

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS RURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA**

**MODELOS MISTOS NO AJUSTE DE CURVAS DE
CRESCIMENTO DE *Ricinus communis* L.**

TESE DE DOUTORADO

Janete Pereira Amador

Santa Maria, RS, Brasil

2010

**MODELOS MISTOS NO AJUSTE DE CURVAS DE
CRESCIMENTO DE *Ricinus communis* L.**

por

Janete Pereira Amador

Tese apresentada ao Programa de Pós-Graduação em Agronomia,
Área de Concentração em Produção Vegetal, Universidade Federal de Santa
Maria (UFSM, RS), como requisito parcial para obtenção do grau de
Doutora em Agronomia.

Orientador: Sidinei José Lopes, Dr.

Santa Maria, RS, Brasil

2010

A481a	Amador, Janete Pereira, 1966- Modelos mistos no ajuste de curvas de crescimento de <i>Ricinus communis</i> L. / Janete Pereira Amador. - 2010. 139 f. ; il. Tese (doutorado) – Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós-Graduação em Agronomia, 2010. “Orientador: Prof. Dr. Sidinei José Lopes” 1. Agronomia 2. Mamona 3. Ajuste de curvas de crescimento 4. Modelos mistos I. Lopes, Sidinei José II. Título CDU: 633.85
-------	---

Ficha catalográfica elaborada por
Patrícia da Rosa Corrêa – CRB 10/1652
Biblioteca Setorial do Centro de Ciências Rurais/UFSM

**Universidade Federal de Santa Maria
Centro de Ciências Rurais
Programa de Pós-Graduação em Agronomia**

A Comissão Examinadora, abaixo assinada,
Aprova a Tese de Doutorado

**MODELOS MISTOS NO AJUSTE DE CURVAS DE
CRESCIMENTO DE *Ricinus communis* L.**

elaborada por
Janete Pereira Amador

Como requisito parcial para obtenção do grau de
Doutora em Agronomia

COMISSÃO EXAMINADORA

Sidinei José Lopes, Dr. (UFSM)
Presidente/Orientador

Roselaine Ruviaro Zanini, Dra. (UFSM)

Paulo Henrique de Oliveira, Dr. (UFRGS)

José Henrique Souza da Silva, Dr. (UFSM)

José Pedro Pereira Trindade, Dr. (CPPSUL/EMBRAPA)

Santa Maria, 16 de julho de 2010

AGRADECIMENTOS

A todas as forças do cosmos que conspiraram a favor da realização deste trabalho.

À minha família, do Departamento de Estatística pela compreensão e estímulo em todas as etapas desta aprendizagem.

Ao Professor Sidinei José Lopes, pela orientação, confiança, incentivo e amizade, meu agradecimento especial.

Aos Professores do Comitê de Orientação e do Programa de Pós-Graduação em Agronomia e Engenharia Florestal, pelos ensinamentos.

Ao Prof. José Henrique Souza da Silva e o Prof. João Eduardo da Silva Pereira pela amizade, confiança e grande sabedoria.

Aos colegas do curso de Pós-Graduação em Agronomia pela amizade, estímulo e colaboração.

**“Monstros existem e fantasmas também.
Vivem dentro de nós e, às vezes, vencem.”**

Stephen King

RESUMO

Tese de Doutorado
Programa de Pós-Graduação em Agronomia
Universidade Federal de Santa Maria

MODELOS MISTOS NO AJUSTE DE CURVAS DE CRESCIMENTO DE *Ricinus communis* L.

AUTORA: JANETE PEREIRA AMADOR
ORIENTADOR: SIDINEI JOSÉ LOPES
Santa Maria, 16 de Julho de 2010.

A demanda energética mundial precisa ser atendida, entretanto, as fontes convencionais estão em vias de esgotamento, além de provocarem efeitos prejudiciais ao ambiente. Neste sentido, têm-se pesquisado fontes energéticas renováveis, como alternativa a agroenergia, que poderá contribuir de maneira substancial na oferta futura de energia. Uma das vertentes da agroenergia é a do biodiesel, um combustível biodegradável alternativo ao diesel de petróleo, criado a partir de fontes renováveis de energia, livre de enxofre em sua composição. No campo das oleaginosas, a mamona (*Ricinus communis* L.) vem se destacando como fornecedora de matéria-prima para a fabricação de biodiesel. No Rio Grande do Sul, para que a ricinocultura, torna-se sustentável em nível de produção em escala como matéria prima para o biodiesel o incremento na produção torna-se fundamental. E este depende do desenvolvimento de tecnologias agrícolas adaptadas às condições do estado. Neste contexto, existe uma escassez de pesquisas que estudem a influência da densidade de plantas no cultivo desta oleaginosa e a relação destas com a forma de crescimento da planta. Já que este fator pode influenciar, de maneira significativa, o desenvolvimento da mamoneira, refletindo na produção. Esta pesquisa objetiva avaliar a influência de três densidades de cultivo 1,0, 1,2 e 1,4 m, mantendo constante a distância entre linha de 1,0 m, sobre o crescimento da mamona. Para explicar esta influencia será utilizada a técnica de análise de dados longitudinais por meio de modelagem de curvas de crescimento utilizando a metodologia de modelos mistos. Constatou-se que esta permite explorar de forma mais completa as informações fornecidas pelos dados, pelo fato de ser possível acomodar a correlação destes, por meio da seleção de estruturas de covariância. Sendo assim, das seis estruturas de covariância testadas para modelar G e R as selecionadas foram UN(1) e VC. A estratégia utilizada tanto para seleção destas estruturas como para os efeitos aleatórios mostrou-se como alternativa mais simples nesta parte da modelagem. Na análise univariada e multivariada de perfil para escolha dos efeitos fixos do modelo com algumas informações fornecidas pelas duas técnicas, é possível verificar qual a que melhor se adéqua aos dados. Modelou-se duas curvas uma densidades 1 e 1,4 m que proporcionaram a maior altura de planta, não diferindo significativamente. Para a densidade de cultivo de 1,2 m, a planta apresentou um menor crescimento, seu comportamento de crescimento foi modelado por outra curva. Mesmo considerando a complexidade computacional dos procedimentos utilizados, espera-se, através das rotinas do SAS compiladas neste material, contribuir para uma maior facilitação no uso destes procedimentos.

Palavras-chave: mamona, ajuste de curvas de crescimento, modelos mistos.

ABSTRACT

Doctoral Dissertation

Graduate Program of Agronomy
Universidade Federal de Santa Maria

MIXED MODEL BY ADJUSTMENT OF GROWTH CURVES IN *Ricinus communis* L.

AUTHOR: JANETE PEREIRA AMADOR
ADVISOR: SIDINEI JOSÉ LOPES
Santa Maria, July 16th, 2010.

The world energy demand must be met, however, the conventional sources are nearing exhaustion, and cause harmful effects to the environment. In this sense, it has been researched renewable energy sources as an alternative to bioenergy, which could contribute substantially in the future supply of energy. One of the aspects of bioenergy is that of biodiesel, a biodegradable fuel alternative to diesel oil, created from renewable energy sources, free of sulfur in its composition. In the field of oil seed, castor bean (*Ricinus communis* L.) has stood out as a supplier of raw material to manufacture biodiesel. In Rio Grande do Sul, for the ricinoculture becomes sustainable production-level scale as raw material for biodiesel production in the increment becomes crucial. And this depends on the development of agricultural technologies adapted to the conditions of the state. In this context, there is a paucity of research to study the influence of density on the cultivation of oilseeds and their connection with the form of plant growth. Since this factor can influence significantly the development of castor bean, reflecting production. This research aims to evaluate the influence of three planting densities 1.0, 1.2 and 1.4 m, keeping constant the distance between the row of 1.0 m, on the growth of castor bean. To explain this influence will be used the technique of analysis of longitudinal data using growth curve modeling using mixed model methodology. It was found that this allows to explore more fully the information provided by the data, because it is possible to accommodate the correlation of these through the selection of covariance structures. Thus, the six covariance structures tested to model the G and R were selected UN (1) and VC. The strategy used to select both of these structures as for the random effects proved to be simpler alternative in this part of modeling. In univariate analysis of profile to choose the fixed effects model with some information provided by both techniques, you can check which one best fits the data. Shaped curve is a two densities 1 and 1.4 m that provided the greatest plant height did not differ significantly. For the cultivation density of 1.2 m, the plant showed a lower growth, its growth behavior was modeled by another curve. Even considering the computational complexity of the procedures used, it is hoped, through the SAS routines compiled this material, contribute to greater facilitation in the use of these procedures.

Keywords: castor bean. mixed models, growth curves

LISTA DE FIGURAS

- FIGURA 1 – Diferenças verticais entre as curvas de tratamentos. Santa Maria, RS, 2010... 45
- FIGURA 2 – Curvas não são constantes ao longo do tempo. Santa Maria, RS, 2010..... 46
- FIGURA 3 – Curvas diferem entre os tratamentos ao longo do tempo. Santa Maria, RS, 2010. 47
- FIGURA 4 – Diagrama paralelo de dispersão dos perfis individuais dos dados de altura de planta de mamona para as doze unidades experimentais. Santa Maria, RS, 2010.....89
- FIGURA 5 – Perfis médios de resposta da variável altura de planta de mamona, submetida a três densidades de cultivo e oito medidas no tempo no ciclo. Santa Maria, RS, 2010..... 100
- FIGURA 6 – Gráfico de probabilidade normal dos resíduos para os modelos estimados para a altura de plantas de mamona entre 50 e 150 dias de avaliação para as 12 unidades experimentais. Santa Maria, RS, 2010 105
- FIGURA 7 – Curvas de crescimento ajustadas à altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades de cultivo. Santa Maria, RS, 2010. 111
- FIGURA 8 – Modelagem final das duas curvas de crescimento para a variável altura de planta de mamona entre 50 e 150 dias de avaliação. Santa Maria, RS, 2010. 111
- FIGURA 9 – Gráfico de probabilidade normal dos resíduos para o modelo ajustado à altura de plantas de mamona entre 50 e 150 dias de avaliação nas densidades de cultivo de 1,0m ou 1,4m. Santa Maria, RS, 2010. 112

FIGURA 10 – Gráfico de probabilidade normal dos resíduos para o modelo ajustado à altura de plantas de mamona entre 50 e 150 dias de avaliação nas densidades de cultivo de 1,2m. Santa Maria, RS, 2010..... 113

LISTA DE TABELAS

TABELA 1 – Configuração dos dados para uma estrutura balanceada e completa de dados longitudinais. Santa Maria, RS, 2010.	37
TABELA 2 – Tabela da análise de variância com estatísticas para os testes de hipótese do modelo de parcelas subdivididas em um delineamento inteiramente casualizado. Santa Maria, RS, 2010.	44
TABELA 3 – Critérios e estatísticas utilizadas para testar a hipótese linear geral. Santa Maria, RS, 2010.	59
TABELA 4 – Média (cm), desvio padrão (S) e coeficiente de variação (CV) da variável altura de planta de mamona, coletadas em quatro repetições, de acordo com os dias de avaliação e tratamento. Santa Maria, RS, 2010.....	90
TABELA 5 – Resultado da análise univariada de perfil para os dados da variável altura de planta de mamona, submetida a três densidades de cultivo e oito medidas no tempo igualmente espaçadas. Santa Maria, RS, 2010.	91
TABELA 6 – Análise univariada com correção dos números de graus de liberdade para os efeitos intra-indivíduos, da variável altura de planta de mamona, submetida a três densidades de cultivo e oito medidas no tempo igualmente espaçadas. Santa Maria, RS, 2010.....	94
TABELA 7 – Resultado da análise multivariada da variável altura de planta de mamona, submetida a três densidades de cultivo e oito medidas no tempo igualmente espaçadas. Santa Maria, RS, 2010.....	98

TABELA 8 – Teste <i>t-Student</i> para verificar a igualdade de intercepto e inclinação das retas de regressão para as três densidades de cultivo (tratamentos –T) da variável altura de planta no período de 10 a 50 dias. Santa Maria, RS, 2010.....	96
TABELA 9 – Teste de hipótese de tendência de variação da altura de plantas média de mamona entre 50 e 150 dias de avaliação para as três densidades de cultivo (Tratamento). Santa Maria, RS, 2010.....	97
TABELA 10 – Estimativas dos parâmetros das curvas ajustadas às médias da altura de planta em função do tempo, com seus respectivos erros padrões, e nível de significância de cada parâmetro, entre parênteses, para cada tratamento (densidade) no período de 50 a 150 dias. Santa Maria, RS, 2010.	98
TABELA 11 – Estimativas individuais dos coeficientes das regressões polinomiais, com o respectivo nível mínimo de significância para a altura de plantas de mamona entre 50 e 150 dias de avaliação para as 12 unidades experimentais. Santa Maria, RS, 2010.	100
TABELA 12 – Estruturas de variância-covariância testadas para a matriz G (variação entre indivíduos) e seleção de efeitos aleatórios para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.....	103
TABELA 13 – Estruturas de variância-covariância testadas para a R (variação dentro de indivíduos) para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.	105
TABELA 14 – Testes para efeitos fixos do modelo completo utilizando as estruturas de matriz de covariância (UN(1) para G e VC para R para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.....	106
TABELA 15 – Teste de Tukey – Kramer para três densidades de semeadura, com o modelo completo, utilizando as estruturas de matriz de covariância UN(1) para G e VC para R. Santa Maria, RS, 2010.....	106
TABELA 16 – Estimativas dos parâmetros dos modelos das curvas crescimento ajustadas à altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.....	107

LISTA DE APÊNDICES

APÊNDICE A – Dados da altura de planta (cm) referente aos três tratamentos nos oito períodos de avaliação.....	125
APÊNDICE B – Modelos testados para matriz G.	126
APÊNDICE C – Modelos testados para matriz R.	130
APÊNDICE D – Testes para efeitos fixos e ajuste de curvas	135
APÊNDICE E – Estruturas de variância-covariância testadas para a matriz G (variação entre indivíduos) e seleção de efeitos aleatórios para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.....	136
APÊNDICE F – Estruturas de variância-covariância testadas para a R (variação dentro de indivíduos) para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.....	137

LISTA DE QUADROS

QUADRO 1 – Variância-covariância em negrito, variâncias em destaque, e correlações residuais para variável altura de plantas de mamona em oito avaliações (dias) no ciclo. Santa Maria, RS, 2010.....	90
---	----

SUMÁRIO

1 INTRODUÇÃO	15
2 REVISÃO BIBLIOGRÁFICA	17
2.1 Biocombustíveis	17
2.2 Biodiesel	20
2.3 A cultura de mamona	23
2.3.1 Caracterização botânica.....	23
2.3.2 Frutos, sementes e cultivares	24
2.3.3 Aspectos gerais da cultura	25
2.3.4 A cultura no Rio Grande do Sul	26
2.3.5 População de plantas	28
2.4 Caracterização de dados Longitudinais	30
2.5 Metodologias de análise para dados longitudinais	34
2.5.1 Análise univariada de perfil.....	37
2.5.2 Análise multivariada de perfil	47
2.5.3 Ajuste de curvas de crescimento por modelo linear misto	56
2.5.3.1 Estimação de β e b do modelo linear misto	60
2.5.3.2 Estimação dos parâmetros de covariância em G e R.....	63
2.5.3.3 Propriedades dos estimadores de G e R.....	66
2.5.3.4 Estruturas de variância-covariância.....	67
2.5.3.5 Seleção de modelos e testes.....	74
2.5.3.6 Inferência sobre os efeitos fixos e aleatórios.....	76
3 MATERIAL E MÉTODOS	80
3.1 Instalação do experimento	80
3.2 Semeadura e Manejo da Cultura	80
3.3 Variável resposta e coleta de dados	81
3.4 Procedimentos estatísticos	82

3.4.1 Seleção dos efeitos fixos.....	82
3.4.2 Identificação dos efeitos aleatórios	83
3.4.3 Estimação e comparação de modelos	83
4 RESULTADOS E DISCUSSÃO	86
4.1 Altura de planta	86
4.1.1 Seleção dos efeitos fixos.....	86
4.1.1.2 Análise Univariada de Perfil.....	88
4.1.1.3 Análise Multivariada de Perfil.....	92
4.1.2 Identificação dos efeitos aleatórios	100
4.1.3 Seleção de modelos e testes.....	102
5 CONCLUSÕES.....	111
BIBLIOGRAFIA	112
APÊNDICES.....	126

1 INTRODUÇÃO

A necessidade de inserir na matriz energética atual, alicerçada no petróleo, novas alternativas de fontes energéticas, tornou a produção de matéria-prima de biocombustíveis, objeto de estudo em Universidades e Centros de Pesquisas.

Entre as fontes de biocombustíveis, o biodiesel desempenha um papel importante na substituição do óleo diesel oriundo do petróleo. No campo das oleaginosas, a mamona (*Ricinus communis* L.) vem se destacando como fornecedora de matéria-prima para a fabricação de biodiesel, pelo fato de produzir óleo, com características químicas, que o torna único na natureza.

No Rio Grande do Sul, para que a ricinocultura, torne-se sustentável em nível de produção em escala, como matéria-prima para o biodiesel, o incremento na produção torna-se fundamental. Sendo assim, é imprescindível estabelecer sistemas de produção adequados. Neste sentido, a correta determinação da densidades de plantas para o cultivo é muito importante, pois influencia no crescimento da mamoneira refletindo, assim, na produtividade da planta.

A densidade de cultivo é um dos fatores que influenciam na capacidade produtiva da mamona e esta pode ser estudada realizando-se experimentos com medidas repetidas. Estes experimentos descrevem o comportamento das variáveis respostas, quando estas são medidas em vários tempos e, geralmente, sob diferentes tratamentos. Um caso especial dos experimentos com medidas repetidas constituem os planejamentos longitudinais.

O que distingue a análise de dados longitudinais é a estrutura de covariância dos dados observados. Os modelos de análise de variância tradicionais assumem erros independentes. Assumir erros independentes significa que todas as observações de uma unidade experimental em um planejamento de dados longitudinais são igualmente correlacionadas. Entretanto, não é isto que ocorre, porque duas medidas tomadas em tempos adjacentes são normalmente mais correlacionadas do que duas medidas tomadas em tempos mais distantes.

O uso de técnicas de análise que levem em consideração a correlação serial dos dados é importante, e pode ser realizada pela seleção de estruturas de variância-covariância adequadas aos dados. Desta forma, uma das alternativas para a análise dos dados é a

utilização da técnica de análise de curvas de crescimento ou modelagem de curvas de crescimento por meio de modelos mistos.

A técnica de modelagem de curvas de crescimento por modelos mistos possibilitam o uso de diferentes tipos de estruturas para as matrizes de variâncias-variância-covariância, optando-se por aquela que melhor representa a estrutura de correlação dos dados. Para se estruturar um modelo misto é necessário ter um modelo para os efeitos fixos, que deve descrever o comportamento da curva média; um modelo para os efeitos aleatórios, que junto com os efeitos fixos, devem descrever as curvas individuais; e, um modelo para dispersão condicional, que deve descrever as variância-covariância das observações individuais ao longo do tempo (SUYAMA, 1995).

A estimação dos parâmetros neste tipo de modelo é baseada na verossimilhança dos dados e, a escolha das estruturas de covariância mais adequada aos dados, é feita utilizando-se o Teste de Razão de Verossimilhança, que serve para comparar o ajuste de modelos encaixados e os Critérios de Informação de Akaike (AIC) e de Schuartz (BIC).

O presente trabalho tem como objetivo principal apresentar aplicabilidade das técnicas de análise de dados longitudinais, com o ajuste de curvas de crescimento, por meio de modelos mistos, para a variável altura de planta de mamona, submetidas a três densidades de cultivo. Para tanto, pretende-se abordar algumas questões que podem ser consideradas como objetivos específicos:

- quanto à metodologia de análise, qual a que melhor se adéqua na escolha dos efeitos fixos do modelo;
- quais as estratégias a serem utilizadas na seleção dos efeitos aleatórios do modelo e das estruturas de covariância;
- qual (ou quais) a (s) estrutura (s) de covariância que melhor representa (m) a variabilidade dos dados;
- quanto aos procedimentos disponíveis no Software estatístico SAS[®], quando e como utilizar cada procedimento;
- quanto a densidade de cultivo, qual proporcionou maior altura de planta para a Cultivar Vinema T1.

2 REVISÃO BIBLIOGRÁFICA

2.1 Biocombustíveis

O que se apresenta, no cenário mundial, é a progressiva redução das fontes de combustíveis fósseis. A quantidade de petróleo descoberto nas novas reservas é cada vez menor e o consumo mundial cada vez maior. A extração de petróleo bruto nas antigas reservas e nas novas descobertas está cada vez mais difícil. O petróleo é de pior qualidade, o que eleva o custo da extração e do beneficiamento, além da instabilidade política nas regiões petrolíferas (Oriente Médio), refletindo nos preços do barril de petróleo.

O consumo atual de petróleo está em torno de 84 milhões de barris por dia e cresce à medida que aumenta a demanda de energia pelas novas economias, como a China e a Índia (Oil Analytics, 2008). O consumo crescente de petróleo provoca uma redução no EROEI, *Energy Return Over Energy Invested*, ou seja, com a redução das reservas de fácil exploração, torna-se necessário explorar petróleo de difícil acesso ou com menor concentração de hidrocarbonetos, a um maior custo (PEREIRA, 2009).

Além das questões relacionadas à escassez e ao elevado custo de exploração, em termos ambientais, esses combustíveis contribuem de maneira significativa para degradação do meio ambiente através do efeito estufa. O uso intensivo da energia petrolífera provoca a emissão descontrolada de gases poluentes e a destruição indiscriminada dos recursos naturais que resultam em alterações na variabilidade do clima e na frequência e intensidade de alguns fenômenos climáticos extremos, no aumento do efeito estufa e no crescimento da temperatura da terra.

O efeito estufa é causado pelo aumento da concentração dos gases na atmosfera, tais como: o gás carbônico (CO_2), o metano (CH_4), o óxido nitroso (N_2O) e os clorofluorcarbonos (CFC) (Metzger, 2001). Esses gases capturam parte da radiação infravermelha que a terra devolve para o espaço, fazendo com que ocorra um aumento na temperatura da superfície terrestre, o qual pode levar ao degelo das calotas polares e a elevação das marés, causando prejuízos às regiões costeiras.

Diante destes aspectos, surge a necessidade de introduzir na matriz energética atual, alicerçada no petróleo, novas alternativas de fontes energéticas e que estas não acarretem consequências nocivas ao meio ambiente. Neste sentido, pesquisas a respeito de fontes energéticas renováveis – biocombustíveis - estão sendo realizadas em universidades e centros de pesquisas. De acordo com Conceição *et al.* (2007), biocombustíveis são combustíveis de origem biológica. São fabricados a partir de produtos agrícolas, tais como: milho, cana-de-açúcar, plantas oleaginosas, biomassa florestal e outras fontes de matéria orgânica.

No Brasil, o modelo de industrialização esteve atrelado ao desenvolvimento automotivo. O que faz com que, atualmente, a participação do transporte na matriz energética seja alta. Em 2004, cerca de 29% de toda a energia consumida foi nesse setor (Ministérios MAPA, MCT, MME e MDIC, 2005). Em se tratando disto, Goldemberg; Vilanueva (2003) destacam o uso de combustíveis alternativos, dentre as soluções técnicas para reduzir a emissão de poluentes no setor de transporte. Na União Européia, o Parlamento estabeleceu em 2003, as diretrizes para a promoção e uso dos combustíveis renováveis no setor de transporte e essa medida faz parte do conjunto de ações que visam ao cumprimento das metas estabelecidas no Protocolo de Kyoto, além de contribuir para a garantia de suprimento de energia no médio e longo prazo (EUROPEAN UNION, 2003).

Antes do advento dos biocombustíveis, algumas outras fontes alternativas de energia foram exploradas com sucesso, como é o caso da energia nuclear, hidrelétrica, eólica ou solar (PEREIRA, 2009). Porém, devido à restrição dos recursos naturais ou tecnológicos, essas fontes talvez não sejam suficientes para substituir toda a energia fornecida pelo petróleo e para atender o aumento na demanda. Neste contexto, então, a fotossíntese das plantas surge como alternativa renovável de captação de energia que pode ser armazenada na forma de biocombustíveis ou agrocombustíveis.

No Brasil, a alternativa energética, do que é hoje denominado de biocombustíveis, iniciou-se, não por uma questão ambiental, mas, sim por um estratagema energético e econômico, desencadeada por duas guerras, Árabe- Israelense em 1973 e Irã-Iraque, em 1979. Estes dois conflitos foram determinantes para o País perceber o risco na dependência do petróleo árabe. Conforme Melo; Fonseca (1981), no primeiro conflito, o País importava 76,9% do petróleo usado na demanda interna de combustíveis. Durante os dois conflitos, o Brasil adotou políticas que visavam, por uma parte, a construção de usinas nucleares e hidroelétricas, e por outra, a auto-suficiência de combustíveis líquidos (PUERTO, 2007).

Nesta conjuntura, iniciou-se o Programa Nacional de Álcool (PROALCOOL) através do Decreto 76.593 de 1975, com a perspectiva de que o álcool produzido daria saída aos excedentes de açúcar e quando misturado a gasolina reduziria as importações de petróleo. Somente trinta anos após a sua implantação é que se funde ao PROALCOOL o argumento de combustível ecologicamente limpo.

Conforme as diretrizes de política de agroenergia dos Ministérios (MAPA, MCT, MME e MDIC, 2005), no curto e médio prazo, a função dos biocombustíveis será a de propiciar uma transição mais tranquila rumo a matriz energética com maior participação da energia renovável, inclusive ampliando o horizonte de uso das atuais fontes de carbono fóssil. Subsidiariamente, o desenvolvimento dos biocombustíveis, no Brasil, promoverá importante aumento de investimentos, empregos, renda e desenvolvimento tecnológico, e será uma oportunidade para atender parte da crescente demanda mundial por combustíveis de reduzido impacto ambiental. Essa visão de futuro é plenamente aplicável ao Brasil, que poderá se constituir no maior provedor individual de energia renovável no mercado internacional de biocombustíveis.

A produção de energia, a partir de produtos agrícolas, é classificada em: etanol, biogás, biodiesel, florestas e resíduos. Em relação à produção de etanol, pretende-se obter totalmente a partir de cana-de-açúcar. O biogás é uma fonte de energia produzida de restos de matéria orgânica em fase de decomposição. O biodiesel é uma fonte de energia obtida a partir do processamento de determinados grãos, tais como: mamona, dendê, girassol, babaçu, amendoim e soja. Os óleos derivados desses vegetais podem ser usados integralmente ou agregados ao diesel (fóssil) em quantidades variadas (BRASIL ESCOLA, 2010).

No contexto das diretrizes dos biocombustíveis, pode-se dizer que esta abrange quatro vertentes principais: álcool; biodiesel; florestas energéticas cultivadas; e resíduos agrofloretais. É importante ressaltar que existem inter-relações entre esses segmentos, como o uso do etanol para a produção de biodiesel, a co-geração de energia elétrica com resíduos da produção de álcool, ou o aproveitamento de resíduos de biomassa florestal (MAPA, MCT, MME e MDIC, 2005).

Mundialmente, a evolução das questões ambientais com a forte participação da energia renovável é um fator positivo da matriz energética do Brasil. O crescimento da participação do álcool e do biodiesel nesta matriz vem colaborando com o desafio internacional da redução

dos gases causadores do efeito estufa, dado que sua queima mais limpa, e com sustentabilidade, por ser um recurso de base renovável florestal (MAPA, MCT, MME e MDIC, 2005).

2.2 Biodiesel

Biodiesel é o nome genérico dos combustíveis biodegradáveis feitos com óleos ou gorduras de origem vegetal ou animal (PIRES *et al.*,2004). Submetidas a reações químicas, essas substâncias são convertidas em um produto que, puro ou misturado ao diesel de petróleo, faz girar os motores de ciclo diesel sem necessidade de conversão. Enquanto que, o uso de outros combustíveis limpos, como o gás natural ou biogás, requerem adaptação dos motores. Sendo assim, o biodiesel configura-se como uma alternativa capaz de atender toda a frota já existente movida a óleo diesel (SALES *et al.*,2004).

Tecnicamente, biodiesel é definido como um éster alquílico de ácidos graxos obtido de reação de transesterificação, que ocorre entre qualquer triglicerídeo (óleos e gorduras animais ou vegetais) com álcool de cadeia curta, na presença de um catalisador ácido ou básico. Como resultado, obtém-se o éster metílico ou etílico –Biodiesel - conforme o álcool utilizado (DANTAS *et al.*,2004).

O biodiesel etílico possui propriedades combustíveis iguais as do metílico produzido na Europa e EUA, com a vantagem de tratar-se de um combustível obtido totalmente a partir de fontes renováveis (RAMPIN, 2007). Com o domínio da produção do biodiesel pela rota etílica, pode-se dizer que o Brasil encontra-se mais uma vez, na frente dos processos de produção dos combustíveis renováveis, podendo destacar-se ainda mais, pela utilização de recursos como sua vasta extensão territorial e clima tropical, propícios ao cultivo de oleaginosas para a produção de biodiesel.

O primeiro registro, do que hoje conhecemos como biodiesel, trata-se de uma patente belga de 31 de agosto de 1937. O seu autor, Charles Chavanne descreve a utilização de óleo vegetal de palma e metanol no emprego de reação de transesterificação, utilizando ácido sulfúrico como catalisador na produção de ésteres metílicos a serem utilizados como combustíveis em motores diesel (NÓBREGA, 2008).

No Brasil, com o lançamento do programa PROBIODIESEL, em 2002, este passa assumir lugar ímpar no planejamento estratégico da matriz energética brasileira. O Programa Nacional de Produção e uso de Biodiesel (PNPB) ou PROBIODIESEL é um programa interministerial do Governo Federal que objetiva (PUERTO, 2007):

- i) a implantação de forma sustentável, tanto técnica, como econômica, da produção e uso de um combustível renovável;
- ii) a diversidade da matriz energética;
- iii) a redução das importações de diesel e petróleo;
- iv) a inclusão social e o desenvolvimento regional, via geração de empregos e renda;
- v) fixar as famílias no campo e expandir a agricultura familiar sustentável;
- vi) o uso de diferentes fontes oleaginosas, dependendo da região, e aproveitando solos inadequados para a produção de culturas alimentares.

O programa prevê a adoção gradual do óleo vegetal na matriz brasileira de combustíveis, começando com a mistura de 2% (B2, como foi batizada) em 2008, até chegar a 100% (B100) em 2013 (OLIVEIRA et al.,2004). Com este programa, o Governo Federal definiu o marco regulatório para a produção e distribuição do biodiesel no Brasil, publicando os atos legais que definem: o percentual de mistura do biodiesel ao óleo diesel, as especificações, o regime tributário com diferenciação por região de plantio, por oleaginosa e por categoria de produção (agronegócio e agricultura familiar), criando ainda o selo Combustível Social e a figura do produtor de biodiesel, além de estruturar a cadeia de comercialização (SANTOS, 2007). Estima-se que o Brasil reduza em 33% suas importações de diesel, gerando uma economia anual de US\$ 350 milhões, além de um grande número de empregos diretos e indiretos.

A utilização do biodiesel traz uma série de vantagens ambientais, econômicas e sociais (Conceição et al.,2007). Isto porque a substituição do óleo diesel mineral pelo biodiesel resulta em reduções de emissões de: 20% de enxofre, 9,8% de anidrido carbônico, 14,2% de hidrocarbonetos não queimados, 26,8% de material particulado e 4,6% de óxido de nitrogênio. Os benefícios ambientais podem, ainda, gerar vantagens econômicas. O País poderia enquadrar o biodiesel nos acordos estabelecidos no Protocolo de Kyoto e nas diretrizes dos mecanismos de desenvolvimento limpo, já que existe a possibilidade de venda de cotas de carbono através do Fundo Protótipo de Carbono, pela redução das emissões de

gases poluentes e também créditos de “sequestro de carbono”, através do Fundo Bio de Carbono, administrado pelo Banco Mundial.

Existem algumas limitações na utilização de determinados óleos vegetais como matéria prima na produção de biodiesel. Estas limitações se devem as características indesejáveis que estes óleos apresentam, tais como: elevado grau de instauração e alta viscosidade, as quais são transferidas para o biodiesel (RAMOS, 2003). Sendo assim, tomando como base as informações da Companhia Nacional de Abastecimento (CONAB), as matérias-primas oleaginosas que mais se destacam, quer seja pelo teor de óleo, qualidade do óleo ou pela produtividade são: a mamona, o girassol, o amendoim, o gergelim, a canola, o dendê, a soja, o algodão e o babaçu (SOUZA, 2005).

A mamona vem se destacando como uma das oleaginosas fornecedoras de matéria prima para a fabricação de biodiesel. Isso se deve ao fato de suas sementes conter óleo com características químicas que lhe tornam único na natureza. O principal componente do óleo da mamona é o ácido ricinoléico ($C_{18}H_{34}O_3$), cujas moléculas têm propriedade e estrutura incomum entre os ácidos graxos existentes nos óleos vegetais (MOSHKIN, 1986). São estas características peculiares ao óleo da mamona que permitem seu uso em mais de 400 processos industriais, tais como: na produção de anticongelantes de combustível de aviões e espaçonaves, revestimento e paredes e poltronas de aviões (na queima não libera gases tóxicos), componentes de automóveis, lubrificantes, tintas, resinas, cosméticos (NÓBREGA, 2008). Nos últimos anos, novos produtos têm sido desenvolvidos, como por exemplo: próteses ósseas e medicamentos diversos (VIEIRA; LIMA, 2008).

Por não ser comestível, por ser solúvel em álcool e também devido a outras propriedades físicas e químicas, o óleo presente nas sementes de mamona é ideal para produção de biodiesel. A solubilidade e a viscosidade, por exemplo, são características que permitem menor consumo de energia na transformação. Além destes fatores econômicos, características agrônômicas e sociais têm sido consideradas na escolha deste óleo e não de outros, como exemplo cita-se a cultura apresenta resistência à seca e a fixação do pequeno agricultor no campo (PARENTE, 2003).

2.3 A cultura de mamona

2.3.1 Caracterização botânica

A mamona (*Ricinus communis L.*) é uma planta da família das *Euphorbiaceae*, e popularmente chamada de mamona, carrapateira, rícino e palma-de-cristo. Conforme Moshkin (1986), a espécie *Ricinus communis L* é a única do gênero, entretanto autores como Savy Filho (1999), dentre outros, consideram a ocorrência de subespécies e formas, que cruzam entre si e produzem descendentes férteis.

Grande variabilidade é observada em plantas de mamona nos caracteres morfológicos e agronômicos, tais como: hábito de crescimento, cor das folhas, do caule, dos ramos e dos frutos, tamanho das sementes, teor de óleo, altura das plantas, o que permite distinguir um material *genético* de outro com relativa facilidade (NÓBREGA, 2008). Alguns tipos são perenes e se desenvolvem como pequenas árvores e outras, como anuais anãs, e uma enorme graduação entre os extremos é observada (WEISS, 1983). Botanicamente, a mamona é considerada perene (SAVY FILHO, 1999), mas é explorada comercialmente como uma cultura anual (BALDANZI et al.,2003)

A mamona é uma espécie de crescimento indeterminado, no sentido da emissão de inflorescências de várias ordens e idades fisiológicas, que traz problemas na colheita mecanizada, em especial em cultivares que têm frutos que abrem na maturidade (SILVA et al.,2005).

