathbf{G}_1 - \mathbf{G}_2(\mathbf{G}_2'\mathbf{S}\mathbf{G}_2)^{-1}\mathbf{G}_2'\mathbf{S}\mathbf{G}_1] \text{ e}$$

$$\hat{Var}(\hat{\xi}) = \frac{n-g-1}{(n-g)-\nu-1} * \frac{1}{(n-g)-\nu} *[\mathbf{G}_1'\mathbf{S}\mathbf{G}_1 - \mathbf{G}_1'\mathbf{S}\mathbf{G}_2(\mathbf{G}_2'\mathbf{S}\mathbf{G}_2)^{-1}\mathbf{G}_2'\mathbf{S}\mathbf{G}_1]\otimes (\mathbf{X}'\mathbf{X})^{-1},$$

onde ν é o número de variáveis concomitantes (= número de colunas de $\mathbf{Y}_2$).

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

$\mathbf{E} = \mathbf{U}'\mathbf{Y}_1'[\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{Y}_1\mathbf{U}$,

$\mathbf{H} = (\mathbf{C}\hat{\xi}\mathbf{U} - \Gamma)'(\mathbf{C}\mathbf{R}\mathbf{C}')^{-1}(\mathbf{C}\hat{\xi}\mathbf{U} - \Gamma)$,

onde, $\mathbf{R}$ é o menor principal de $(\mathbf{D}'\mathbf{D})^{-1}$, sendo:

$\mathbf{R}_{gxg} = (\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\,[\mathbf{S}^{-1} - \mathbf{S}^{-}1\mathbf{G}1'(\mathbf{G}_1\mathbf{S}^{-1}\mathbf{G}_1')^{-}1\mathbf{G}1\mathbf{S}^{-1}]\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ proposto por Khatri (1966), ou ainda, por Morrison (1976),

$\mathbf{R}_{gxg} = (\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{S}^{-1}\,\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - \hat{\xi}(\mathbf{G}_1\mathbf{S}^{-1}\mathbf{G}_1')\hat{\xi}$.

A hipótese $H_{0:}\ C\xi U = \Gamma$ pode ser testada por alguma das estatísticas apresentadas em 5.3. Os parâmetros utilizados serão:

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

$m_1 = \frac{|c - u_1 - 1|}{2}$ e

$m_2\ \frac{n - g - (p - q) - u_1 - 1}{2}$.

5.5 Desenvolvimento proposto

Considerando-se as adversidades dos experimentos agronômicos, no presente estudo serão contemplados conjuntos de respostas longitudinais regulares (as condições de avaliação quanto ao intervalo entre duas medidas consecutivas devem ser constantes); balanceados (as medidas devem ser obtidas nos mesmos instantes de avaliação em todas as unidades de avaliação) e completos (sem falhas de observação).

Uma das dificuldades encontrada no estudo de Curvas de Crescimento é o procedimento analítico dos dados, pois os modelos estabelecidos para o estudo em discussão impõem operações matriciais complexas.

5.5.1 Modelo de curvas de crescimento polinomial

No estudo longitudinal, o modelo de crescimento deve descrever adequadamente o comportamento da variável ao longo dos momentos observados.

Fitzhugh Jr. (1976) cita os seguintes critérios para a comparação de modelos de crescimento:

a) o melhor ajuste dos dados, que se deve ao menor desvio residual (aderência);
b) convergência dos dados (dificuldade computacional);

c) interpretação biológica dos parâmetros obtidos para cada modelo.

Uma pressuposição para o estudo de ajuste de curvas polinomiais é considerar que as matrizes de covariâncias dos grupos experimentais não se diferem entre si. A homocedasticia é comprovada pelo Teste de Bartlet Generalizado (Morrison, 1976), considerando-se a seguinte hipótese da nulidade:

H_0: $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_g$

A estatística utilizada no teste é dada por MC^{-1}, onde:

$$\mathbf{M} = (n - g) \ln | S | - \sum_{i=1}^{g} (n_i - 1) \ln | S_i |;$$

$n = n_1 + n_2 + \ldots + n_g$;

g = número de grupos;

$$\mathbf{S} = S_{POOL} = \frac{(n_1 - 1)S_1 + ... + (n_g - 1)S_g}{n - g};$$

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p + 1)(g - 1)} \left(\sum_{i=1}^{g} \frac{1}{n_i} - \frac{1}{\sum_{i=1}^{g} n_i} \right).$$

Sob a veracidade de H_0, $MC^{-1} \sim \chi^2_{[\frac{1}{2}(g-1)p(p+1)]}$, onde χ^2 representa o quantil de ordem 100(1 - α)% da distribuição Qui-quadrado com parâmetro $\frac{1}{2}(g - 1)p(p + 1)$ graus de liberdade. A regra de decisão do teste estatístico é a habitual.

As matrizes de covariâncias dos grupos são dadas por $\mathbf{S}_i = [\, s_{jj}^{(i)} \,]$, onde,

$s_{jj}^{(i)} = s_j^2 = \frac{\sum_{i=1}^{n_i} x^2_{ij} - n_i * \bar{x}^2_{.j}}{n_i - 1}$ (Variância da j-ésima variável resposta, para j = 1, ..., p, relativa ao i-ésimo grupo, i = 1, ..., g).

$$s_{jj'}^{(i)} = \frac{\sum_{i=1}^{n_i} x_{ij} * x_{ij'} - n_i * \overline{x_{.j}} * \overline{x_{.j'}}}{n_i - 1}$$ (Covariância entre j-ésima e j´-ésima variável resposta, relativa ao i-ésimo grupo, i = 1, ..., g).

O resultado do Teste de Bartlet dos grupos experimentais será apresentado ao usuário por meio do P-valor, que decidirá se a análise dos dados prosseguirá ou não.

O modelo proposto será:

$$\beta_{ik} = \xi_{i0} + \xi_{i1}\, t_k + \xi_{i2}\, t_k^2 + \ldots + \xi_{iq-1}\, t_k^{q-1}$$

onde t_k corresponde ao valor dos momentos do planejamento experimental.

Sabendo-se que as observações foram avaliadas ao longo de p momentos, o grau máximo do polinômio será λ ($\lambda \leq$ p-1).

O estimador de máxima verossimilhança (o qual coincide com o estimador de mínimos quadrados) de ξ, sob as considerações do modelo em estudo, será:

$$\hat{\xi} = (\mathbf{X'X})^{-1}\, \mathbf{X'Y}\, \mathbf{S}^{-1}\, \mathbf{G'}(\mathbf{GS}^{-1}\mathbf{G'})^{-1}.$$

O estimador da matriz de variância-covariância de $\hat{\xi}$ será:

$$\hat{Var}(\hat{\xi}) = \frac{(n-g)-1}{(n-g)-(p-q)-1} * \frac{1}{(n-g)-(p-q)} * (GS^{-1}G')^{-1} \otimes (X'X)^{-1}.$$

Sob as considerações do modelo condicional proposto por Rao, tem-se, basicamente, têm-se três situações com relação a utilização da estrutura das variáveis concomitantes:

i) não usar variáveis concomitantes, conduzindo ao estimador não ponderado;

ii) considerar somente um subconjunto de $\mathbf{Y}_2$, ou seja, somente aquelas que tiverem um resultado significativo na seleção;

iii) considerar todas as variáveis de $\mathbf{G}_2$ como concomitantes, conduzindo ao estimador de máxima verossimilhança.

No primeiro caso, reporta-se ao caso de Potthoff & Roy, com $\Delta = \mathbf{I}$, onde informações contidas nos dados podem estar sendo ignoradas, pois as variáveis associadas aos graus mais elevados dos polinômios estão sendo negligenciadas. No segundo, quando somente alguns vetores de $\mathbf{Y}_2$ serão selecionados como variáveis

concomitantes, basta eliminar da matriz $\mathbf{G}_2$, as colunas correspondentes àquelas variáveis cujas correlações não foram significativas.

A terceira situação trata-se de uma abordagem mais geral, que envolve todas as variáveis contidas em $\mathbf{G}_2$ (graus mais altos do polinômio), cujas correlações são significativas ou não como variáveis concomitantes.

Todos os casos conduzem a estimativas não viciadas da variância dos respectivos estimadores, diferenciando-se pela eficiência dos estimadores.

No programa computacional apresentado no presente estudo será considerada a terceira situação, todas as variáveis do espaço erro ($\mathbf{Y}_2$) serão consideradas como concomitantes.

5.5.2 Testes estatísticos

Sob as considerações descritas para o MLM (5.3), a hipótese linear geral é dada por:

$$H: \mathbf{C}\,\beta\,\mathbf{U} = \mathbf{0}$$

onde **C** e **U** são matrizes de constantes conforme o interesse do teste de hipóteses, com postos c e u, respectivamente.

5.5.2.1 Teste para o ajuste do MLMC

O primeiro teste que deve ser realizado é o do ajuste do modelo, sendo a hipótese de nulidade do teste escrita da forma geral:

H_0: $\mathbf{C}\beta\mathbf{U} = \varnothing$, com

$\mathbf{C} = \mathbf{I}$ e $\mathbf{U} = \mathbf{G}_2$ e $\beta = \xi\mathbf{G}$.

A partir da matriz **G**, considere as matrizes $\mathbf{G}_1$ e $\mathbf{G}_2$, definidas anteriormente, que serão empregadas no teste do ajustamento do modelo.

Para verificar se o conjunto de dados se ajusta a um modelo polinomial especificado, ou seja, se pode ser descrito por um polinômio de grau $l \leq p-1$, realiza-se o procedimento do “descarte”. Isto é, a hipótese relativa a $\mathbf{G}_2$ (H_0: $\mathbf{X}\xi\mathbf{G}\mathbf{G}_2 = \mathbf{0}$) é avaliada por

meio da hipótese linear geral H_0: $\mathbf{C\beta U} = \mathbf{0}$, com $\mathbf{C} = \mathbf{I}_{gxg}$ e $\mathbf{U} = \mathbf{G}_{2\ (px(p-q))}$. Utilizando-se o critério da Razão de Verossimilhança Generalizada de Wilks e verifica-se quais linhas de **G** serão descartadas, ou seja, as linhas referentes aos graus polinomiais mais altos (os especificados em $\mathbf{G}_2$).

Para a utilização desse critério, as matrizes de somas de quadrados e produtos cruzados devidas aos resíduos e hipótese são calculadas, respectivamente, por:

$\mathbf{E}_1 = \mathbf{G}_2'\mathbf{S}\,\mathbf{G}_2 \sim W\ (n - g, \mathbf{G}_2'\Sigma\mathbf{G}_2, \varnothing)$ central.

$\mathbf{H}_1 = \mathbf{G}_2'\mathbf{Y}'\mathbf{X}(\mathbf{X}'\,\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}\mathbf{G}_2 \sim W\ [g, \mathbf{G}'2\,\Sigma\mathbf{G}_2, \mathbf{G}'_2\,\xi^{*}{}'\mathbf{X}'\mathbf{X}\xi^{*}\mathbf{G}_2]$ não central.

A partir dessas duas matrizes, calcula-se Λ, que é expressa por:

$$\Lambda = \frac{|E_1|}{|H_1 + E_1|} = \prod_{\upsilon=1}^{s}(1-\theta_\upsilon) = \prod_{\upsilon=1}^{s}(1+\lambda_\upsilon)^{-1}.$$

Os parâmetros considerados são:

s = min (g, p-q);

$m_1 = \frac{|g-(p-q)|-1}{2}$ e

$m_2 = \frac{n-g-(p-q)-1}{2}$.

- Se s = 1, então,

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

- Se s = 2, então,

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

Para outros casos, mas quando o tamanho da amostra é grande, utiliza-se a aproximação de Bartlett representada por RV (Razão de Verossimilhança):

$$RV = -\ [n - g - \frac{1}{2}(p\text{-}q - g + 1)]\ \ln \Lambda \sim \chi^2\ [g^*(p\text{-}q)],$$

Dada a matriz **G** de ordem p x p, o resultado do procedimento do "descarte" será apresentado numa tabela, como a seguir e deve-se considerar a primeira linha desta matriz como zero no procedimento de descarte.

Tabela1. Resultados do procedimento estatístico de "descarte" do polinômio (grau mais alto).

Linhas de **G** utilizadas na construção de $\mathbf{G}_2$	Grau do polinômio cujo ajuste é testado	Valor da estatística do teste	Nível descritivo do teste (P-valor)
Última	p – 2	RV (p-2)	$P_{(p-2)}$
Duas últimas	p – 3	RV (p-3)	$P_{(p-3)}$
M	M	M	M
(q-1) últimas	1	RV (1)	$P_{(1)}$

A interpretação dos resultados da Tabela1 deverá ser enfocada sempre nos graus descartados da matriz **G**. Para a análise da Tabela1, deve-se observar o valor do nível descritivo para cada grau do polinômio. Se $P > \alpha$, o referido grau pode ser descartado. Se algum grau for considerado significativo ($P < \alpha$), tem-se o grau adequado do polinômio. Deve ser destacado que α refere-se ao nível de significância do teste (probabilidade de erro tipo I), fixado “a priori”.

O próximo passo é estimar $\hat{\xi}$ e $Var\hat{(}\hat{\xi})$, descritos no item 5.3..

A melhor opção para escolha do grau do polinômio é o de menor grau possível por causa da facilidade de interpretação biológica dos dados.

O programa computacional proposto no presente trabalho apresenta o modelo polinomial a partir do resultado deste procedimento de “descarte”, considerando-se o menor grau possível.

5.5.2.2 Teste de coincidência das curvas

A coincidência das curvas de crescimento é testada considerando-se a hipótese:

H_{02}: $\mathbf{C\xi U} = \mathbf{0}$, onde as matrizes de quadrados e produtos cruzados relativas aos resíduos e hipótese, são conforme descritos em por Rao (5.4.3):

$\mathbf{E}_2 = \mathbf{U'S_1\,U}$, onde $\mathbf{S_1 = (G_1\,S^{-1}\,G_1')^{-1}}$,

$\mathbf{H_2} = \mathbf{(C\,\hat{\xi}\,U)'\,(CRC)^{-1}\,(C\,\hat{\xi}\,U)}$, sendo,

$\mathbf{R_{gxg}} = \mathbf{(X'X)^{-1}} + \mathbf{(X'X)^{-1}X'YS^{-1}\,Y'X(X'X)^{-1}} - \hat{\xi}\,\mathbf{(G_1S^{-1}G_1')}\,\hat{\xi}$.

$$C_{(g-1)xg} = \begin{bmatrix} 1 & -1 & 0 & \Lambda & 0 \\ 1 & 0 & -1 & \Lambda & 0 \\ M & M & M & \Lambda & M \\ 1 & 0 & 0 & \Lambda & -1 \end{bmatrix} \quad \text{e} \quad \mathbf{U}_2 = \mathbf{I}_q.$$

Os parâmetros serão:

s = min (g-1, q);

$m_1 = \frac{|g-1-q|-1}{2}$ e

$m_2 = \frac{n-g-(p-q)-1}{2}$.

O critério utilizado para o teste de coincidências das curvas e dos coeficientes da curva polinomial de cada grupo será a Razão de Verossimilhança Generalizada de Wilks, descrito no item 5.2.1. Os resultados da análise serão apresentados na Tabela2, descrita a seguir:

Tabela2. Resultados do Teste de Coincidências das Curvas Polinomiais e dos coeficientes da curva polinomial.

Critério	Resultado da estatística	Distribuição da estatística	Nivel descritivo (P-valor)
Wilks	RV	$F_{[s(2m_1+s+1),s(2m_2+s+1)]}$	P_W

A regra de decisão dos testes estatísticos é a usual, ou seja, se $F_{calculado} > F_{crítico}$ ($P < \alpha$) rejeita-se a hipótese de coincidência das curvas polinomiais.

5.5.2.3 Medida de Adequacidade do Modelo (MAM)

Propõe-se verificar o ajuste das respostas experimentais ao polinômio que descreve a curva de crescimento adotado, a Medida de Adequacidade do Modelo (MAM), expressa pela razão entre a soma de quadrados devido ao ajuste polinomial e a soma de quadrados total dos dados observados, para cada um dos grupos experimentais.

Neste sentido, tem-se, para o grupo G_i, i = 1, ..., g, a seguinte expressão da Medida de Adequacidade:

$$\text{MAM} = 100 * \frac{SQDevido}{SQTotal}\%, \text{ com i = 1,..., g, com,}$$

$$\text{SQDev } (G_i) = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\sum_{k=1}^{p}\left(\hat{y}_{ijk} - \overline{y}_{i..}\right)^2 \text{ e SQTot } (G_i) = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\sum_{k}^{p}\left(y_{ijk} - \overline{\overline{y}_{i..}}\right)^2 .$$

5.5.2.4 Intervalos de confiança para contrastes de coeficientes polinomiais de grupos

O fato de uma hipótese de igualdade de efeitos médios de g grupos experimentais ser rejeitado, não significa que é possível identificar quais grupos diferem entre si, nem o momento em que isso ocorre. A simples rejeição de uma hipótese não é suficiente para uma conclusão mais satisfatória do efeito dos tratamentos ao longo dos momentos de avaliação.

Os intervalos de confiança bilaterais propiciam ao pesquisador o estudo comparativo do parâmetro entre os grupos experimentais e identificar o momento em que a possível diferença ocorrer.

Para o presente estudo, será utilizado o método de Scheffé para a construção dos intervalos de confiança para os contrastes:

$$\text{IC } (\xi_{i\lambda} - \xi_{i'\lambda}): \ (\hat{\xi}_{i\lambda} - \hat{\xi}_{i'\lambda}) \pm \sqrt{(g-1) * F_{(\alpha, g-1, n-g)} * \left(\frac{\hat{Var}(\hat{\xi}_{il})}{n_i} + \frac{\hat{Var}(\hat{\xi}_{i'l)}}{n_{i'}}\right)}, \text{i, i´= 1, ..., g.}$$

O intervalo de confiança será construído com 95% de confiança e a interpretação será considerada da seguinte forma: grupos com intervalos de confiança com valores comuns, não se diferem entre si, caso contrário, os parâmetros comparados são distintos.

5.6 Programa Computacional CRESYS

5.6.1 Descrição do programa

Na elaboração do programa computacional sobre a metodologia utilizou-se a linguagem Delphi, uma ferramenta para programação visual estruturada baseada

em Object Pascal, que por sua vez é uma linguagem muito robusta e fortemente indicada para o desenvolvimento de sistemas de grande porte. A plataforma utilizada no desenvolvimento foi a Windows XP, tornando assim, o sistema compatível com as plataformas Windows 98 e Windows XP, que são sistemas operacionais mais utilizados no momento, e de aplicação em diversas áreas científicas.

Para o armazenamento dos dados utilizou-se o banco de dados Firebird, que possui um excelente desempenho e pode armazenas um grande volume de dados. O Firebird é um Banco de Dados Relacional e de gratuito.