Nos ramos surgem as inflorescências em cachos, contendo flores masculinas na parte inferior e flores femininas na parte superior que, quando fecundadas, se desenvolvem em frutos espinhosos, compostos de três lojas. A produtividade da plantas está condicionada ao número de cachos, número de sementes por cacho e peso de sementes. Quanto maior o período de tempo disponível para a planta com condições ambientais favoráveis, maior será o número de cachos e, portanto, maior a produtividade (AVELAR et al.,2006).

O desenvolvimento das ramificações é importante fator de produção, uma vez que cada ramo vai formar um rácemo de mamona. O florescimento da mamoneira é chamado

botanicamente de simpodial, com determinado intervalo entre a emissão das inflorescência primárias e secundárias (BALDANZI et al.,2003).

O desenvolvimento da planta depende das condições locais e da variedade utilizada. Em geral, o sistema radicular é vigoroso, do tipo pivotante, profundo, com o desenvolvimento de poucas raízes laterais, porém, de aspecto robusto, assumindo a forma do sistema radicular dos pequenos arbustos. Há forte emissão de radículas ao longo das raízes, conferindo grande área de absorção de umidade e nutrientes no solo (SILVA et al.,2005).

2.3.2 Frutos, sementes e cultivares

O fruto da mamona é uma cápsula (ovário fecundado e desenvolvido) lisa ou com estruturas semelhantes a espinhos, podendo ser deiscente ou indeiscente (BELTRÃO; AZEVEDO, 2007).

A deiscência é a abertura da cápsula seca com a liberação, na mamona, de três sementes. A colheita dos frutos de cultivares deiscentes deve ser feita antes da ocorrência desse fenômeno, com a complementação da secagem do fruto sob controle (PEIXOTO, 1973).

No melhoramento genético de cultivares deiscentes, buscam-se as semi-deiscentes, ou seja, sementes, que permanecem por um período maior na planta, sem se desprender do fruto, o que permite a coleta dos ramos com maior percentagem de frutos secos, conseqüentemente reduzindo o tempo de secagem natural no terreno (SAVY FILHO, 2005).

Os cultivares de frutos deiscentes são mais indicados para pequenos e médios produtores, visto que todas as operações, manuais e mecanizadas, podem ser feitas durante o desenvolvimento da cultura. Esta exige colheita parcelada dos ramos, à medida que vão secando, na média de quatro repasses de colheita, variando com o ciclo da cultivar. A secagem é natural, no terreno, sendo as sementes liberadas do fruto após um período de 3 a 5 dias. Algumas cultivares que possuem estas características são: IAC-80, BRS-149 Nordestina, BRS-188 e Paraguaçu (SAVY FILHO *et al.*,1999; INSTITUTO CULTURAL PADRE JOSINO, 2008).

Os cultivares de frutos indeiscentes não liberam a semente depois de seco, nem na planta e nem no terreno, mantêm a semente em seu interior. A colheita é única e não ocorre perda por deiscência (SAVY FILHO, 2003). Essas cultivares necessitam ser descascadas mecanicamente, com máquina específica para mamona. Algumas cultivares que possuem estas características são: Guarani, IAC 266, Al Guarany 2002 (SAVY FILHO et al., 1999; INSTITUTO CULTURAL PADRE JOSINO, 2008).

Quanto às sementes, apresentam grande variabilidade de: cor, forma, tamanho, peso, proporção do tegumento, presença ou ausência de carúncula e maior ou menor aderência do tegumento ao endosperma (MAZZANI, 1983). A semente de mamona, em média, é constituída por 35% de casca e 65% de amêndoa. Sua composição química varia com a cultivar e a região de cultivo. Entre as frações químicas componentes da semente, o óleo responde por cerca de 50% do grão (AZEVEDO; LIMA 2001).

As sementes apresentam grande variação de tamanho e densidade, o que acarreta em dificuldades para a regulação de semeadoras e desuniformidade na emergência das plantas (SHEPENTINA; SEVAST'YANOVA, 1986), o que resulta em populações desiguais, levando a queda na produtividade. Apresentam ainda, problemas de dormência, além de baixa qualidade fisiológica das sementes disponíveis no mercado. Santos; Barros (2003) recomendam adquirir sementes de produtor idôneo e detentor de registro no Ministério da Agricultura.

No armazenamento das sementes, estas são acondicionadas em sacos de algodão ou fibra de *nylon* para comercialização. A umidade aceitável das sementes deve estar entre 6 e 10%, uma vez que neste nível, a proliferação de fungos é inibida, assim como as atividades enzimáticas inerente ao metabolismo das sementes (INSTITUTO CULTURAL PADRE JOSINO, 2008).

2.3.3 Aspectos gerais da cultura

A origem da mamona é muito discutida, já que existem relatos, em épocas bastante longínquas, de seu cultivo na Ásia e na África (CHIERICE; NETO, 2001). No Brasil, a

mamona foi trazida pelos portugueses com a finalidade de utilizar seu óleo para iluminação e lubrificação dos eixos das carroças.

A mamoneira é cultivada comercialmente em mais de 15 países, sendo os principais produtores: a Índia, a China e o Brasil (VIEIRA; LIMA, 2008). O Brasil é o terceiro produtor mundial e tem a capacidade de aumentar rapidamente sua participação no mercado, pois dispõe de área para aumentar o plantio (NÓBREGA, 2008). Atualmente a Índia lidera a produção de óleo de mamona, sendo responsável por mais de 60% de todo o óleo comercializado no mundo (INDIAMART, 2009).

Entre os aspectos agronômicos mais importantes da cultura, destacam-se: a grande adaptabilidade, rusticidade, a resistência ao déficit hídrico (WEISS, 1983), a possibilidade de cultivo consorciado com outras culturas, a utilização da torta como fertilizante ou na alimentação animal e o excelente rendimento (45 a 48%) e qualidade do óleo produzido.

A maioria da produção é obtida pela família do pequeno e do médio agricultor. O tamanho médio da lavoura de mamona pode ser estimado entre um a dez hectares, desenvolvido com a mão de obra familiar. Geralmente é consorciada com culturas alimentícias, tais como: feijão, milho, arroz e abóbora, para otimizar o uso da terra, obtendo-se dessa forma receita de dois cultivos (INSTITUTO CULTURAL PADRE JOSINO, 2008).

A comercialização da mamona pode ser feita tanto na forma bruta e de pouco valor agregado (mamona em baga), quanto em formas intermediárias (óleo bruto ou refinado) ou, através da exploração de seus derivados de alto valor agregado (ácido graxo destilado de óleo de mamona desidratado, óleo de mamona hidrogenado, entre outros) (KOURI; SANTOS, 2006).

2.3.4 A cultura no Rio Grande do Sul

A mamona é uma planta de clima tropical e por isso prefere locais onde a temperatura do ar varia entre 20 e 30 °C, com precipitações pluviais de pelo menos 500 mm, elevada insolação e umidade relativa do ar baixa (menor do que 60%) durante maior parte do seu ciclo (BELTRÃO et al., 2003).

A produção brasileira de mamona está concentrada na região nordeste principalmente no Estado da Bahia. Dados do IBGE (safra 2003/2004), indicam que este Estado é o principal produtor nacional de mamona, com cerca de 149,5 mil hectares plantados (90% da área cultivada do País) e uma produção estimada em de 134,9 mil toneladas (89% da produção nacional) (KOUR; SANTOS, 2006).

O cultivo de mamoneira pode ser realizado em todo o Brasil, excluindo-se alguns ecossistemas de baixa altitude e locais muito frios. Embora a maior parte da produção nacional concentre-se na região nordeste, as regiões sul e sudeste apresentam elevadas produtividades de grãos, devido, principalmente, à disponibilidade de água e adequado manejo cultural, bem como, pela utilização de variedades de porte anão e colheita única (BELTRÃO et al.,2003).

A mamona exige uma estação quente e úmida para favorecer a fase vegetativa e uma estação seca para permitir condições favoráveis na colheita. Tais exigências climáticas são verificadas em alguns meses do ano (setembro a outubro) em Estados de clima subtropical, como o Rio Grande do Sul, possibilitando seu cultivo, com a obtenção de elevadas produtividades (SÃO PAULO, 1977).

O Rio Grande do Sul possui condições climáticas adequadas para a produção de mamona, por estar localizado nos paralelos 27° e 33° do hemisfério sul, o que proporciona condições semelhantes aos Países de melhor e maior produtividade (Índia e China). No Estado, ao longo de 23 anos de experimentos demonstrativos, foram observados médias de produtividade de 2700 kg ha⁻¹(COMISSÃO ESPECIAL DE BIOENERGIA-RS, 2006). Neste sentido, Kouri; Santos (2006) argumentam que nas regiões Sudeste e Sul, para garantir a competitividade da cultura de mamona com outros produtos concorrentes, tornou-se necessário a geração de técnicas que facilitassem a mecanização e o desenvolvimento de variedades mais rentáveis. Deste modo, tornou-se possível cultivar variedades de porte anão e médio (indeiscentes), cuja maturação ocorre aproximadamente ao mesmo tempo em todas as bagas o que permite colheita mecânica única anual.

Devido a aspectos como rusticidade e tolerância à seca, a mamona se adapta tanto em climas tropicais como em subtropicais. Silva et al., (2004) consideram o Rio Grande do Sul, mesmo na região não havendo tradição no cultivo de mamona, com grande potencial para o cultivo, devido aos seus índices produtivos serem altos, quando comparados com regiões

tradicionais como nordeste. Isso se justifica por que, mesmo o Estado da Bahia com 90% da área plantada, a média de produtividade é baixa, em torno de 780 Kg ha⁻¹. No estado do Rio Grande do Sul, esta média é maior, aproximadamente, 1500 a 2000 kg ha⁻¹ (SILVA et al.,2007).

Com relação às cultivares desenvolvida para atender as características climáticas do Rio Grande do Sul, pesquisadores da Embrapa Clima Temperado relatam que até agora, não existem cultivares de mamona desenvolvidas por um programa específico para o Rio Grande do Sul. O Programa de Melhoramento de Mamona da Embrapa Clima Temperado é recente. As cultivares que estão sendo utilizadas por agricultores e órgãos de pesquisa são introduzidas dos programas de melhoramento da CATI/SP, Embrapa Algodão, IAC/SP e Sementes Armani S/A (SILVA et al., 2007)

Quanto à época de semeadura, no Rio Grande do Sul, é feita em período livre de geadas, buscando o período do ano com temperatura elevadas, enquanto ocorre situação inversa na região nordeste do País, que deve semear a mamona em regiões de altitude, onde a temperatura é mais amena (WREGGE et al.,2006). Ainda, conforme o autor, nas regiões com temperaturas do ar acima de 40°C, pode ocorrer a senescência das flores, prejudicando a produção de frutos e, ainda, provocar a reversão sexual das flores femininas para flores masculinas e reduzir o teor de óleo das sementes.

Apesar dos índices produtivos no Rio Grande do Sul serem maiores do que nos Estados com tradição na ricinocultura, para uma produção em grande escala (biodiesel) será necessário um incremento na produção. Neste sentido, torna-se imprescindível encontrar sistemas de produção adequados, com a correta determinação da população de plantas para otimizar o uso da terra, pois essa poderá influenciar no crescimento da mamoneira, refletindo, assim, na produtividade da planta.

2.3.5 População de plantas

A definição da população ótima é um passo tecnológico básico, fundamental para o bom desempenho de uma cultura em determinada região. A população ótima de uma lavoura depende, basicamente, de quatro fatores: cultivar, umidade, fertilidade do solo e necessidade

do tráfego de animais ou máquinas, para o controle de pragas, de doenças e de plantas daninhas (AZEVEDO et al., 2006).

No plantio da mamona, a definição adequada da população de plantas assume relevada importância, não envolve custos elevados e desempenha grande efeito sobre a produtividade. A população de plantas é definida pelo espaçamento entre linha e a distância da planta dentro da linha, normalmente chamada de densidade. O plantio na configuração correta é uma boa prática agrícola e que resulta em ganhos de produtividade e redução de custos com controle de doenças e pragas (FERREIRA et al., 2006). Plantas muito adensadas, em condições de maior umidade relativa, têm menor ventilação interna na copa e favorecem o ataque de mofo cinzento.

Savy Filho (2005) aconselha não ultrapassar a população de 15.000 plantas ha^{-1} (espaçamento entre linhas de 0,80 a 1,35m e a densidade de plantio de 0,70 a 0,80m). No entanto, alguns híbridos e variedades anãs podem ser cultivados em populações tão elevadas quanto 50.000 plantas ha^{-1} , desde que se tenha o controle do fornecimento da água. Espaçamentos de 0,70 a 1,0m por 0,35 a 0,90m são recomendados com esse tipo de mamoneira. Alguns técnicos têm verificado produtividades de 900 a 1800 kg ha^{-1} no plantio de híbridos, no cerrado do Mato Grosso, no espaçamento de 0,90 a 1,00m entre linhas e 0,35 m entre plantas (RANGEL et al., 2003).

Moraes et al. (2006) conduziram um experimento para verificar a influência de cinco espaçamentos entre linhas, variando de 1,5 a 3,5m, sobre a altura das plantas e a produtividade de uma variedade de mamoneira de porte médio. Os autores concluíram que nenhuma das características avaliadas foi influenciada pelo espaçamento.

O espaçamento de 3m entre linhas tem sido recomendado para as cultivares BRS Nordestina, para diversos ambientes, mas há indícios de que o adensamento populacional possa proporcionar o aumento de produtividade (SEVERINO et al., 2006).

A capacidade produtiva de uma cultura é avaliada através de fatores que estão diretamente ligados ao crescimento. Dentre estes, a densidade de cultivo assume relevada importância, influenciando na capacidade produtiva da planta. A relação da densidade sobre o crescimento da planta pode ser estudada através da metodologia de análise de dados longitudinais, por meio de construção de curvas de crescimento.

2.4 Caracterização de dados Longitudinais

Na pesquisa com experimentos, a metodologia de análise estatística dos dados é selecionada de acordo com a estrutura destes, a qual poderá apresentar-se na estrutura univariada ou multivariada. A estrutura univariada é obtida quando para cada unidade experimental somente uma variável resposta é avaliada. Quando a estrutura é multivariada tem-se a avaliação de mais de uma variável resposta numa mesma unidade experimental. Outra forma de estrutura multivariada é obtida quando para a mesma unidade experimental avalia-se uma variável, em diferentes condições de avaliação, constituindo dessa forma a estrutura de dados denominada de dados longitudinais.

Estudos longitudinais têm sido considerados em diversas áreas, sob diferentes denominações: estudo de *coortes*, em pesquisas sociológicas ou biomédicas, onde também é denominado estudo do tipo *follow-up* e *painel* em amostragem (DWYER; FEINLEIB, 1992).

De uma forma geral, estudos longitudinais envolvem observações de um conjunto de unidades de investigação (animais, por exemplo) classificados em diferentes subpopulações, segundo um ou mais fatores ou tratamentos (tais como: raça, procedência, tipo de ração, entre outros), ao longo de diversas condições de avaliação (tais como: tempo, dose, entre outros), que representam as unidades de observação. Nesse contexto, as unidades de investigação constituem as unidades naturais para a análise e, as unidades de observação correspondentes podem ser essencialmente encaradas como perfis multivariados de resposta associados às mesmas (RIBOLDI, 1999).

Estudos longitudinais constituem um caso especial dos estudos de medidas repetidas, que abrange o delineamento conhecido como parcelas subdivididas e crossover. Os estudos envolvendo dados longitudinais têm como objetivo descrever as mudanças nas respostas individuais à medida que o tempo passa. Além de investigar a influência de outros fatores ou covariáveis no padrão temporal da resposta (WALD, 2000). Resende et al. (1999) complementam que a interação época x tratamento é, possivelmente, o mais importante aspecto destes planejamentos, e quando significativa, indica uma tendência não paralela da resposta em relação ao tempo para diferentes níveis de tratamentos.

A característica distinta de estudos longitudinais é a dimensão ordenada com que os dados são coletados e o fato que as observações repetidas para um indivíduo tendem a ser correlacionadas. Tal correlação pode ser modelada através de uma estrutura de covariância dos dados observados, sendo que, para outros tipos de dados, é usual assumir que os erros sejam independentes (COSTA, 2003).

Riboldi (1999) entende como medidas repetidas, os dados gerados por um número de unidades de investigação repetidamente sobre diferentes condições de avaliação, assumindo-se que as unidades de investigação constituem uma amostra aleatória de uma população de interesse. O autor complementa relatando que um tipo comum de medidas repetidas são os dados longitudinais. Mais especificamente, dados longitudinais podem ser definidos como medidas repetidas, em que as observações, dentro das unidades de investigação, não foram ou não podem ter sido atribuídas aleatoriamente para as diferentes condições de avaliação (usualmente tempo ou posição no espaço).

O termo medidas repetidas refere-se aos conjuntos de dados em que são feitas múltiplas medidas de uma variável resposta na mesma unidade experimental. Grande parte do desenvolvimento da metodologia de medidas repetidas ocorreu na área da psicologia humana. Como resultado, as unidades experimentais são frequentemente chamadas de indivíduos, mas o sujeito pode se referir a um animal, uma planta, uma amostra de laboratório ou um pedaço de equipamento industrial (LITTELL et al., 2006).

Os estudos com dados longitudinais têm como principais vantagens os seguintes fatos (VENEZUELA, 2003):

- i. requerem menos unidades experimentais do que planejamentos completamente casualizados;
- ii. proporcionam condições mais adequadas para o estudo de covariáveis que possam ter influencia na variável resposta;
- iii. melhoram, em geral, a precisão de contrastes associados às diferenças entre os valores médios da variável resposta de diferentes tratamentos; e,
- iv. permitem o estudo da mudança do comportamento da resposta média da unidade experimental nos diferentes tratamentos (incorpora informação sobre a variação individual da análise).

Os dados associados a cada unidade experimental podem ser expressos em termos de um vetor, com valores observados das variáveis respostas em cada instante e, de uma matriz, com os correspondentes valores das m variáveis explanatórias ou covariáveis. O vetor com as p_i respostas da i -ésima unidade experimental ($i= 1, \dots, n$), chamado de perfil individual, pode ser expresso como $Y_i = (y_{i1}, \dots, y_{ip_i})^t$ em que y_{ik} representa o valor observado da variável resposta da i -ésima unidade experimental no t -ésimo instante de observação ($t=1, \dots, p_i$). A matriz de variáveis explanatórias, de dimensões $p_i \times m$ pode ser expressa como $X_i = (x_{i1}, \dots, x_{im})$, em que: $x_{ij} = (x_{ij1}, \dots, x_{ijp_i})^t$; com x_{ijt} , o valor da j -ésima variável explanatória ($j= 1, \dots, m$) para a i -ésima unidade experimental no t -ésimo instante de observação (ARAÚJO, 2007; SINGER et al., 2008).

Uma das características mais importantes para análise de dados longitudinais é o método de obtenção dos dados, isto é, o planejamento da pesquisa, visto que a metodologia estatística é, em parte, determinada por este planejamento. Os dados devem ser obtidos de forma que cada unidade experimental seja avaliada várias ocasiões e sob diferentes condições, definidas no planejamento do estudo. Baseado no número de condição de avaliação em que cada unidade experimental é observada, as três principais categorias de planejamento são (ANDRADE; SINGER, 1986; CASTRO, 1997):

- i. transversal ou *cross-sectional*: em que se planeja observar cada unidade experimental em uma única condição, o que irá gerar um vetor de respostas de dimensão igual a um.
- ii. longitudinal: em que se planeja observar cada unidade experimental em todas as condições de avaliação para as quais o estudo foi dimensionado. Desta forma irá gerar um vetor de resposta maior do que um.
- iii. misto ou *linked cross-sectional*: em que se planejam observar diferentes grupos de unidades experimentais em diferentes subconjuntos das condições de avaliação para as quais o estudo foi planejado.

Devido à natureza dos estudos longitudinais, que podem durar meses ou até anos, e requerem observações nas mesmas unidades amostrais em múltiplas ocasiões, torna-se frequente a existência de dados incompletos, não balanceados, e irregulares no tempo (LITTELL et al., 2006)

Os dados longitudinais são regulares no tempo se o intervalo entre duas medidas consecutivas quaisquer for constante ao longo do estudo, caso contrário, serão irregulares no tempo (SUYAMA, 1995). Em relação ao balanceamento, são ditos balanceados com relação ao tempo se as observações forem nos mesmos instantes de tempo em todas as unidades amostrais, de outro modo, serão não balanceados em relação ao tempo. Quando houver observações perdidas, então a estrutura dos dados é incompleta e se não houver observações perdidas, então a estrutura dos dados é completa.

Em situações em que se têm dados longitudinais em uma estrutura balanceada e completa tem-se a configuração dos dados da Tabela 1 (CASTRO, 1997; RIBOLDI, 1999).

Tabela 1 – Configuração dos dados para uma estrutura balanceada e completa de dados longitudinais. Santa Maria, RS, 2010.

Tratamentos ou grupos	Unidade experimental	Condições de avaliação		
		1	2	t
1	1	Y_{111}	Y_{121}	Y_{11t}
1	2	Y_{121}	Y_{122}	Y_{12t}
.
.
.
1	n_1	Y_{1n1}	Y_{1n2}	Y_{1nt}
2	1	Y_{211}	Y_{212}	Y_{21t}
2	2	Y_{221}	Y_{222}	Y_{22t}
.
.
.
2	n_2	Y_{2n21}	Y_{2n22}	Y_{2n2t}
.
.
.
g	1	Y_{g11}	Y_{g12}	Y_{g1t}
g	2	Y_{g21}	Y_{g22}	Y_{g2t}
.
.
.
g	n_g	Y_{gn_g1}	Y_{gn_g2}	Y_{gn_gt}

Em que: y_{ijk} é o valor da variável resposta da j -ésima unidade experimental dentro do i -ésimo tratamento, sob a k -ésima condição de avaliação para $i = 1, 2, \dots, g$; $j = 1, 2, \dots, n_i$; $k = 1, 2, \dots, t$;

$$\sum_{i=1}^g n_i = N.$$

2.5 Metodologias de análise para dados longitudinais

A análise de dados longitudinais é um conjunto de técnicas estatísticas em que os históricos individuais, interpretados como realizações de um processo estocástico, são os focos primários de interesse (ANDRADE; SINGER, 1986). As unidades experimentais (pessoas, animais ou parcelas de campo) podem ser classificadas em diferentes grupos, segundo um ou mais fatores ou tratamentos, tais como: espécie, sexo, tipo de adubação, entre outros. As respostas podem envolver variáveis categorizadas (presença ou ausência de um sinal clínico, por exemplo) ou contínuas (fitomassa seca, peso corporal, área foliar, entre outros).

De uma maneira geral, a análise de dados longitudinais tem os seguintes objetivos (RIBOLDI, 1999):

- i. comparação das diferentes subpopulações quanto ao padrão de variação das respectivas distribuições de respostas ao longo das diferentes condições de avaliação, isto é, verificar se os perfis são paralelos, ou seja, a interação entre os tratamentos e o tempo é nula.
- ii. comparação das diferentes subpopulações quanto às respectivas distribuições das médias (em relação às diversas condições de avaliação) de respostas, isto é verificar se os perfis são coincidentes, ou seja, verificação de efeitos de tratamentos.
- iii. comparação das diferentes subpopulações quanto às respectivas distribuições das médias (em relação às diversas subpopulações) de respostas, isto é, verificar se os perfis são horizontais, ou seja, verificação de efeitos das condições de avaliação.
- iv. ajuste de modelos para explicar a variação das respostas médias como função do fator que define as condições de avaliação (ajuste de curvas).

Vários métodos têm sido propostos para análise de experimentos com dados longitudinais, os quais diferenciam-se: no esquema de aplicação dos tratamentos, nos objetivos da pesquisa, no enfoque quanto a estrutura dos dados e no modelo estatístico adotado.

Os modelos para dados longitudinais podem ser classificados como incondicionais ou condicionais (SINGER *et al.*, 2008). Os primeiros são aqueles em que o valor esperado da variável resposta, $E(y_{ik})$, é expresso exclusivamente em termos de variáveis explicativas (x_{i1k}, \dots, x_{imk}). Se, entre elas o tempo é tomado como uma variável discreta que indica a ordem em que a resposta é observada em cada unidade amostral sob um planejamento balanceado, os modelos correspondentes são conhecidos como modelos de perfis e, são equivalente aqueles costumeiramente considerados em análise de variância (ANOVA) e análise de covariância (ANCOVA). Nos casos em que o tempo é considerado uma variável contínua, isto é, em que o interesse recai na sua relação funcional com a variável resposta, os modelos correspondentes são denominados modelos de crescimento. O autor finaliza, explicando que os modelos condicionais, são aqueles em que a relação entre a variável resposta e as variáveis explicativas num certo instante é condicionada a valores prévios da resposta. As técnicas clássicas de análise de dados longitudinais são geralmente dirigidas para caso de dados completos e balanceados em relação ao tempo, destacando-se (LIMA, 1996):

- i. Análise Multivariada de Perfis e Análise Univariada de Perfis;
- ii. Análise de Curvas de Crescimento (que também pode ser realizada através de modelos uni e multivariado e modelos mistos).

A análise de perfil é uma técnica clássica de análise de dados longitudinais que pode ser realizada utilizando-se técnicas multivariadas ou univariadas e tem como principal objetivo testar hipóteses sobre os valores médios da variável resposta nas diversas ocasiões de observação e compará-los entre os diferentes grupos de unidades experimentais, ou tratamentos (AUBIN, 1984).

Quando o objetivo da pesquisa é estudar o crescimento, o modelo matemático usado nessas situações consiste na descrição do comportamento da variável em função do tempo, através de uma curva, geralmente de forma polinomial. Comparações entre efeitos de tratamentos diferentes podem ser estudadas através de comparações dos parâmetros das respectivas curvas (CASTRO, 1999; SINGER, 1977).

O ajuste de curvas de crescimento pode ser realizado através da análise de perfis, utilizando um modelo univariado, de acordo com um delineamento de parcelas subdivididas, que impõe forte restrição quanto à matriz de variâncias-variância-covariância. Alternativamente, pode ser usado um modelo multivariado, que utiliza uma matriz de variâncias-variância-covariância sem restrições, chamada não-estruturada (SCHEINER; GUREVITCH, 2001).

A análise de curvas de crescimento são técnicas de análise de dados longitudinais nas quais descreve-se o comportamento da variável de interesse em função do tempo, através de curvas, e compara-se o efeito dos tratamentos através da comparação dos parâmetros das respectivas curvas (ANDRADE; SINGER, 1986). Isto é, a diferença entre dois tratamentos quaisquer pode ser avaliada entre os dois coeficientes angulares e lineares, respectivamente. Além disso, o ajuste de curvas possibilita fazer previsões para o comportamento médio das respostas ao longo do tempo, utilizando, em geral, poucos parâmetros.

Em modelos de curvas de crescimento, ou polinomiais, parte-se do princípio que existe uma relação funcional entre os efeitos de tratamentos e o tempo de aplicação, e que esta relação possa ser modelada. A função pode ser aproximada por um polinômio, devendo os coeficientes dessa representação polinomial, bem como, variâncias e variância-covariância, serem estimadas através dos dados (XAVIER, 2000).

Lima (1988) destaca algumas das vantagens no ajuste de curvas:

- i. possibilitar que sejam feitas previsões sobre o comportamento médio da resposta, para ocasiões diferentes daquelas para as quais o estudo foi planejado;
- ii. possibilitar que sejam feitas comparações mais sensíveis entre as subpopulações, em função do número reduzido de parâmetros;
- iii. possibilitar a aquisição de um maior conhecimento sobre o processo de variação temporal da resposta.

A análise de curvas de crescimento pode ser realizada comparando-se curvas individuais (coeficientes aleatórios) ou curvas médias. Normalmente, consideram-se as observações em indivíduos sob o mesmo tratamento como repetições, ajustando-se o modelo para a média (BRANDÃO, 1996). Este é o caso do modelo proposto por Potthoff; Roy (1964), que utilizam técnicas de análise de variância multivariada, descrevendo o comportamento da variável resposta através de modelos lineares.

Outra forma de ajustar curvas de crescimento, é através de modelos mistos lineares ou não-lineares, que possibilitam o uso de diferentes tipos de estruturas para as matrizes de variâncias-variância-covariância, optando-se por aquela que melhor representa a estrutura de correlação dos dados, e também permitem descrever o comportamento dos perfis médios através de curvas.

Um dos problemas no ajuste de modelos mistos lineares a um conjunto de observações é a metodologia mais adequada para a estimação dos efeitos fixos. Dado que em medidas repetidas a estrutura de variância-covariância entre tempos é uma estrutura dos efeitos fixos, usa-se informá-la na estrutura de erros, como pode ser visto em VERBEKE; MOLENBERGHS (1997) e LITTELL et al. (1998).

Para a estimação dos efeitos fixos ou o modelo que descreve o comportamento da curva média, as técnicas clássicas de análise são (SUYAMA, 1995): Análise univariada (ou Modelo Univariado de Perfis), em que os dados longitudinais são considerados observações tomadas em subdivisões das parcelas, geralmente exigindo que a variância da resposta seja constante nas ocasiões de avaliação e que as variância-covariância entre as respostas de diferentes ocasiões sejam iguais; e, Análise multivariada (ou Modelo Multivariado de Perfis), em que admite-se que estas variâncias e variância-covariância possam ser distintas.

Desta forma, a seguir será apresentada uma revisão sobre as metodologias para análise de dados longitudinais, objetivando o ajuste de curvas de crescimento através de modelos mistos.

2.5.1 Análise univariada de perfil

A análise univariada de perfil é uma das técnicas mais difundidas na análise de dados longitudinais, baseia-se no número total de observações e requer algumas pressuposições na estrutura de variâncias e covariância entre as observações. Esta técnica utiliza o modelo de parcelas subdivididas, em que os dados longitudinais representam as observações tomadas em subdivisões das parcelas, considerando os tratamentos e o fator tempo, como efeito de parcela e subparcela, respectivamente. Sendo assim, os tratamentos e o tempo (as condições de

avaliação) são considerados como dois fatores fixos e cruzados e, as unidades experimentais, como um fator aleatório hierárquico dentro dos tratamentos.

Aspectos básicos da análise de perfis por técnicas univariadas foram estudados por vários autores. Alguns sugerem o uso do esquema em parcelas subdivididas, considerando o tempo como subparcelas, o que pode causar problemas, pois, como se sabe, este esquema pressupõe que a estrutura da matriz de variância-covariância (Σ) satisfaça a condição de esfericidade, o que nem sempre acontece. O que se encontra na literatura é que medidas repetidas em uma mesma unidade experimental ao longo do tempo são, no geral, correlacionadas e que essas correlações são maiores para tempos mais próximos (MALHEIROS, 1999).

Quando um experimento é sistematicamente arranjado, sem aleatorização, ao se analisar no esquema de parcelas subdivididas no tempo, pode-se inflacionar a probabilidade de falsamente rejeitar a hipótese nula (erro tipo I) (FERNANDEZ, 1991).

Para o modelo univariado, no esquema de parcelas subdivididas, são feitas pressuposições de que, tanto o erro da parcela, que engloba o fator tratamentos ou grupos, como o erro da subparcela, onde são alocados os tempos e a interação tratamentos x tempos, tenham distribuição normal, sejam independentes e identicamente distribuídos, com variâncias constantes. O erro da parcela também é conhecido como erro entre indivíduos, e o erro da subparcela como intra-indivíduos (XAVIER; SANTOS, 2001).

As pressuposições descritas anteriormente dizem que os vetores de respostas $Y_{ij} = (y_{ij1}, y_{ij2}, \dots, y_{ijp})'$, $i=1,2,\dots,g$, $j=1,2,\dots,n_i$, são independentes, seguem uma distribuição normal p-variada, com vetor de médias $\mu(p \times 1)$ e matriz de covariância (Σ) igual a:

$$\Sigma_{(p \times p)} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix} \quad (1)$$

$$\text{Em que: } \sigma^2 = \sigma_y^2 + \sigma_e^2 \quad \text{e} \quad \rho = \sigma_y^2 / (\sigma_y^2 + \sigma_e^2)$$

Sendo ρ uma constante que representa a correlação entre dois elementos quaisquer dentro do mesmo vetor de resposta. A matriz (1) é conhecida como uniforme, e segue o

padrão de uniformidade ou tem a forma de simetria composta, em que a variância das respostas em qualquer um dos tempos é igual a $\sigma_y^2 + \sigma_\varepsilon^2$ e, a covariância entre dois tempos quaisquer é igual a σ_y^2 (LIMA, 1996).

Respeitando as pressuposições citadas por Xavier; Santos (2001), e seguindo a estrutura da matriz de covariância dada em (1), o modelo em parcelas subdivididas ou Split-Plot pode ser empregado para analisar dados longitudinais de acordo com a estrutura apresentada na tabela 1, com o modelo matemático (STEEL; TORRIE, 1988):

$$Y_{ijk} = \mu + \tau_i + \pi_{j(i)} + \beta_k + \tau\beta_{ik} + \varepsilon_{ijk}, \quad (2)$$

para $i=1,2,\dots,g$; $j=1,2,\dots,n_i$; $k=1,2,\dots,p$, em que:

Y_{ijk} = efeito observado na j-ésima unidade experimental submetida ao i-ésimo

tratamento medido na k-ésima condição de avaliação;

μ = média geral;

τ_i = efeito do i-ésimo tratamento;

$\pi_{j(i)}$ = efeito aleatório da j-ésima unidade experimental dentro do i-ésimo tratamento;

β_k = efeito da k-ésima condição de avaliação;

$\tau\beta_{ik}$ = efeito da interação entre o i-ésimo tratamento e a k-ésima condição de avaliação;

ε_{ijk} = representa o erro aleatório da observação Y_{ijk} .

Para os efeitos aleatórios têm-se as seguintes pressuposições:

$$\pi_{j(i)} \cap NIID(0, \sigma_\pi^2) \text{ e } \varepsilon_{ijk} \cap NIID(0, \sigma_\varepsilon^2), \quad (3)$$

Em que: IID significa independente e identicamente distribuída.

As restrições impostas para o modelo são as seguintes:

$$\sum_{i=1}^g \tau_i = \sum_{k=1}^p \beta_k = \sum_{i=1}^g \beta_{ik} = \sum_{k=1}^p \tau\beta_{ik} = 0 \quad (4)$$

Steel; Torrie (1960) sugerem que, quando o delineamento for aleatorizado em blocos, a interação entre blocos x tempos, além do efeito de blocos, seja isolada e acrescentada ao modelo (2), por considerar que existe grande probabilidade dessa interação ser significativa. Para Lima (1996), geralmente os blocos são construídos somente para controlar a heterogeneidade entre unidades experimentais e não se espera uma interação significativa.

Considerando que o modelo (2) é adequado aos dados e que o número de repetições é a mesma para todos os tratamentos. As hipóteses de nulidade associadas aos efeitos principais de tratamento (variação entre indivíduos) e tempo e da interação tratamento x tempo (variação intra-indivíduos), podem ser testados por meio das estatísticas: F_1 , F_2 e F_3 , apresentadas na Tabela 2. Sob as respectivas hipóteses nulas, as estatísticas, F_1 , F_2 e F_3 , seguem distribuições exatas F centrais com $(g-1)$ e $(n-g)$; $(p-1)$ e $(n-g)(p-1)$; $(g-1)(p-1)$ e $(n-g)(p-1)$ graus de liberdade, respectivamente (STEEL; TORRIE, 1988; LIMA, 1996).

As comparações múltiplas podem ser obtidas através dos métodos usuais de análise de variância univariada, porém a variância de cada comparação pode envolver combinações diferentes dos dois resíduos (LIMA, 1996).

Maiores detalhes com relação a estas comparações podem ser encontradas em Steel; Torrie (1988) e Milliken; Johnson (1992).

Tabela 2 – Tabela da análise de variância com estatísticas para os testes de hipótese do modelo de parcelas subdivididas em um delineamento inteiramente casualizado. Santa Maria, RS, 2010.

Fontes de Variação	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	Estatísticas F
Tratamento	$g-1$	SQ_1	$SQ_1/g-1$	$F_1 = \frac{(n-g)SQ_1}{(g-1)SQ_2}$
Erro (a)	$n-g$	SQ_2	$SQ_2/n-g$	
Condições de avaliação	$p-1$	SQ_3	$SQ_3/p-1$	$F_2 = \frac{(n-g)SQ_3}{SQ_5}$
Tratamento x Condições de avaliação	$(g-1)(p-1)$	SQ_4	$SQ_4/(g-1)(p-1)$	$F_3 = \frac{(n-g)SQ_4}{(g-1)SQ_5}$
Erro (b)	$(n-g)(p-1)$	SQ_5	$SQ_5/(n-g)(p-1)$	
Total	$np-1$	SQ_6		

Em que:

$$SQ_1 = \left[\frac{1}{rp} \sum_{i=1}^g y_{i..}^2 \right] - FC; \quad SQ_2 = \left[\frac{1}{p} \sum_{i=1}^g \sum_{j=1}^r y_{i.}^2 \right] - \left[\frac{1}{rp} \sum_{i=1}^g y_{i..}^2 \right];$$

$$SQ_3 = \left[\frac{1}{n} \sum_{k=1}^p y_{..k}^2 \right] - FC; \quad SQ_4 = \left[\frac{1}{r} \sum_{i=1}^g \sum_{k=1}^p y_{i.k}^2 \right] - FC - SQ_1 - SQ_3;$$

$$SQ_5 = SQ_6 - SQ_1 - SQ_2 - SQ_3 - SQ_4;$$

$$SQ_6 = \left[\sum_{i=1}^g \sum_{j=1}^r \sum_{k=1}^p y_{ijk}^2 \right] - FC; \quad FC = \frac{1}{np} y_{...}^2, \text{ onde } n = rg$$

Neste caso, utilizando-se a análise univariada de perfil, as hipóteses de interesse são:

a) a razão entre $F_1 = \frac{(n-g)SQ_1}{(g-1)SQ_2}$, testa a hipótese:

$$H_0 = \tau_1 = \tau_2 \dots = \tau_g = 0$$

$$H_1 = \text{pelo menos um } \tau_i \neq 0$$

Quando a hipótese $H_0 = \tau_1 = \tau_2 \dots = \tau_g = 0$ é rejeitada, indica que existem diferenças verticais entre os tratamentos. Sendo assim, a interpretação geométrica pode ser visualizada na Figura 1.

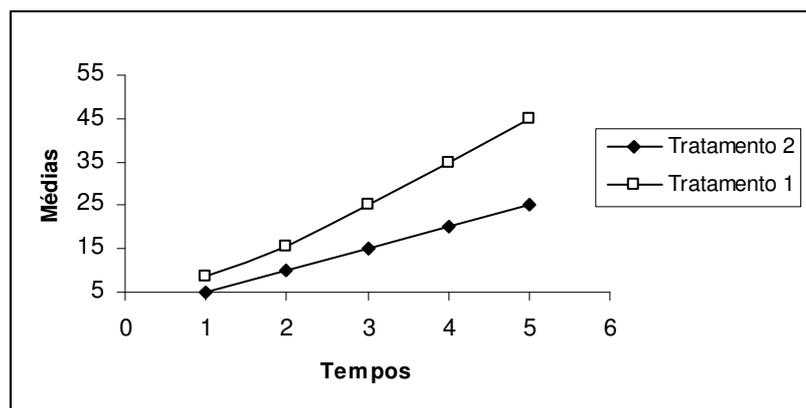


Figura 1 – Diferenças verticais entre as curvas de tratamentos. Santa Maria, RS, 2010.

b) a razão entre $F_2 = \frac{(n-g)SQ_3}{SQ_5}$, testa a hipótese:

$$H_0 = \beta_1 = \beta_2 \dots = \beta_p = 0$$

$$H_1 = \text{pelo menos um } \beta_k \neq 0$$

Quando a hipótese $H_0 = \beta_1 = \beta_2 \dots = \beta_p = 0$ é rejeitada, indica que as curvas não são constantes no tempo. A interpretação geométrica pode ser visualizada na Figura 2.

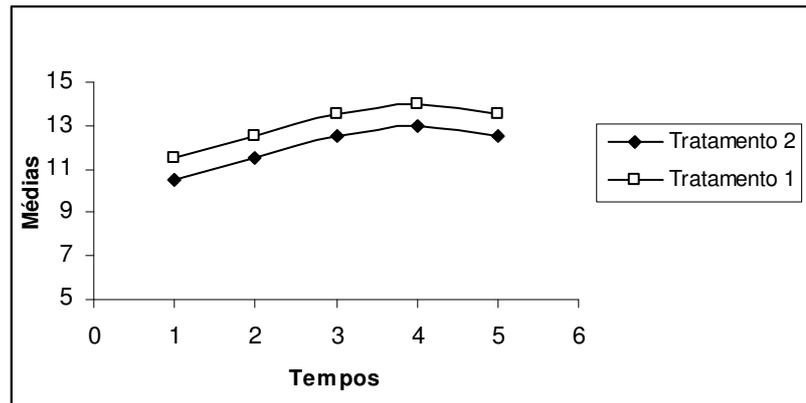


Figura 2 – Curvas não são constantes ao longo do tempo. Santa Maria, RS, 2010.

c) $F_3 = \frac{(n-g)SQ_4}{(g-1)SQ_5}$, testa a hipótese:

$$H_0 = (\tau\beta)_{11} = (\tau\beta)_{12} = (\tau\beta)_{ik} = 0$$

$$H_1 = \text{pelo menos um } (\tau\beta)_{ik} \neq 0$$

Quando a hipótese $H_0 = (\tau\beta)_{11} = (\tau\beta)_{12} = (\tau\beta)_{ik} = 0$ é rejeitada, indica que a forma das curvas difere entre os tratamentos ao longo do tempo. Na Figura 3 encontra-se a interpretação geométrica.

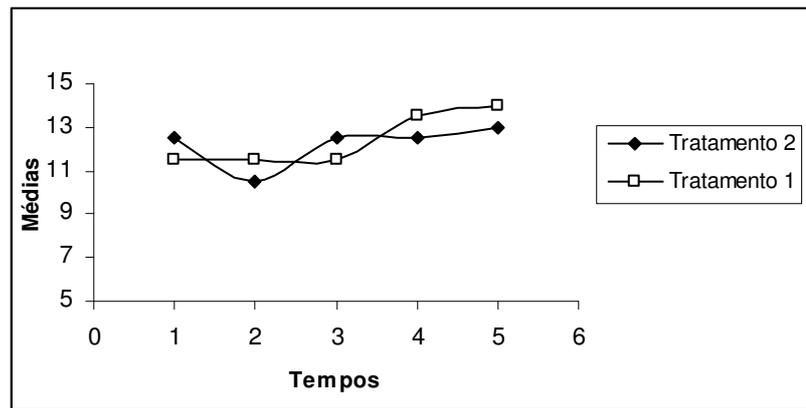


Figura 3 – Curvas diferem entre os tratamentos ao longo do tempo. Santa Maria, RS, 2010.

Geralmente, considera-se 0,05 como nível mínimo para a rejeição da hipótese H_0 , ou seja, sempre que o valor da probabilidade do teste F for menor ou igual a 0,05, aceita-se que há diferença entre os níveis dos fatores. A interpretação dos testes deve ser iniciada pelas interações, primeiramente a dupla. Caso a interação não seja significativa, consideram-se os testes para os efeitos principais.

O primeiro passo do procedimento univariado deve ser a verificação da pressuposição de homocedasticidade, isto é, verificar se a matriz de variâncias e variância-covariância é a mesma para todos os tratamentos. Para isto, pode-se aplicar o teste de homogeneidade de variância de Bartlett. Em seguida, verifica-se a condição de uniformidade da matriz de variâncias e variância-covariância, para justificar o uso deste procedimento. Salienta-se que a análise univariada é válida para testes de efeitos de tratamentos independentemente da uniformidade da matriz de variâncias e variância-covariância, ou seja, aceita-se a pressuposição de homocedasticidade, e a estatística F_1 pode ser utilizada mesmo que a condição de uniformidade não seja satisfeita. A condição de uniformidade interfere apenas nas estatísticas F_2 e F_3 (MILLIKEN; JOHNSON, 1992; CASTRO, 1997).

Castro (1999) ressalta que o padrão de uniformidade, ou simetria composta para a matriz de variância e covariância (Σ), é uma condição suficiente para que as estatísticas F_2 e F_3 tenham distribuição exata F central, porém ela não é uma condição necessária.

A condição de simetria composta implica que a variável aleatória seja igualmente correlacionada e tenha variâncias iguais, considerando as diferentes ocasiões. Uma condição mais geral da forma de Σ é descrita por Huynh; Feldt (1970). Essa condição, denominada

de HUYNH-FELDT (H-F) ou condição de esfericidade (circularidade), especifica que os elementos da matriz da Σ sejam expressos, para um $\lambda > 0$, como:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \frac{(\sigma_1^2 + \sigma_2^2) - \lambda}{2} & \frac{(\sigma_1^2 + \sigma_3^2) - \lambda}{2} & \frac{(\sigma_1^2 + \sigma_4^2) - \lambda}{2} \\ \frac{(\sigma_1^2 + \sigma_2^2) - \lambda}{2} & \sigma_2^2 & \frac{(\sigma_2^2 + \sigma_3^2) - \lambda}{2} & \frac{(\sigma_2^2 + \sigma_4^2) - \lambda}{2} \\ \frac{(\sigma_3^2 + \sigma_1^2) - \lambda}{2} & \frac{(\sigma_3^2 + \sigma_2^2) - \lambda}{2} & \sigma_3^2 & \frac{(\sigma_3^2 + \sigma_4^2) - \lambda}{2} \\ \frac{(\sigma_4^2 + \sigma_1^2) - \lambda}{2} & \frac{(\sigma_4^2 + \sigma_2^2) - \lambda}{2} & \frac{(\sigma_4^2 + \sigma_3^2) - \lambda}{2} & \sigma_4^2 \end{bmatrix} \quad (5),$$

Em que: λ é a diferença entre a média das variâncias e a média das variância-covariância.

A condição H-F é necessária e suficiente para que o teste F da análise de variância usual, no esquema de parcelas subdivididas no tempo, seja válido. Esta condição é equivalente a especificar que variâncias da diferença entre pares de erros sejam iguais, e se as variâncias são todas iguais, então, a condição é equivalente a simetria composta (XAVIER, 2000).

Para verificar a condição de circularidade, Mauchy (1940) apresenta o teste de esfericidade. O teste verifica se uma população multivariada possui variâncias iguais e correlações nulas. A população que apresente essa simetria será denominada de esférica.

O teste de Mauchy é o mais popular para verificar a condição de circularidade. Entretanto, os autores alertam para o fato de que, como todos os testes que envolvem variâncias e variância-covariância, o teste de Mauchy é bastante sensível a não normalidade dos dados (CROWDER; HAND, 1990). Vonesh; Chinchilli (1997) mostram que o teste de esfericidade não é muito poderoso para amostras pequenas e não é robusto quando há violação da suposição de normalidade.

O teste de esfericidade utiliza a condição H-F para a matriz de variância-covariância das t medidas dos indivíduos requeridos nos $(t-1)$ contrastes ortogonais normalizados, para as medidas repetidas não correlacionadas com variâncias iguais.

Pode-se dizer que dois contrastes são ortogonais, quando a soma dos pares de produtos dos coeficientes dos contrastes for igual a zero (contrastos perpendiculares) (CASTRO, 1999; XAVIER; SANTOS, 2001). A ortogonalidade dos contrastes garante que:

- i. cada contraste é associado a uma única porção da variabilidade explicada pelo efeito que se está testando;
- ii. está sendo testado o número máximo de hipóteses, em que cada hipótese é associada a uma única porção da variabilidade explicada pelo modelo;
- iii. o teste é aproximadamente independente.

Para t tempos existem mais de um conjunto de $(t-1)$ contrastes ortogonais, sendo que um contraste ortogonal será normalizado quando for dividido pela sua norma.

Seja a matriz Σ , a matriz de variâncias-variância-covariância das medidas repetidas. A condição requerida por H-F para as variância-covariância dos contrastes é

$$C_{(t-1)t} \sum_{(txt)} C'_{t(t-1)} = \lambda I_{(t-1)(t-1)} \quad (6)$$

Em que:

C : é a matriz dos coeficientes dos contrastes ortogonais normalizados que representa o total de hipóteses nulas;

Σ : é a matriz de variâncias-variância-covariância;

λ = é um escalar maior do que zero; e

I : é a matriz identidade.

Se a condição (7) for satisfeita, a matriz de variância-covariância (Σ) será dita esférica. Segundo Box (1954), Geisser; Greenhouse (1963) e Huynh; Feldt (1976), ainda que a matriz Σ não satisfaça a condição de esfericidade, a distribuição F central poderá ser usada, de maneira aproximada, desde que efetuada uma correção dos graus de liberdade associados às causas de variação que envolvem o fator tempo. As correções nos graus de liberdade dessas fontes de variação são feitas multiplicando-se os graus de liberdade originais por um fator ϵ . Quando a matriz Σ é uniforme, o valor de $\epsilon = 1$. A medida que a matriz Σ se afasta da condição de circularidade, o valor de ϵ vai decrescendo até chegar a $(t-1)^{-1} \leq \epsilon \leq 1$, em que t é o número de tempos.

A utilização destes valores extremos de ϵ pode tornar o teste muito liberal (quando ϵ for igual a 1) ou muito conservador (quando ϵ for igual a $(t-1)^{-1}$). Desta forma, Greenhouse;

Geiser (1959) propuseram uma solução intermediária através de um procedimento de três estágios:

- i. utiliza-se o teste exato F liberal ($\varepsilon = 1$), ou seja, sem qualquer redução nos graus de liberdade. Se o teste não for significativo, aceita-se a hipótese de nulidade e o procedimento é encerrado; caso contrário passa-se para o segundo estágio;
- ii. se o teste exato F liberal ($\varepsilon = 1$) for significativo, utiliza-se o teste exato F conservador ($\varepsilon = (t - 1)^{-1}$), o qual utiliza a maior redução no número de graus de liberdade. Se o teste for significativo rejeita-se a hipótese de nulidade e o procedimento é encerrado; caso o contrário passa-se para o terceiro estágio;
- iii. se o teste exato F conservador não for significativo, utiliza-se o teste aproximado F com ε estimado a partir da matriz de variâncias e variância-covariância amostrais.

A vantagem deste procedimento é que, em muitas situações, apenas no primeiro e segundo estágio, já é possível tomar uma decisão, sem a necessidade de estimar o valor de ε (CASTRO, 1997).

Para Freitas (2007), a correção do número de graus de liberdade deve ser feita apenas em estatísticas que envolvam comparações dentro de sujeitos (fator tempo e interação tempo x tratamento). A estatística envolvendo comparações entre sujeitos não precisam de correções nos graus de liberdade, pois sempre terá distribuição exata F central.

Com a correção do número de graus de liberdade, obtêm-se testes mais conservadores, que são limitados a assegurar que α esteja abaixo de um certo nível, isso para casos em que um teste aproximado não é desejável, e casos no qual a matriz de varianci-variância-covariância é diferente de tratamento para tratamento (MULLER; BARTON, 1989).

Quando a suposição de esfericidade é satisfeita, os fatores de correção $\tilde{\varepsilon}$ e $\hat{\varepsilon}$ são iguais a 1, caso contrário, são menores, mas devem ter um mínimo de $\frac{1}{(t-1)}$, em que t significa o número de medidas repetidas no tempo (KIRK, 2005).

Barcikowski; Robey (1984) recomendam utilizar a correção de GEISSER-GREENHOUSE (ϵ^*) sempre que $\epsilon^* < 0,75$, e a correção de HUYNH-FELDT (ϵ^*), no caso contrário, isto é, sempre que $\epsilon^* \geq 0,75$.

Quanto à escolha de qual correção para o número de graus de liberdade usar, Muller; Barton (1989), depois de vários estudos com simulações, verificando o poder dos testes quando as correções são utilizadas, sugerem que a correção de Geisser-Greenhouse seja utilizada, já que o teste produz aceitável controle do erro tipo I, enquanto maximiza o poder. Porém, segundo Huynh; Feldt (1976), a correção de Geisser-Greenhouse tem a desvantagem de superestimar o verdadeiro nível de significância.

Sendo assim, a análise univariada é recomendada, mesmo que a condição de H-F, para a matriz de variância-covariância, não seja satisfeita, porém, utilizando-se a correção de Huynf-Feldt, desde que o teste de esfericidade seja significativo, com um nível de probabilidade entre 0,01 e 0,05.

Devido ao procedimento univariado ser mais exigente no que se relaciona a estrutura de variância-covariância entre as observações dentro de cada unidade experimental, a análise univariada de perfil, em muitos casos não pode ser utilizada. Desta forma, as técnicas multivariadas são indicadas, pois este tipo de solução é aplicável para uma matriz de variância e covariância (Σ) qualquer. A única exigência do procedimento multivariado é que a matriz Σ deva ser comum a todos os tratamentos.

A seguir serão apresentados alguns aspectos importantes sobre a análise multivariada de perfil.

2.5.2 Análise multivariada de perfil

Devido à natureza essencialmente multivariada dos vetores de respostas, nos estudos envolvendo dados longitudinais, a técnica de análise de variância multivariada, conhecida também como Análise Multivariada de Perfis, constitui uma alternativa natural para o problema em questão (WALD, 2000). A análise multivariada de perfil é bastante discutida na literatura (MORRISON, 1981; ANDRADE; SINGER, 1986; LIMA, 1996, dentre outros).

A análise multivariada de perfis é uma das técnicas estatísticas utilizadas para analisar observações provenientes de experimentos com dados longitudinais. Essa técnica baseia-se tanto no número de unidades experimentais, quanto no tamanho da amostra (CASTRO, 1997).

Diferente do modelo da análise univariada de perfis, o modelo para análise multivariada de perfis não requer que a variância das medidas repetidas ou que a correlação entre pares de medidas repetidas permaneça constante ao longo do tempo. Todavia, ambos os modelos requerem que as variâncias e correlações sejam homogêneas a cada instante do tempo (VIEIRA, 2006).

As técnicas usuais para análise de dados longitudinais impõem pressuposições de que todas as observações sejam independentes. No entanto, a suposição de independência não se aplica aos dados longitudinais, em que as observações feitas no mesmo indivíduo usualmente são correlacionadas. Sendo assim, tratar as observações de forma multivariada seria a maneira adequada para a análise (VONESH; CHINCHILLI, 1997).

Colle; Grizzle (1966) utilizam o procedimento de análise de variância multivariada, segundo a formulação de Smith et al. (1962) e comentam a sua versatilidade para a construção de testes de hipóteses específicos de interesse, que podem ser obtidos como casos particulares do procedimento do teste de hipótese linear geral multivariado. Referem que tais hipóteses podem ser testadas através de três critérios alternativos, todos eles dependentes das raízes características de funções da matriz, devida a hipótese e da matriz, devida ao erro. Os critérios são: da raiz característica máxima, do produto das raízes (critério da razão de verossimilhança) e da soma das raízes. Os autores ilustram a aplicação da análise de variância multivariada a medidas repetidas e mostram que as informações que se desejam obter destes experimentos podem ser formuladas em termos das seguintes hipóteses de nulidade:

- i. não há efeitos principais de “condições de medidas” (ocasiões);
- ii. não há efeitos de tratamentos;
- iii. não há interação entre tratamentos e ocasiões.

A análise de variância multivariada é um instrumento poderoso para analisar dados longitudinais, mas se a hipótese de uniformidade da matriz de variâncias e variância-covariância não for rejeitada, a análise univariada deve ser adotada. No entanto, se a matriz de variâncias e variância-covariância, para os dados repetidos, tem a estrutura de correlação

serial, deve-se utilizar um método de análise que leve em conta a estrutura dessa matriz para se ter um aumento do poder dos testes. Desta forma a análise de variância multivariada torna-se a mais conveniente, senão, a única apropriada, entre os procedimentos disponíveis (SMITH et al., 1962; COLLE; GRIZZLE, 1966)

A análise multivariada de perfil tem como principal vantagem o fato de permitir a adoção de um modelo bastante geral para representar a estrutura de variância-covariância, admitindo que as variâncias das respostas em cada tempo e as variância-covariância das respostas entre tempos distintos sejam diferentes (LIMA, 1996).

Greenhouse; Geiser (1959), estudando métodos de análise de dados longitudinais, observaram que os quocientes entre os quadrados médios obtidos na análise de variância para o modelo misto univariado, só terão distribuição de probabilidade F exata se as observações no tempo forem normalmente distribuídas, com variâncias iguais e mutuamente independentes ou igualmente correlacionadas. Com estas pressuposições restritas, os autores preferem considerar as observações no tempo como um vetor de amostras de uma distribuição normal multivariada com matriz de variâncias e variância-covariância arbitrária. Sendo assim, o enfoque multivariado, apresentado por Morrison (1981), permite a adoção de um modelo geral para representar a estrutura de covariância das observações. Neste caso, a matriz de variância-covariância é conhecida como não-estruturada, onde todas as variâncias e variância-covariância podem ser diferentes e, de acordo com Andreoni (1989), só é aplicável quando:

- i. não existirem bases teóricas ou empíricas para estabelecer algum padrão para esta matriz;
- ii. não houver necessidade de extrapolar o modelo além das ocasiões de observações consideradas.

A quantidade de parâmetros associados à matriz não estruturada, que precisa ser estimada, é proporcional ao número de condições de avaliação. Em situações em que o número é grande (quando o número de unidades experimentais é pequeno em relação ao número de ocasiões de avaliação ou quando há presença de muitas observações incompletas), a eficiência dos estimadores dos parâmetros pode ser afetada. Em alguns casos é impossível estimar os parâmetros dessa matriz de variância-covariância (WALD, 2000). A desvantagem da análise multivariada é a falta de poder para estimar os parâmetros da matriz de variância-

covariância, isto quando t (número de ocasiões de medidas ou tempos) é grande e n é pequeno (MEREDITH; STEHMAN, 1991).

Stuker (1986) comenta da restrição da análise de variância multivariada em que o número de unidades experimentais menos o número de tratamentos deva ser maior que o número de observações realizadas em cada unidade experimental, pois, em caso contrário, a matriz devida ao erro requerida para os testes é singular.

As restrições para a aplicação da técnica de análise multivariada de perfil ocorrem pelo fato da necessidade de perfis individuais de respostas completos e ao baixo poder dos testes de hipóteses, devido à excessiva parametrização. No entanto, salvo estas restrições, na maioria dos casos em estudos com dados longitudinais, o procedimento de análise de variância multivariada será o mais conveniente, senão o único apropriado entre os disponíveis (TIMM, 1980).

O modelo sugerido por Vonesh; Chinchilli (1997) e utilizado por Lima (1996), Xavier (2000); Freitas (2007), em suas pesquisas aplicadas em experimento com delineamento de parcelas subdivididas, com medidas repetidas no tempo corresponde:

$$Y_{ijk} = \mu + \beta_i + \tau_j + \gamma_k + (\tau\gamma)_{jk} + \varepsilon_{ijk}, \quad (7)$$

para $i=1,2,\dots,b$; $j=1,2,\dots,g$; $k=1,2,\dots,t$, em que:

Y_{ijk} = efeito observado para variável resposta medida no k -ésimo tempo para o j -ésimo tratamento no i -ésimo bloco;

μ = é uma constante a todas as observações;

β_i = efeito do i -ésimo bloco;

τ_j = efeito do j -ésimo tratamento;

γ_k = efeito do k -ésimo tempo;

$(\tau\gamma)_{jk}$ = efeito da interação entre o j -ésimo tratamento com o k -ésimo tempo;

ε_{ijk} = representa o erro aleatório da observação Y_{ijk} , suposto homocedástico

Para que o modelo tenha posto completo, é necessário impor algumas restrições conforme segue:

$$\sum_{i=1}^b \beta_i = \sum_{j=1}^g \tau_j = \sum_{k=1}^t \gamma_k = \sum_{j=1}^g (\tau\gamma)_{jk} = 0, \quad (8)$$

Para $k=1,\dots,t$ e $\sum_{k=1}^t (\tau\gamma)_{jk} = 0$, com $j=1,\dots,g$ e com vetor aleatório $\varepsilon_{ijk} = [e_{ij1}, \dots, e_{ijt}] \sim i.i.d.N(0, \Sigma)$ em que Σ é uma matriz $t \times t$, positiva definida, como uma estrutura geral.

Neste modelo, a configuração das observações na forma matricial é a seguinte:

$$Y_{(n \times t)} = \underbrace{\begin{bmatrix} y_{111} & y_{112} & \cdots & y_{11t} \\ y_{121} & y_{122} & \cdots & y_{12t} \\ \vdots & \vdots & & \vdots \\ y_{1n_1 1} & y_{1n_1 2} & \cdots & y_{1n_1 t} \\ \vdots & \vdots & \vdots & \vdots \\ y_{g11} & y_{g12} & \cdots & y_{g1t} \\ \vdots & \vdots & & \vdots \\ y_{gn_g 1} & y_{gn_g 2} & \cdots & y_{gn_g t} \end{bmatrix}}_{\text{matriz de dados}} = \underbrace{\begin{bmatrix} y'_{11} \\ y'_{12} \\ \vdots \\ y'_{1n_1} \\ \vdots \\ y'_{g1} \\ \vdots \\ y'_{gn_g} \end{bmatrix}}_{\text{perfil de resposta da unidade de observação (ij)}} \quad (9)$$

O modelo linear usado na análise multivariada de perfis pode ser representado na forma matricial da Análise de Variância Multivariada (MANOVA), sendo este:

$$Y = X\beta + \Psi \quad (10)$$

em que: $Y_{(n \times t)} = [y_{111}, \dots, y_{gn_g t}]$ é a matriz dos dados observados de t respostas para os n indivíduos, conforme apresentado em (9).

$y_{ij} = [y_{ij1}, y_{ij2}, y_{ij3}, \dots, y_{ijt}]$ é o perfil de resposta da unidade experimental (i,j) , de acordo com (9).

$X_{(gb \times (g+b+1))}$ é a matriz delineamento, conhecida de postos completo. Um dos elementos importantes na definição do modelo para o vetor de médias é a matriz delineamento X (ANDRADE; SINGER, 1986). Suas colunas podem estar associadas com:

- i. os fatores que definem a estrutura das subpopulações, isto é, servem para identificar diferenças entre tratamentos;
- ii. o fator tempo, e servem para identificar a forma das curvas a serem ajustadas;
- iii. possíveis covariáveis, servem para identificar possíveis efeitos de outras variáveis secundárias no modelo.

Em geral, as colunas associadas com efeitos de tratamentos são constituídas de 1's e 0's ou de 1's, 0's e -1's, dependendo da parametrização adotada:

$\beta_{((g+b+1)xt)}$ é a matriz de parâmetros dos efeitos desconhecidos (cada linha é constituída por t elementos que representam os valores esperados da variável observada em cada um dos instantes considerados para um determinado grupo.

$\psi_{(gbxt)}$ é a matriz do erro experimental.

Em que:

$$X_{(gbx(g+b+1))} = [X_1 \vdots X_2 \vdots X_3], \text{ com } X_1 = [1_{gb}], X_2 = \begin{bmatrix} I_b \\ I_b \\ \vdots \\ I_b \end{bmatrix} \text{ e } X_3 = \begin{bmatrix} 1_b & 0 & \cdots & 0 \\ 0 & 1_b & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1_b \end{bmatrix} \text{ sendo}$$

1_{gb} um vetor com gb elementos iguais a um; $I_{(bxb)}$ é uma matriz identidade; e, $1_{(b \times 1)}$ é um vetor de uns.

A matriz dos parâmetros corresponde:

$$\beta_{((g+b+1)xt)} = \begin{bmatrix} \mu_1 & \mu_2 & \cdots & \mu_t \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1t} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{b1} & \beta_{b2} & \cdots & \beta_{bt} \\ \tau_{11} & \tau_{12} & \cdots & \tau_{1t} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{g1} & \tau_{g2} & \cdots & \tau_{gt} \end{bmatrix} \text{ para } \mu = [\mu_1 \ \mu_2 \ \cdots \ \mu_t]', \beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1t} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{b1} & \beta_{b2} & \cdots & \beta_{bt} \end{bmatrix},$$

$$\tau = \begin{bmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1t} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{g1} & \tau_{g2} & \cdots & \tau_{gt} \end{bmatrix}, \text{ sendo a matriz de erro experimental } \psi = [\varepsilon_{11} \ \varepsilon_{12} \ \cdots \ \varepsilon_{gb}], \text{ com}$$

$\varepsilon_{ij} = [\varepsilon_{ij1} \ \varepsilon_{ij2} \ \cdots \ \varepsilon_{ijt}]$, que é o vetor de erro experimental associado ao i -ésimo indivíduo, sendo que as linhas ε_{ij} , da matriz ψ , são supostas independentes e identicamente distribuídas com distribuição normal t -variada com média 0 e matriz de covariância comum, Σ . Assim, enquanto as linhas de ψ são independentes, os elementos das suas colunas podem ser correlacionados. Por extensão, os y_{ij} são independentes e seguem distribuições

normais *t-variadas*, com média $E(y_{ij}) = X\beta$ e a matriz de variância-covariância é não estruturada ou completamente parametrizada e tem $t(t+1)/2$ parâmetros representada por

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1t} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1t} & \sigma_{2t} & \cdots & \sigma_t^2 \end{bmatrix}, \text{ portanto: as } y_{ij} \sim i.i.d.N_t(X\beta; \Sigma). \text{ Também é preciso admitir}$$

que todos os grupos (tratamentos) devem ter a mesma matriz de variância-covariância, Σ .

As hipóteses testadas em (11) correspondem (ANDRADE; SINGER, 1986; SUYANA, 1995):

- i.* H_{0I} = os perfis médios de respostas correspondentes aos diversos tratamentos são paralelos, ou seja, não existe interação entre os fatores que definem os tratamentos e as condições de avaliação. Essa hipótese é fundamental, pois sua validade pode determinar a forma de expressão das outras duas hipóteses de interesse. Em termos de parâmetros do modelo, a hipótese H_{0I} pode ser expressa como:

$$H_{0I} = \begin{bmatrix} \mu_{11} & - & \mu_{12} \\ \mu_{12} & - & \mu_{13} \\ \vdots & & \vdots \\ \mu_{1(t-1)} & - & \mu_{1t} \end{bmatrix} = \begin{bmatrix} \mu_{21} & - & \mu_{22} \\ \mu_{22} & - & \mu_{23} \\ \vdots & & \vdots \\ \mu_{2(t-1)} & - & \mu_{2t} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{g1} & - & \mu_{g2} \\ \mu_{g2} & - & \mu_{g3} \\ \vdots & & \vdots \\ \mu_{g(t-1)} & - & \mu_{gt} \end{bmatrix}$$

- ii.* H_{0G} = os perfis médios resposta dos tratamentos são coincidentes, ou seja, não existe efeito dos tratamentos que definem os grupos. Esta hipótese é representada em termos de parâmetro do modelo por:

$$H_{0G} = \sum_{k=1}^t \mu_{1k} = \sum_{k=1}^t \mu_{2k} = \dots = \sum_{k=1}^t \mu_{gk}, \text{ na forma matricial tem-se:}$$

$$H_{0G} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1t} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2t} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{g1} \\ \mu_{g2} \\ \vdots \\ \mu_{gt} \end{bmatrix};$$

- iii.* H_{0T} = os perfis médios de resposta correspondentes aos tratamentos são paralelos ao eixo das abscissas, ou seja, não existe efeito dos fatores que definem as condições de avaliação. Em termos de parâmetros do modelo,

representa-se a hipótese H_{0T} como: $H_{0T} = \sum_{k=1}^g \mu_{i1} = \sum_{k=1}^g \mu_{i2} = \dots = \sum_{i=1}^g \mu_{it}$, a

mesma hipótese em termos matricial é:

$$H_{0T} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{g1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{g2} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{gt} \end{bmatrix}$$

Para efeito de unificação dos resultados estatísticos, é conveniente expressar todas as hipóteses apresentadas acima na forma da hipótese linear geral:

$H : G\beta T = 0$ em que $G_{((g-1) \times (g+b+1))}$ e $T_{(t \times (t-1))}$ são matrizes de constantes conhecidas com postos g e t , respectivamente. Na matriz G , ocorre a comparação entre tratamentos que são as linhas da matriz β . A matriz T é responsável por comparações entre as ocasiões de observação (tempo), colunas da matriz β . Explicitamente, têm-se as seguintes correspondências:

$$H_{0I} : G_{1(g-1) \times (g+b+1)} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & 1 & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 & 0 & \dots & -1 \end{bmatrix} \text{ e } T_{1(t \times (t-1))} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & -1 \end{bmatrix};$$

$H_{0G} : G_1$ e $T_2 = 1_t$ ainda $H_{0T} : G_2 = 1'_g$ e T_1 em que 1_t e $1'_g$ são vetores de uns com dimensões t e g , respectivamente, G_1 e T_1 , já definidos anteriormente. Salienta-se que estas representações não são únicas.

Para testar a hipótese linear geral, os testes são realizados considerando as magnitudes das somas de quadrados e produtos de tratamento e resíduo pela variância generalizada (JHONSON; WICHERN, 1998). Os testes para o modelo linear geral podem ser obtidos através de diversos critérios heurísticos (ANDRADE; SINGER, 1986); em geral, as estatísticas de teste correspondente são funções das raízes características da matriz HE^{-1} , em que: $H = T'((X'X)^{-1} X'Y)G'[G(X'X) - G']^{-1}G((X'X)^{-1} X'Y)T$ é a matriz de somas de quadrados

de produtos cruzados, devido a hipótese nula. E a matriz de somas de quadrados e produtos cruzados devido ao erro é: $E = T'Y'[I - X(X'X)^{-1}X']YT$.

Dentre as várias estatísticas de teste disponíveis, as mais difundidas são obtidas através dos princípios de união – interseção de Roy e da razão de verossimilhança de Wilks. As outras duas estatísticas muito utilizadas são Lawley – Hotelling e Pillai. Estas estatísticas são apresentadas na Tabela 3.