A instalação do programa computacional CRESYS é simples e o manual de instruções encontra-se no Apêndice 3. A análise de curvas de crescimento é realizada para no mínimo dois (dez, no máximo) grupos experimentais avaliados ao longo de no mínimo três momentos (dez, no máximo).

Os perfis médios dos grupos podem ser observados no gráfico construído a partir do banco de dados.

A partir da análise de homogeneidade das variâncias dos grupos (Teste de Bartlet), os resultados do procedimento do “descarte” poderão ser verificados e o próprio programa computacional selecionará o grau adequado ao polinômio a ser ajustado.

Na seqüência, as matrizes dos parâmetros estimados ($\hat{\xi}$), dos desvios-padrão e os intervalos de confianças dos parâmetros poderão ser obtidos e apresentados em forma de relatório.

5.6.2 Entrada de dados

O programa apresenta opção de entrada de dados via teclado. Após o fornecimento das informações, a análise é salva em banco de dados para que posteriormente o pesquisador possa abrir, editar valores, imprimir gráfico e realizar os cálculos novamente.

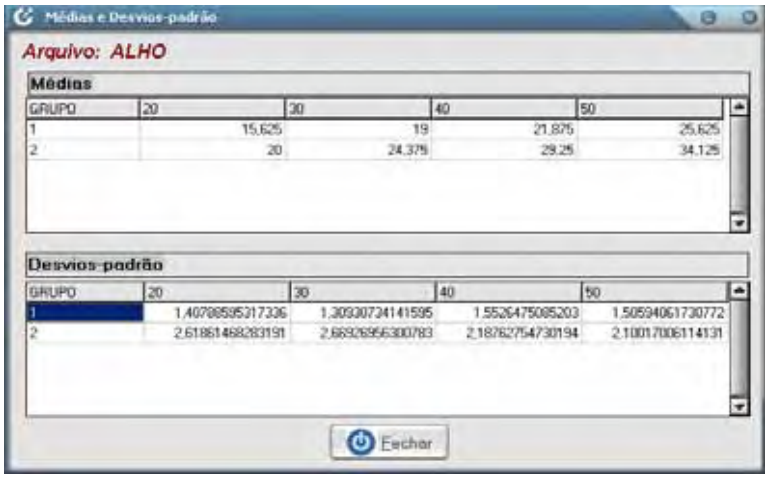
5.7 Exemplo ilustrativo

Foram utilizadas as alturas, em centímetros, de plantas de alho do tipo caçador, cultivadas em Santa Juliana, Minas Gerais. Dois grupos de plantas, testemunha e tratamento, com nove plantas em cada um, foram avaliados em quatro momentos distintos: aos 20, 30, 40 e 50 dias após a plantação. No grupo tratamento foi acrescentado adubo nitrogenado para a verificação de seu efeito na altura das plantas no período de 20 a 50 dias.

Os dados da altura das plantas de alho encontram-se no Apêndice 1.

6. RESULTADOS E DISCUSSÃO

Os resultados obtidos na análise dos dados das alturas das plantas de alho, em centímetros, pelo programa computacional Cresys, são apresentados a seguir. As médias e os respectivos desvios-padrão das alturas dos dois grupos podem ser visualizados na tela:



Esses valores podem ser organizados numa tabela como a apresentada a seguir:

Tabela3. Médias e desvios-padrão das alturas, em centímetros, das plantas de alho caçador dos grupos Testemunha e Tratamento, em quatro momentos.

Grupo	Momentos (horas)			
	20	30	40	50
Testemunha	15,63 ± 1,41	19,00 ± 1,31	21,98 ± 1,55	25,63 ±1,51
Tratamento	20,00 ± 2,62	24,38 ± 2,67	29,25 ± 2,19	34,16 ± 2,10

O comportamento médio da altura das plantas de alho de cada grupo experimental pode ser visualizado na figura a seguir:

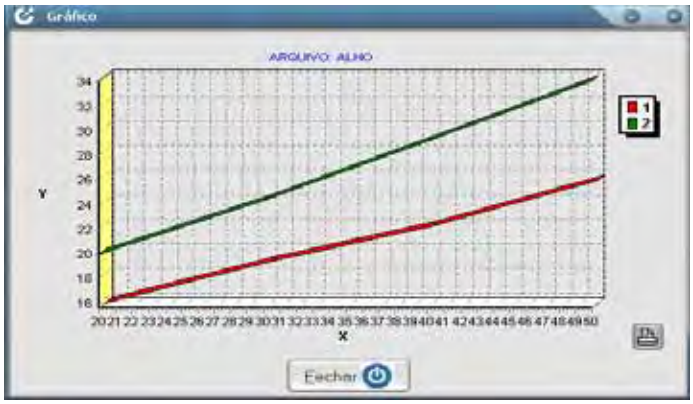


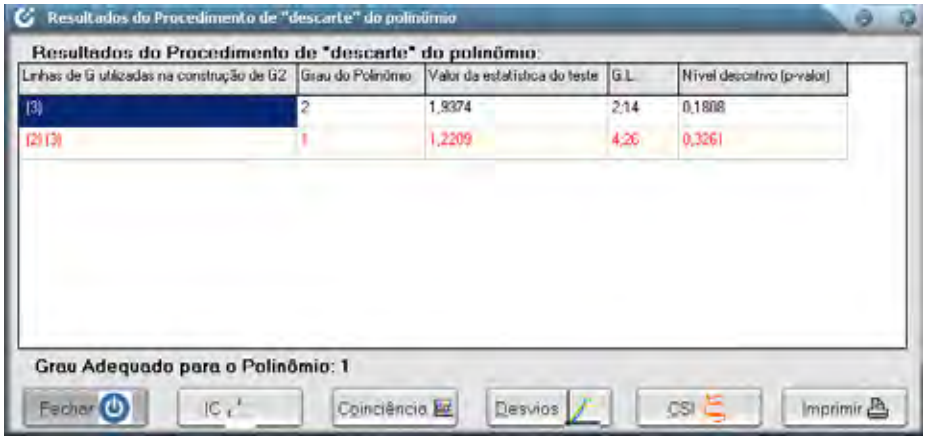
Figura 1. Curva de crescimento médio da altura das plantas alho do tipo Caçador, dos grupos Testemunha e Tratamento, em centímetros, aos 20, 30, 40 e 50 dias após a plantação.

O resultado do teste de Bartlet é apresentado na tela:

Com o resultado obtido, p-valor = 0,5798, não é rejeitada a hipótese da homocedastia, ou seja, as matrizes de covariâncias não se diferem.

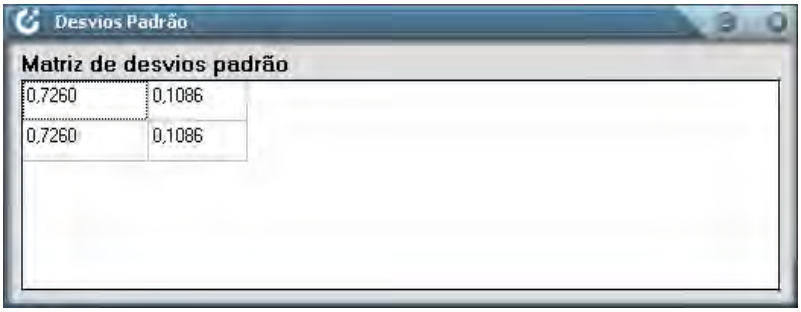
Prosseguindo-se com a análise estatística dos dados, obtem-se o grau adequado do polinômio das curvas de crescimento ajustadas aos dados observados.

Quadro3. Resultados do Procedimento de descarte do grau do polinômio da altura das plantas de alho.



Com os resultados obtidos no procedimento de descarte, modelo que melhor descreve as alturas médias das plantas é o linear.

Os valores estimados dos coeficientes do polinômio e os respectivos desvios-padrão estimadas de $\hat{\xi}$ são:



Esses parâmetros podem ser organizados numa tabela como a seguir:

Tabela 4. Estimativas de $\hat{\xi}$, e seus respectivos desvios-padrão, dos polinômios das curvas de crescimento que descrevem o comportamento médio da porcentagem de ganho de peso dos frutos de Araribá segundo o grupo.

Grupo	$\hat{\xi}_0$	$\hat{\xi}_1$	DP($\hat{\xi}_0$)	DP($\hat{\xi}_1$)
Testemunha	8,5585	0,3770	0,7260	0,1086
Tratamento	10,7358	0,4704	0,7260	0,1086