Apesar dessa abordagem ser bastante atraente, ela tem o inconveniente de incluir a interação entre bloco e tempo no modelo (LIMA, 1996). Embora existam casos em que esta inclusão é necessária, geralmente, os blocos são construídos somente para controlar a heterogeneidade entre as unidades experimentais e não se espera uma interação significativa entre blocos e tempo. Além disso, a inclusão desta interação pode provocar um aumento de até $(b-1)t$ parâmetros no modelo e, caso este número seja grande, muitos graus de liberdade serão utilizados para estimar parâmetros de pouca importância, provocando uma diminuição no poder dos testes relativos aos parâmetros de maior interesse.

Uma estratégia para a análise de perfis para a situação de não existência de interação entre blocos e tempos, que depende do tipo da matriz T que define as hipóteses de interesse e que pode ser implementada através de programas convencionais de análise de variância multivariada (SINGER; ANDRADE, 1994).

Tabela 3 – Critérios e estatísticas utilizadas para testar a hipótese linear geral. Santa Maria, RS, 2010.

Critério	Estatística
Lambda de Wilks	$\Lambda = \prod_{i=1}^S (1 - \theta_i)$
Raiz máxima de Roy	$\theta_s = \max(\theta_i)$
Traço de Hotelling-Lawley	$T = \sum_{i=1}^S \theta_i (1 - \theta_i)^{-1}$
Traço de Pillai	$P = \sum_{i=1}^S \theta_i$

Em que: $\theta_i = \lambda_i(1 + \lambda_i)^{-1}$ e λ_i é a *i*-ésima raiz característica de HE^{-1} , e S é o número de raízes características não nulas sendo $S = \min(t-1; g-1)$; t são os tempos; e, g os tratamentos.

As distribuições exatas dessas estatísticas, sob a hipótese nula, dependem unicamente dos parâmetros: $m_1 = (|(t-1) - (g-1)| - 1) / 2$; $m_2 = (n - g - (t-1) - 1) / 2$; e, s já definido.

Os quatro testes multivariados podem produzir diferentes níveis descritivos. Em geral, a ordem de preferência em termos de poder é: Traço de Pillai, Lambda de Wilks, Traço de Lawley-Hotelling e Roy. Entretanto, Lambda de Wilks é o teste mais comumente usado. Os testes não requerem a condição de H-F, pois são baseados em uma matriz de covariância sem estrutura. As estimativas destas matrizes podem ser obtidas pelo método dos momentos (XAVIER, 2000).

O princípio de união interseção de Roy é o mais adequado, pois permite calcular os limites de confiança simultâneos para todas combinações lineares dos elementos da matriz $H : G\beta T = 0$, com uma probabilidade conjunta $(1 - \alpha)$ (MORRISON, 1981).

Muitos autores recomendam utilizar o critério de Wilks como referência, por se tratar de um teste baseado na razão de verossimilhança. Outros recomendam que a hipótese nula deva ser rejeitada se, pelo menos, três dos quatro critérios forem significativos em um nível nominal de significância previamente adotado (FERREIRA, 1996).

2.5.3 Ajuste de curvas de crescimento por modelo linear misto

O estudo de curvas de crescimento foi intenso na primeira metade do século XX (OSHIIWA; PADOVANI, 2005). Neste sentido, destacam-se as contribuições de Wishart (1938); Box (1954); Elstson; Grizzle (1962), entre outros. No entanto, o modelo proposto por Potthoff; Roy (1964), “Modelo Linear Multivariado de Crescimento (MLMC)”, foi considerado um paradigma nessa área. Esse modelo corresponde a uma generalização do modelo usual de análise de variância Multivariada (11), no sentido de permitir a expressão dos perfis médios de resposta através de uma forma polinomial. Lima (1988) relaciona ao fato, de ser aplicado uma transformação de variáveis, baseada em uma matriz de ponderação (

Δ), arbitrária, simétrica, positiva definida e não estocástica, os autores conseguiram reduzir o modelo generalizado ao modelo de análise de variância multivariada.

Na utilização de modelos multivariados para resolver problemas práticos, surgem algumas dificuldades de estimação e de eficiência das estimativas dos parâmetros das curvas. Isto se deve ao grande número de parâmetros associados à matriz de variância-covariância não estruturada, apontando a necessidade de se considerar algum tipo de estrutura parcimoniosa para essa matriz, reduzindo o número de parâmetros necessário para modelá-la (LIMA, 1996).

Neste sentido, Laird; Ware (1982) e Fairclough; Helms (1986) estudaram uma classe de estruturas de variância-covariância estimadas através de especificação de Modelos de Efeitos Aleatórios (ou mistos), que proporcionam uma maior versatilidade na aplicação da técnica de análise de curvas de crescimento para dados longitudinais. A metodologia de modelos mistos torna-se uma ferramenta poderosa na modelagem de curvas de crescimento, permitindo a inclusão de efeitos fixos e aleatórios no processo de modelagem e apresenta as seguintes vantagens (LAIRD; WARE, 1982)

- i. permitem a modelagem da matriz de variância-covariância com um número menor de parâmetros que a de modelos multivariados gerais;
- ii. podem ser usados quando os dados não são balanceados em relação ao tempo;
- iii. possibilitam o ajuste de curvas polinomiais de graus diferentes para cada grupo de unidades experimentais.

A metodologia de modelos mistos foi inicialmente, proposta por Henderson (1949, 1975), e tornou-se conhecido nas últimas décadas devido o desenvolvimento de suportes computacionais robustos. Antes desses avanços, as análises de modelos mistos eram executadas adaptando-se métodos para modelos de efeitos fixos. Isso trazia limitações à aplicabilidade, porque as estruturas de covariância não eram modeladas, como é o caso das análises realizadas pelo *proc GLM* no SAS. As versões recentes do SAS incluem o *proc mixed*, que permitem a modelagem da estrutura de variância-covariância dos dados e o cálculo de estimativas eficientes dos efeitos fixos e de seus respectivos erros padrões (LITTELL et al., 2006).

O modelo misto estende o modelo linear geral, proporcionando uma especificação mais flexível para a matriz de covariância de $\boldsymbol{\epsilon}$ (resíduo). Possibilita correlações e variâncias

heterogêneas para os erros, acomodando, por exemplo, medidas repetidas, dados espaciais e temporais e, permitindo efeitos aleatórios no modelo (WALD, 2000; RIBOLDI, 2007).

Uma abordagem mais atual na modelagem de curvas de crescimento, consiste no uso de modelos lineares mistos, que baseiam-se em três aspectos fundamentais: estimação e teste de hipóteses sobre os parâmetros de efeito fixo; predição dos parâmetros de efeito aleatório; e, estimação dos componentes de variância. O sucesso do procedimento de modelagem está fortemente associado ao exame dos efeitos aleatórios e a possibilidade de se introduzir, no modelo, estruturas de variâncias e variância-covariância (CAMARINHA FILHO, 2002).

O nome modelo misto vem do fato que o modelo contém parâmetros de efeitos fixos, β , e parâmetros de efeitos aleatórios, b . Estes modelos são usados para modelar a parte aleatória através da inclusão de uma matriz de variância-covariância (LITTELL et al., 2006).

O modelo misto simplifica e unifica diversas análises estatísticas comuns. Além da análise de dados espaciais e de modelos de efeitos aleatórios, o modelo misto é útil para: análise de planejamento de experimentos longitudinais; planejamento em parcelas subdivididas; meta análise e modelos de variância heterogênea. Além disso, extensões úteis de modelo misto estão disponíveis para modelos lineares generalizados e modelos não lineares (PASCUAL, 2000).

Existem dois tipos de modelos mistos úteis para a análise de dados longitudinais: o modelo padrão de covariância e o modelo de coeficientes aleatórios. O modelo padrão de covariância é utilizado quando os dados são balanceados; em contrapartida, o modelo de coeficientes aleatórios é apropriado para dados desbalanceados ou se há interesse particular na relação entre a variável resposta e o tempo, caso típico das curvas de crescimento. Neste caso, curvas de regressão são ajustadas para cada sujeito e são admitidos coeficientes de regressão variando aleatoriamente entre os sujeitos. Essa variação ocorre, de forma simples, nos interceptos ou, de forma mais complexa, também nas inclinações. O modelo de coeficientes aleatório é um caso especial do modelo misto, pois os coeficientes de regressão das curvas individuais serão considerados efeitos aleatórios e os coeficientes de regressão das curvas médias, efeitos fixos (HAUSER et al., 2009).

Para se estruturar um modelo misto é necessário ter um modelo para os efeitos fixos, que deve descrever o comportamento da curva média; um modelo para os efeitos aleatórios, que junto com os efeitos fixos devem descrever as curvas individuais e um modelo para

dispersão condicional, que deve descrever as variância-covariância das observações individuais ao longo do tempo (SUYAMA, 1995),

O modelo misto em dois estágios, proposto por: Laird; Ware (1982) e Diggle et al. (1998), Searle (1987), Searle et al. (1992), entre outros, foi baseado no trabalho de Harville (1977). Os autores consideram os efeitos fixos no primeiro estágio para a obtenção da curva polinomial média e, no segundo estágio, permitem diferentes curvas para cada indivíduo.

O modelo linear misto é denotado pela seguinte forma matricial:

$$Y = X \beta + Zb + \varepsilon \text{ em que:} \quad (11)$$

Y = é o vetor de observações;

X ($n \times p$) é uma matriz de posto $p < n$, conhecida e de especificação, associada ao vetor β ($p \times 1$) de parâmetros de efeitos de tratamentos desconhecidos;

Z ($n \times q$) é uma matriz conhecida e de especificação, de posto coluna completo e associada ao vetor de efeitos aleatórios b ($q \times 1$) de diferença individuais em torno dos valores populacionais (efeitos aleatórios desconhecidos); e,

ε ($n \times 1$) é o vetor de erros aleatórios não observáveis;

em que: n corresponde ao número total de observações de todas as unidades experimentais; p é o número de parâmetros; e, q o número de efeitos aleatórios.

Pinheiro (1994) comenta que a estrutura geral do modelo linear de efeitos mistos, atualmente, parece ser aceita pela maioria dos pesquisadores. A literatura referente aos modelos lineares de efeitos mistos que tem sido publicada após Harville (1977) e Laird; Ware (1982) referem-se mais a generalizações do modelo (13) e/ou diferentes abordagens para a estimação, do que reformulações na estrutura básica do modelo. Neste sentido, Lima (1996) argumenta que em situações onde o objetivo da análise é ajustar curvas de crescimento, pode-se dizer que os modelos mistos assumem a existência de curvas subpopulacionais fixadas ($X \beta$), em torno das quais, existem variações aleatórias (Zb) das curvas individuais e também, existem variações aleatórias de medidas (ε) em torno dessas curvas individuais. Essas variações aleatórias (ε) são independentes dos efeitos aleatórios individuais (b).

A matriz Z pode conter quaisquer covariáveis que afetam diretamente as unidades experimentais, e a forma de especificação de (X) é bastante similar àquela utilizada nos modelos de regressão (LIMA, 1996). Suas colunas podem estar associadas:

- i. aos fatores que definem a estrutura das subpopulações (tratamentos);
- ii. ao fator tempo, indicando, por exemplo, a forma da curva a ser ajustada; e,
- iii. as covariáveis, cujos efeitos na resposta média deseja-se pesquisar.

Uma suposição básica para o modelo (13) é que b e \mathcal{E} são independentes, normalmente distribuídos, ou seja, $\mathcal{E} \sim N(0, R)$ e $b \sim N(0, G)$. As esperanças e variâncias são dadas pelas matrizes:

$$E \begin{bmatrix} b \\ \mathcal{E} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \text{Var} \begin{bmatrix} b \\ \mathcal{E} \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

A variância de Y é $V = ZGZ' + R$. Modelar-se V através da matriz de planejamento dos efeitos aleatórios Z e especificando-se as estruturas de covariância para G e R . Assume-se ainda que V é não singular, e a $E(Y) = X\beta$, assim, $Y \sim N(X\beta, ZGZ' + R)$.

É importante ressaltar que os elementos do vetor b são adicionados ao modelo com objetivo de promover melhor ajuste. O vetor \hat{b} é composto de parâmetros individuais que representam as diferenças entre a curva do tratamento e a curva de cada indivíduo. As estimativas dos parâmetros do modelo ($\hat{\beta}$) e a predição dos efeitos aleatórios (\hat{b}) podem ser obtidas pela solução do sistema de equações do modelo propostas por Henderson (1984) que será apresentado a seguir.

2.5.3.1 Estimação de β e b do modelo linear misto

Nos modelos mistos tem-se, em geral, interesse na: estimação dos efeitos fixos, predição dos efeitos aleatórios e estimação dos componentes de variância e covariância. Assim, quando se tem um modelo misto como dado em (13), com matriz de variâncias e variância-covariância, $V = ZGZ' + R$, a estimação dos efeitos fixos $\hat{\beta}$ e predição dos efeitos

aleatórios \hat{b} , podem ser obtidas utilizando-se o Método de Estimação dos Mínimos Quadrados Generalizados (GLSE: *Generalized Least Square Estimation*), que conduz ao sistema de equação normais do modelo misto (15). Desta forma, combinando os conceitos de predição e de melhor estimador linear não viesado (BLUP: *Best Linear Unbiased Predictor*) e BLUE (*Best Linear Unbiased Estimator*), são obtidas as equações normais para modelos mistos.

Para estimar parâmetros e prever efeitos aleatórios, a estrutura especial da distribuição normal para b e \mathcal{E} permite dividir o desenvolvimento teórico em duas etapas. Primeiro, considera-se G e R conhecidas, obtendo-se solução para β e b . Depois, estimam-se os parâmetros não conhecidos em G e R . Computacionalmente, a segunda etapa é executada primeiro (RIBOLDI, 2007).

A estimação dos parâmetros β e b é mais complexa no modelo misto que no modelo linear geral. Não somente faz com que se tenha β como no modelo linear geral, mas se tem parâmetros desconhecidos em b , G , e R , a serem estimados (RIBOLDI, 2007). O método de mínimos quadrados não é o melhor método. Mínimos quadrados generalizados (GLS) é mais apropriado, pois contempla qualquer estrutura não singular de V , o que leva a minimizar a equação $(y - X\beta)'V^{-1}(y - X\beta)$.

A estimação dos efeitos fixos e a predição dos efeitos aleatórios são realizadas através da derivação das equações do modelo misto (MME). Essa derivação pode ser feita pela maximização da soma dos quadrados dos resíduos ou pela maximização da função densidade de probabilidade conjunta de Y e b , sendo esta última mais adotada:

$$f(y) = \frac{1}{(2\pi)^{n/2} (ZGZ' + R)^{1/2}} e^{-1/2[(y-X\beta)(ZGZ'+R)^{-1}(y-X\beta)]}$$

A função densidade de probabilidade conjunta de Y e b pode ser compreendida como o produto entre a função densidade condicional de Y , dado b , e a função densidade de probabilidade de b , sendo: $f(y, b) = f(y | b).f(b)$. Desta forma temos:

$$\frac{1}{(2\pi)^{n/2} |R|^{1/2}} e^{-1/2[(y-X\beta-Zb)(R)^{-1}(y-X\beta-Zb)]} \cdot \frac{1}{(2\pi)^{g/2} |G|^{1/2}} e^{-1/2[(b-0)(G)^{-1}(b-0)]}, \quad \text{sendo } |R| \text{ e } |G| \text{ os}$$

determinantes das matrizes de variância-covariância.

Para se proceder à maximização de $f(y, b)$, pode-se usar a transformação por logaritmo. Assim, o logaritmo da função de verossimilhança é:

$$l(\beta, b | y) = \frac{1}{2} 2n \log(2\pi) - \frac{1}{2} (\log R + \log G) - \frac{1}{2} (y' R^{-1} y - 2y' R^{-1} X\beta - 2y' R^{-1} Zu + 2\beta' X' R^{-1} Zu + \beta' X' R^{-1} X\beta + v' Z' R^{-1} Zu + u' G^{-1} u).$$

Derivando-se $l(\beta, b | y)$ em relação a β e b e, tornando-se tais derivadas identicamente nulas, obtêm-se:

$$\begin{bmatrix} \frac{\partial l(\beta, b | y)}{\partial \beta} \\ \frac{\partial l(\beta, b | y)}{\partial b} \end{bmatrix} = \begin{bmatrix} -X' R^{-1} y + X' R^{-1} X\hat{\beta} + X' R^{-1} Z\hat{b} \\ -Z' R^{-1} y + Z' R^{-1} X\hat{\beta} + Z' R^{-1} Z\hat{b} + G^{-1}\hat{b} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} X' R^{-1} X\hat{\beta} + X' R^{-1} Z\hat{b} \\ Z' R^{-1} X\hat{\beta} + Z' R^{-1} Z\hat{b} + G^{-1}\hat{b} \end{bmatrix} = \begin{bmatrix} X' R^{-1} y \\ Z' R^{-1} y \end{bmatrix} \quad \text{e então,}$$

$$\begin{bmatrix} X' R^{-1} X & X' R^{-1} Z \\ Z' R^{-1} X & Z' R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X' R^{-1} y \\ Z' R^{-1} y \end{bmatrix} \quad (12)$$

Desta forma, obtêm-se as equações do modelo misto (EMM) de Henderson, com as quais são fornecidas soluções para os efeitos fixos ($\hat{\beta}$) e predição para os efeitos aleatórios (\hat{b}). As EMM são equações normais estendidas ou também, equações de mínimos quadrados generalizados.

As soluções são dadas por

$$\begin{bmatrix} \hat{\beta} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X' R^{-1} X & X' R^{-1} Z \\ Z' R^{-1} X & Z' R^{-1} Z + G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X' R^{-1} y \\ Z' R^{-1} y \end{bmatrix} \quad (13)$$

que podem ser escritas como

$$\hat{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} y \quad (14)$$

com a mesma solução de mínimos quadrados generalizados, e

$$\hat{b} = GZV^{-1}(y - X(X' V^{-1} X)^{-1} X' V^{-1} y) = GZV^{-1}(y - X\hat{\beta}) \quad (15)$$

O estimador de β e o preditor de b , (14) e (15), são, respectivamente, o estimador de Mínimos Quadrados Generalizados (GLSE), o BLUE e o melhor preditor linear imparcial (BLUP).

O termo predição refere-se aos fatores aleatórios e a melhor predição linear não viesada pode ser, resumidamente, definida como resultado da regressão dos efeitos de um fator aleatório (b) em função das observações (y) corrigidas para os efeitos dos fatores fixos ($X\beta$) (MARTINS et al.,1993).

2.5.3.2 Estimação dos parâmetros de covariância em G e R

A solução das EMM (15) depende do conhecimento da matriz de variâncias e variância-covariância V , cuja estrutura é conhecida, porém, via de regra, seus componentes não o são. A matriz $V = ZGZ' + R$ esta associada aos componentes de variância (G e R) que se deseja estimar. Considera-se θ o vetor de parâmetros em V . Desse modo, torna-se necessário substituí-lo por suas estimativas, ou seja, estimar θ . Existem vários métodos de estimação de componentes de variâncias (θ), dentre os quais se destacam: Método da ANOVA ou dos momentos (Tipo 1, 2 e 3); Estimador Quadrático Não-Viesado de Variância Mínima (MIVQUE) ou de Norma Mínima (MINQUE); Máxima Verossimilhança (ML); e, Máxima Verossimilhança Restrita (REML).

O método de estimação por quadrados mínimos (LS), na situação de dados desbalanceados, apresenta a propriedade de não vício, mas suas demais propriedades estatísticas são pouco conhecidas (RESENDE et al.,1996). Outra grande desvantagem do método LS, é a possibilidade de obtenção de estimativas negativas de componentes de variância. Assim, posterior ao ano de 1967, outros processos têm sido considerados para a estimação de componentes de variância, notadamente aqueles ligados à máxima verossimilhança.

Hartley; Rao (1967) derivaram um procedimento de máxima verossimilhança para estimação de variâncias no contexto dos modelos lineares mistos. Este foi, posteriormente, modificado e aplicado por melhoristas animais (HENDERSON, 1973). Os estimadores de máxima verossimilhança apresentam as seguintes propriedades desejáveis (KENNEDY, 1981): translação invariante (não afetados por mudanças nos efeitos fixos), suficiência, consistência e eficiência. Outra vantagem da ML é a geração de estimativas não negativas de componentes de variância. Por outro lado, os estimadores ML são viciados em decorrência de

dois aspectos (SHAW, 1987): imposição de restrição de não negatividade e perda de graus de liberdade devido à estimação dos efeitos fixos.

Os efeitos fixos devem ser conhecidos para a estimação ML de componentes de variância (RESENDE, 2007). No entanto, eles não são conhecidos e são substituídos por suas estimativas obtidas por ML. Porém, na estimação dos componentes de variância, o método ML não considera a perda de graus de liberdade devido a estimação desses efeitos fixos, causando então o vício. Este vício conduz a subestimativas dos parâmetros de variância e, portanto, podem conduzir a inferências incorretas.

Visando eliminar a fonte de vício, relacionada a perda de graus de liberdade, Patterson; Thompson (1971) apresentaram uma modificação ao método ML, o qual passou a ser denominado de máxima verossimilhança restrita ou "residual" (REML). Além do aperfeiçoamento de natureza estatística, o método REML mantém as demais propriedades do ML.

Atualmente, de acordo com Resende (2000), o procedimento padrão de estimação de componentes de Variância é o da Máxima Verossimilhança Restrita (REML), desenvolvido por Patterson; Thompson (1971), sendo que o procedimento ótimo de predição é o BLUP.

Os melhores estimadores lineares não-viesados (BLUE – Best Linear Unbiased Estimator), obtidos pelo uso do método de máxima verossimilhança e máxima verossimilhança restrita, pressupondo normalidade, são os mesmos fornecidos pelo uso do método de quadrados mínimos. Entretanto, apenas os resultados obtidos com o uso do método de máxima verossimilhança restrita coincidem com aqueles obtidos com o método de quadrados mínimos (LOPES et al.,1998).

No método REML, somente a porção da verossimilhança que é invariante aos efeitos fixos é maximizada. Assim, o REML mantém as demais propriedades do ML, é não viciado e permite também a imposição de restrições de não negatividade. Dessa forma, o REML é o procedimento ideal de estimação de componentes de variância com dados desbalanceados. Com o REML, os componentes de variância são estimados sem serem afetados pelos efeitos fixos do modelo e, os graus de liberdade referentes à estimação dos efeitos fixos são considerados, produzindo estimativas não viciadas (Mac CULLOCH; SEARLE, 2001; RESENDE, 2007).

No REML, cada observação é dividida em duas partes independentes, uma referente aos efeitos fixos e outra aos aleatórios, de maneira que a função de densidade de probabilidade das observações é dada pela soma das funções de densidade de probabilidade de cada parte (RIBOLDI, 2007). A maximização da função densidade de probabilidade da parte referente aos efeitos aleatórios, em relação aos componentes de variância, elimina o viés resultante da perda de graus de liberdade na estimação dos efeitos fixos do modelo. O *proc mixed* do SAS minimiza estas funções usando o algoritmo iterativo de Newton-Raphson.

Em geral, o método mais importante para estimar parâmetros de covariância é o da máxima verossimilhança restrita (REML), que é método padrão dos procedimentos MIXED e GLIMMIX do SAS para modelos mistos.

Freitas (2007) e Macedo (2007) apresentam, para os métodos da máxima verossimilhança (ML) e máxima verossimilhança restrita (REML), os log das funções de verossimilhanças, sendo estes:

i. log da função de verossimilhança para o método ML:

$$l(G, R) = -\frac{1}{2} \log|V| - \frac{n}{2} \log(y - X(X'V^{-1}y)'V^{-1}(y - X(X'V^{-1}X)X'V^{-1}y)) - \frac{n}{2} \left[1 + \log\left(\frac{2\pi}{n}\right) \right] \quad (16)$$

ii. log da função de verossimilhança para o método REML:

$$l_R(G, R) = -\frac{1}{2} \log|V| - \frac{1}{2} \log|X'V^{-1}X| - \frac{n-p}{2} \log(y - X(X'V^{-1}X)X'V^{-1}y)'V^{-1}(y - X(X'V^{-1}X)X'V^{-1}y) - \frac{n-p}{2} \left[1 + \log\left(\frac{2\pi}{n-p}\right) \right] \quad (17)$$

em que: p é o posto da matriz X .

A expressão (16) é a função objetivo para ML utilizada pelo *proc mixed* do SAS. Minimizando-se essa função, sobre todos os parâmetros desconhecidos, obtém-se um sistema de equações cuja solução fornece as estimativas ML. Assim como para a expressão (16), o *proc mixed* também fornece as estimativas para REML. Essas equações são não lineares e são resolvidas numericamente, em geral por processos iterativos como o algoritmo de Newton-Raphson. O processo é repetido até que o critério de convergência adotado seja satisfeito (PERRI; IEMMA, 1999). A principal diferença entre os métodos ML e REML é que o ML usa a função de verossimilhança de y ou o logaritmo desta função, enquanto o REML adota a

função de verossimilhança de $K'y$, um vetor de combinações lineares das observações (com esperança nula) que representa efetivamente as observações ajustadas para os efeitos fixos.

Maiores detalhes sobre as diferenças entre os métodos ML e REML podem ser encontrados em: Perri; Iemma (1999) e Verbeke; Molenberghs (2000). Além destes, Searle et al. (1992) apresentam uma obra extensa sobre a estimação de componentes de variância e análise de modelos mistos, tanto para dados balanceados quanto para desbalanceados. Nessa obra, os autores apresentam detalhes no que se refere à teoria dos métodos, além de apresentar ilustrações sobre cada método.

Com relação aos algoritmos para minimizar as funções: $l(G, R)$ e $l_R(G, R)$, desta forma, estimar componentes de variâncias (θ), Lindstrom; Bates (1988) comentam que um bom algoritmo deve realizar cálculos iterativos que não consumam muito tempo, e deve ter rapidez e consistência de convergência. Neste sentido, o algoritmo de Newton-Raphson possui boas características como algoritmo de otimização, assim como, Scoring de Fisher e, sendo superior ao algoritmo EM (Expectation – Maximization) (PASCUAL, 2000).

2.5.3.3 Propriedades dos estimadores de G e R

Considerando a variância V conhecida, do vetor de observações Y , conseqüentemente, as matrizes G e R também o são, sendo assim, $\hat{\beta}$ é o melhor estimador linear imparcial (BLUE) de β e, \hat{b} é o melhor preditor linear imparcial (BLUP) de b (SEARLE, 1971; HARVILLE, 1990).

No entanto, G e R são usualmente desconhecidas e estimadas utilizando um dos métodos de estimação de componentes de variância mencionados em 2.5.4.2. Estes estimadores são substituídos na expressão (14) para se obter as equações de modelos mistos:

$$\begin{bmatrix} X' \hat{R}^{-1} X & X' \hat{R}^{-1} Z \\ Z' \hat{R}^{-1} X & Z' \hat{R}^{-1} Z + \hat{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X' \hat{R}^{-1} y \\ Z' \hat{R}^{-1} y \end{bmatrix}$$

Utilizando-se as matrizes de variância-covariância estimadas \hat{G} e \hat{R} , o termo empírico ou estimado é acrescentado a BLUE e BLUP, obtendo-se EBLUE (empirical best

linear unbiased estimator) e EBLUP (empirical best linear unbiased predictor) (LITTELL et al., 2006).

Na solução das equações dos modelos mistos com matrizes de variância-covariância estimadas \hat{G} e \hat{R} , tem-se a matriz de variância-covariância estimada \hat{C} dada por:

$$\hat{C} = \begin{bmatrix} X' \hat{R}^{-1} X & X' \hat{R}^{-1} Z \\ Z' \hat{R}^{-1} X & Z' \hat{R}^{-1} Z + \hat{G}^{-1} \end{bmatrix}^{-1}, \text{ que pode ser exata ou aproximada.}$$

Henderson (1984) e McLean; Sanders (1988) mostram que \hat{C} também pode ser escrita como:

$$\hat{C} = \begin{bmatrix} \hat{C}_{11} & \hat{C}'_{21} \\ \hat{C}_{21} & \hat{C}_{21} \end{bmatrix}$$

Em que:

$$\begin{aligned} \hat{C}_{11} &= (X' \hat{V}^{-1} X)^{-1} \\ \hat{C}_{21} &= -\hat{G} Z' \hat{V}^{-1} X \hat{C}_{11} \\ \hat{C}_{22} &= (Z' \hat{R}^{-1} Z + \hat{G}^{-1})^{-1} - \hat{C}_{21} X' \hat{V}^{-1} Z \hat{G} \end{aligned}$$

Observa-se que \hat{C}_{11} é uma fórmula conhecida para estimar a matriz de variância e covariância do estimador de mínimos quadrados generalizados (GLS) de β , isto é:

$$\hat{V}(\hat{\beta}) = \hat{C}_{11} = (X' \hat{V}^{-1} X)^{-1} \text{ (RIBOLDI, 2007; FREITAS, 2007).}$$

2.5.3.4 Estruturas de variância-covariância

Conforme foi verificado anteriormente, a análise dos dados obtidos em estudo longitudinais pode ser efetuada através da análise de perfil, que pode abordar três aspectos: univariado, com análise de variância para experimentos em parcelas subdivididas no tempo (*split plot in time*); multivariado; e, análise de curvas de crescimento. No caso univariado, é exigido que a estrutura da matriz de covariância satisfaça a condição H-F. Quando a abordagem for multivariada, a matriz de covariância é não estruturada, assume uma forma mais geral, permitindo diferentes variâncias e variância-covariância. Caso semelhante ocorre com o Modelo Linear Multivariado de Crescimento, cuja eficiência das estimativas dos

parâmetros das curvas está associada à matriz de covariância não estruturada, em que surge a necessidade de se considerar algum tipo de estrutura parcimoniosa para essa matriz.

Quando se trata de modelos mistos, é possível modelar a estrutura da matriz de covariância da forma que melhor represente os dados, respeitando a sua natureza, sejam eles independentes, dependentes, ou correlacionados ou que apresentem qualquer outro tipo de relação, que não se consiga explicar utilizando a estrutura de simetria composta ou não estruturada. A vantagem de se trabalhar com modelos mistos, que envolvam curvas de crescimento ou polinomiais, é a possibilidade de poder optar pela estrutura de variância-covariância que melhor represente os dados (XAVIER, 2000).

A escolha da melhor estrutura da matriz de variância-covariância visa obter uma estrutura parcimoniosa, que explique bem a variabilidade dos dados nas diversas ocasiões e a correlação entre essas medidas com um número pequeno de parâmetros, o que pode melhorar a eficiência das inferências sobre os parâmetros do modelo proposto para os valores médios nas diversas ocasiões (BARBOSA, 2009).

A escolha da estrutura de covariância afeta as estimativas e erros padrões de efeitos fixos, diagnósticos, interpretações e extrapolações. A escolha depende de informação empírica, da estrutura dos dados e de programas computacionais disponíveis com a estrutura selecionada (CASTRO, 1999).

No modelo misto, a variância V , do vetor de observações Y , pode ser modelada através da especificação das estruturas de covariância mais adequadas para as matrizes G e R . A seguir serão apresentadas algumas das estruturas de covariância, para as matrizes G e R , mais utilizadas e já implementadas no SAS (VERBEKE; MOLENBERGS, 2000; XAVIER, 2000; LITTEL et al., 2006).

Todas as estruturas apresentadas representam $n_i = 4$ ocasiões de medidas:

I. Componente de Variância (VC)

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}, \text{ esta estrutura exige variâncias iguais em todas as ocasiões}$$

de medidas e observações independentes e envolve um único parâmetro.

II. Simetria Composta (CS)

$$\Sigma = \begin{bmatrix} (\sigma^2 + \sigma_1^2) & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & (\sigma^2 + \sigma_1^2) & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & (\sigma^2 + \sigma_1^2) & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & 0 & (\sigma^2 + \sigma_1^2) \end{bmatrix}, \text{ impõe variâncias iguais em todas as}$$

ocasiões de medidas e mesmas covariâncias entre medidas feitas em ocasiões distintas e envolve dois parâmetros. Esta estrutura pode ser representada conforme apresentado em (1). Em análise univariada de perfil, item 2.5.1, encontra-se detalhamentos sobre a matriz de variância e covariância CS.

III. Sem Estrutura (UN)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}, \text{ todas as variâncias e as covariâncias podem ser desiguais, ou}$$

seja, as variâncias podem ser diferentes para cada uma das n_i ocasiões e variância-covariância diferentes entre medidas feitas em ocasiões distintas. As variâncias são restritas a valores não negativos e as covariâncias não têm restrições. Envolve $n_i(n_i + 1)/2$ parâmetros.

IV. Auto Regressiva de Primeira Ordem (AR(1))

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}, \text{ esta estrutura, assim como a CS, indica que todos os níveis de}$$

tempo têm a mesma variância. Indica, também, que a correlação entre as medidas tomadas no tempo vai decrescendo à medida que aumenta a distância entre os níveis de tempo. Isto quer dizer que as medidas tomadas em tempos mais distantes têm menor correlação que as medidas

tomadas em tempos mais próximos. Na estrutura AR(1), o número de parâmetros envolvidos é igual a dois.

V. Auto Regressiva de Primeira Ordem Heterogênea (ARH(1))

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix} \text{ é uma generalização da estrutura AR(1),}$$

impondo variâncias e variância-covariância desiguais. A estrutura ARH(1) somente é viável se os dados são igualmente espaçados no tempo e envolve $n_i + 1$ parâmetros.

VI. Simetria Composta Heterogênea (CSH)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}, \text{ impõe parâmetros de variâncias diferentes para}$$

cada elemento da diagonal principal e, a raiz quadrada desses parâmetros nos elementos fora da diagonal principal, sendo: σ_i^2 , o i-ésimo parâmetro de variância e ρ o parâmetro de correlação satisfazendo $|\rho| < 1$. Envolve $n_i + 1$ parâmetros.

A estrutura CSH e a estrutura ARH(1) são generalizações das estruturas básicas CS e AR(1). Elas são obtidas por permitirem variâncias distintas ao longo da diagonal principal. Para a estrutura CSH, o parâmetro de correlação ρ permanece constante, e para a estrutura ARH(1), ρ é um parâmetro auto-regressivo, o que significa que as correlações vão decrescendo exponencialmente (CASTRO, 1999).

VII. Toeplitz (TOEP)

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}, \text{ é a estrutura de covariância de um processo estocástico de médias}$$

móveis de ordem $q = (n_i - 1)$, neste exemplo $q = 3$.

VIII. Auto Regressiva de Primeira Ordem Médias Móveis (ARMA(1,1))

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\rho & \lambda\rho^2 \\ \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho & \gamma & 1 & \gamma \\ \lambda\rho^2 & \gamma\rho & \gamma & 1 \end{bmatrix}, \text{ é a estrutura associada aos dados de séries temporais com}$$

parâmetro auto-regressivo ρ , componente de médias móveis γ , sendo σ^2 a variância residual. O número de parâmetros envolvidos é igual a três. Conforme Castro (1999), a estrutura ARMA(1,1) é um caso particular dos modelos Auto Regressivos de Média Móvel. Estes modelos são particularmente atraentes quando existem poucas unidades experimentais, um número relativamente grande de observações por unidade experimental e estas observações são igualmente espaçadas no tempo (o que comumente chamam-se séries temporais).

IX. Toeplitz Heterogênea (TOEPH)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & 0 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_1\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ 0 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{bmatrix}, \text{ dados de séries temporais igualmente}$$

espaçados, com parâmetros de variâncias diferentes para cada elemento da diagonal, sendo os elementos fora da diagonal principal, funções de variâncias e do k-ésimo parâmetro de autocorrelação ($|\rho_k| < 1$) para cada defasagem q-1, e zeros para as últimas defasagens.

X. Ante-Dependência de Ordem 1 (ANTE(1))

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \sigma_1\sigma_4\rho_1\rho_2\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 \\ \sigma_3\sigma_1\rho_2\rho_1 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_3\sigma_4\rho_3 \\ \sigma_4\sigma_1\rho_3\rho_2\rho_1 & \sigma_4\sigma_2\rho_3\rho_2 & \sigma_4\sigma_3\rho_3 & \sigma_4^2 \end{bmatrix}, \text{ imp\~{o}e par\~{a}metros de vari\~{a}ncia}$$

diferentes para cada elemento da diagonal, sendo os elementos fora da diagonal principal, fun\c{c}oes de vari\~{a}ncias e do k-ésimo par\~{a}metro de autocorrela\c{c}ao, satisfazendo ($|\rho_k| < 1$).

Envolvem $2n_i + 1$ par\~{a}metros.

Esta estrutura permite que as vari\~{a}ncias sejam diferentes e é aplic\~{a}vel em estudos longitudinais em que as condi\c{c}oes de avalia\c{c}ao n\~{a}o s\~{a}o igualmente espa\c{c}adas, e apresentam heterogeneidade de vari\~{a}ncia e correla\c{c}ao serial (BARBOSA, 2009).

XI. Correla\c{c}ao sem Estrutura (UNR)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{21} & \sigma_1\sigma_3\rho_{31} & \sigma_1\sigma_4\rho_{41} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_2\sigma_3\rho_{31} & \sigma_2\sigma_4\rho_{42} \\ \sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_3\sigma_4\rho_{43} \\ \sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2 \end{bmatrix}, \text{ especifica uma matriz de vari\~{a}ncia-}$$

covari\~{a}ncia completamente geral em termos de vari\~{a}ncias e correla\c{c}oes. Essa estrutura ajusta o mesmo modelo que o tipo UN, mas com diferente parametriza\c{c}ao. Sendo σ_i^2 o i-ésimo par\~{a}metro de vari\~{a}ncia e ρ_{jk} é a correla\c{c}ao entre a j-ésima e a k-ésima medida satisfazendo $|\rho_{jk}| < 1$.

XII. Espacial (SP(POW)(c-list)

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{12}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{13}} & \rho^{d_{23}} & 1 & \rho^{d_{34}} \\ \rho^{d_{14}} & \rho^{d_{24}} & \rho^{d_{34}} & 1 \end{bmatrix}, \text{ as covari\~{a}ncias s\~{a}o fun\c{c}oes da dist\~{a}ncia Euclidiana}$$

entre os vetores especificados pelas coordenadas. O par\~{a}metro c-list associado a essa estrutura espacial, corresponde aos nomes das vari\~{a}veis numéricas no conjunto de dados. Essas vari\~{a}veis s\~{a}o usadas como coordenadas das observa\c{c}oes no espa\c{c}o. A estrutura de covari\~{a}ncia espacial tem grande utilidade quando os dados longitudinais est\~{a}o tomados no espa\c{c}o e n\~{a}o no tempo, pois considera as dist\~{a}ncias euclidianas entre medidas repetidas (CASTRO, 1999).

O expoente d_{ij} é a distância Euclidiana entre o nível i e o nível j do fator de repetição, que, neste caso, é espacial.

XIII. Diagonal Principal “Banded” (UN(1))

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}, \text{ é caracterizada por variâncias desiguais e covariâncias nulas.}$$

XIV. Huynh-Feldt (H-F)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \frac{(\sigma_1^2 + \sigma_2^2)}{2} - \lambda & \frac{(\sigma_1^2 + \sigma_3^2)}{2} - \lambda & \frac{(\sigma_1^2 + \sigma_4^2)}{2} - \lambda \\ \frac{(\sigma_1^2 + \sigma_2^2)}{2} - \lambda & \sigma_2^2 & \frac{(\sigma_2^2 + \sigma_3^2)}{2} - \lambda & \frac{(\sigma_2^2 + \sigma_4^2)}{2} - \lambda \\ \frac{(\sigma_3^2 + \sigma_1^2)}{2} - \lambda & \frac{(\sigma_3^2 + \sigma_2^2)}{2} - \lambda & \sigma_3^2 & \frac{(\sigma_3^2 + \sigma_4^2)}{2} - \lambda \\ \frac{(\sigma_4^2 + \sigma_1^2)}{2} - \lambda & \frac{(\sigma_4^2 + \sigma_2^2)}{2} - \lambda & \frac{(\sigma_4^2 + \sigma_3^2)}{2} - \lambda & \sigma_4^2 \end{bmatrix}, \text{ é similar à estrutura}$$

CSH, pois tem a mesma heterogeneidade não estruturada ao longo da diagonal principal e o mesmo número de parâmetros desconhecidos, entretanto, a estrutura HF constrói os elementos fora da diagonal, calculando uma média aritmética. Castro (1999) comenta que os testes estatísticos para a estrutura esférica têm servido como uma regra chave na decisão entre os tradicionais procedimentos univariado e multivariado para dados longitudinais. Entretanto, dificilmente se ajusta esta estrutura para os dados, pois as análises padrão exploram o fato de que um conjunto de contrastes ortogonais de dados com esta estrutura de covariância tem uma estrutura diagonal.

Além destas estruturas de variância-covariância, outras estão disponibilizadas no *proc MIXED*, o problema surge em qual estrutura a ser testada em G e R? Wald (2000) sugere que para o modelo linear misto (13), a estrutura geral de variância e variância-covariância, geralmente, é obtida pela especificação de uma estrutura de variância-covariância geral para G e, qualquer uma das várias estruturas comumente utilizada para R. Neste sentido, Lima (1996) argumenta que geralmente assume-se que G é uma matriz de dispersão não estruturada (UN) ou, em alguns casos, pode ser considerada como auto-regressiva de primeira ordem.

Vonesch; Chinchilli (1997) relatam que as estruturas mais comuns para R são a de componente de variância, simetria composta e auto-regressiva de primeira ordem.

Littel et al. (2000) sugerem que para o ajuste das estruturas de variância-covariância deve se iniciar com estruturas mais parcimoniosas, tais como: a simetria composta, não estruturada e auto-regressiva. Além disso, os autores sugerem que a comparação dessas estruturas seja feita por meio dos critérios de informação de Akaike (AIC) e do Baysiano de Schwarz (BIC).

2.5.3.5 Seleção de modelos e testes

Selecionar o melhor modelo significa, não só selecionar a melhor estrutura para as médias (parte fixa), como também a melhor estrutura de covariância. Já que modelo misto é construído em três etapas: seleção dos efeitos fixos; identificação dos efeitos aleatórios; e, estimação e comparação dos modelos.

Existe uma série de técnicas gráficas e analíticas que auxiliam na escolha das estruturas das matrizes de variância dos modelos lineares mistos, e para estudos longitudinais é razoável utilizar informações sobre o comportamento da resposta ao longo das ocasiões de avaliação na modelagem da estrutura de covariância intra-unidades amostrais (ROCHA, 2001).

Para as variadas estruturas de variância-covariância propostas para G e R, um critério objetivo de seleção deve ser seguido. Neste sentido, Kincaid (2010) argumenta que depois de ter os efeitos aleatórios determinados, então seleciona-se a estrutura de covariância. Para a escolha da estrutura de covariância pode-se incluir: o número de parâmetros, a interpretação da estrutura, os resultados de diagnóstico, e os efeitos sobre os efeitos fixos.

Muitos métodos já foram desenvolvidos visando facilitar a escolha da estrutura de covariância que melhor explique o comportamento da variabilidade e da correlação entre as medidas repetidas e, dessa forma, selecionando o modelo adequado (FLORIANO et al., 2006). Dentre aqueles métodos, desatacam-se o AIC e o BIC, que são, na verdade, valores

para os logaritmos das funções de verossimilhanças do modelo, que dependem do número de observações e do número de parâmetros do modelo (COSTA, 2003).

O critério de AIC, baseando-se no logaritmo de máxima verossimilhança (ML) e máxima verossimilhança restrita (REML), respectivamente, são calculados por $AIC = -2L(\hat{\theta}) + 2d$, em que: d representa o número total de parâmetros de efeito fixo e aleatório estimado no modelo. Considera-se o melhor modelo, o que apresentar menor valor de AIC.

O critério de BIC é calculado por: $BIC = -2L(\hat{\theta}) + \ln(N)d$, em que: $N = \sum n_i$ é o somatório de todas as n observações. O melhor modelo será o que apresentar o menor valor de BIC.

Os critérios de informação tendem a apresentar termos de penalização que são uma função do número de parâmetros (d) e alguma medida do tamanho da amostra (n) (COSTA, 2003).

Keselman et al. (1998) investigaram o uso dos critérios de AIC e BIC sobre condições, tais como: heterogeneidade de covariância, não normalidade e tamanho dos grupos desiguais. Concluíram que o critério AIC selecionou a estrutura correta em apenas 47% dos casos testados, e o critério BIC, em apenas 35% dos casos. Recentemente, Ferron et al. (2002) realizaram um estudo sobre a utilização desses dois critérios para modelar curvas de crescimento, encontrando uma taxa de sucesso para o critério AIC 70% e 45%, para o critério BIC. West et al. (2007) alertam para o fato de que alguns softwares calculam os valores de AIC e BIC utilizando equações diferentes, dependendo se a estimação é feita por ML ou REML.

Outros dois testes para comparar estruturas de covariância são: a estatística Z de Wald e o Teste Assintótico de Verossimilhança. A estatística Z de Wald é calculada com o parâmetro estimado dividido por seu erro padrão assintótico (CAMARINHA FILHO, 2002). Os erros padrões assintóticos são obtidos a partir da inversa da matriz de derivada segunda da verossimilhança, em relação a cada um dos parâmetros de efeito aleatório. A estatística Z de Wald é válida para grandes amostras, mas ela pode ser incerta para pequenos conjuntos de dados e para parâmetros, tais como, componentes de variância, que apresentam uma distribuição assimétrica ou distribuição amostral limite. Complementando, Camarinha Filho (2002) relata que uma alternativa melhor é a razão de verossimilhança χ^2 . Essa estatística

compara dois modelos de covariância, um como caso especial do outro. Para realizar esse teste, ajusta-se o modelo completo e o modelo reduzido e, então, subtrai-se os valores correspondentes a -2 vezes o log verossimilhança. Pode-se usar o ML ou REML para construir esta estatística, que testa se o modelo completo é melhor do que o modelo reduzido.

O teste assintótico de verossimilhança compara dois modelos estimados por máxima verossimilhança, em que um dos modelos é uma versão restrita do outro, ou seja, um modelo tem p parâmetros adicionais (XAVIER, 2000).

A estatística log-verossimilhança ($-2\log Ver_{ss}$) é baseada no logaritmo da razão entre duas verossimilhanças do modelo mais simples $l(\tilde{\theta})$, e o modelo mais complexo, $l(\hat{\theta})$ a qual segue uma *Qui-Quadrado*, com r graus de liberdade, ou seja $-2\log Ver_{ss} = 2[\log(l(\tilde{\theta})) - \log(l(\hat{\theta}))] \sim \chi_r^2$, a hipótese testada H_0 : o modelo mais simples é adequado. Se a hipótese nula for aceita, os valores da função de verossimilhança avaliadas em $\tilde{\theta}$ e $\hat{\theta}$ devem estar próximos, indicando que os dados se ajustam ao modelo com menor número de parâmetros (BARBOSA, 2009). A desvantagem desse teste, embora seja bastante eficaz, é que só pode ser usado para comparar dois modelos de cada vez, sendo que um desses modelos é sempre um caso especial do outro (GUIMARÃES, 1994).

Quando os modelos são ajustados por REML, os valores de AIC e BIC e log-verossimilhança somente podem ser comparados entre modelos com mesma estrutura de efeitos fixos. Já quando os modelos são ajustados por ML, os valores de AIC e BIC podem ser comparados com quaisquer modelos ajustados para os mesmos dados. Neste último caso, a qualidade do ajuste pode ser avaliada para diferentes especificações de efeitos fixos ou diferentes especificações de efeitos aleatórios ou ambos (BATES; PINHEIRO, 2000).

2.5.3.6 Inferência sobre os efeitos fixos e aleatórios

Finalmente, para inferências relativas aos parâmetros de efeitos fixos e aleatórios no modelo misto, consideram-se combinações lineares estimáveis da seguinte forma:

$$\omega = L \begin{bmatrix} \beta \\ b \end{bmatrix}$$

Funções desta natureza são chamadas funções estimáveis, se na parte fixa β satisfaça a exigência de estimabilidade, uma vez que, qualquer combinação linear de b é estimável. Tal como uma formulação em termos de uma matriz geral L contempla vários procedimentos inferenciais comuns, como aqueles empregados nos testes Tipo I e III e médias de mínimos quadrados (RIBOLDI, 2007).

A inferência em efeitos fixos é o foco e, neste caso, a porção b associada a L é suposta conter todos os elementos nulos, ou seja, igual a zero. Inferências estatísticas, portanto, podem ser obtidas testando as hipóteses:

$$H_0 : L \begin{bmatrix} \beta \\ u \end{bmatrix} = 0 \text{ ou, construindo estimativas por pontos e/ou por intervalos.}$$

Quando L consiste em uma única linha, uma estatística t geral pode ser construída como segue:

$$t = \frac{L \begin{bmatrix} \hat{\beta} \\ u \end{bmatrix}}{\sqrt{L\hat{C}L'}} \quad (18)$$

Sob as pressuposições de normalidade de b e \mathcal{E} sendo que t tem uma distribuição t exata, somente para dados que exibem certos tipos de balanceamentos e para alguns casos desbalanceados especiais. Em geral, t é somente aproximadamente t distribuído, e seus graus de liberdade devem ser estimados, por vários métodos, dentre os quais os de Satterthwaite e Betwithin (MacLEAN; SANDERS, 1988; FAI; CORNELIUS, 1996; KENWARD; ROGER, 1997).

Se $\hat{\nu}$ forem os graus de liberdade associados com $\hat{\omega}$, o intervalo de confiança associado a $\hat{\omega}$ é o seguinte:

$$L \begin{bmatrix} \hat{\beta} \\ u \end{bmatrix} \pm t_{\hat{\nu}, \alpha/2} \sqrt{L\hat{C}L'} \quad (19)$$

Em que: $t_{\hat{\nu}, \alpha/2}$ é o percentil $(1 - \alpha/2)100\%$ da distribuição $t_{\hat{\nu}}$.

Quando o posto de L é maior do que 1, a seguinte estatística geral F deve ser considerada:

$$F = \frac{\begin{bmatrix} \hat{\beta} \\ u \end{bmatrix}' L' (L' \hat{C} L)^{-1} L \begin{bmatrix} \hat{\beta} \\ u \end{bmatrix}}{\text{posto}(L) \hat{\sigma}^2} \quad (20)$$

Analogamente para t , F em geral tem uma distribuição aproximada F , com posto (L), como graus de liberdade do numerador e \hat{v} graus de liberdade do denominador. Para amostras grandes, a equação (18) tem distribuição normal e a equação (20) tem uma distribuição χ^2 .

As estatísticas t e F permitem que se façam inferências sobre os efeitos fixos levando em conta o modelo de variância-covariância selecionado. Uma alternativa para grandes amostras é a estatística χ^2 associada com o teste da razão de verossimilhança. Esta estatística compara dois modelos de efeitos fixos, um é caso especial do outro (aninhado ou encaixado), que somente é calculada quando comparamos diferentes modelos de covariância, embora, se deva usar ML e não REML, por que falta o termo associado com verossimilhança restrita que depende da especificação dos efeitos fixos (CAMARINHA FILHO, 2003; RIBOLDI, 2007).

Na verdade, o teste Wald pode ser utilizado para avaliar a significância dos efeitos fixos e o teste de razão de verossimilhança, para avaliar tanto os efeitos fixos como os aleatórios em modelos encaixados, por intermédio de comparação de modelos mais simples com mais gerais. Quando os modelos não são encaixados, pode-se utilizar o critério de informação de Akaike (CAMARINHA FILHO, 2002).

2.5 O SAS na análise de modelos mistos

As análises com medidas repetidas no tempo, por longa data eram realizadas através do GLM no SAS (LI, 2010). No entanto, com o pacote de ferramentas para a análise de modelo misto (*proc mixed*), introduzido no SAS (versão 6.09) em 1992, essas análises passaram a ser melhor elaboradas, já que torna-se possível modelar as estruturas de variância-covariância. As estruturas de covariância, para as matrizes G e R podem ser estimadas através de vários métodos, através do comando `METHOD` contido no *proc mixed*.

Pelo SAS, define-se o método de estimação, que no presente estudo será o REML. O *proc mixed* permite também informar a estrutura da matriz G , através do comando

RANDOM, e a dos erros R, através do comando REPEATED. O comando REPEATED modela a variação dentro da unidade experimental. Define também a estrutura de covariância de ε , isto é, R, que é definida como bloco diagonal, com cada bloco correspondendo a uma unidade experimental (RIBOLDI, 1999).

Dessa forma, a maneira correta para a análise de dados com medidas repetidas é o uso do procedimento MIXED do SAS, que emprega um modelo misto e considera a variação entre as unidades experimentais, que são decorrentes das diferenças entre os níveis de tratamentos e a variação dentro das unidades experimentais, que são atribuídas as mensurações em diferentes tempos (CROWDER; HAND, 1990; LITTELL et al., 1998).

Portanto, no decorrer deste estudo serão apresentados os principais comandos do SAS com o proc mixed, objetivando o ajuste de curvas de crescimento.

3 MATERIAL E MÉTODOS

3.1 Instalação do experimento

O Experimento com a cultura de mamoneira foi conduzido no ano agrícola 2007/2008. O local foi na área Experimental do Departamento de Fitotecnia, Centro de Ciências Rurais, da Universidade Federal de Santa Maria, Santa Maria, RS (latitude: 29°41'S, longitude: 53°41'W e altitude: 95m), que está localizada na Região da Depressão Central do Estado do Rio Grande do Sul. O clima da região, segundo a classificação de Köeppen, é Cfa Subtropical úmido sem estação seca definida com verões quentes (MORENO, 1961).

A cultivar utilizada foi a Vinema T1, produzida pela Vinema Multióleos Vegetais. Apresenta porte médio, com alta porcentagem de flores femininas e frutos indeiscentes, muito ramificada, o que aumenta a produtividade. Quanto ao ciclo vegetativo, pode variar de 90 a 150 dias.

3.2 Semeadura e Manejo da Cultura

Três meses antes da instalação do experimento em campo, foi realizada a análise de solo no Laboratório de Análises de Solos da UFSM, na qual se verificou a necessidade de adição de calcário. Neste momento, foram adicionadas cinco toneladas por hectare de calcário Filler, sendo a incorporação do mesmo, realizada através de uma lavração com arado de discos. Posteriormente, foram realizadas três gradagens, com grade de discos (10/09/2007, 10/10/2007 e 18/10/2007), com o objetivo de nivelamento e destorroamento do solo. Conforme o período recomendado (de 21 de setembro a 31 de dezembro) para a semeadura da cultura (WREGGE et al., 2007), na região de Santa Maria, a semeadura foi realizada de forma manual, no dia 20 de outubro de 2007. As covas foram abertas com auxílio de enxadas, sendo colocadas três sementes por covas, a uma profundidade entre 5 a 8 cm. Vinte dias após a semeadura (DAS), foi realizado o desbaste manual.

No momento da semeadura, foi realizada a adubação de base, de acordo com resultados de análise do solo, realizada duas semanas antes da semeadura, e seguindo as recomendações de adubação e calagem para a cultura de mamona no Rio Grande do Sul (SCIVITTARO; PILLON, 2006). A adubação de base foi constituída de: 15kg ha⁻¹ de N (33,33kg ha⁻¹ de uréia), 45kg ha⁻¹ de P₂O₅ (98kg ha⁻¹ de Supertriplo) e 15kg ha⁻¹ de K₂O (26kg ha⁻¹ de Cloreto de potássio). Em cobertura foi aplicado apenas nitrogênio (80kg ha⁻¹) na forma de uréia (178kg ha⁻¹).

Foram semeadas três parcelas com nove fileiras cada, com espaçamento fixo entre linhas de 1m. O espaçamento entre plantas (densidade de cultivo) foi de 1,0; 1,2; 1,4 m. A área da parcela foi de 22 m de comprimento por 8m de largura. Desta forma a área total do experimento foi de 176 m².

O controle de plantas daninhas, durante todo ciclo da cultura, foi realizado através de capinas manuais. O controle de pragas e doenças foi efetuado segundo recomendações de Embrapa (2006), Soares et al. (2007) e Araújo; Coutinho (2007).

3.3 Variável resposta e coleta de dados

O modelo matemático foi inteiramente casualizado, composto por três densidades de cultivo, mantendo-se o espaçamento entre linhas de 1,0 m. As densidades correspondem:

1, 2 m;

1, 4 m;

1,0 m.

A variável resposta foi altura de planta, medida em centímetros (cm), com a coleta de dados iniciando 10 dias após a emergência, ou seja, a primeira coleta foi dia 9 de novembro de 2007. Foram realizadas oito coletas com intervalos igualmente espaçados de 20 dias, até as plantas completarem 150 dias de ciclo. Durante cada coleta eram utilizadas quatro plantas por densidade. Os dados da variável altura de planta (Tabela A) encontram-se no Apêndice 1.

3.4 Procedimentos estatísticos

Na construção das curvas de crescimento, foi utilizada uma abordagem do modelo linear misto (13), sendo esta técnica detalhada em 2.5.3.

O procedimento adotado constou de três etapas:

- i. um modelo para os efeitos fixos ($X\beta$), que descreve o comportamento das curvas médias de resposta;
- ii. um modelo para os efeitos aleatórios (Zb), que juntamente com os efeitos fixos, descreve as curvas individuais;
- iii. um modelo para matriz de dispersão condicional (R), que descreve as variância-covariância das observações ao longo do tempo.

3.4.1 Seleção dos efeitos fixos

As informações necessárias para a definição de um modelo para efeitos fixos foram fornecidas utilizando a técnica de análise univariada de perfil (2.5.1) e análise multivariada de perfil (2.5.2), respectivamente. Sendo os candidatos a efeito fixos, os seguintes fatores: densidades, tempos e suas interações.

A estrutura dos dados para a análise de dados longitudinais encontra-se na Tabela 1 (em 2.4). O modelo matemático (2) para a análise univariada em esquema de parcelas subdivididas no tempo está apresentado em 2.5.1. Para a análise multivariada de perfil foi utilizado o modelo matemático (8), sem o componente de bloco.

Quando a hipótese de não efeito da interação densidade*tempo for rejeitada, serão realizados testes sobre tendências de variação das respostas médias das densidades. Estes testes foram utilizados para determinar o grau da curva polinomial que melhor explique o comportamento da variável altura de planta.

3.4.2 Identificação dos efeitos aleatórios

Para a identificação dos efeitos aleatórios foram realizadas análises de regressão polinomial para cada perfil individual de respostas e selecionou-se os efeitos aleatórios a partir do teste *t-Student*, para verificar a significância dos coeficientes. Foram considerados como efeitos aleatórios, aqueles coeficientes que se apresentarem significativamente diferentes de zero. A adequação dos modelos, quanto à normalidade, foi verificada através da análise de resíduos.

3.4.3 Estimação e comparação de modelos

Após a identificação dos efeitos fixos ($X\beta$) e aleatórios (Zb), do modelo linear misto de crescimento (13), através das técnicas citadas em 3.4.1 e 3.4.2 e revisadas em 2.5.1 e 2.5.2, respectivamente. Determinou-se uma estrutura parcial para este modelo, definindo-se as colunas da matriz delineamento X e da matriz de incidência Z . Sendo assim, as colunas da matriz X estão associadas às densidades que tem efeito significativo na resposta e ao tempo. Este último indicou o grau da curva polinomial a ser ajustada às densidades, que possuem interação significativa com o tempo. Já a matriz Z é constituída por potências dos níveis de tempo, cujos coeficientes foram identificados como efeitos aleatórios.

Considerando que os efeitos fixos, selecionados para a altura de planta, sejam o efeito principal de densidade e da interação densidade x tempo. Ainda, o comportamento das respostas médias associadas às densidades possa ser explicado por curvas quadráticas, sendo o intercepto, o coeficiente linear e o coeficiente quadrático aleatórios.

O modelo linear misto de crescimento (13), para cada unidade experimental, “ y_i ”, pode ser escrito como: $Y_i = X_i \beta + Z_i b_i + \varepsilon_i$, em que cada parâmetro do modelo foi interpretado da mesma forma que em (13). Sendo assim, para a primeira unidade experimental (y_1) e a última unidade (y_{12}), as matrizes correspondem:

– Para “ y_1 ”

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \\ y_{115} \\ y_{116} \\ y_{117} \\ y_{118} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & | & 10 & 0 & 0 & | & 100 & 0 & 0 \\ 1 & 0 & 0 & | & 30 & 0 & 0 & | & 900 & 0 & 0 \\ 1 & 0 & 0 & | & 50 & 0 & 0 & | & 2500 & 0 & 0 \\ 1 & 0 & 0 & | & 70 & 0 & 0 & | & 4900 & 0 & 0 \\ 1 & 0 & 0 & | & 90 & 0 & 0 & | & 8100 & 0 & 0 \\ 1 & 0 & 0 & | & 110 & 0 & 0 & | & 12100 & 0 & 0 \\ 1 & 0 & 0 & | & 130 & 0 & 0 & | & 16900 & 0 & 0 \\ 1 & 0 & 0 & | & 150 & 0 & 0 & | & 22500 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \end{bmatrix} + \begin{bmatrix} 1 & 10 & 100 \\ 1 & 30 & 900 \\ 1 & 50 & 2500 \\ 1 & 70 & 4900 \\ 1 & 90 & 8100 \\ 1 & 110 & 12100 \\ 1 & 130 & 16900 \\ 1 & 150 & 22500 \end{bmatrix} \begin{bmatrix} b_{(1.1)^0} \\ b_{(1.1)^1} \\ b_{(1.1)^2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{113} \\ \varepsilon_{114} \\ \varepsilon_{115} \\ \varepsilon_{116} \\ \varepsilon_{117} \\ \varepsilon_{118} \end{bmatrix}$$

– Para “ y_2 ”

$$\begin{bmatrix} y_{341} \\ y_{342} \\ y_{343} \\ y_{344} \\ y_{345} \\ y_{346} \\ y_{347} \\ y_{348} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & | & 0 & 0 & 10 & | & 0 & 0 & 100 \\ 0 & 0 & 1 & | & 0 & 0 & 30 & | & 0 & 0 & 900 \\ 0 & 0 & 1 & | & 0 & 0 & 50 & | & 0 & 0 & 2500 \\ 0 & 0 & 1 & | & 0 & 0 & 70 & | & 0 & 0 & 4900 \\ 0 & 0 & 1 & | & 0 & 0 & 90 & | & 0 & 0 & 8100 \\ 0 & 0 & 1 & | & 0 & 0 & 110 & | & 0 & 0 & 12100 \\ 0 & 0 & 1 & | & 0 & 0 & 130 & | & 0 & 0 & 16900 \\ 0 & 0 & 1 & | & 0 & 0 & 150 & | & 0 & 0 & 22500 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \end{bmatrix} + \begin{bmatrix} 1 & 10 & 100 \\ 1 & 30 & 900 \\ 1 & 50 & 2500 \\ 1 & 70 & 4900 \\ 1 & 90 & 8100 \\ 1 & 110 & 12100 \\ 1 & 130 & 16900 \\ 1 & 150 & 22500 \end{bmatrix} \begin{bmatrix} b_{(3.4)^0} \\ b_{(3.4)^1} \\ b_{(3.4)^2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{341} \\ \varepsilon_{342} \\ \varepsilon_{343} \\ \varepsilon_{344} \\ \varepsilon_{345} \\ \varepsilon_{346} \\ \varepsilon_{347} \\ \varepsilon_{348} \end{bmatrix}$$

A três primeiras colunas das matrizes X_1 e X_{12} estão relacionadas aos interceptos das três curvas (β_1 , β_2 e β_3); as três colunas intermediárias correspondem aos coeficientes lineares (β_4 , β_5 e β_6); e, as últimas três colunas referem-se aos coeficientes quadráticos (β_7 , β_8 e β_9). Quanto às três colunas de Z_1 e Z_{12} descrevem: os interceptos, coeficientes lineares e coeficientes quadráticos individuais (aleatórios).

Definindo os p elementos do vetor β e os q elementos que irão constituir b_i e construída as matrizes X_i e Z_i , ou seja, quando têm-se todas as unidades experimentais, o modelo pode ser escrito na forma (13). Em que: $Y_{(96 \times 1)}$, $X_{(96 \times p)}$, $\beta_{(p \times 1)}$, $Z_{(96 \times 12q)}$, $b_{(54q \times 1)}$ e $\varepsilon_{(96 \times 1)}$.

Dando sequência aos procedimentos, o próximo passo foi selecionar as estruturas mais adequadas para as matrizes de covariância G e R , (item 2.5.3.4) a fim de modelar a variância ($V = ZGZ' + R$) do vetor de observações y . Para a estimação foi utilizado o método ML ou REML e os testes de razão de verossimilhança e os critérios de informação de Akaike (AIC) e de Schwarz (BIC).

Realizada a seleção de uma estrutura para as matrizes de covariância, a estimação do modelo final para os efeitos fixos foi realizada pelo método REML, através de testes da razão da verossimilhança e dos critérios de Akaike e Schwarz, e testes de interesse serão realizados pelo teste Tukey-Kramer ou de Wald. O diagnóstico dos modelos ajustados é feito através de métodos gráficos e pelo teste de Shapiro-Wilks.

4 RESULTADOS E DISCUSSÃO

4.1 Altura de planta

4.1.1 Seleção dos efeitos fixos

Os dados da variável altura de planta encontram-se no Apêndice A. Na Figura 4 são apresentados os diagramas paralelos individuais de perfil para as três densidades. Na Tabela 4 encontram-se os valores médios, desvio padrão (S) e coeficiente de variação (CV) da variável altura de planta para cada dia de avaliação e densidade utilizada

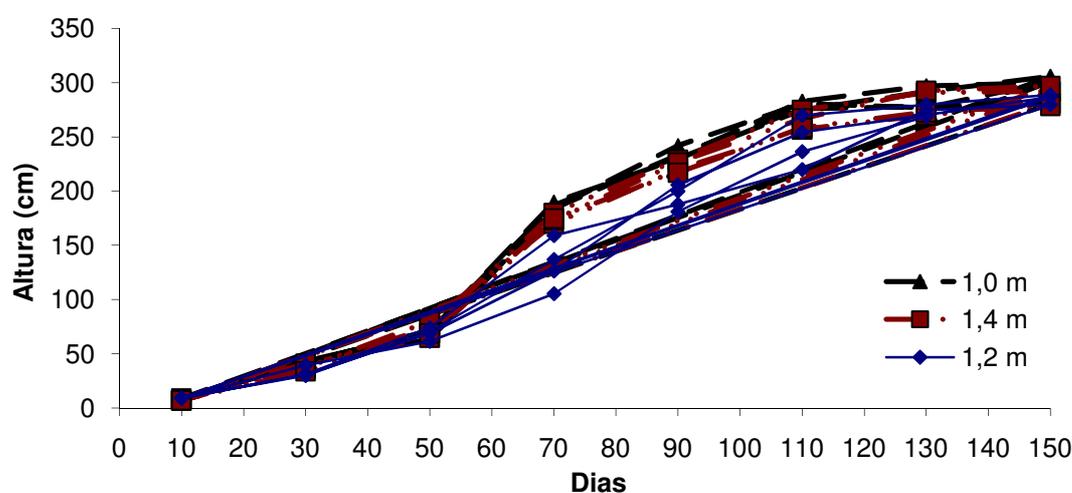


Figura 4 – Diagrama paralelo de dispersão dos perfis individuais dos dados de altura de planta de mamona para as doze unidades experimentais. Santa Maria, RS, 2010.

Observa-se que a altura de planta de mamona aumenta até, aproximadamente, 110 dias, quando então diminui a taxa de crescimento, e que até os 50 dias, as densidades são muito semelhantes. Além disso, não são verificados perfis discrepantes entre as densidades. Percebe-se que a densidade 1,0 m influencia de forma mais marcante no crescimento da

planta; já a densidade 1, relacionado à densidade de cultivo de 1,2 m, apresenta menor influência.

Tabela 4 – Média (cm), desvio padrão (S) e coeficiente de variação (CV) da variável altura de planta de mamona, coletadas em quatro plantas por densidade, de acordo com os dias de avaliação e densidade. Santa Maria, RS, 2010.

Dias	Densidades								
	1,0 m			1,2 m			1,4 m		
	Média	S (±)	CV(%)	Média	S (±)	CV(%)	Média	S (±)	CV(%)
10	7,83	0,90	11,54	9,30	0,14	1,52	7,70	0,84	10,87
30	42,85	0,81	1,88	32,98	4,39	13,31	37,73	2,97	7,88
50	68,91	3,72	5,39	68,79	5,32	7,74	71,98	6,47	8,99
70	186,34	2,91	1,56	131,98	22,27	16,87	174,60	4,15	2,38
90	234,28	5,36	2,29	193,83	11,21	5,78	223,13	5,52	2,47
110	278,83	3,79	1,36	245,30	21,70	8,85	264,45	8,34	3,15
130	286,10	10,12	3,54	273,58	5,00	1,83	281,65	11,86	4,21
150	296,98	11,15	3,75	281,55	8,43	2,99	287,23	8,76	3,05

A densidade de cultivo de 1,0m proporcionou as maiores médias de altura de planta em, praticamente, todos os dias de avaliação seguidas da densidade 1,4 m (Tabela 4). Quanto à variabilidade dos dados, os desvios padrão e coeficientes de variação oscilam entre valores baixos e altos, conforme o período avaliado. Este comportamento é esperado para variáveis medidas no tempo, que se caracterizam por heterocedasticidade de variâncias.

Tendo como base o diagrama paralelo de dispersão dos perfis individuais foram testadas as hipóteses de perfis: paralelos (a interação entre as densidades e os dias de avaliação é nula); coincidentes (verificação de não efeitos de densidades) e horizontais (verificação de não efeitos dos dias de avaliação). Estas hipóteses foram testadas através das análises univariada e multivariada de perfil, detalhadas em 2.5.1 e 2.5.2.