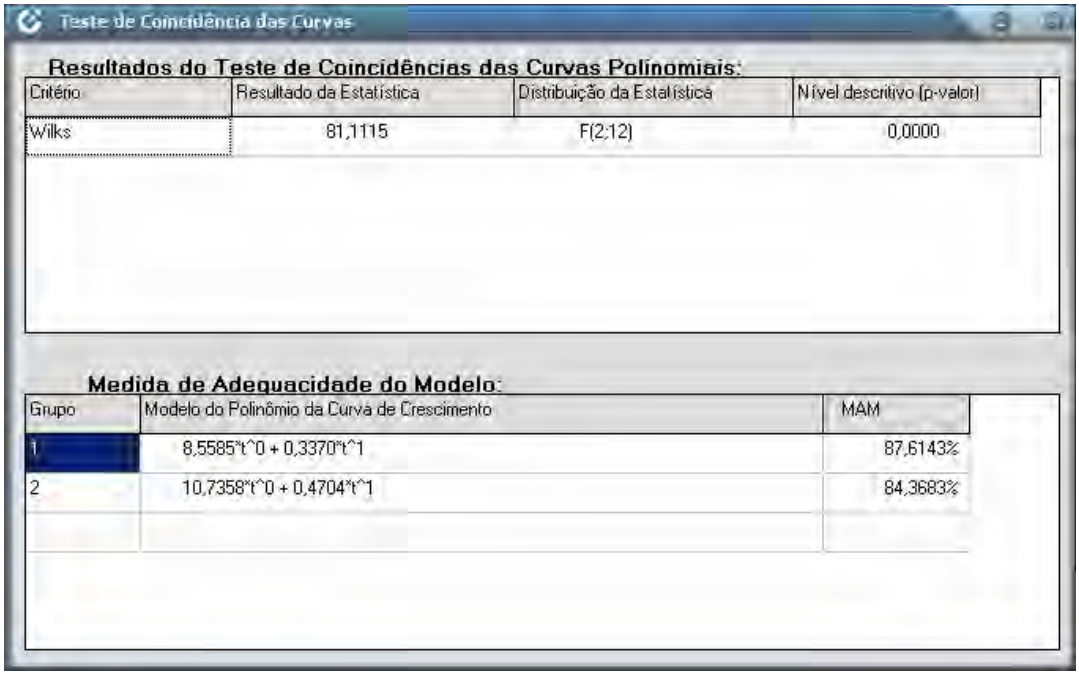
As curvas de crescimentos dos grupos são descritos pelos polinômios:

Grupo Testemunha: $\beta_1 = 8{,}5585 + 0{,}3770\ t$;

Grupo Tratamento: $\beta_2 = 10{,}7358 + 0{,}4704\ t$.

Conclui-se que para o período considerado, as plantas do grupo Testemunha cresceram, em média, 3,77 centímetros de uma avaliação para outra, enquanto que as plantas que receberam adubo nitrogenado, cresceram 4,70 centímetros, em média.

Os resultados do Teste das Coincidências das Curvas e as Medidas de Adequacidade do Modelo, para cada grupo e total, são apresentados pela tela:


Teste de Coincidência das Curvas

Resultados do Teste de Coincidências das Curvas Polinomiais:

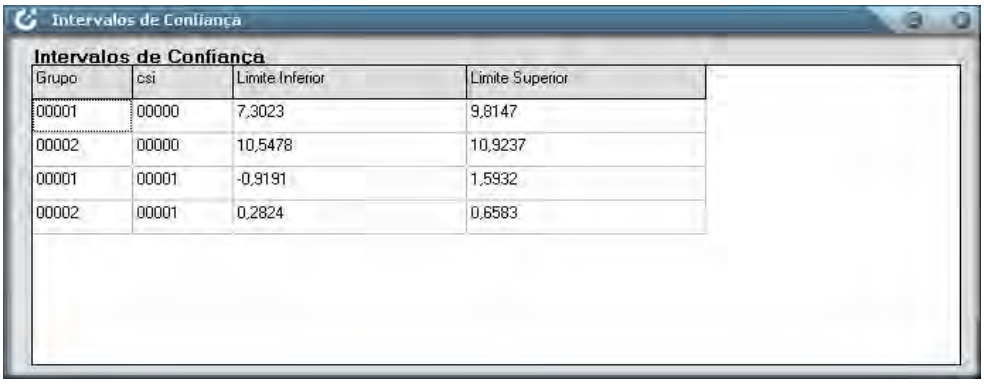
Critério	Resultado da Estatística	Distribuição da Estatística	Nível descritivo (p-valor)
Wilks	81,1115	F(2;12)	0,0000

Medida de Adequacidade do Modelo:

Grupo	Modelo do Polinômio da Curva de Crescimento	MAM
1	8,5585*t^0 + 0,3370*t^1	87,6143%
2	10,7358*t^0 + 0,4704*t^1	84,3683%

Esses resultados mostram que as curvas de crescimento que descrevem a altura média de cada grupo não são coincidentes, ou seja, as alturas médias das alturas das plantas de alhos são explicadas pela mesma família de função matemática, no caso linear, mas não pelo mesmo polinômio.

Os Intervalos de Confiança dos parâmetros estimados foram:


Intervalos de Confiança

Intervalos de Confiança

Grupo	csi	Limite Inferior	Limite Superior
00001	00000	7,3023	9,8147
00002	00000	10,5478	10,9237
00001	00001	-0,9191	1,5932
00002	00001	0,2824	0,6583

Os resultados dos Intervalos de Confiança mostram que os coeficientes lineares das duas restas se diferem, enquanto que os coeficientes angulares não se diferem, com 95% de confiança.

7. CONCLUSÕES

A revisão de literatura mostrou várias situações experimentais onde a estrutura de dependência existente nas parcelas é desconsiderada nos procedimentos estatísticos. A metodologia apresentada destaca a importância dessa estrutura não apenas no procedimento estatístico, mas também, para a melhoria da informação agronômica contida nos dados.

O ajuste de curvas de crescimento permite estabelecer uma relação funcional entre o efeito dos grupos experimentais (tratamentos) e o tempo (momentos de avaliação), que possibilita estimar esse efeito de tratamento para um momento não observado experimentalmente. As funções polinomiais permitem explicar, com consistência, o comportamento do conjunto de dados ao longo dos momentos quando existe adequação do modelo, simplifica a interpretação de resultados, possibilitando um maior conhecimento da variação da variável resposta no período experimental, propiciando interessantes conclusões práticas ao pesquisador.

O programa computacional CRESYS, de acesso livre e de fácil manuseio, realiza o estudo da homogeneidade dos grupos experimentais, condição fundamental para o ajuste de curvas de crescimento; representa graficamente o comportamento médio dos grupos experimentais nos momentos de avaliação; calcula os valores das médias e dos respectivos desvios-padrão; estabelece o grau adequado ao conjunto de dados observados por meio do “teste do descarte”; calcula os valores dos coeficientes do polinômio e os respectivos desvios-padrão; calcula a medida de adequacidade do modelo e verifica a possível coincidência das curvas ajustadas.

8. REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, L., BERGAMIN FILHO, A. Análise comparativa de curvas de progresso do carvão em quatro variedades de cana-de-açúcar com o uso de Análise de Variância Multivariada. **Fitopatol.Bras**, v.17, n. 1, p. 42-48, 1992.

AUBIN, E.C.Q. **Análise de experimentos com medidas repetidas**. São Paulo, 1984. 164p. Dissertação (Mestrado em Estatística/Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo.

BAKER,L.A., REYNOLDS, C., PHELPS,E. Biometrical analysis of individual growth curves. **Behav Genet**. v.22, n. 2, p. 253-64, 1992.

BARROSO, P.A., et al. Efeito do cultivo in vitro na presença de Na Cl em plantas de abacaxizeiro na fase de aclimatação. **R.Bras.Eng.Agric.Ambiental**. Campina Grande, v.7, n.3, p.473-477, 2003.

BOVI, M. L.A., SPIERING, S.H. Estimativa da área superficial de frutos de pupunheira por relações alométricas. **Sci. Agric**. (Piracicaba), vol.59, no.4, p.717-721, 2002.

BRACCINI NETO, J., et al. Análise de curvas de crescimento de galinhas de postura. **Revista Brasileira de Zootecnia**, Viçosa, v.25, n.6, p. 1062-1073, 1996.

CARVALHO, L.R., PINHO, S. Z. Método para comparação de curvas de crescimento logísticas com a presença de autocorrelação nos resíduos. Energia na Agricultura, v.11, n.1, 172p., 1996.

CASTRO, E.M., et al. Crescimento e anatomia foliar de plantas jovens de Mikania glomerata Sprengel (guaco) submetidas a diferentes fotoperíodos. **Ciênc.Agrotec**., Lavras, v.27, n.6, p.1293-1300, nov./dez., 2003.

CECCHETTI, D. **Poder e taxas de erro tipo I de quatro critérios multivariados para o teste de igualdade de efeitos de tratamentos avaliados por meio do Método de Monte Carlo**. Lavras, 1999. 62p. Dissertação (Mestrado em Agronomia/Estatística e Experimentação Agronômica) - Universidade Federal de Lavras.

COLDEBELLA, A., et al. Análise de curvas de crescimento através de modelos mistos: uma opção para comparar sistemas de alimentação para frangos de corte coloniais. In: REUNIÃO DA RBRAS, 48, SEAGRO, 10, Lavras. **Anais**... Lavras,2003. p. 964 – 969.

CROCCI, A.J. **Modelo auto-regressivo para análise de experimentos com vacas em lactação**. Piracicaba, 1984. 86p. Tese (Doutorado em Agronomia/Estatística e Experimentação Agronômica) – ESALQ, Universidade de São Paulo.

DIAS-FILHO, M.B., Physiological response of solanum crinitum lam. To contrasting light environments. **Pesq.Agropec.Bras**. vol.32, n.8, 1997.

http://atlas.sct.embrapa.br/pab/pab.nsf/FrAnual. Acesso em: 17 agosto 2004.

DUTRA, I., et al. Determinação do fator de cobertura do melão cultivado sob diferentes lâminas e salinidades da água de irrigação. **R.Bras.Eng.Agri.Ambiental**, Campina Grande, v.4, n.2, p.146-151, 2000.

DUTRA JR,W.M., et al. Predição de curvas de crescimento de tecidos de fêmeas suínas por intermédio da função alométrica estendida. **Revista Brasileira de Zootécnica**. Viçosa, v. 30, n. 3, 2001. Disponível em: http://www.scielo.br. Acesso em: 09 setembro 2002.

ELIAS, A.M. **Análise de curvas de crescimento de vacas das raças Nelore, Guzerá e Gir**. Piracicaba, 1998. 128p. Dissertação (Mestrado em Agronomia/Ciência Animal e Pastagens). Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo.