4.1.1.2 Análise Univariada de Perfil

Na análise de variância univariada, por meio do *proc GLM*, em esquema de parcelas subdivididas no tempo, os testes envolvendo as subparcelas, ou efeitos dentro de sujeitos (dias e interação Densidades x Dias), podem não ser válidos, caso a matriz Σ não satisfaça a condição de esfericidade. Assim como não seja efetuada a correção dos graus de liberdade associados às causas de variação que envolva o fator dias. Já para o fator entre indivíduos, envolvendo as parcelas, relacionado a efeitos de densidades, é realizado corretamente quando se especifica o termo de erro que deva ser utilizado para testar as densidades. Sendo assim, o programa deve seguir os seguintes passos:

```
proc glm data=univariado;
class dens dias rep;
model A=dens dias dias*dens rep(dens);
test h=dens e=rep(dens);
run; quit;
```

A opção *test* especifica que o fator densidade (h=dens) deve ser testado com o termo repetição dentro de densidade (e=rep(dens)) que é o erro da parcela. O resultado da análise univariada de perfil para o esquema em parcelas subdivididas encontra-se na Tabela 5.

Tabela 5 – Resultado da análise univariada de perfil para os dados da variável altura de planta de mamona, submetida a três densidades de cultivo e oito medidas no tempo igualmente espaçadas. Santa Maria, RS, 2010.

Causas de Variação	GL ¹	SQ ²	QM ³	F ⁴	Pr>F ⁵
Modelo	32	1102683,699	34458,866	512,850	<0,0001
Resíduo	63	4233,069	67,192		
Total Corrigido	95	1106916,768			
Densidade	2	7066,383	3533,191	21,020	0,0004
Dias	7	1087831,503	155404,500	2312,860	<0,0001
Densidades x Dias	14	6272,791	448,057	6,670	<0,0001
Repetição(Densidade)	9	1513,021	168,113	2,500	0,0163
R ² =0,9962; CV=4,93%; Média da Altura de planta (cm)=166,16					

1- graus de liberdades; 2- soma de quadrados; quadrados médio; 4 - teste F de Snedecor; 5- nível de significância para a hipótese nula.

A análise de variância, para as diferenças entre densidades (ou entre indivíduos), testa a hipótese nula de não haver efeito de densidades, sendo esta rejeitada ($<0,0001$). O que significa, neste caso, que a altura de planta apresenta comportamento diferenciado conforme a densidade de cultivo. Quanto aos testes de hipóteses para efeito dentro de sujeitos, testado nas subparcelas, observa-se, em um primeiro momento, que o efeito Dias assim como a interação Densidades x Dias são significativos. Porém, estes resultados serão válidos somente se realizado o teste de esfericidade e, caso necessário, as correções dos graus de liberdade. O programa para realização destes procedimentos encontra-se a seguir e os resultados das análises podem ser visualizados na Tabela 6.

```
proc glm data=multivariado;
class dens;
model A10 A30 A50 A70 A90 A110 A130 A150= dens/nouni;
repeated dias 8 polynomial/summary printm printe;
run; quit;
```

O comando *class* informa que as variáveis especificadas são classificatórias. No comando *model*, especifica-se o modelo que será analisado (variáveis dependentes e variáveis independentes). Em *nouni*, especifica-se que não se deve executar as análises unidimensionais, dentro de cada tempo. Com o comando *repeated* é que se determina que a análise é de variância com medidas repetidas. Este comando executa análises uni e multivariadas, incluindo o teste de esfericidade para matriz de variância e covariância e as correções de graus de liberdade sugeridas na literatura.

As opções: *polynomial* gera contrastes de polinômios ortogonais para os níveis do fator tempo; *summary* apresenta a tabela da análise de variância dos contrastes; *printm* imprime a matriz dos contrastes (denominada *m*). O teste de esfericidade da matriz de variância-covariância, além da correção para os graus de liberdade, é gerado com a opção *printe*.

No teste de esfericidade, realizado com 27 graus de liberdade, critério de Mauchly (0,001062) e aproximação de *Qui-Quadrado* (44,179), constatou-se que a condição de esfericidade foi violada em nível de significância de 0,00198. Isso significa que a matriz de variância-covariância não pode ser considerada do tipo H-F, uma vez que a suposição de esfericidade não foi atendida. Quanto à esfericidade, têm-se três situações, que indicam ao pesquisador o teste a ser utilizado (FERNANDEZ, 1991):

- i. se a condição de H-F para a matriz de variância-covariância for satisfeita (teste de esfericidade não significativo) o teste univariado pode ser utilizado;
- ii. se a condição H-F para a matriz de variância-covariância não for satisfeita, e o nível de significância do teste de esfericidade estiver entre 0,05 e 0,01, poderão ser utilizados a correção para os números de graus de liberdade ou os testes multivariados; e,
- iii. se a condição de H-F para a matriz de variância-covariância for rejeitada, em nível de significância menor do que 0,01, somente testes multivariados deverão ser utilizados.

As estimativas das covariâncias e das correlações entre os níveis de tempo (Quadro 1), obtidas a partir da matriz de somas de quadrados e produtos dos erros da variável altura, mostram correlações positivas e negativas muito diferentes, e variâncias (diagonal principal) e covariância totalmente distintas. A simples observação da matriz pode indicar a não satisfação da suposição de esfericidade para altura, o que foi confirmada pelo teste de Mauchly.

Dias	10	30	50	70	90	110	130	150
10	0,5119	0,5337	-0,5846	1,4244	-0,2786	1,6219	4,0111	0,3769
30	0,2049	9,5860	-9,3100	-6,1600	-6,9398	-6,7632	0,0276	-0,0741
50	-0,1544	-0,5686	27,9990	0,4991	-0,5784	-0,9386	0,9771	0,3306
70	0,1510	-0,6006	0,5562	173,8460	15,8940	-33,7332	41,3946	41,0643
90	-0,0497	-0,2855	-0,0139	0,1536	61,6200	74,9863	1,2795	24,4480
110	0,1667	-0,1606	-0,0130	-0,1882	0,7025	184,9020	47,0666	63,6499
130	0,5932	0,0732	0,0195	0,3322	0,1615	0,3663	89,3150	62,5055
150	0,0553	-0,0172	0,0066	0,3271	0,3271	0,4916	0,6946	90,6740

Quadro 1 – Covariância em negrito, variâncias em destaque, e correlações residuais para variável altura de plantas de mamona em oito avaliações (Dias) no ciclo. Santa Maria, RS, 2010.

Considerando a hipótese de corrigir o número de graus de liberdade dos fatores dentro de sujeitos (ou intra-indivíduos), ou seja, do fator Dias e da interação Densidades x Dias, o resultado da análise univariada encontra-se na Tabela 6.

Tabela 6 – Análise univariada com correção dos números de graus de liberdade para os efeitos entre e intra indivíduos, da variável altura de planta de mamona, submetida a três densidades de cultivo e oito medidas no tempo igualmente espaçadas. Santa Maria, RS, 2010.

Causas de variação	TESTE PARA OS FATORES ENTRE INDIVÍDUOS						
	GL	SQ	QM	F	Pr>F		
Densidade	2	7066,383	3533,191	21,020	0,0004		
Erro	9	1513,021	168,113				
Causas de variação	TESTE PARA OS FATORES INTRA INDIVÍDUOS						
	GL	SQ	QM	F	Pr>F	G-G	H-F
Dias	7	1087831,503	155404,5	2312,86	<,0001	<,0001	<,0001
Densidades x Dias	14	6272,791	448,057	6,67	<,0001	0,0002	<,0001
Erro(dias)	63	4233,069	67,192				

G-G = Greenhouse-Geisser $\hat{\epsilon} = 0,4400$

H-F= Huynh-Feldt $\tilde{\epsilon} = \frac{0,4400}{0,56} = 0,8436$

A hipótese de perfis coincidentes, que é verificada através do teste para fator entre indivíduos, ou seja, entre densidades, foi rejeitada ($p < 0,0004$) indicando que as densidades possuem um efeito diferenciado sobre a variável altura de planta Tabela 6.

Para os fatores intra indivíduos, relacionados ao efeito de dias de avaliação, testou-se a hipótese de perfis horizontais, que foi rejeitada ($p < 0,0001$). Isto denota que os perfis médios de resposta correspondentes as densidades são influenciados pelas medidas repetidas no tempo. A hipótese de perfis paralelos, testada através da interação Densidades x Dias, não foi aceita ($p < 0,0001$), sugerindo que a combinação de diferentes densidades com os diferentes dias de avaliação interage de forma a interferir na altura de planta de mamona.

Como a condição de esfericidade foi rejeitada, para a tomada de decisão em relação às hipóteses testadas, os níveis mínimos de significância em negrito, das estatísticas, Greenhouse-Geisser ($\hat{\epsilon} = 0,4400$) e Huynh-Feldt ($\tilde{\epsilon} = \frac{0,4400}{0,56} = 0,8436$), foram consideradas.

Singer; Andrade (1984) sugerem, a respeito do uso das correções para os graus de liberdade, usar $\hat{\epsilon}$, sempre que não se tem idéia do verdadeiro valor de ϵ ou que $\epsilon < 0,75$; e, usar $\tilde{\epsilon}$ como estimador para ϵ , somente quando $\epsilon \geq 0,75$. No entanto, Malheiros (2001)

avaliou a precisão dos testes F, quando as correções dos graus de liberdade propostas por Greenhouse-Geisser e Huynh-Feldt são realizadas. Concluiu que, tanto para Dias como para a interação Densidades x Dias, com a correção de Huynh-Feldt, os testes F da análise de variância são imprecisos e com a correção de Greenhouse-Geisser são precisos. Esses resultados concordam com Littell et al. (1998), que sugerem a correção de Greenhouse-Geisser.

Além da forte rejeição ($p < 0,00198$) da estrutura de esfericidade da matriz de Σ , verificou-se que o modelo univariado em parcelas subdividas no tempo não é o mais indicado para a análise dos dados, mesmo com o uso das correções para os graus de liberdade, que apresentaram restrições. Desta forma, a metodologia de análise adequada é a análise multivariada de perfil.

4.1.1.3 Análise Multivariada de Perfil

Os resultados da análise multivariada de perfil utilizando o proc GLM com o comando *manova/printe*, é adicionado obedecendo a sequência de comandos:

```
proc glm data=multivariado;
```

```
class dens;
```

```
model A10 A30 A50 A70 A90 A110 A130 A150= trat/nouni; manova/printe; repeated grau 8
```

```
polynomial/summary printm printe;
```

```
run; quit;
```

Com estes comandos, são fornecidas as matrizes *E* (usando a opção *printe*), de soma de quadrados e produtos cruzados dos resíduos (SSPC) e *R*, das correlações parciais dos resíduos, com os respectivos níveis descritivos associados aos testes das correlações parciais. Estas matrizes podem ser verificadas a seguir:

$$E = \begin{bmatrix} & \text{Dia10} & \text{Dia30} & \text{Dia50} & \text{Dia70} & \text{Dia90} & \text{Dia110} & \text{Dia130} & \text{Dia150} \\ \text{Dia10} & 4,61 & 4,80 & -5,26 & 1,28 & -2,51 & 14,60 & 36,10 & 3,39 \\ \text{Dia30} & 4,80 & 86,27 & -83,83 & -220,68 & -62,46 & -60,87 & 19,27 & -45,60 \\ \text{Dia50} & -5,26 & -83,83 & 251,99 & 349,21 & -5,21 & -84,48 & 8,79 & 2,98 \\ \text{Dia70} & 1,28 & -220,68 & 349,21 & 1564,61 & 143,05 & -303,59 & 372,51 & 369,58 \\ \text{Dia90} & -2,51 & -62,46 & -5,21 & 143,05 & 554,58 & 674,88 & 107,85 & 386,23 \\ \text{Dia110} & 14,60 & -60,87 & -48,48 & -303,59 & 674,88 & 1664,12 & 423,60 & 572,85 \\ \text{Dia130} & 36,10 & 19,27 & 8,79 & 372,51 & 107,85 & 423,60 & 803,84 & 562,55 \\ \text{Dia150} & 3,39 & -45,60 & 2,98 & 369,58 & 386,23 & 572,85 & 562,55 & 816,07 \end{bmatrix}$$

Existe um fator limitante, relacionado à álgebra da matriz E. Isto ocorre quando o número de medidas repetidas no tempo for maior que o número de unidades experimentais. Assim, a matriz E torna-se inversível, ou seja, singular, e nestas situações as estatísticas multivariadas não poderão ser calculadas.

$$R = \begin{bmatrix} & \text{Dia10} & \text{Dia30} & \text{Dia50} & \text{Dia70} & \text{Dia90} & \text{Dia110} & \text{Dia130} & \text{Dia150} \\ \text{Dia10} & 1 & 0,2409 & -0,1544 & 0,0151 & -0,0496 & 0,1667 & 0,5932 & 0,0553 \\ & & \mathbf{0,5025} & \mathbf{0,6702} & \mathbf{0,9670} & \mathbf{0,8918} & \mathbf{0,6453} & \mathbf{0,0707} & \mathbf{0,8793} \\ \text{Dia30} & 0,2409 & 1 & -0,5686 & -0,6006 & -0,2855 & -0,1606 & 0,0732 & -0,1719 \\ & & \mathbf{0,5025} & \mathbf{0,0863} & \mathbf{0,0663} & \mathbf{0,4239} & \mathbf{0,6575} & \mathbf{0,8408} & \mathbf{0,6350} \\ \text{Dia50} & -0,1544 & -0,5686 & 1 & 0,5561 & -0,0139 & -0,1305 & 0,0195 & 0,0066 \\ & & \mathbf{0,6702} & \mathbf{0,0863} & & \mathbf{0,0950} & \mathbf{0,9695} & \mathbf{0,7194} & \mathbf{0,9573} & \mathbf{0,9856} \\ \text{Dia70} & 0,0151 & -0,6006 & 0,5561 & 1 & 0,1536 & -0,1881 & 0,3322 & 0,3271 \\ & & \mathbf{0,9670} & \mathbf{0,0663} & \mathbf{0,0950} & & \mathbf{0,6719} & \mathbf{0,6027} & \mathbf{0,3484} & \mathbf{0,3563} \\ \text{Dia90} & -0,0496 & -0,2855 & -0,0139 & 0,1536 & 1 & 0,7025 & 0,1615 & 0,5741 \\ & & \mathbf{0,8918} & \mathbf{0,4239} & \mathbf{0,9695} & \mathbf{0,6719} & & \mathbf{0,0235} & \mathbf{0,6557} & \mathbf{0,0826} \\ \text{Dia110} & 0,1667 & -0,1606 & -0,1305 & -0,1881 & 0,7025 & 1 & 0,3663 & 0,4916 \\ & & \mathbf{0,6453} & \mathbf{0,6575} & \mathbf{0,7194} & \mathbf{0,6027} & \mathbf{0,0235} & & \mathbf{0,2979} & \mathbf{0,1490} \\ \text{Dia130} & 0,5932 & 0,0732 & 0,0195 & 0,3322 & 0,1615 & 0,3663 & 1 & 0,6946 \\ & & \mathbf{0,0707} & \mathbf{0,8408} & \mathbf{0,9573} & \mathbf{0,3484} & \mathbf{0,6557} & \mathbf{0,2979} & & \mathbf{0,0258} \\ \text{Dia150} & 0,0553 & -0,1719 & 0,0066 & 0,3271 & 0,5741 & 0,4916 & 0,6946 & 1 \\ & & \mathbf{0,8793} & \mathbf{0,6350} & \mathbf{0,9856} & \mathbf{0,3563} & \mathbf{0,0826} & \mathbf{0,1490} & \mathbf{0,0258} \end{bmatrix}$$

A matriz R mostra as correlações entre as observações tomadas nos oito Dias de medição, do dia 10 ao dia 150 para a variável altura de planta. A simples observação da matriz R contraria as hipóteses apresentadas por Littell et al. (1998) e Xavier; Santos (2001),

de que medidas em tempos mais próximos são mais correlacionadas do que as medidas verificadas em tempos mais distantes.

Por meio dos testes multivariados testou-se a hipótese nula de perfis horizontais e perfis paralelos. Na tabela 7, em que se encontram os resultados da análise multivariada de perfil, mostram que na primeira parte foram usadas as matrizes H da soma de quadrados e produtos cruzados para o fator Dias de avaliação, e E da soma de quadrados e produtos cruzados para os resíduos. Além disso, utilizaram-se os parâmetros S , m_1 e m_2 , que são empregados nos cálculos que envolvem a determinação dos valores de F calculado e tabelado, sendo que, de acordo com Morrisson (1981), estes parâmetros proporcionam a distribuição exata da estatística do teste sob a hipótese nula. Segundo Xavier (2000), os testes multivariados não requerem a condição H-F da matriz de Σ , pois são baseados em uma matriz de Σ sem estrutura e as estimativas dessas matrizes podem ser obtidas pelo método do momento.

Tabela 7 – Resultado da análise multivariada da variável altura de planta de mamona, submetida a três densidades de cultivo e oito medidas no tempo igualmente espaçadas. Santa Maria, RS, 2010.

H=Matriz de soma de quadrados e produtos cruzados para o fator Dias de Avaliação					
E=Matriz de soma de quadrados e produtos cruzados para os resíduos					
$S=1 \quad m_1=2,5 \quad m_2=0,5$					
Estatística	Valor	F	Num.GL ¹	Den.GL ²	Pr>F
Wilks' Lambda	0,0003	1475,83	7	3	<,0001
Pillai's Trace	0,9997	1475,83	7	3	<,0001
Hotelling-Lawley Trace	3443,6118	1475,83	7	3	<,0001
Roy's Greatest Root	3443,6118	1475,83	7	3	<,0001
H=Matriz de soma de quadrados e produtos cruzados para o fator Densidades x Dias					
E=Matriz de soma de quadrados e produtos cruzados para os resíduos					
$S=2 \quad m_1=2 \quad m_2=0,5$					
Estatística	Valor	F	Num.GL ¹	Den.GL ²	Pr>F
Wilks' Lambda	0,0111	3,64	14	6	0,0603
Pillai's Trace	1,4327	1,44	14	8	0,3073
Hotelling-Lawley Trace	49,1433	12,64	14	2,78	0,0366
Roy's Greatest Root	48,3150	27,61	7	4	0,0031

1: Grau de liberdade do numerador; 2: Grau de liberdade do denominador.

A hipótese de perfis horizontais, testada pelos dias de avaliação, foi rejeitada (<,0001) para os quatro critérios Wilks' Lambda, Pillai's Trace, Hotelling-Lawley Trace e Roy's Greatest Root. Verifica-se também que para a interação Densidades x Dias, que testa a hipótese de paralelismo, foi rejeitada pelos critérios de Hotelling-Lawley Trace e Roy's Greatest Root. Porém, nem sempre os resultados destes testes são equivalentes Reis (1997) e Xavier (2000) justificam que os testes multivariados podem produzir diferentes níveis descritivos. Em geral, a ordem de preferência em termos de poder é: Traço de Pillai, Lambda de Wilks, Traço de Lawley-Hotelling e Roy's Greatest Root. Entretanto, Lambda de Wilks é o teste mais comumente usado.

Após, identificou-se os componentes fixos do modelo, sendo estes os dias, densidades e suas interações. Agora, o interesse é o estudo de uma representação funcional simples para descrever o comportamento médio da variável altura de plantas ao longo do tempo. Esta

função é obtida através da verificação da tendência de variação da resposta média da altura de plantas. Sendo assim, será determinado o grau do polinômio que melhor descreva esta variação, ou seja, o grau da curva média ajustada às densidades. Os polinômios são obtidos utilizando a opção *polynomial/ summary printm*.

As opções: *polynomial* gera contrastes de polinômios ortogonais para os níveis do fator tempo; *summary* apresenta a tabela da análise de variância dos contrastes; *printm* imprime a matriz dos contrastes (denominada *m*). Esta matriz é utilizada para analisar as tendências da função a ser ajustada.

De acordo com Littel et al. (1991), quando é usado o *proc GLM* para análise de medidas repetidas é necessário que seja indicada uma transformação (tipo de contraste), já que um conjunto de contrastes pode ser usado para analisar tendências sobre o fator de medidas repetidas, e realizar comparações entre os níveis desse mesmo fator. Conforme Xavier (2000); Freitas (2007), os dados originais de cada indivíduo são transformados em um novo conjunto de dados, obtidos através de um conjunto de $t-1$ contrastes, onde t é o número de medidas repetidas. Tais transformações (contrastos) são utilizadas na tentativa de amenizar a influência de algumas estruturas de variância-covariância, na análise univariada de medidas repetidas, que podem invalidar os resultados dos testes.

Entretanto, antes de gerar os contrastes polinomiais é importante grafar um diagrama de dispersão do perfil médio da variável altura de planta, para as três densidades, em função dos dias de avaliação. Com este diagrama de dispersão é possível observar se as densidades possuem o mesmo comportamento em todos os tempos; além disso, tem-se uma idéia prévia do grau do polinômio que irá representar a curva média. Na Figura 5 encontra-se o diagrama de dispersão dos perfis médios da variável.

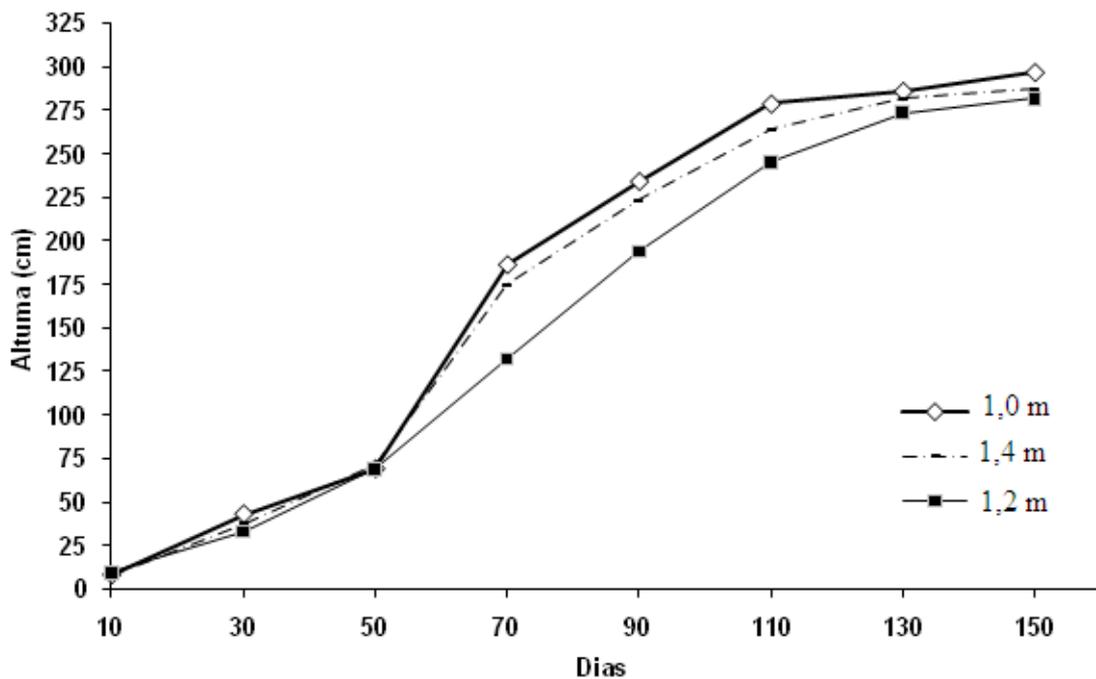


Figura 5 – Perfis médios de resposta da variável altura de planta de mamona, submetida a três densidades de cultivo e oito medidas no tempo no ciclo. Santa Maria, RS, 2010.

É possível observar que todos os perfis médios apresentam um comportamento aparentemente quadrático. No entanto, o que ocorre é que, dos 10 aos 50 dias, a variável altura de planta apresenta um crescimento linear e não diferenciando entre as densidades. Este comportamento foi, previamente, evidenciado nos diagramas paralelos de dispersão dos perfis individuais (Figura 4). Assim, realizou-se o teste *t-Student*, sendo testada a hipótese de igualdade de intercepto e inclinação das retas (NETER; WASSERMAN, 1974). O resultado do teste encontra-se na Tabela 8.

Tabela 8 – Teste *t-Student* para verificar a igualdade de intercepto e inclinação das retas de regressão para as três densidades de cultivo da variável altura de planta no período de 10 a 50 dias. Santa Maria, RS, 2010.

Densidades (m)	Intercepto		Inclinação	
	Estatística <i>t-Student</i>	Pr>t	Estatística <i>t-Student</i>	Pr>t
1,2 – 1,4	0,2020	0,8586	0,6317	0,5922
1,2 – 1,0	0,9489	0,4428	1,3368	0,3131
1,4 – 1,0	0,2702	0,8124	0,9993	0,4229

Nota-se que, para todas as densidades, foi aceita a hipótese de igualdade de intercepto e igualdade de inclinação, confirmando o diagrama de dispersão dos perfis médios na Figura 5. Assim, para confirmar a tendência de crescimento quadrático, a partir dos 50 dias de avaliação, foi realizado o teste de tendência, utilizando os contrastes polinomiais. Para a utilização da técnica, a programação (descrita em análise univariada de perfil) sofreu uma pequena alteração, o comando model ficará: “*model A50 A70 A90 A110 A130 A150= trat/nouni*” e o comando repeated mudará para: “*repeated grau 6 polynomial/summary printm printe*”. O resultado do teste encontra-se na Tabela 9.

Tabela 9 – Teste de hipótese de tendência de variação da altura de plantas média de mamona entre 50 e 150 dias de avaliação para as três densidades de cultivo. Santa Maria, RS, 2010.

Grau 1					
Causas de Variação	G.L	S.Q	Q.M.	F	Pr>F
Média	1	379389,8539	379389,8539	3662,10	<,0001
Densidade	2	294,5286	147,2643	1,42	0,2907
Erro	9	932,3910	103,5990		
Grau 2					
Causas de Variação	G.L	S.Q	Q.M.	F	Pr>F
Média	1	46757,7330	46757,7330	834,70	<,0001
Densidade	2	2040,3835	1020,1917	18,21	0,0007
Erro	9	504,1550	56,0172		
Grau 3					
Causas de Variação	G.L	S.Q	Q.M.	F	Pr>F
Média	1	632,0394	632,0394	13,34	0,0020
Densidade	2	1889,6611	944,8305	0,44	0,6563
Erro	9	637,3638	70,8182		
Grau 4					
Causas de Variação	G.L	S.Q	Q.M.	F	Pr>F
Média	1	108,3244	108,3244	0,79	0,3979
Densidade	2	172,0654	86,0327	0,63	0,5567
Erro	9	1237,6781	137,5197		
Grau 5					
Causas de Variação	G.L	S.Q	Q.M.	F	Pr>F
Média	1	437,4107	437,4107	18,97	0,0018
Densidade	2	191,4624	95,7312	4,15	0,0528
Erro	9	207,4882	23,0542		

Para o grau do polinômio, dois testes são apresentados: um para a média e um para as densidades. No teste para a média, verifica-se a significância do contraste para a média de todas as densidades (perfil médio). No teste para densidades, testa se a média do contraste é a

mesma para cada nível das densidades, ou seja, se existe interação entre as densidades e os dias de avaliação.

Verifica-se que a variação da altura de plantas, para as três densidades de cultivo, pode ser explicada por um polinômio de grau 2, já que foi o único grau que apresentou significância tanto para densidade quanto para a interação entre densidade e dias. Outra confirmação dos resultados sobre tendência de variação pode ser obtida com os valores das médias das alturas de plantas ajustadas ao longo do tempo (Tabela 10).

Tabela 10 – Estimativas dos parâmetros das curvas ajustadas às médias da altura de planta em função do tempo, com seus respectivos erros padrões, e nível de significância de cada parâmetro, entre parênteses, para cada densidade no período de 50 a 150 dias. Santa Maria, RS, 2010.

Densidade (m)	Intercepto	Linear	Quadrático	Erro padrão modelo	R ²
1,2	-172,66 (0,040)	4,364 (0,001)	-0,0194 (0,004)	5,92	0,9970
erro padrão	22,68	0,490	0,002		
1,4	-219,149 (0,006)	6,306 (0,001)	-0,037 (0,004)	8,59	0,9935
erro padrão	32,92	0,771	0,003		
1,0	-261,0,18 (0,010)	7,229 (0,004)	-0,04104 (0,010)	13,29	0,9859
erro padrão	50,91	1,09	0,005		

Evidencia-se um bom ajuste para o modelo quadrático com todos os parâmetros significativos e coeficientes de determinação superiores a 98%. Com estes resultados é possível confirmar a identificação do modelo polinomial do segundo grau para a curva média dos efeitos fixos para a densidade no período de 50 a 150 dias.

A sequência de comandos para obter as curvas ajustadas com as médias da altura em função dos dias de avaliação (50 a 150 dias), de acordo com as densidades, é dado por:

```
proc sort; by dens dias; proc means; by dens dias; var A;
output out=polim mean=A;
proc print data=polim;
data polim1; set polim;
if dias<=30 then delete;
```

```

dias2=dias*dias;
proc glm;
model A= dias dias2; by dens;
run;
quit;

```

Com os comandos: *proc sort; by dens dias; proc means; by dens dias; var A;* é informado que a variável A (altura de planta) deve ser ordenada por densidades e dias e devem ser calculadas as médias das densidades para a altura de acordo com os dias de coletas. Estes comandos foram utilizados para estruturar as estatísticas da Tabela 4. Após, retirou-se os dados referentes às medições dos 10 dias e 30 dias, para dar sequência às análises no período de 50 a 150 dias.

Com as instruções: *output out=polim mean=A;* é solicitado ao sistema para arquivar na memória (arquivo chamado *polim*) em que contém as médias de A previamente calculadas com o comando anterior. Com a informação *proc print data=polim*, o sistema imprime o arquivo *polim*; este comando é opcional, e serve para conferência dos dados. Já com os comandos: *data polim1; set polim;* cria-se um arquivo chamado *polim1* a partir do arquivo *polim*. A instrução: *if dias<=30 then delete* informa que deverão ser ignoradas, para os próximos procedimentos, os dados referentes aos dias de avaliação 10 e 30. Com o comando: *dias2=dias*dias* cria-se o componente quadrático dos modelos. E, finalmente, com os comandos *model A= dias dias2; by dens;* são estimados os modelos de acordo com as densidades.

4.1.2 Identificação dos efeitos aleatórios

A identificação dos efeitos aleatórios foi realizada através de análise de regressão polinomial para cada perfil individual de resposta; estas análises encontram-se na Tabela 11. Em negrito, estão apresentados os coeficientes submetidos ao teste *t-Student*, que foram significativamente diferentes de zero. Os comandos para realização das análises são:

```

data polue; set univariado;
if dias<=30 then delete;
dias2=dias*dias;

```

```
proc sort data=polue;
by UE;
proc glm data=polue;
model A= dias dias2/ss3; by ue;
```

Para criar os polinômios para cada unidade experimental (UE), cria-se o arquivo *data polue*, que irá utilizar os dados do arquivo univariado; além disso, o arquivo *polue* deverá ser ordenado por UE (*proc sort data=polue; by UE;*).

Tabela 11 – Estimativas individuais dos coeficientes das regressões polinomiais, com o respectivo nível mínimo de significância para a altura de plantas de mamona entre 50 e 150 dias de avaliação para as 12 unidades experimentais. Santa Maria, RS, 2010.

Unidade Experimental	Coeficientes e nível de significância					
	Intercepto	P>t	Linear	P>t	Quadrático	P>t
1	-232,152	0,027	7,656	0,0085	-0,027	0,0208
2	-278,391	0,0260	8,673	0,0096	-0,033	0,0203
3	-272,954	0,0087	8,676	0,0029	-0,034	0,0059
4	-260,576	0,0103	8,313	0,0034	-0,031	0,0078
5	-118,365	0,1317	4,649	0,0333	-0,013	0,1213
6	-216,583	0,0189	6,733	0,0068	-0,022	0,0208
7	-203,298	0,0156	6,346	0,0055	-0,021	0,0182
8	-192,399	0,0426	5,727	0,0184	-0,017	0,0649
9	-245,727	0,0045	7,970	0,0014	-0,029	0,0033
10	-192,844	0,0061	6,765	0,0015	-0,024	0,0038
11	-250,860	0,0053	7,853	0,0018	-0,028	0,0046
12	-227,084	0,0122	7,439	0,0037	-0,027	0,0088

Verifica-se que o intercepto, os componentes lineares e quadráticos podem ser considerados como efeitos aleatórios. Para verificar a adequação dos modelos, na Figura 6 é apresentado o gráfico de probabilidade normal para os resíduos gerados para cada modelo.

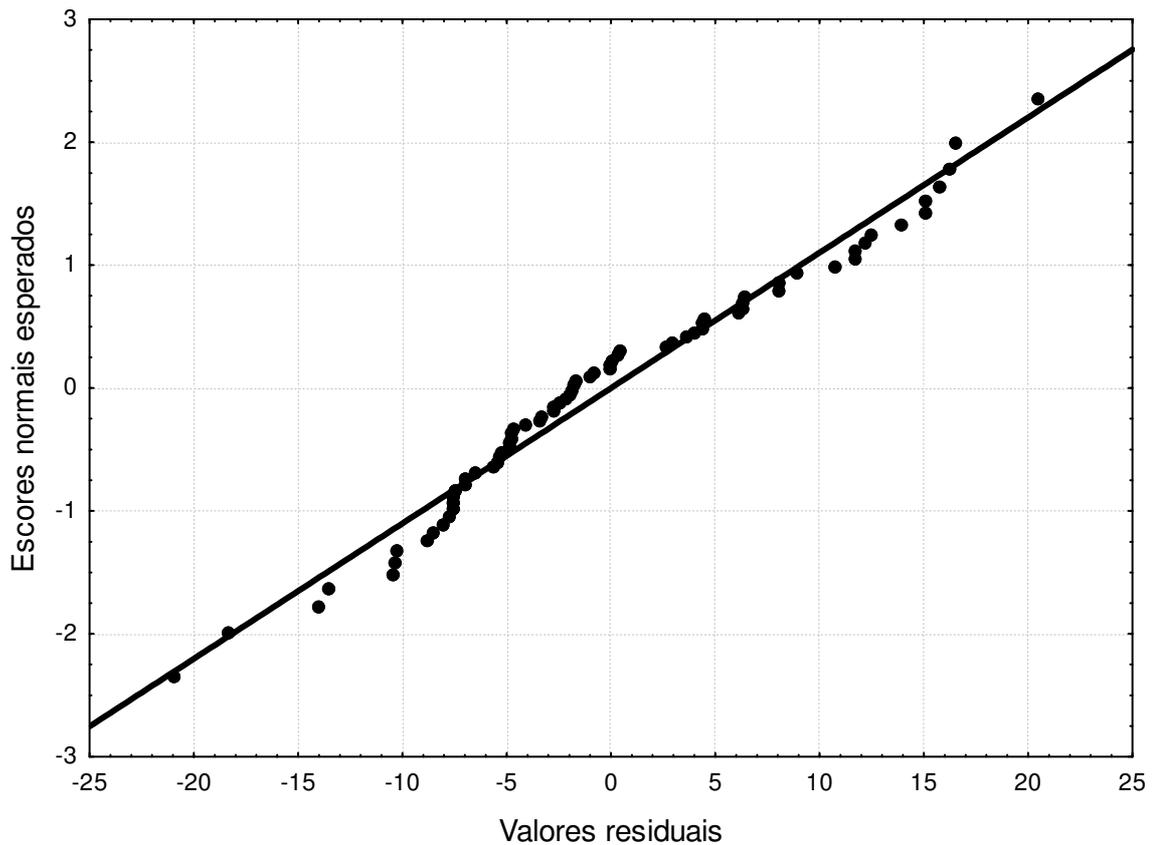


Figura 6 – Gráfico de probabilidade normal dos resíduos para os modelos estimados para a altura de plantas de mamona entre 50 e 150 dias de avaliação para as 12 unidades experimentais. Santa Maria, RS, 2010.

Outra confirmação da normalidade foi obtida pelo teste de Shapiro-Wilks, cujo valor da estatística foi igual 0,9742 ($p=0,145165$), confirmando o bom ajuste para os modelos. Desta forma, conclui-se que o modelo normal é aceitável com a inclusão dos três coeficientes como aleatórios.

4.1.3 Seleção de modelos e testes

Após identificado um modelo polinomial do segundo grau para os efeitos fixos e um modelo para os efeitos aleatórios, que inclui o intercepto e os coeficientes lineares e quadráticos, o próximo passo para estruturação da curva de crescimento para a variável altura de planta de mamona é a busca de uma estrutura parcimoniosa para a matriz de covariância. Preferencialmente, a estrutura deverá ter um número de parâmetros a estimar menor do que o

número da estrutura de covariância não estruturada (UN), apresentada em 2.5.3.3. Em se tratando disso, Littell et al. (2006), sugerem o ajuste de um modelo inicial saturado (maximal), tanto para os efeitos fixos quanto para a estrutura de covariância. Sugerem, ainda, o ajuste de estrutura de variâncias mais parcimoniosas, tais como: a uniforme, uniforme heterogênea e auto-regressiva.

Conforme Marcelino; Iemma (2000), no *proc MIXED* encontram-se disponíveis cerca de quarenta estruturas de variância e variância-covariância. Além disso, se não for especificado nenhuma estrutura de covariância para R, o SAS usa por *default* a estrutura $\sigma^2 I$, conhecida como componente de variância (VC). Com esta estrutura, assume-se que os erros intra-indivíduos são independentes e homocedásticos (PERRI; IEMMA, 1996).

As estruturas de variância e variância-covariância testadas para modelar a matriz de efeitos aleatórios, relacionada a G e R, (definidas em 2.5.3.3) foram: UN (sem estrutura), VC (componente de variância), AR(1) (auto-regressiva de primeira ordem), CSH (simetria composta heterogênea), ARH (1) (auto regressiva de primeira ordem heterogênea) e UN(1). De acordo com as considerações de autores, tais como: Vonesch; Chinchilli (1997), Malheiros (2001) e Johnson (2010), a maioria dos problemas práticos é resolvida utilizando estas estruturas.

Para iniciar a modelagem com a matriz G, Wald (2000) aconselha o uso de uma estrutura geral ou qualquer uma das várias estruturas de variância-covariância comumente utilizadas para R. Sendo assim, inicialmente ajustou-se, por REML, um modelo com todos os efeitos fixos e aleatórios sugeridos em 4.1.1 e 4.1.2 (modelo maximal), admitindo-se as estrutura VC para R e variando as estruturas de G.

Após testado o modelo maximal (efeitos fixos =dias, densidades e interações; efeitos aleatórios= intercepto, dias, dias², deu-se início a retirada de cada efeito aleatório do modelo. A estratégia proposta é pela retirada de um efeito por vez e gerando novamente o modelo reduzido. A ordem de retirada se deu primeiro pelo intercepto, após, dias, subsequentemente, dias² e sem efeito aleatório. Os melhores resultados para cada estrutura G foram obtidos considerando dias e dias² como aleatório. Os modelos foram selecionados utilizando os critérios de AIC, BIC e 2 vezes o log, sendo o critério mais importante o AIC, conforme comentado no item 2.5.3.4 (Seleção de modelos e testes). A estratégia utilizada para a seleção

dos efeitos e as diferentes estruturas testadas em G pode ser verificada na Tabela 12. A programação para o ajuste das estruturas G e R encontra-se nos Apêndices B e C.

Tabela 12 – Estruturas de variância-covariância testadas para a matriz G (variação entre indivíduos) e seleção de efeitos aleatórios para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.

Grupos	Modelos	Estruturas		Critérios			Número de parâmetros
		R	G	-2Res Log	AIC	BIC	
A	3	VC	UN	580,7	584,7	585,7	2
	4	VC	UN	583,2	585,2	585,7	1
B	1	VC	VC	580,7	584,7	585,7	4
	2	VC	VC	580,9	582,7	585,7	3
	3	VC	VC	582,2	584,7	584,9	2
	4	VC	VC	583,2	585,2	585,7	1
C	3	VC	AR(1)	580,7	585,7	586,2	3
	4	VC	AR(1)	583,2	586,2	588,7	1
D	2	VC	CSH	580,7	583,7	585,2	4
	3	VC	CSH	580,2	586,7	585,7	3
	4	VC	CSH	583,2	587,2	588,7	1
E	2	VC	ARH(1)	580,7	584,7	585,7	4
	3	VC	ARH(1)	580,6	585,7	584,7	3
	4	VC	ARH(1)	583,2	586,2	585,7	1
F	2	VC	UN(1)	571,7	564,7	575,2	4
	3	VC	UN(1)	580,0	583,2	585,7	2
	4	VC	UN(1)	683,2	585,2	585,7	1

Conforme comentado anteriormente, foram selecionando dias e dias² como efeitos aleatórios, para cada uma das estruturas testada em G, dentro de cada um dos grupos, ou seja, o modelo de número 2, da maioria dos grupos, foi o que apresentou valores reduzidos de AIC e BIC e - 2 vezes o log verossimilhança, além do número reduzido de parâmetros de covariância.

Nota-se ainda que, para os modelos de alguns grupos como A, modelos 1 e 2, utilizando a estrutura UN, relacionados ao modelo maximal e dias e dias² como efeito aleatórios, não houve convergência. O mesmo comportamento ocorreu com o grupo C, modelos 1 e 2, que utiliza a estrutura (AR1). Já nos grupos D, E e F não houve convergência para o modelo maximal com as estruturas CSH, ARH(1) e UN(1), respectivamente. Portanto, as melhores estruturas em ordem de prioridade para G, devido aos menores valores de AIC, foram: UN(1), VC, CSH, UN, ARH(1), AR(1).

Selecionado o modelo que ainda mantém todos os efeitos fixos e retirando apenas o intercepto dos efeitos aleatório, o próximo passo é escolher a melhor estrutura de covariância para R. Para tanto, foi utilizado o método de estimação por máxima verossimilhança restrita (REML) e a estratégia para os testes encontra-se na Tabela 13. A configuração completa das Tabelas 12 e 13 encontram-se nos Apêndices E e F.

Tabela 13 – Estruturas de variância-covariância testadas para a R (variação dentro de indivíduos) para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.

Grupos	Modelos	Estruturas		Critérios			Parâmetros de Covariância
		R	G	-2Res Log	AIC	BIC	
A							
B	1	VC	VC	580,7	584,7	588,7	3
	3	AR(1)	VC	580,7	584,7	585,1	4
C							
D	1	VC	CSH	580,7	584,7	585,7	4
	3	AR(1)	CSH	580,7	588,7	590,6	5
E	1	VC	ARH(1)	580,7	584,7	585,7	4
	3	AR(1)	ARH(1)	580,7	588,7	590,6	5
F	1	VC	UM(1)	580,7	582,7	584,7	4
	3	AR(1)	UM(1)	580,7	586,7	588,1	5

No grupo A e C não houve convergência de nenhum modelo para a estrutura UN e AR (1) em G, respectivamente, quando R foi submetido as seis estruturas testadas (VC, UM, AR(1), CSH, ARH(1) e UN(1)). Nos outros grupos alguns modelos também não convergiram.

Sendo assim, a estrutura escolhida para R é a VC (componente de variância), apresentando os menores valores de AIC e BIC. Esta estrutura se caracteriza por variâncias iguais e observações independentes. Portanto, do grupo F, modelo 1, foi selecionado para modelar as estruturas de R e G, sendo que a estrutura UN(1), relacionada a G, obteve os melhores valores nos critérios de seleção (Tabela 12).

Selecionado os efeitos aleatórios das estruturas para G e R, a próxima etapa é a verificação dos testes para efeitos fixos (Tabela 14) e, caso algum efeito não seja significativo, deverá ser retirado do modelo. No Apêndice D encontra-se a programação para testar estes efeitos.

Tabela 14 – Testes para efeitos fixos do modelo completo utilizando as estruturas de matriz de covariância (UN(1) para G e VC para R para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.

Causas de variação	GL. do numerador	GL. do denominador	F	Pr>F
Densidade	2	54	3,40	0,0408
Dias	1	54,5	734,14	<0,0001
Densidades x Dias	2	54,5	7,37	0,0015
Densidades x Dias ²	3	54	130,39	<0,0001

Verifica-se que todos os efeitos fixos são significativos; portanto, realizou-se o teste de Tukey-Kramer, para comparação de efeitos densidades. O teste encontra-se na tabela 15.

Tabela 15 – Teste de Tukey – Kramer para três densidades de semeadura, com o modelo completo, utilizando as estruturas de matriz de covariância UN(1) para G e VC para R. Santa Maria, RS, 2010.

Médias (cm)	Densidades (m)	Diferenças	Erro Padrão	GL	Valor de T	Pr>T
(1,2m) = 199,17	1,2 – 1,4	-18,0017	4,5350	45	-3,97	<0,0007
(1,4m)= 217,17	1,2 – 1,0	26,0683	4,5350	45	5,75	<0,0001
(1,0m) = 225,24	1,4 – 1,0	8,0667	4,5350	45	8,0667	0,1882

Pelo teste de Tukey – Kramer, as comparações entre as densidades de cultivo de 1,4 e 1,0 m não são significativas, sugerindo que essas apresentam o mesmo efeito sobre o crescimento da planta. Com a densidade 1,2 m, as plantas apresentaram o menor crescimento (199,17 cm), diferindo significativamente das densidades 1,4 e 1,0 m. Os parâmetros das curvas, para cada densidade no modelo completo (Tabela 14), considerando dias e dias² como efeito aleatório e as estruturas de matriz de covariância UN(1) para G e VC para R, podem ser visualizados na Tabela 16.

Tabela 16 – Estimativas dos parâmetros dos modelos das curvas crescimento ajustadas à altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.

Efeitos	Densidade de cultivo (m)	Estimativas	Erro padrão	GL	Valor t	Pr > t
Intercepto		-229,13	21,3807	45	-10,72	<,0001
Dens1.0	1	-31,8894	30,2368	45	-1,05	0,2972
Dens1.2	1,2	46,4677	30,2368	45	1,54	0,1313
Dens1.4	1,4	0,0000
dias		7,5067	0,4624	9	16,23	<,0001
dias* Dens1.0	1	0,8226	0,6539	9	1,26	0,2401
dias* Dens1.2	1,2	-1,6427	0,6539	9	-2,51	0,0332
dias* Dens1.4	1,4	0,0000
dias*dias* Dens1.0	1	-0,03105	0,002285	9	-13,59	<,0001
dias*dias* Dens1.2	1,2	-0,01832	0,002285	9	-8,02	<,0001
dias*dias* Dens1.4	1,4	-0,02726	0,002285	9	-11,93	<,0001

Os modelos aplicados aos dados podem ser visualizados na Figura 7, e foram os seguintes:

$$\text{Densidade 1,2} \quad \hat{y}_1 = -182,36 + 5,864 * dias - 0,018232 * dias^2$$

$$\text{Densidade 1,4} \quad \hat{y}_2 = -229,13 + 7,567 * dias - 0,02726 * dias^2$$

$$\text{Densidade 1,0} \quad \hat{y}_3 = -228,3074 + 8,3293 * dias - 0,03105 * dias^2$$

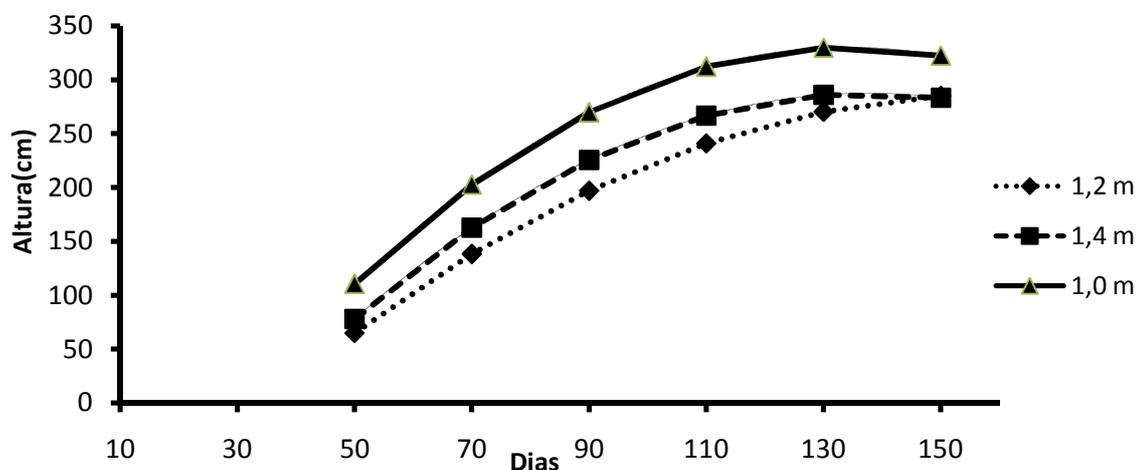


Figura 7 – Curvas de crescimento ajustadas à altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades de cultivo. Santa Maria, RS, 2010.

Visualizou-se, por meio das curvas, os resultados obtidos pelos testes de Tukey-Kramer, sendo que a densidade de cultivo de 1,0 m proporcionou maior altura de plantas, entretanto, não difere significativamente da densidade 1,4 m. Já a densidade de 1,2 m destaca-se por proporcionar menor crescimento em altura diferindo significativamente das densidades 1,0 e 1,4 m. Sendo assim, pode-se ajustar duas curvas para representar a altura de plantas. Na Figura 8 é apresentada a modelagem final das duas curvas de crescimento para a variável altura de planta de mamona.

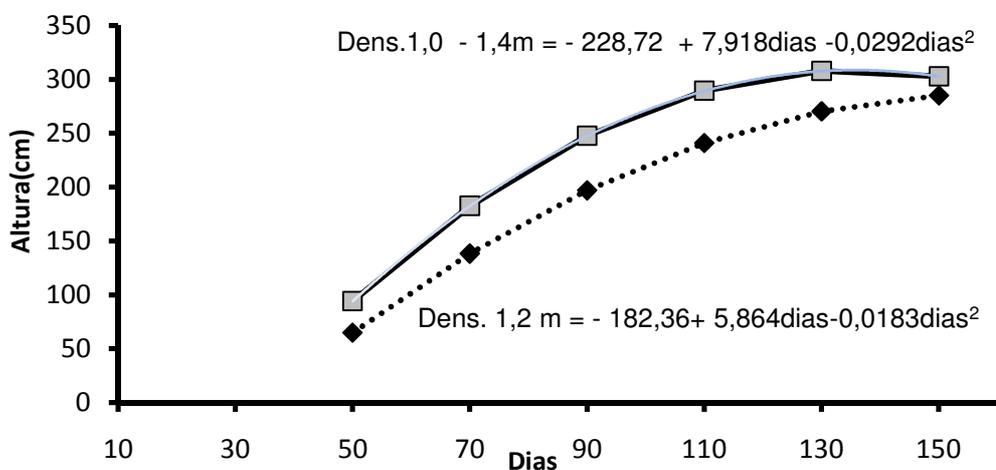


Figura 8 – Modelagem final das duas curvas de crescimento para a variável altura de planta de mamona entre 50 e 150 dias de avaliação. Santa Maria, RS, 2010.

Observa-se que as plantas apresentam semelhanças no perfil de crescimento dos 50 aos 150 dias. Dos 130 aos 150 dias, para a curva Dens 1,0 - 1,4m, o crescimento se mantém aproximadamente constante. Já para o Dens 1,2, nestes mesmos períodos, ocorre a desaceleração do crescimento, e aos 150 dias se dá o pico máximo. Isto já era esperado, pois o crescimento da Variedade Vinema T1 vai até os 150 dias, iniciando-se a colheita dos frutos.

Finalizando-se a modelagem, a fim de verificar a qualidade do ajuste das curvas estimadas (Figuras 7 e 8), fez-se o gráfico de probabilidade normal dos resíduos gerados para as duas curvas de crescimento.

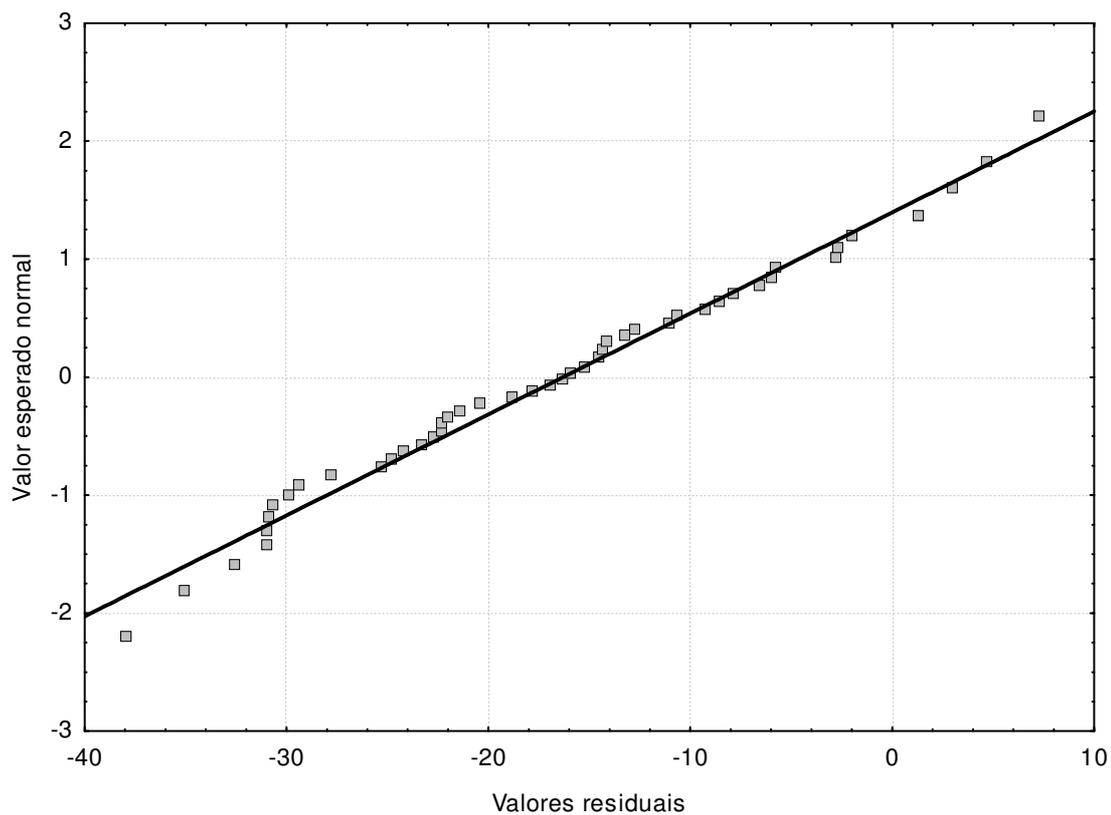


Figura 9 - Gráfico de probabilidade normal dos resíduos para o modelo ajustado à altura de plantas de mamona entre 50 e 150 dias de avaliação nas densidades de cultivo de 1,0m ou 1,4m. Santa Maria, RS, 2010.

Verifica-se que o modelo apresentou um bom ajuste, não apresentando *outliers*. A confirmação do bom ajuste se deu através do teste de Shapiro-Wilks, cujo valor da estatística foi igual 0,9797 ($p=0,5662$).

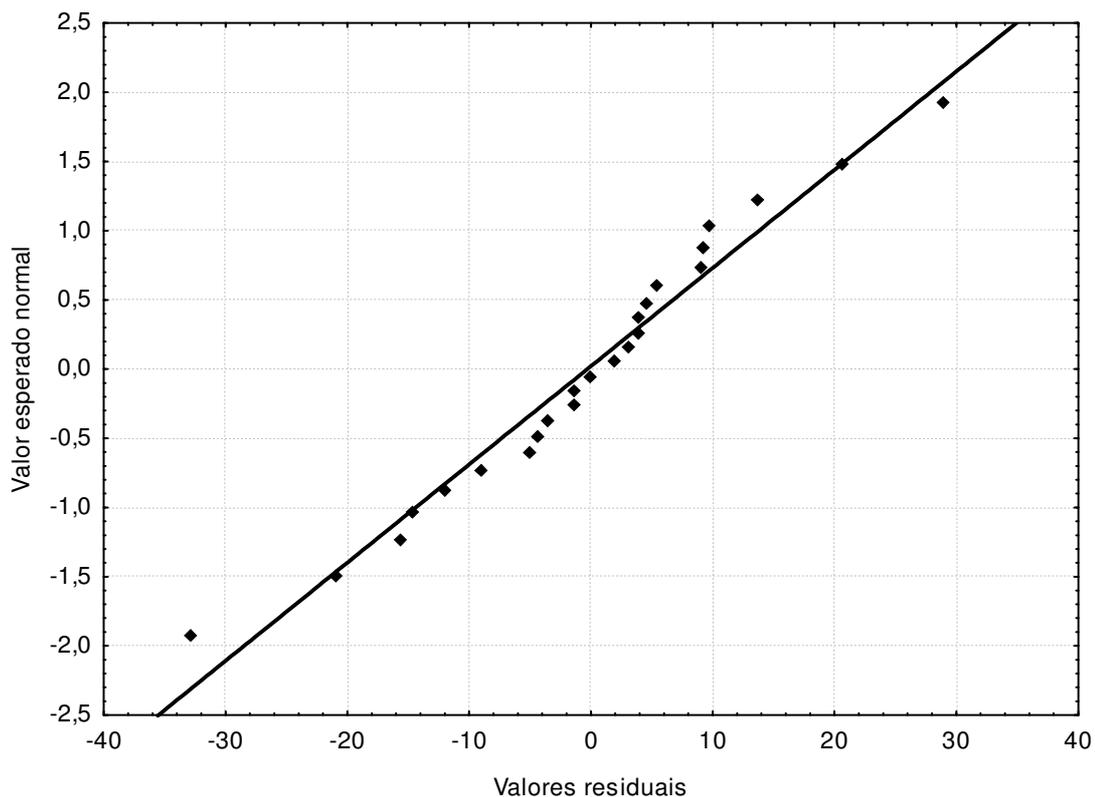


Figura 10 - Gráfico de probabilidade normal dos resíduos para o modelo ajustado à altura de plantas de mamona entre 50 e 150 dias de avaliação nas densidades de cultivo de 1,2m. Santa Maria, RS, 2010.

O modelo para a densidade 1,2 m pode ser considerado com um ajuste razoável, sendo que se identifica apenas um ponto que apresenta um pequeno afastamento da linha. Com o teste de Shapiro-Wilks, com valor da estatística igual 0,9801 ($p=0,8976$), confirma-se a adequação do modelo aos dados.

5 CONCLUSÕES

A metodologia do ajuste de curvas de crescimento por modelos mistos apresenta a grande vantagem de poder ser aplicada em situações nas quais os métodos tradicionais de análise não se aplicam. Além disso, a aplicação desta metodologia está sustentada no fato de se poder explorar de forma mais completa as informações fornecidas pelos dados. Quanto à estratégia utilizada para seleção dos efeitos aleatórios, do modelo e das estruturas de covariância, mostrou-se como alternativa prática e simples para o ajuste das curvas. Além disso, a identificação dos efeitos aleatórios por meio de contrastes polinomiais, gerados pelo *proc GLM*, foi fundamental para identificação do grau do polinômio médio a ser ajustado aos dados, já que o *proc MIXED* não apresenta este subsídio.

Como sugestões para estudos futuros é a implementação da análise, quando o comportamento temporal das respostas de diferentes tratamentos não puder ser explicado por curvas de mesmo grau, assim como, a utilização de modelos de crescimento não-lineares para identificação dos efeitos fixos e aleatórios.

E finalizando, para garantir a competitividade da cultura da mamona no Rio Grande do Sul, sugere-se a realização de mais estudos sobre: técnicas de cultivo, manejo, doenças, cultivares, entre outros.

BIBLIOGRAFIA

ANDRADE, D. F.; SINGER, J da M. Análise de dados longitudinais. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 7. 1986. Campinas. USP. Minicurso. Campinas: SINAPE. **Minicurso**, 106 p.

ANDREONI, S. **Modelos de efeitos aleatórios para a análise de dados longitudinais não balanceados em relação ao tempo**. 1989. 142 f. Dissertação (Mestrado em Estatística). Instituto de Matemática - Universidade de São Paulo, 1989.

ARAÚJO, A. M. **Modelos não lineares para dados de contagem longitudinais**. 2007. 197f. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz” – Universidade de São Paulo, 2007.

ARAÚJO, A. E. de.; SUASSUNA, N. D.; COUTINHO, W. M. Doenças e seu manejo. In: AZEVEDO, D. M. P. de; BELTRÃO, N. E. de M. (Ed.). **O Agronegócio da Mamona no Brasil**, 2. ed. Campina Grande: Embrapa Algodão, Brasília: Embrapa Informação Tecnológica, 2007. cap. 12, p.281-332.

AUBIN, E. C. Q. **Análise de Experimentos com Medidas Repetidas**. 1984. 164f. Dissertação (Mestrado em Estatística) – Instituto de Matemática, Estatística. Universidade de São Paulo, São Paulo, 1984.

AVELAR, R. C. et al. Produção de Mudas de Mamona (*Ricinus communis* L.) em Tubetes de Diferentes Tamanhos. In: BIODIESEL O NOVO COMBUSTÍVEL DO BRASIL, __, 2006, **Anais....__2006.__**: 1 CD-RON.

AZEVEDO, D. M. P. de. et al. População de Plantas no consórcio mamoneira/sorgo. In: CONGRESSO BRASILEIRO DA MAMONA, 2., 2006. **Anais....__2006.__**: 1 CD-RON.

AZEVEDO, D. M. P.; LIMA, E. F. **O agronegócio da mamona no Brasil**. Embrapa Informação Tecnológica, Brasília, 2001. 350 p.

BALDANZI, M.; FAMBRINI, M; PUGLEIESEI, C. Redesign of the castor bean plant body plan for optional combining harvesting. **Applied Biology**, Warwickshire, v. 142, n.3, p. 299 – 306, 2003.

BARBOSA, M. Uma abordagem para a análise de dados com medidas repetidas utilizando modelos lineares misto. 2009. 112f. Dissertação (Mestrado Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz” – Universidade de São Paulo. São Paulo, 2009.

BARCIKOWSKI, R. S.; ROBEY, R. R. Decisions in single group repeated measures analyses: Statistical test and three computer pack-pages. **The American Statistician**, 38, 148-150, 1984.

BATES, D.; PINHEIRO, J.C. **Imer for SAS PROC MIXED users**. [s.l.: s.n.], 2000. Disponível em: <<http://bioestat.hitcock.org/FacultyandStaff/OnlineManuals/PDF%20Files/Imesas.pdf>>. Acesso em: 10 janeiro 2010.

BELTRÃO, N. E. de M. et al. **Mamona: Árvore do conhecimento e sistemas de produção para o Semi-Árido Brasileiro**. Campina Grande: Embrapa Algodão, 2003. 19p. (Circular Técnica, 70).

BELTRÃO, N. E. de M; AZEVEDO, D. M. P. de. Fisiologia. In: AZEVEDO, D. M. P. de; BELTRÃO, N. E. de M. (Ed.). **O Agronegócio da Mamona no Brasil**, 2. ed. Campina Grande: Embrapa Algodão, Brasília: Embrapa Informação Tecnológica, 2007. cap. 6, p.140-150.

BOX, G. E. Some theorems on quadratic forms applied in the study of variance problems. **Mathematical Statistics**, v. 25, p. 290-302, 1954.

BRANDÃO, A. L. S. **Modelos não lineares para a análise de dados longitudinais**. 1996. 137f. Dissertação (Mestrado em Estatística) – Instituto de Matemática, Estatística e Computação. Universidade Estadual de Campinas, São Paulo, 1996.

BRASIL ESCOLA. **Biocombustíveis**. Disponível em: <<http://www.brasilecola.com/geografia/biocombustivel.htm>>. Acesso em 20 janeiro de 2010.

CAMARINHA FILHO, J. A. **Modelos lineares Mistos: Estruturas de Matrizes de variância e covariância e seleção de modelo**. 2002. 74f. Tese (Doutorado em Agronomia) – Escola Superior de agricultura “Luis de Queiros”- Universidade de São Paulo.

CAMARINHA FILHO, J. **Nota Metodológica sobre Modelos Lineares Mistos. Paraná: UFP- DEPARTAMENTO DE ESTATÍSTICA**, 2003. 32p.

CASTRO, S. M. de J. **A metodologia de análise de dados longitudinais**. 1997. 119f. Monografia (Bacharel em Estatística) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 1997.

CASTRO, S. M. de J. **A metodologia de análise de dados longitudinais em plantas forrageiras no contexto de estruturas de covariância**. 1999. 156f. Dissertação (Mestrado em Zootecnia) – Universidade Federal do Rio Grande do Sul, 1999.

COLLE, J. W. L. GRIZZLE, J. E. Applications of Multivariate Analysis of Variance to Repeated Measurements Experiments. **Biometrics**, v. 22, p. 810 – 828, 1966.

COMISSÃO ESPECIAL DA BIOENERGIA-RS. **Mamona petróleo verde - uma alternativa de verdade**. Porto alegre: [s.n.], 2006. 53p. (Documentos, 1).

CONCEIÇÃO, M. M. et al. Degradação térmica de biodiesel de mamona. In: CONGRESSO BRASILEIRO DA MAMONA, I. 2004, Campina Grande. **Anais eletrônicos...** Campina Grande: EMBRAPA, 2004. Disponível em: <<http://www.cnpa.embrapa.br/produtos/mamona/publicacoes.pdf>>. Acesso em: 15 maio. 2007.

COSTA, da. S. C. **Modelos lineares generalizados mistos para dados longitudinais**. 2003. 110f. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz” – Universidade de São Paulo. São Paulo, 2003.

CROWDER, M. J.; D. J. HAND. **Analysis of repeated measures**. Chapman and Hall: London, 1990. 257p.

DANTAS, H. J. et al. Comportamento reológico de biodiesel de mamona. In: CONGRESSO BRASILEIRO DA MAMONA, 1., 2004, Campina Grande:____, **Anais...**Campina Grande:____,2004. 1CD-ROM.

DIGGLE, P. J. et al. **Analysis of longitudinal data**. 4 ed. New York: Oxford University Press, 1998, 253p.

DWYER, J. H.; FEINLEIB, M. Introduction to statistical models for longitudinal observations. In. DWYER, J. H *et al* (eds). **Statistical models for longitudinal studies of health**. New York: Oxford University Press, p.3-48, 1992.

ELSTON, R. C; GRIZZIE, J. E. Time response curves and their confidence bands. **Biometrics**, v.18. p. 148-159, 1962.

EMBRAPA. Centro Nacional de Pesquisa em Algodão. **Mamona: 500 perguntas 500 respostas**. Brasília: Embrapa Informação Tecnológica, 2006. 248p.

EUROPEAN UNION. Directive 2003/30/EC of the European Parliament and of the Council on the promotion of the use biofuels or other renewable fuels transport of 08 may, 2003. **Official Journal of the European Union**. European Union, 17/05/2003. L123/42.

FAI, A. H. T.; CORNELIUS, P. L. Approximate F-tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-plot Experiments. **Statistical Computation and Simulation**, v.54, p. 363–378, 1996.

FAIRCLOUGH, D. L.; HELMS, R. W. A Mixed Linear Model with Linear Covariance Structure: A Sensitivity Analysis of Maximum Likelihood Estimators. **J. Statist. Comput. Simul**, v.25, n.2 , p. 205-236, 1986

FERNANDEZ, G.C.J. Repeated measure analysis of line-source sprinkler experiments. **HortScience**, v.26, p.339-342, 1991.

FERREIRA, D. F. **Análise multivariada**. Lavras: UFLA, 1996. 389p.

FERREIRA, G. B. et al. Variação do crescimento vegetativo e produtivo de alguns genótipos de mamona em diferentes populações de cultivo. In: CONGRESSO BRASILEIRO DA MAMONA, 2., 2006, **Anais....__2006.__**: 1 CD-RON.

FERRON, J., et al. Effects of misspecifying the firstlevel error structure in two level models of change. **Multivariate Behavioral Research**, v. 37, p. 379-403, 2002.

FLORIANO, E. P. *et al.* Ajuste e seleção de modelos tradicionais para série temporal de dados de altura das árvores. **Ciência Florestal**, v.16, n.2, p.177-199, 2006.

FREITAS, G. E de. **Análise de dados longitudinais em experimentos com cana-de-açúcar**. 2007. 74f. Dissertação (Mestrado em Agronomia) – Curso de Pós – graduação em Agronomia. Escola Superior de agricultura “Luis de Queiros” - Universidade de São Paulo, 2007.

GEISSER, S.; GREENHOUSE, S. W. An extension of Box's results on the use of the F distribution in multivariate analysis. **Mathematical Statistics**, v. 29, p.885-891, 1963.

GOLDEMBERG, J.; VILANUEVA, L. D. **Energia, meio ambiente & desenvolvimento**. EDUSP: São Paulo, 2003. 145 p.

GREENHOUSE, S. W.; GEISSER, S. On methods in the analysis of profile data. **Psychometrika**, v. 24, p. 95-112, 1959.

GUIMARÃES, P. R. B. **Modelo linear misto de Laird-Ware: predição de efeitos aleatórios e estimação de parâmetros via filtro de Kaiman**. 1994. 135f. Dissertação (Mestrado em Estatística) – Universidade Estadual de Campinas.

HARTLEY, H.O.; RAO, J.N.K. Maximum likelihood estimation for the mixed analysis of variance model. **Biometrika**, v.54, n.1/2, p.93-108, 1967.

HARVILLE, D. A. BLUP (Best Linear Unbiased Prediction), and beyond. In **Advances in Statistical Methods for Genetic Improvement of Livestock**. New York, Springer-Verlag, 239-276. 1990.

HARVILLE, D. A. Maximum likelihood approaches to variance component estimation and to related problems. **J. Am. Stat. Assoc.**, v.72, p.320-340, 1977.

HARVILLE, D.A. Maximum likelihood approaches to variance component estimation and to related problems. **Journal of the American Statistics Association**, v.72, p.320-328, 1977.

HAUSER, et al. Curva de crescimento usando modelo misto: Uma aplicação na progressão da doença de Machado-Joseph. **Revista do Hospital de Clinicas de Porto Alegre**, v.29, n.1, p. 05-17, 2009

HENDERSON, C. R. **Applications of linear models in animal breeding**. Guelph: University of Guelph. 1984. 462 p.

HENDERSON, C. R. Estimation of changes in herd environment. **Journal of Dairy Science**, v.32, p. 706-711, 1949.

HENDERSON, C. R. The best linear unbiased estimation and prediction under a selection model. **Biometrics**, v.31, p. 423-449. 1975.

HENDERSON, C.R. Sire evaluation and genetic trends. In: ANIMAL BREEDING AND GENETICS SYMPOSIUM IN HONOR OF J.LUSH. III. American Society of Animal Science, Champaign, 1973, **Anais...**1973: p.10-41.