FARO, L.E., ALBUQUERQUE, L.G. Comparação de alguns modelos matemáticos para o ajuste às curvas de lactação individuais de vacas da raça Caracu**. Arq. Bras. Med. Vet. Zootec**., vol.54, no.3, p.295-302, 2002.

FITZHUGH Jr., H.A Analysis of growth curves and strategies for altering their shapes. **Journal of Animal Science**, v.42. n.4, p.1036-1051, 1976.

GANTMACHER, F.R. **The Theory of Matrices**. Chelsea Publishinng Company, 2ed., p.276, 1989.

GHOST, D., DEISHER, T.A., ELLSWORTH, J.L. Statistical methods for analysing repeated measures. **J Pharmacol Toxicol**, v. 42, p. 157-162, 1999.

GIOLO, S.R, et al. Análise de dados longitunais em um estudo sobre a evaporação de caldas no controle de insetos em saúde pública. http://www.est.ufpr.br/rt. Acesso em: 22 agosto 2004.

GODOI, C.R.M. Análise Estatística Multidimensional. SEAGRO,1, RBRAS, 30. Piracicaba, 187p, 1985.

GUEDES, M.H.P. et al. Estudo das curvas de crescimento de cordeiros das raças Santa Inês e Bergamácia considerando a heterogeneidade das variâncias. **Cienc.agrotec**. , Lavras, v.28, n.2, p.383-390, 2004.

KHATRI, C.G. **A note on a MANOVA model applied to problems in growth curve**. Ann. Inst. Statist. Math.. v.18, p.75-86, 1966.

KROLL, L. B. **Estudo do crescimento de vacas leiteiras através de modelo com autocorrelação nos erros**. Botucatu, 1990. 97p. Tese (Doutorado em Agronomia/Energia na Agricultura) – FCA, Universidade Estadual Paulista.

LIMA, C.G. **Análise de curvas de crescimento de aves – um enfoque multivariado**. Piracicaba, 1988. 69p. Dissertação (Mestrado em Agronomia/Estatística e Experimentação Agronômica). ESALQ, Universidade de São Paulo.

LIMA, C.G. **Análise de dados longitudinais provenientes de experimentos em blocos casualizados**. Piracicaba, 1996. 126p. Tese (Doutorado em Agronomia/Estatística e Experimentação Agronômica). Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo.

LIMA, C.G. Ajuste de curvas de crescimento em peso de frangos de corte sob um enfoque multivariado. In: REUNIÃO DA RBRAS, 48; SEAGRO, 10. **Resumos**... Lavras: Universidade Federal de Lavras, p. 40-45, 2003.

LITTELL, R.C., HENRY, P.R., AMMERMAN, C.B. Statistical analysis of repeated measures data using SAS procedures. **J.Anim.Sci**., v.76, p. 1216-1231, 1998.

MALHEIROS, E. B. Estudo da precisão da análise de dados longitudinais, com diferentes estruturas para a matriz de variâncias e covariâncias, quando se utiliza o esquema em parcelas

subdivididas. In: REUNIÃO DA RBRAS, 44: SEAGRO, 8, 2001. **Resumos**... Botucatu: UNESP, 1999. p.88.

MARTINELLI, S., BITTAR, R.E., ZUGAIB, M. Proposta de nova curva de altura uterina para gestações entre a 20ª e a 42ª Semana. **Rev. Bras. Ginecol. Obstet**., v.23, n.4, p.235-241, 2001.

MAZZINI, AR.**A Análise da curva de crescimento de machos Hereford considerando heterogeneidade de variâncias e autocorrelação dos erros**. Lavras: UFLA, 2001. p.94. Dissertação (Mestrado em Agronomia/Estatística e Experimentação Agropecuária) – UFLA, Universidade Federal de Lavras.

MILANI, S., BOSSI A., MARUBINI, E. Individual growth curves and longitudinal growth charts between 0 and 3 years. **Acta Paediatr Scand Suppl**., , v. 350, p. 95-104, 1989.

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

NETER, J., WASSERMAN, W., KUTNER, M.H. **Applied linear statistical models: regression, analysis of variance, and experimental designs**. 2.ed. Homewood: Richard D. Irwin, INC., p.1127, 1985.

PERESSIN, V.A., et al. Acúmulo de matéria seca na presença e na ausência de plantas infestantes no cultivar de mandioca SRT 59- Branca de Santa Catarina. **Bragantia**, v. 57 n.1, p. 135-148, 1998.

POTTHOFF, F.B., ROY, S.N. A generalized multivariate of variance model useful especially for growth curve problems. **Biometrika**, v. 51, pp. 313-326, 1964.

RAO, C.R. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. **Biometrika**, v.52, n.3-4. p. 447-458., 1965.

RAO, C.R. **Linear Statistical Inference an its Applications**, 2 ed. New York: Wiley, 1973.

ROCHA, P.E.D. **Aspectos ecofisiológicos da germinação de smentes de Araribá** (centrolobium tomentosum Guill. Ex Benth. – FABACEAE). Botucatu, 1993. 103p. Dissertação (Mestrado em Ciências Biológicas) – Instituto de Biociências – Uiversidade Estadual Paulista.

ROSA, G.J.M., PADOVANI, C.R. Estudo do crescimento de alevinos de carpa através de alguns procedimentos de análise estatística. In: JORNADA CIENTÍFICA DA

ASSOCIAÇÃO DOS DOCENTES DE BOTUCATU, 18, 1993. **Anais**... Botucatu, 1993. p.246.

ROSA, G.J.M. **Desenvolvimento de aplicativo para a análise de perfil na experimentação zootécnica**. Botucatu, 1994, 72p. Dissertação (Mestrado em Zootécnica/Nutrição e Produção Animal). Faculdade de Medicina Veterinária e Zootécnica da Universidade Estadual Paulista.

ROSA, G.J.M. et al. Assesment of alternative linear statistical models for studying growth in crossbred lambs. **J. Anim, Sci**., v.78, s.1, p.145, 2000.

ROSA, G.J.M., GIANOLA, D., PADOVANI, C.R. Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. **Journal of Applied Statistics**, v.31,n.7, p.855-873, 2004.

SHINZATO, A., SAÑUDO, A., SINGER, J.M. Curvas percentuais de peso fetal estimado para gestações gemelares. In: REUNIÃO DA RBRAS, 45. **Resumos**... São Paulo: UFSCar, p.164-167, 2000.

SILVA, E. M., et al. Comparação de modelos matemáticos para o traçado de curvas granulométricas. **Pesq.Agropec.Bras**., Brasília, v.39, n.4, p.363-370, 2004.

SILVA, R.G. Manual de procedimentos em análises por quadrados mínimos. Jaboticabal, FUNEP, 1993. 169p.

SILVEIRA JR, P., et al. Método de obtenção de estimativas preliminares de parâmetros de modelos de crescimento sazonais. **Pesq.Agropec.Bras**. Brasília, v.27, n.12, p.1597-1605, 1992.

SINGER, J.M. **Análise de curvas de crescimento**, 1977. 113p. Dissertação (Mestrado em Estatística). Universidade de São Paulo.

SINGER, J.M., ANDRADE, D.F. **Análise de dados longitudinais**. SINAPE, 7. Campinas, 106p, 1986.

SINGER, J.M., ROCHA, F.M.M., NOBRE, J.S. Análise de Medidas Repetidas. JORNADA REGIONAL DE ESTATÍSTICA, 4, Maringá, p.93, 2004.

SOUZA, J. B., GONCALVES, A. L., MUCILLO, G. Longitudinal study of weight gain in preterm infants from birth to the 6th post-term month. **Rev. Saúde Pública**, Oct. 1992, vol.26, no.5, p.350-355.

SOUZA, L. **Curvas de Crescimento: análise de dados longitudinais incompletos com resíduos autocorrelativos**. Piracicaba, 1989. 84p. Tese (Doutorado em Agronomia/Estatística e Experimentação Agronômica). ESALQ, Universidade de São Paulo.

STEEL, R.G.D.; TORRIE, J.H. **Principles and procedures of statistics: a biometrical approach**. 2.ed. New York: McGraw-Hill, 1980. 631p.

STOEL, R.D., WITTENBOER, G. Transforming the time scale in linear multivariate growth curve models. **Historical Social Research**, vol.25, n.3-4, p. 57-74, 2000.

SUYAMA, E. IN: REUNIÃO ANUAL RBRAS, 40; SEAGRO. **Mini Curso**... Modelos de efeitos aleatórios para dados longitudinais. Ribeirão Preto, 1995, 94p.

TEDESCHI, L.O. **Determinação dos parâmetros da curva de crescimento de animais da raça Guzerá e seus cruzamentos alimentados a pasto, com e sem suplementação**. Piracicaba, 1996. 140p. Dissertação (Mestrado em Agronomia/Ciência Animal e Pastagens). ESALQ, Universidade de São Paulo.

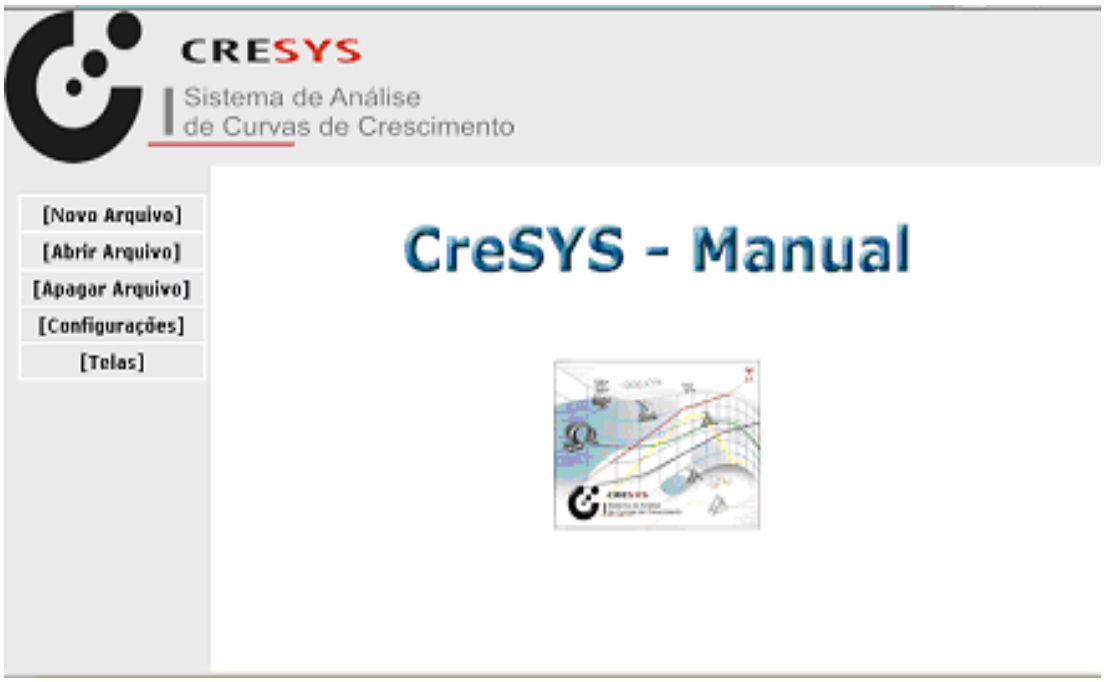
APÊNDICE 1

Altura, em centímetros, dos grupos Testemunha e Tratamento de plantas de alho do tipo Caçador, avaliados em quatro momentos

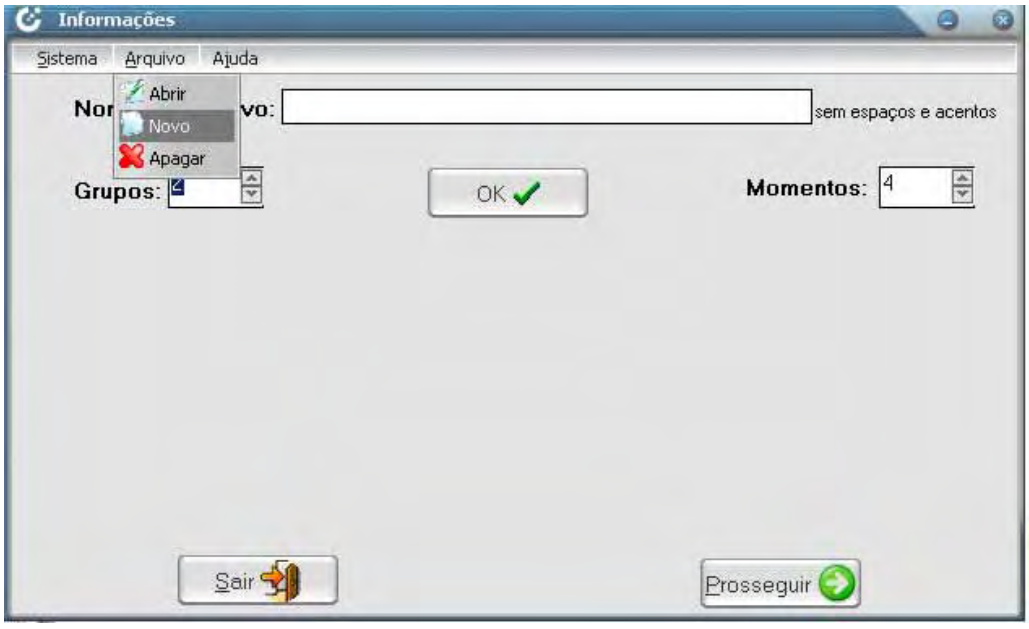
Grupo	Unidade	Momentos (dias)			
		20	30	40	50
1	1	16	19	21	25
1	2	15	18	22	26
1	3	17	20	23	25
1	4	14	18	22	25
1	5	15	19	21	26
1	6	16	20	23	27
1	7	18	21	24	28
1	8	14	17	19	23
1	9	15	19	22	27
2	1	19	24	30	36
2	2	17	21	26	30
2	3	20	26	30	35
2	4	23	28	32	36
2	5	22	26	30	34
2	6	23	25	29	35
2	7	16	20	26	32
2	8	20	25	31	35
2	9	22	26	30	34

APÊNDICE 2

Manual do usuário



Novo Arquivo



Para criar um novo arquivo o pesquisador deve fornecer o nome do arquivo na caixa de texto correspondente, indicar a quantidade de grupos e períodos da análise e em seguida clicar sobre o botão OK. Na seqüência, o pesquisador deverá fornecer os valores das médias e dos respectivos desvios-padrão dos grupos e períodos.

Nesta tela o pesquisador deve fornecer as informações sobre a Análise, tais como:

- Nome da Análise: é o nome único que identifica análise;

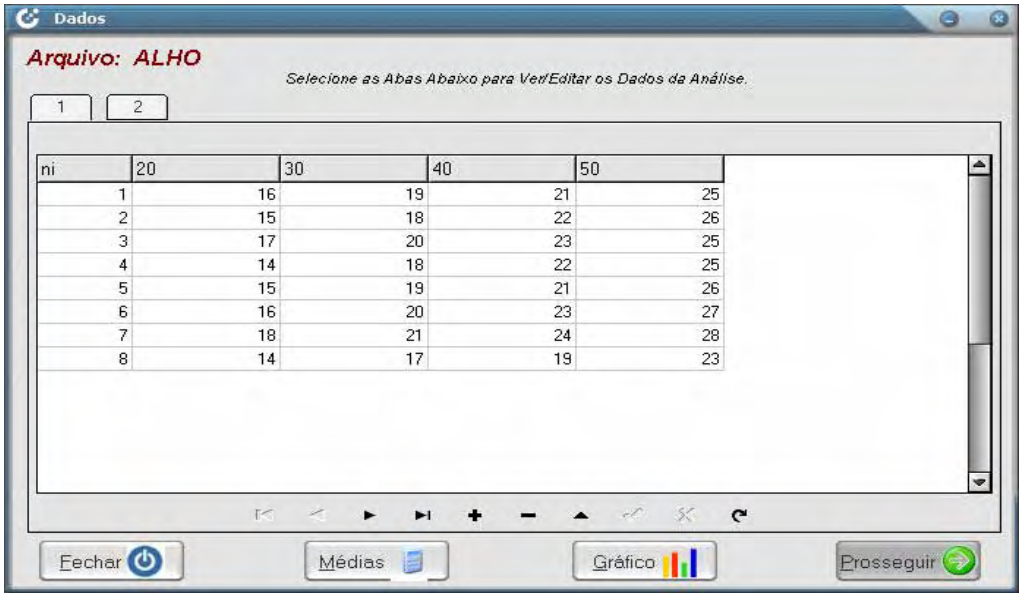
- Grupos: é a quantidade de grupos existentes na análise (mínimo de dois e máximo de dez grupos);

- Períodos: significa a quantidade de momentos utilizados na análise (mínimo de três e máximo de dez);

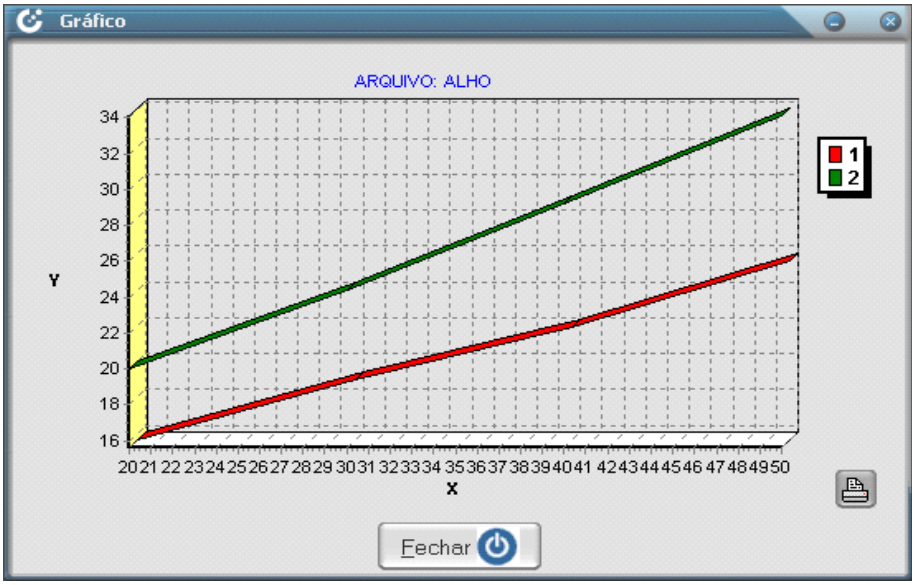
- Nome dos grupos: o pesquisador deve fornecer o nome de cada grupo utilizado na análise, pode ser números e/ou letras;

- Valor dos Períodos: estes valores correspondem aos valores dos momentos da análise e devem ser inteiros.