HUYNH, H.; FELDT, L. S. Condition under which mean square ratios in repeated measurements designs have exact F-distributions. **J. Am. Stat. Assoc.**, v.72, p.320-40, 1970.

HUYNH, H.; FELDT, L. S. Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs. **J. Educ. Assoc.**, v.1, n.1, p.69-82, 1976.

INDIAMART. **Castor oil and seed.** Disponível em: <<http://www.finance.indiamart.com/markets/commodity/castor>>. Acesso em 6 fevereiro de 2009.

INSTITUTO CULTURAL PADRE JOSINO. **Agricultura camponesa e as energias renováveis:** um guia técnico. 1. ed. Porto Alegre: Padre Josino, 2008. 230 p.

JHONSON, R. A. WICHERN, D. W. **Applied multivariate statistical analysis**, 2 ed. New York: Prentice Hall, 1988, 607 p.

KESELMAN, et al. Comparison of Two Approaches For Selecting Covariance Structures in The Analysis of Repeated Measurements, 1998, <http://home.cc.umanitoba.ca/~kesel/cis1998.pdf>

KENNEDY, B.W. Variance component estimation and prediction of breeding values. **Canadian Journal of Genetics and Cytology**, v.23, n.4, p.565-578, 1981.

KENWARD, M. G.; ROGER, J. H.. Small sample inference for fixed effects from restricted maximum likelihood. **Biometrics**, v. 53, p. 983-997. 1997

KINCAID. C. Guidelines for Selecting the Covariance Structure in Mixed. COMSYS Information Technology Services, Inc..., Portage, ML. Disponível em: < <http://cran.r-project.org/web/packages/SASmixed/vignettes/Usinglmer.pdf>>. Acesso em: 21 de abril. 2010.

KIRK, R. E. **Experimental design: procedures for the behavioral**. 3 ed. New York: Wadsworth Publishing, 2005. 928p.

KOURI, J.; SANTOS, R. F. dos. **Panorama mundial do agronegócio da mamona.** In: CONGRESSO BRASILEIRO DA MAMONA, 2., 2006, **Anais....__2006.__**: 1 CD-RON.

LAIRD, N. M.; WARE, J. H. Random effects models for longitudinal data. **Biometrics**, v.38, p. 963-974, 1982.

LI, D. (DONGGUANG). A SAS Macro to Automate The Covariance Structure Determination in Repeated Measures Data Analysis with Mixed Procedure. **NCIC-CTG at Queen's University, Kingston, ON, Canada.** Disponível em: <http://www.lexjansen.com/pharmasug/2003/statisticspharmacokinetics/sp013.pdf>. Acesso em: 21 de abril. 2010.

LIMA, C. G de. **Análise de dados longitudinais provenientes de experimento em blocos casualizados.** 1996. 126f. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz” – Universidade de São Paulo, 1996.

LIMA, C. G. **Análise de curvas de crescimento de aves – um enfoque multivariado.** 1988. 69f. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz” – Universidade de São Paulo, 1988.

LINDSTROM, M. J.; BATES, D. M. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. **Journal of the American Statistics Association**, v.83, p.1014-1022, 1988.

LITTELL, R.C. et al. **SAS System for Mixed Models.** 1 ed. North Carolina: SAS Institute Inc., Cary, 2000, 615 p.

LITTELL, R.C. et al. **SAS System for Mixed Models.** 2 ed. North Carolina: SAS Institute Inc., Cary, 2006, 814p.

LITTELL, R.C. et al. Statistical analysis of repeated measures data using SAS procedures. **Journal Animal Science**, v.76, p. 1216-1231, 1998.

LOPES, P. S et al. **Estimação de componentes de variância.** Viçosa: UFV, 1998. 61p. (Cadernos didáticos, 39).

MACEDO, J. O. **Modelos de regressão aleatória usando como bases as funções polinomiais de Legendre, de Jacobi modificadas e trigonométricas, com uma aplicação na análise genética dos pesos de bovinos da raça nelore.** 2007. 100f. Tese (Doutorado em Agronomia) – Escola Superior de agricultura “Luis de Queiros”- Universidade de São Paulo.

MALHEIROS, E.B. Precisão da análise de dados longitudinais, com diferentes estruturas para a matriz de variâncias e variância-covariância, quando se utiliza o esquema em parcelas subdivididas. **Revista de Matemática e Estatística**, v.17, p.229-239, 1999.

MAPA, MCT (Ministério da Ciência e Tecnologia), MME (Ministério de Minas e Energia) & MDIC (Ministério Desenvolvimento Indústria e Comércio). **Diretrizes de Política de Agroenergia 2006-2011**. Brasília, 2005. 33 p.

MARTINS,E.N., et al. **Modelo linear misto**. UFV: Imprensa Universitária, 1993, 46p.

MAUCHLY, J. W.. Significance test for sphericity of a normal n-variate distribution. **Math. Stat.**, v.11, p.204-209, 1940.

MAZZANI, B. Almacenamiento y conserción de semillas oleaginosas. In: MAZZANI, B. **Cultivo y mejoramiento de plantas oleaginosas**. Caracas. Fondo Nacional de Investigación Agropecuárias, 1983. p. 73-93.

Mac CULLOCH, C. E; SEARLE, C. R. **Genaralized, linear, and mixed models**. 2 ed. New York: J. Wiley, 2001. 325 p.

Mac LEAN, R. A.; SANDERS, W. L. Approximating degrees of freedom for standard errors. In **Proceedings of the Statistical Computing Section**. New Orleans: American Statistical Association., p. 50–59,1988.

MELO, F. H.; FONSECA, E. G. Proálcool, energia e transportes. Estudos Econômicos – FIPE. Pioneira: São Paulo. 1981. 250 p.

MEREDITH, M.P.; STEHMAN, S.V. Repeated measures experiments in forestry: focus on analysis of response curves. **Canadian Journal of Forest Research**, v.21, p.957-965, 1991.

METZGER, J. O. Organic reactions without organic solvents and oils as renewable raw materials for the chemical industry. **Chemosphere**, v 43, p.83-87, 2001.

MILLIKEN, G.A.; JOHNSON, D.E. **Analysis of messy data: designed experiments**. v.1. New York: Chapman & Hall, 1992, 473p.

MORAES, C. R. de A. et al. Produção e teor de óleo da mamoneira de porte médio plantada em diferentes espaçamentos. In: CONGRESSO BRASILEIRO DA MAMONA, 2., 2006. **Anais....__2006.__**: 1 CD-RON.

MORRISON, O. F. **Multivariate Statistical Methods**. 2 ed, 3rd reprint. Tokyo: McGraw-Hill, 1981, 415 p.

MOSHKIN, V. A. Economic importance and regions of cultivation of castor. In: _____ **Castor**. New Delhi: Oxanium Press, 1986. p. 1-3.

MULLER, K.; BARTON, C. N. Approximate power for repeated-measures ANOVA lacking sphericity. **Journal of the American Statistical Association**, v.84, p.549-555, 1989.

NETER, J.; WASSERMAN, W. **Applied models. regressions, analysis of variance and experimental designs**. Illinois: Rechar D. Irwin, 1974. 842 p.

NÓBREGA, M. B. de. M. **Avaliação de genótipos de mamona (*Ricinus communis* L) em cruzamentos dialélicos parciais**. 2008. 78f. Tese. (Doutorado em Agronomia) – Escola Superior de Agricultura Luiz de Queiroz, ESALQ, Piracicaba, 2008.

OIL ANALYTICS. **Net energy**. Disponível em: <<http://web.archive.org/web/20041019220802/http://www.oilanalytics.org/neten/neten.html#measure>>. Acesso em: 21 de dez. 2008.

OLIVEIRA, L. M. et al. Teste de tetrazólio em sementes de mamona. In: CONGRESSO BRASILEIRO DA MAMONA, 1., 2004, Campina Grande: Paraíba. **Anais....Campina Grande:_____**,2004. 1CD-ROM.

OSHIIWA, M.; PADOVANI, C. R. Programa computacional para ajuste de curvas polinomiais em experimentos envolvendo dados longitudinais. **Energia na Agricultura**, v.20, n. 3, p. 113-130, 2005.

PARENTE, E. J.S. **Biodiesel**: uma aventura tecnológica num País engraçado. Unigráfica e Térbio: Fortaleza, 2003. 68 p.

PASCUAL, M. G. C. A. **Metodologia de dados espaciais aplicados a plantas forrageiras**. 2000, 100f. Tese (Doutorado em Zootecnia) – Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre, 2000.

PATTERSON, H.D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v.58, p.545-554, 1971.

PEIXOTO, A. R. **Plantas oleaginosas arbóreas**. São Paulo: Nobel, 1973. 284p.

PEREIRA, L. M. **Modelo de formação de preços de commodities agrícolas aplicado ao mercado de açúcar e álcool**. 2009. 209f. Tese (Doutorado em Administração) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2009.

PERRI, S. H. V.; IEMMA A. F. Procedimento "MIXED" do SAS® para análise de modelos mistos. **Scientia Agricola**.v.56, n.4, p. 959 - 967, 1999.

PINHEIRO, J. C. **Topics in mixed effects models**. Madison: University of Wisconsin. 1994. 194f. Tese (Doctorate in Statistic) – University of Wisconsin, Medison, 1994.

POTTHOFF, R. F.; ROY, S. N. A generalized multivariate analysis of variance model useful especially for growth curve problems. **Biometrics**, v. 31, p. 313-326, 1964.

PUERTO, R. J. A. **Programa de biocombustíveis no Brasil e na Colômbia: Uma análise da implantação resultados e perspectivas**. 2007. 194f. Dissertação (Mestrado em Energia) – Programa Interunidades de Pós-Graduação em Energia, Universidade de São Paulo, São Paulo, 2007.

RAMOS, L. P. A importância e a viabilidade do biodiesel como alternativa para a matriz energética nacional. **Revista Biotecnologia & Desenvolvimento**. 2003. Disponível em <http://www.biotecnologia.com.br>. Acesso em 12 janeiro 2010.

RAMPIN, M, A. **Síntese de ésteres etílicos obtidos a partir dos óleos de mamona e soja utilizando a lipase imobilizada de *Thermomyces lanuginosus* (LIPOZYME TL IM)**. 2007. 263f. Dissertação. (Mestrado em Ciências) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, São Paulo, 2009.

RANGEL, L.E.P. et al. **Mamona: situação atual e perspectiva no Mato Grosso**. Campina Grande: Embrapa Algodão, 2003. 16p. (Embrapa Algodão. Documentos, 106).

RESENDE, M. D. V. **Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes**. Colombo: Embrapa Floresta , n. 47, p.1-101, 2000.

RESENDE, M. D. V. de, et al. Estimção de componentes de variância e predição de valores genéticos pelo método da máxima verossimilhança restrita (REML) e melhor predição linear não viciada (BLUP) em *Pinus*. **Boletim de Pesquisa Florestal**, n.32/33, p.23-42, 1996.

RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Floresta, 2007. 362 p.

REZENDE, D.M.L.C. et al. Comparações de técnicas de análises de experimentos utilizando medidas repetidas no tempo. **Ciência e Agrotecnologia**, v.23, n.4, p.927-937, 1999.

RIBOLDI, J. Análise de medidas repetidas: Perspectivas e Tendências. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA. 44 e SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA. 8., 1999, Botucatu. UNESP. Minicurso. Botucatu: Rbras, SEAGRO. **Minicurso**, 92 p.

RIBOLDI, J. Modelos mistos aplicados. REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA. 52. SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA. 12., 2007, Santa Maria. UFSM. Minicurso. Santa Maria: Rbras, SEAGRO. **Minicurso**, 68 p.

SALES, J. C. et al. O biodiesel produzido a partir da mamona e suas conseqüências para o desenvolvimento do Ceará: Aspectos ambientais, sociais e econômicos. In: CONGRESSO BRASILEIRO DA MAMONA, 2., 2004, **Anais....__2004.__**: 1 CD-RON. SANTOS, dos R. F.; BARROS, M. A. L. Cultivo da mamona. Embrapa Algodão. Sistema de Produção. 4. 2004. 2 p.

SANTOS, dos R. F.; BARROS, M. A. L. Cultivo da mamona. Embrapa Algodão. Sistema de Produção. 4. 2003. 2 p.

SANTOS, M. A. dos. **Inserção do biodiesel na matriz energética brasileira: aspectos técnicos e ambientais relacionado ao seu uso em motores de combustão**. 2007. 118f. Dissertação (Mestrado em Energia) – Programa Interunidades de Pós-Graduação em Energia, Universidade de São Paulo, São Paulo, 2007.

SÃO PAULO. Secretaria da Agricultura. **Zoneamento Agrícola do Estado de São Paulo**. Campinas: CATI. 1997. v.2, 131p.

SAVY FILHO, A. et al. **Variedades de mamona do Instituto Agronômico**. Campinas: Instituto Agronômico, 1999. 12p (Boletim Técnico, 183).

SAVY FILHO, A. Hibridação em mamona. In: BORÉM, A. (Ed.). **Hibridação artificial de plantas**. Viçosa: Ed. da UFV, 1999. p. 331-342.

SAVY FILHO, A. **Mamona**. Campinas: Instituto Agronômico, abril 2003. 4p (folheto).

SAVY FILHO, A. **Mamona**: tecnologia de produção. Campinas: EMOPI, 2005. 105p.

SCHEINER S. M; GUREVITCH J. **Design and Analysis of Ecological Experiments**. 2 ed. Oxford: Oxford University Press, Inc; 2001, 445p.

SCIVITTARO, W. B.; PILLON, C. N. **Calagem e adubação para a cultura da mamona no Sul do Brasil**. Pelotas: Embrapa Clima Temperado, 2006. 45p. (Comunicado Técnico, 150).

SEARLE, S. R. **Linear models for unbalanced data**. New York: J John Wiley, 1987. 536p.

SEARLE, S. R. **Linear Models**. New York: John Wiley, 1971, 516 p.

SEARLE, S.R., et al. **Variance components**. New York: Jonh Wiley & Sons, 1992, 501p.

SEVERINO, L. S. et al. Crescimento e produtividade da mamoneira influenciada por plantio em diferentes espaçamento entre linhas. **Ciência Agrônômica**, v. 37, n.1, p.50-54, 2006.

SHAW, R.G. Maximum-likelihood approaches to quantitative genetics of natural populations. **Evolution**, v.41, p.812-826, 1987.

SHEPETINA, F.A.; SEVAST'YANOVA, L.B. **Seed technology**. In: MOSCKIM, V.A. *Castor*. New Delhi: Amerind, 1986. p.175-8.

SILVA, S. D. dos A. et al. **A cultura da mamona no Rio Grande do Sul**. 1.ed. Pelotas: Embrapa Clima Temperado, 2007. v. 1. 115 p.

SILVA, S. D. dos. A. et al. **A cultura da mamona na região de clima temperado: informações preliminares**. Pelotas: Embrapa Clima Temperado, 2005. 33p. (Comunicado Técnico, 149).

SILVA, S. D. dos. A. *et al.* Avaliação de cultivares de mamona em Pelotas – RS, Safra 2003/04. In: CONGRESSO BRASILEIRO DA MAMONA, 1., 2004, Campina Grande:____, **Anais....**Campina Grande:____,2004. 1CD-ROM

SINGER, J. da. M. **Análise de curvas de crescimento**. 1977. 108f. Dissertação (Mestrado em Estatística) Instituto de Matemática e Estatística — Universidade de São Paulo, São Paulo, 1977.

SINGER, J. M. *et al.* **Análise de dados longitudinais**. Caderno Didático (Versão preliminar, setembro/2008). Departamento de Estatística – Universidade de São Paulo. 154p, 2008.

SINGER, J.M.; ANDRADE, D.F. On the choice of appropriate error terms in profile analysis. **The Statistician**, v.43, p. 259-266, 1994.

SMITH, H., R. et al. Multivariate Analysis of Variance. (MANOVA). **Biometrics**, v.2, p. 61 – 67, 1962.

SOARES, J. J. et al. Principais pragas e seu controle. In: AZEVEDO, D. M. P. de; BELTRÃO, N. E. de M. (Ed.). **O Agronegócio da Mamona no Brasil**, 2. ed. Campina Grande: Embrapa Algodão, Brasília: Embrapa Informação Tecnológica, 2007. cap. 11, p.255-280.

SOUZA, M. A. A. Arcabouço regulatório do biodiesel: especificações. In: SEMINÁRIO SOBRE TECNOLOGIA DE COMBUSTÍVEIS LUBRIFICANTES DA UNIFACS. I., 2005, Salvador. **Anais eletrônicos...** Salvador: UNIFACS, 2005. Disponível <<http://www.energia.unifacs.com.br/eventos/recontec>>. Acesso em 12 janeiro 2008.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics**. New York: McGraw - Hill, 1960. 481p.

STEEL, R.G.D.; J.H. TORRIE. 1988. **Bioestadística: principios e procedimientos**. 2 ed. México: McGraw-Hill, 1988, 622p.

STUKER, H. **Análise multivariada para dados onde a característica observada é subdividida em K classes**. 1986. 92f. Dissertação. (Mestrado em Agronomia) Escola Superior de Agricultura “Luiz de Queiros” – ESALQ. São Paulo, 1986.

SUYAMA, E. Modelos de Efeitos Aleatórios para Dados Longitudinais. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA. 40 e SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA. 6., 1995, Ribeirão Preto. USP. Minicurso. Ribeirão Preto: Rbras, SEAGRO. **Minicurso**, 94 p.

TIMM, N.H. Multivariate analysis of variance of repeated measurements. **Handbook of Statistics**, v 1. P. R. Krishnaiah (Ed). North-Holland:New York. p. 41-87, 1980.

VENEZUELA, M. K. **Modelos lineares generalizados para análise de dados com medidas repetidas**. 2003. 95f. Dissertação (Mestrado em Estatística) Instituto de Matemática e Estatística — Universidade de São Paulo. São Paulo, 2003.

VERBEKE, G.; MOLENBERGHS, G. **Linear Mixed Models for Longitudinal Data**. New York, Springer-Verlag. 2000, 569 p.

VERBEKE, G.; MOLENBERGHS, G.. **Linear Mixed Models in Practice: A SAS-Oriented Approach**. New York, Springer-Verlag, 1997. 306 p.

VIEIRA, F. T. P. de A. **Uma Abordagem Multivariada em Experimento Silvopastoril com *Leucaena leucocephala* (Lam.) de Wit. no Agreste de Pernambuco**. 2006. 71f. Dissertação (Mestrado em Biometria). Universidade Federal Rural de Pernambuco, Pernambuco, 2006.

VIEIRA, R. M.; LIMA, E. F.. importância sócio-econômica e melhoramento da mamoneira no Brasil. In: QUEIROS, M. A. de; GOEDERT, C. O.; RAMOS, S. R. R. (Ed.). **Recursos genéticos e melhoramento de plantas para o nordeste brasileiro**. Disponível em: <<http://www.cpatsa.embrapa.br>>. Acesso em 12 janeiro 2008.

VONESH, F.E.; CHINCHILLI, V.M. **Linear and nonlinear models for the analysis of repeated measurements**. New York: Marcel Dekker, 1997. 560p.

WALD, V. B. **A metodologia de modelos mistos não lineares aplicados a análise de dados longitudinais em plantas forrageiras**. 2000. 106f. Dissertação (Mestrado em Zootecnia) – Universidade Federal do Rio Grande do Sul, 2000.

WEISS, E. A. **Oilseed crops**. London: Blackwell Science, 1983. 660 p.

WEST, B. T. et al. **Linear mixed models: a practical guide using statistical software**. New York: Chapman & Hall, 2007. 339 p.

WISHART, J. Growth rate determinations in nutrition studies with the bacon pig and their analysis. **Biometrika**, v.30, p. 16- 38, 1938.

WREGGE, M. S. et al. Época de semeadura da mamona no Rio Grande do Sul com baixo risco de ocorrência de geada. In: CONGRESSO BRASILEIRO DA MAMONA, 2., 2006, **Anais...._2006._**: 1 CD-RON.

XAVIER, L. H. Modelos **univariado e multivariado para análise de medidas repetidas e verificação da acurácia do modelo univariado por meio de simulação**. 2000. 91f. Dissertação (Mestrado Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz” – Universidade de São Paulo. São Paulo, 2000.

XAVIER, L. H.; SANTOS, C. T. D. Acurácia do Modelo Univariado para Análise de Medidas Repetidas por Simulação Multidimensional. **Scientia Agricola**, v.58, n.2, p.241-250, 2001.

APÊNDICES

APÊNDICE A – Dados da altura de planta (cm) referente as três densidade de cultivo nos oito períodos de avaliação.

UE ¹ .	Den ²	Rep ³	Tempo de coleta (dias)							
			10	30	50	70	90	110	130	150
1	1,2	1	7,7	42,21	71,64	190	230	275	292	306
2	1,2	2	6,6	43,6	63,5	184	242	281	277	301
3	1,2	3	8,6	43,5	69,5	184	233,6	276,3	278,1	280,7
4	1,2	4	8,4	42,1	71	187,36	231,5	283	297,3	300,2
5	1,4	1	9,2	31,4	74,16	159	188	220	275	280
6	1,4	2	9,5	30	70,5	137	200	270	280	289
7	1,4	3	9,2	31	69	126,4	206	254,6	270,3	286,9
8	1,4	4	9,3	39,5	61,5	105,5	181,3	236,6	269	270,3
9	1,0	1	7,6	38,9	73,9	180	229	275	291	292
10	1,0	2	7	37	80	170	220	257	270	281
11	1,0	3	8,9	41	65	173,5	226,4	267,2	292,7	297,1
12	1,0	4	7,3	34	69	174,9	217,1	258,6	272,9	278,8

1 – Unidade experimental, 2 – Densidade, 3 – Repetição.

APÊNDICE B - Modelos testados para a matriz G

```
dm 'output; clear; log; clear;';
options formdlim='*' pageno=001 ls=100;
data multivariado (keep=UE Dens Rep A50 A70 A90 A110 A130 A150)
univariado(keep=UE Dens Rep dias A );
input UE Dens Rep A50 A70 A90 A110 A130 A150 ;
output multivariado;
A=A50; dias=50; output univariado;
A=A70; dias=70; output univariado;
A=A90; dias=90; output univariado;
A=A110; dias=110; output univariado;
A=A130; dias=130; output univariado;
A=A150; dias=150; output univariado;
cards;
01 1.0 1 71.64 190 230.0 275.0 292.0 306
02 1.0 2 63.5 184 242.0 281.0 277.0 301
03 1.0 3 69.5 184 233.6 276.3 278.1 280.7

:.....:
11 1.4 3 65 173.5 226.4 267.2 292.7 297.1
12 1.4 4 69 174.9 217.1 258.6 272.9 278.8
;
proc mixed data=univariado;
class Dens rep ;
```

MODELOS PARA TESTADOS PARA A

```
proc mixed data=univariado;
class Dens rep ;
title modelo A1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random intercept dias dias*dias/ type=UN sub=rep(Dens) ;
proc mixed data=univariado;
class Dens rep ;
title modelo A2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=UN;
proc mixed data=univariado;
class Dens rep ;
title modelo A3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc sub=rep(Dens) r;
random dias/subject=rep(Dens) type=UN;
proc mixed data=univariado;
class Dens rep ;
title modelo B4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc sub=rep(Dens) r;
```

MODELOS PARA TESTADOS PARA B

```
proc mixed data=univariado;
class Dens rep ;
title modelo B1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
```

```

random intercept dias dias*dias/ type=vc sub=rep(Dens) ;
proc mixed data=univariado;
class Dens rep ;
title modelo B2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=vc;
proc mixed data=univariado;
class Dens rep ;
title modelo B3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc sub=rep(Dens) r;
random dias/subject=rep(Dens) type=vc;
proc mixed data=univariado;
class Dens rep ;
title modelo B4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc sub=rep(Dens) r;

```

MODELOS PARA TESTADOS PARA C

```

proc mixed data=univariado;
class Dens rep ;
title modelo C1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random intercept dias dias*dias/ type=AR(1) subject=rep(Dens) ;
proc mixed data=univariado;
class Dens rep ;
title modelo C2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=AR(1);
proc mixed data=univariado;
class Dens rep ;
title modelo C3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc sub=rep(Dens) r;
random dias/subject=rep(Dens) type=AR(1);
proc mixed data=univariado;
class Dens rep ;
title modelo C4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc sub=rep(Dens) r;

```

MODELOS PARA TESTADOS PARA D

```

proc mixed data=univariado;
class Dens rep ;
title modelo D1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random intercept dias dias*dias/ type=CSH subject=rep(Dens) ;
proc mixed data=univariado;
class Dens rep ;
title modelo D2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=CSH;
proc mixed data=univariado;

```

```

class Dens rep ;
title modelo D3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random dias/subject=rep(Dens) type=CSH;
proc mixed data=univariado;
class Dens rep ;
title modelo D4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;

```

MODELOS PARA TESTADOS PARA E

```

proc mixed data=univariado;
class Dens rep ;
title modelo E1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random intercept dias dias*dias/ type=ARH(1) subject=rep(Dens) ;
proc mixed data=univariado;
class Dens rep ;
title modelo E2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=ARH(1);
proc mixed data=univariado;
class Dens rep ;
title modelo E3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random dias/subject=rep(Dens) type=ARH(1);
proc mixed data=univariado;
class Dens rep ;
title modelo E4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;

```

MODELOS PARA TESTADOS PARA F

```

proc mixed data=univariado;
class Dens rep ;
title modelo E1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random intercept dias dias*dias/ type=UN(1) subject=rep(Dens) ;
proc mixed data=univariado;
class Dens rep ;
title modelo E2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=UN(1);
proc mixed data=univariado;
class Dens rep ;
title modelo E3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=vc subject=rep(Dens) r;
random dias/subject=rep(Dens) type=UN(1);
proc mixed data=univariado;
class Dens rep ;
title modelo E4;

```

```
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;  
repeated / type=vc subject=rep(Dens) r;  
run;  
quit;
```

APÊNDICE C - Modelos testados para a matriz R

```

dm 'output; clear; log; clear;';
options formdlim='*' pageno=001 ls=100;
data multivariado (keep=UE Dens Rep A50 A70 A90 A110 A130 A150)
univariado(keep=UE Dens Rep dias A );
input UE Dens Rep A50 A70 A90 A110 A130 A150 ;
output multivariado;
A=A50; dias=50; output univariado;
A=A70; dias=70; output univariado;
A=A90; dias=90; output univariado;
A=A110; dias=110; output univariado;
A=A130; dias=130; output univariado;
A=A150; dias=150; output univariado;
cards;
01 1.0 1 71.64 190 230.0 275.0 292.0 306
02 1.0 2 63.5 184 242.0 281.0 277.0 301
03 1.0 3 69.5 184 233.6 276.3 278.1 280.7

:.....:

11 1.4 3 65 173.5 226.4 267.2 292.7 297.1
12 1.4 4 69 174.9 217.1 258.6 272.9 278.8
;
proc mixed data=univariado;
class Dens rep ;

```

MODELOS PARA TESTADOS PARA A

```

title modelo A1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / subject=rep(Dens) type=vc;
random dias dias*dias/subject=rep(Dens) type=un;
proc mixed data=univariado;
class Dens rep ;
title modelo A2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=un sub=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=un;
proc mixed data=univariado;
class Dens rep ;
title modelo A3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=AR(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=un;
proc mixed data=univariado;
class Dens rep ;
title modelo A4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=CSH subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=un;
proc mixed data=univariado;
class Dens rep ;
title modelo A5;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=ARH(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=un;
proc mixed data=univariado;
class Dens rep;
title modelo A6;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;

```

```
repeated / type=UN(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=un;
```

MODELOS PARA TESTADOS PARA B

```
proc mixed data=univariado;
class Dens rep ;
title modelo B1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / subject=rep(Dens) type=VC;
random dias dias*dias/subject=rep(Dens) type=VC;
proc mixed data=univariado;
class Dens rep ;
title modelo B2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=un sub=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=VC;
proc mixed data=univariado;
class Dens rep ;
title modelo B3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=AR(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=VC;
proc mixed data=univariado;
class Dens rep ;
title modelo B4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=CSH subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=VC;
proc mixed data=univariado;
class Dens rep ;
title modelo B5;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=ARH(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=VC;
proc mixed data=univariado;
class Dens rep;
title modelo B6;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=UN(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=VC;
```

MODELOS PARA TESTADOS PARA C

```
title modelo C1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / subject=rep(Dens) type=VC;
random dias dias*dias/subject=rep(Dens) type=AR(1);
proc mixed data=univariado;
class Dens rep ;
title modelo C2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=un sub=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=AR(1);
proc mixed data=univariado;
class Dens rep ;
title modelo C3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=AR(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=AR(1);
```

```

proc mixed data=univariado;
class Dens rep ;
title modelo C4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=CSH subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=AR(1);
proc mixed data=univariado;
class Dens rep ;
title modelo C5;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=ARH(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=AR(1);
proc mixed data=univariado;
class Dens rep;
title modelo C6;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=UN(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=AR(1);

```

MODELOS PARA TESTADOS PARA D

```

title modelo D1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / subject=rep(Dens) type=VC;
random dias dias*dias/subject=rep(Dens) type=CSH;
proc mixed data=univariado;
class Dens rep ;
title modelo D2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=un sub=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=CSH;
proc mixed data=univariado;
class Dens rep ;
title modelo D3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=AR(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=CSH;
proc mixed data=univariado;
class Dens rep ;
title modelo D4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=CSH subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=CSH;
proc mixed data=univariado;
class Dens rep ;
title modelo D5;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=ARH(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=CSH;
proc mixed data=univariado;
class Dens rep;
title modelo D6;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=UN(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=CSH;

```

MODELOS PARA TESTADOS PARA E

```

title modelo E1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / subject=rep(Dens) type=VC;
random dias dias*dias/subject=rep(Dens) type=ARH(1);
proc mixed data=univariado;
class Dens rep ;
title modelo E2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=un sub=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=ARH(1);
proc mixed data=univariado;
class Dens rep ;
title modelo E3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=AR(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=ARH(1);
proc mixed data=univariado;
class Dens rep ;
title modelo E4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=CSH subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=ARH(1);
proc mixed data=univariado;
class Dens rep ;
title modelo E5;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=ARH(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=ARH(1);
proc mixed data=univariado;
class Dens rep;
title modelo E6;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=UN(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=ARH(1);

```

MODELOS PARA TESTADOS PARA F

```

title modelo F1;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / subject=rep(Dens) type=VC;
random dias dias*dias/subject=rep(Dens) type=UN(1);
proc mixed data=univariado;
class Dens rep ;
title modelo F2;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=un sub=rep(Dens) r;
random dias dias*dias/sub=rep(Dens) type=UN(1);
proc mixed data=univariado;
class Dens rep ;
title modelo F3;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=AR(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=UN(1);
proc mixed data=univariado;
class Dens rep ;
title modelo F4;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=CSH subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=UN(1);

```

```
proc mixed data=univariado;
class Dens rep ;
title modelo F5;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=ARH(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=UN(1);
proc mixed data=univariado;
class Dens rep;
title modelo F 6;
model A=Dens dias Dens*dias Dens*dias*dias/ chisq s;
repeated / type=UN(1) subject=rep(Dens) r;
random dias dias*dias/subject=rep(Dens) type=UN(1);
run;
quit;
```

APÊNDICE D – Testes para efeitos fixos e ajuste das curvas

```

dm 'output; clear; log; clear;';
options formdlim='*' pageno=001 ls=100;
data multivariado (keep=UE Dens Rep A50 A70 A90 A110 A130 A150)
univariado(keep=UE Dens Rep dias A );
input UE Dens Rep A50 A70 A90 A110 A130 A150 ;
output multivariado;
A=A50; dias=50; output univariado;
A=A70; dias=70; output univariado;
A=A90; dias=90; output univariado;
A=A110; dias=110; output univariado;
A=A130; dias=130; output univariado;
A=A150; dias=150; output univariado;
cards;
01 1.0 1 71.64 190 230.0 275.0 292.0 306
02 1.0 2 63.5 184 242.0 281.0 277.0 301
03 1.0 3 69.5 184 233.6 276.3 278.1 280.7

:.....:

11 1.4 3 65 173.5 226.4 267.2 292.7 297.1
12 1.4 4 69 174.9 217.1 258.6 272.9 278.8
;
proc mixed data=univariado covtest method=REML;
class Dens rep;
model A=Dens dias Dens*dias Dens*dias*dias/s;
repeated / type=vc subject=rep(Dens) r;
random dias dias*dias/ type=un(1) subject=rep(Dens);
lsmeans Dens/ pdiff adjust=tukey;
run;
quit;

```

APÊNDICE E – Estruturas de variância-covariância testadas para a matriz G (variação entre indivíduos) e seleção de efeitos aleatórios para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.

Grupos	Efeitos Fixos	Modelos	Efeitos Aleatórios	Estruturas		Critérios			Número de parâmetros
				R	G	--2Res Log	AIC	BIC	
A	dias, densidades e interações	1	intercepto, dias, dias ²	VC	UN	Não Convergiu			
		2	dias, dias ²	VC	UN	Não Convergiu			
		3	dias	VC	UN	580,7	584,7	585,7	2
		4	sem	VC	UN	583,2	585,2	585,7	1
B	dias, densidades e interações	1	intercepto, dias, dias ²	VC	VC	580,7	584,7	585,7	4
		2	dias, dias²	VC	VC	580,9	582,7	585,7	3
		3	dias	VC	VC	582,2	584,7	584,9	2
		4	sem	VC	VC	583,2	585,2	585,7	1
C	dias, densidades e interações	1	intercepto, dias, dias ²	VC	AR(1)	Não Convergiu			
		2	dias, dias ²	VC	AR(1)	Não Convergiu			
		3	dias	VC	AR(1)	580,7	585,7	586,2	3
		4	sem	VC	AR(1)	583,2	586,2	588,7	1
D	dias, densidades e interações	1	intercepto, dias, dias ²	VC	CSH	Não Convergiu			
		2	dias, dias²	VC	CSH	580,7	583,7	585,2	4
		3	dias	VC	CSH	580,2	586,7	585,7	3
		4	sem	VC	CSH	583,2	587,2	588,7	1
E	dias, densidades e interações	1	intercepto, dias, dias ²	VC	ARH(1)	Não Convergiu			
		2	dias, dias²	VC	ARH(1)	580,7	584,7	585,7	4
		3	dias	VC	ARH(1)	580,6	585,7	584,7	3
		4	sem	VC	ARH(1)	583,2	586,2	585,7	1
F	dias, densidades e interações	1	intercepto, dias, dias ²	VC	UN(1)	Não Convergiu			
		2	dias, dias²	VC	UN(1)	571,7	564,7	575,2	4
		3	dias	VC	UN(1)	580,0	583,2	585,7	2
		4	sem	VC	UN(1)	683,2	585,2	585,7	1

APÊNDICE F - Estruturas de variância-covariância testadas para a R (variação dentro de indivíduos) para ajuste de modelos para a altura de plantas de mamona entre 50 e 150 dias de avaliação em três densidades. Santa Maria, RS, 2010.

Grupos	Efeitos Fixos	Modelos	Efeitos Aleatórios	Estruturas		Critérios			Parâmetros de Covariância
				R	G	-2