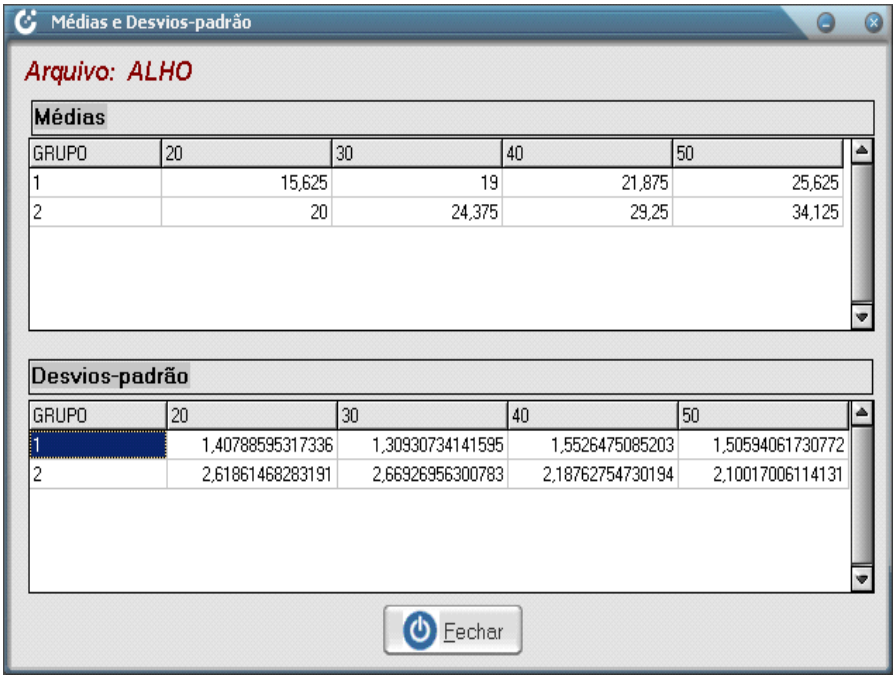
Na janela seguinte, o pesquisador deve selecionar o grupo através das guias (abas) correspondentes e fornecer os dados relativos à seqüência (ni) e períodos.



Após fornecer os dados o pesquisador tem a opção de visualizar o gráfico das médias através do botão que seleciona a opção para gráfico.



Os valores das médias e dos respectivos desvios-padrão são apresentados na seguinte tela:



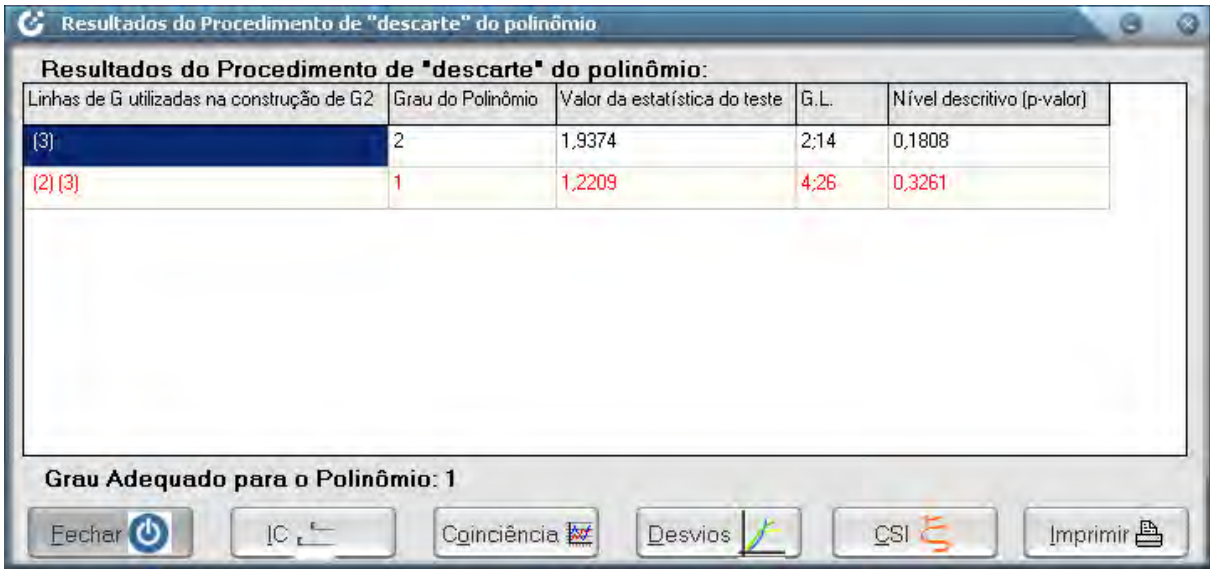
GRUPO	20	30	40	50
1	15,625	19	21,875	25,625
2	20	24,375	29,25	34,125

GRUPO	20	30	40	50
1	1,40788595317336	1,30930734141595	1,5526475085203	1,50594061730772
2	2,61861468283191	2,66926956300783	2,18762754730194	2,10017006114131

O resultado do Teste de Bartlet será apresentado na tela a seguir, onde o usuário poderá decidir, baseado no valor obtido, se prosseguirá com a análise dos dados experimentais.

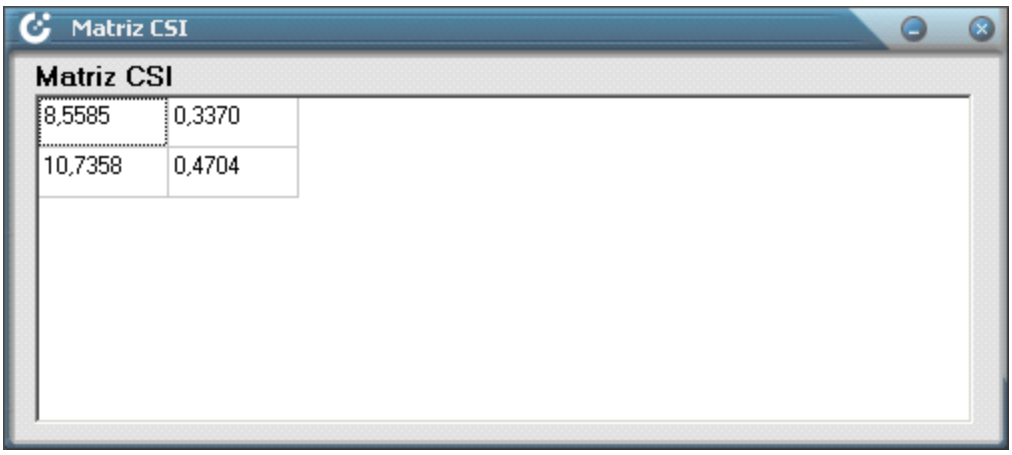
Para prosseguir com a análise o pesquisador deve clicar no botão Yes.

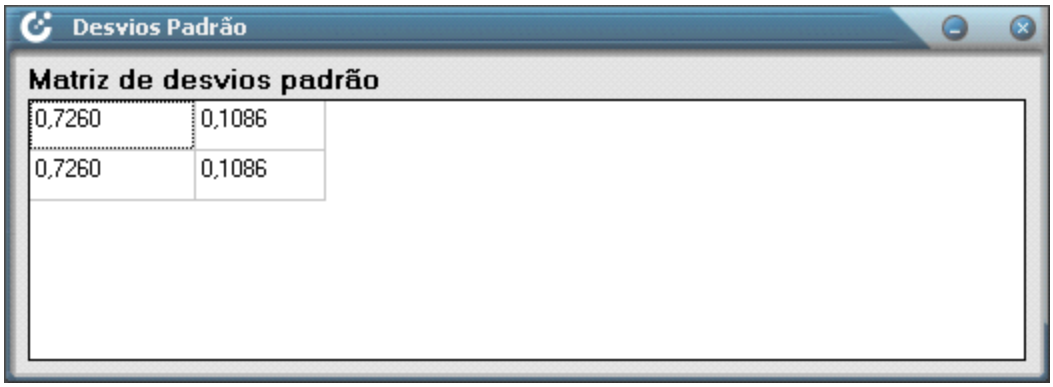
O grau adequado para o polinômio que descreverá o comportamento médio dos dados experimentais é obtido pelo Procedimento de “descarte”:



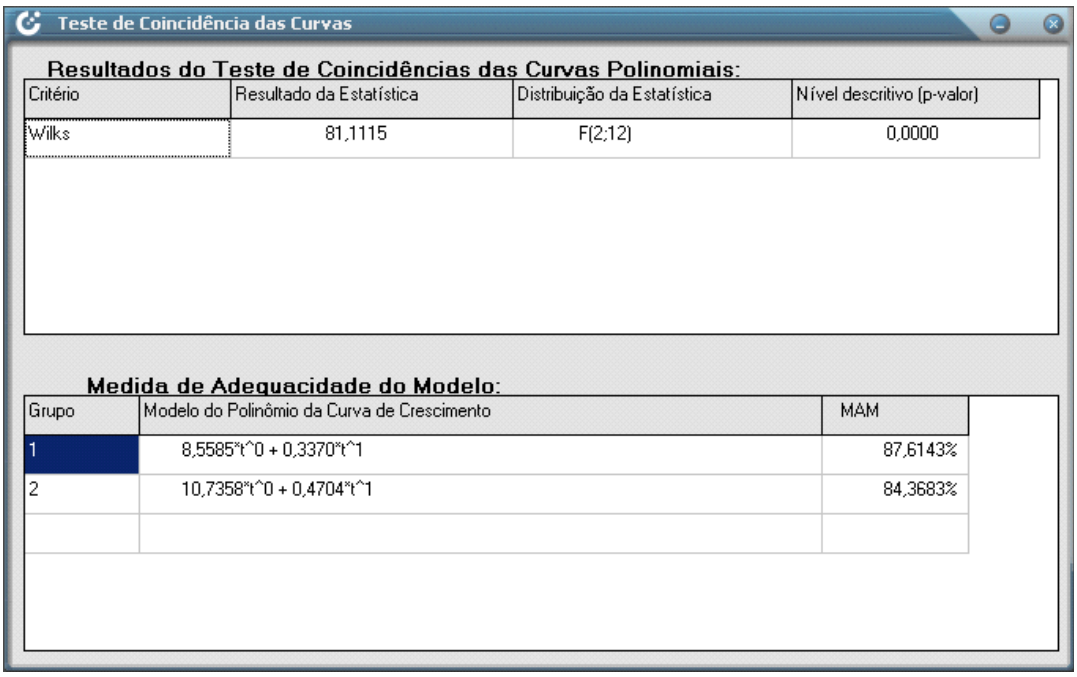
Na seqüência, o usuário poderá obter os valores dos coeficientes dos polinômios das curvas que descrevem cada grupo, selecionando a opção “CSI ξ” e visualizará a tela:

Para a obtenção dos respectivos desvios-padrão, basta selecionar a opção “Desvios”:

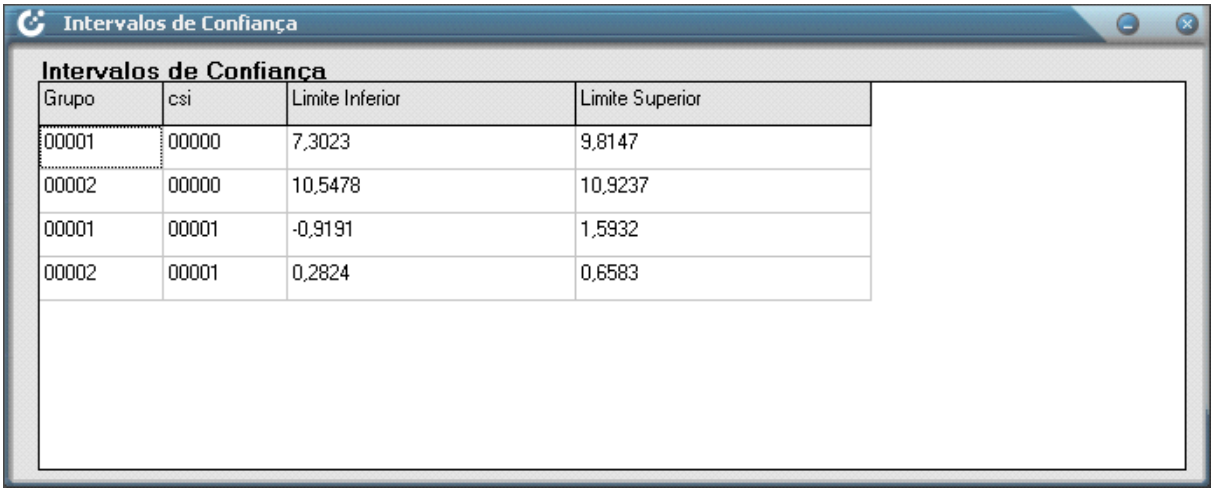




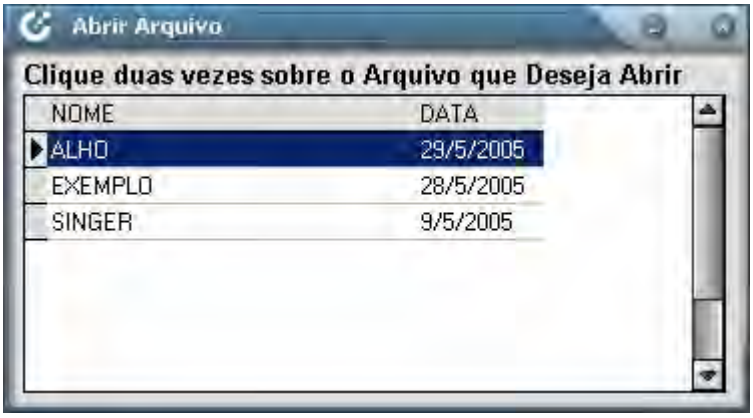
Ao se selecionar a opção “Coincidência”, abre-se uma tela com o resultado do Teste de Coincidências das Curvas de Crescimento e as Medidas de Adequacidade dos Modelos para cada grupo e o total:



Os Intervalos de Confiança para cada coeficiente pode ser encontrado ao se selecionar a opção “IC”:



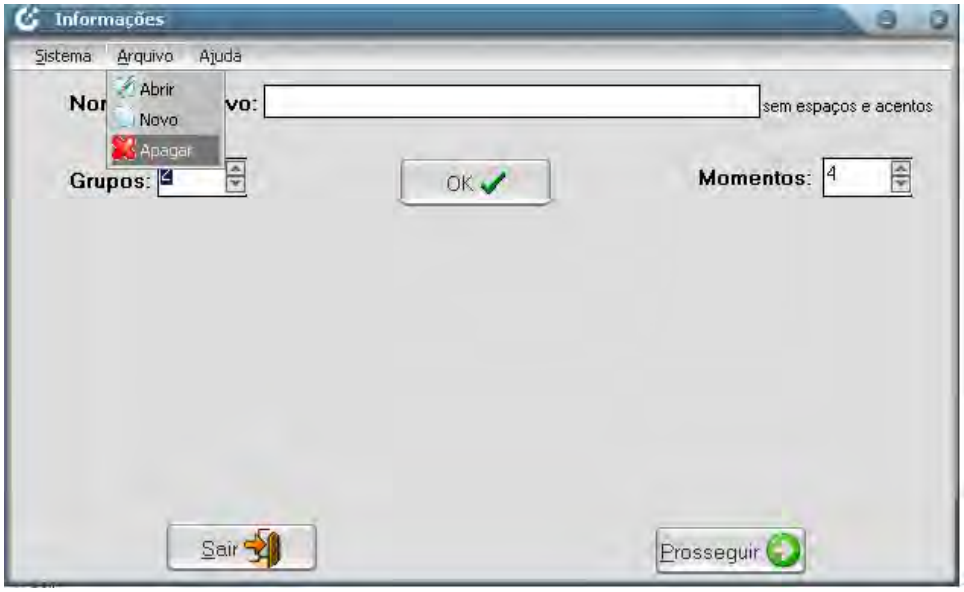
Abrir Arquivo



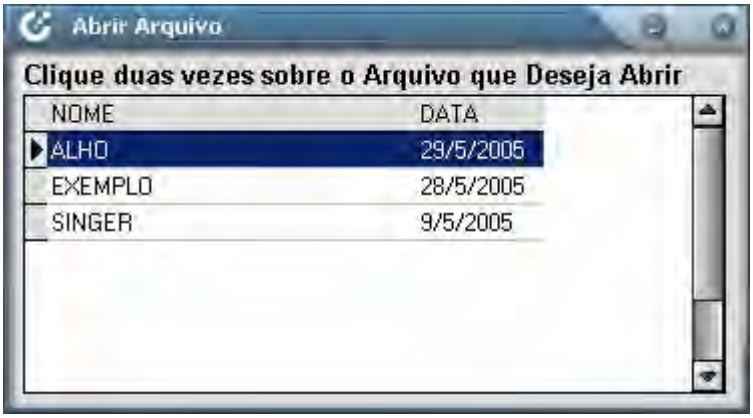
Os procedimentos seguintes seguem-se idênticos aos utilizados quando se cria um novo arquivo. O pesquisador também pode abrir o arquivo referente à uma análise realizada anteriormente clicando no menu Arquivo -> Abrir e clicando duas vezes sobre o arquivo desejado (figura abaixo).

Ao abrir um arquivo o pesquisador pode alterar, incluir e excluir dados da análise.

Apagar Arquivo

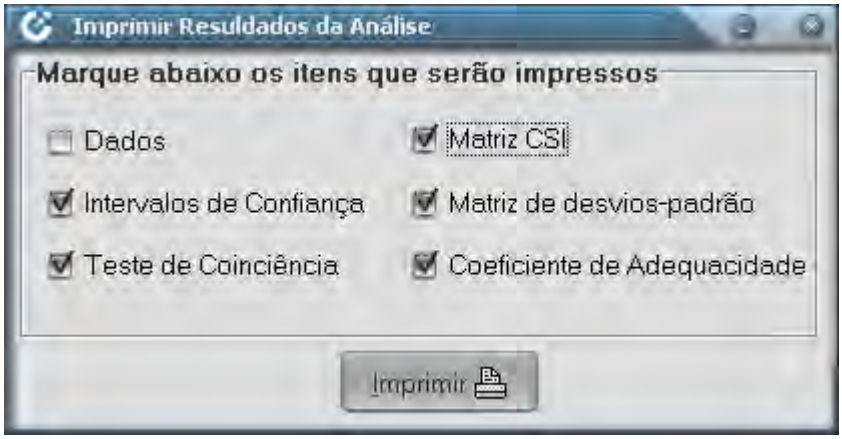


O pesquisador também pode excluir o arquivo referente à uma análise realizada anteriormente clicando no menu Arquivo -> Apagar e clicar duas vezes sobre o arquivo que deseja excluir (figura abaixo).



IMPORTANTE: Após a confirmação de exclusão NÃO será possível a recuperação das informações.

Ao se selecionar a opção “Imprimir”, um relatório da Análise será impressa com os valores selecionados pelo usuário.



Os resultados do estudo de curvas de crescimento poderão ser impressos no Relatório da Análise: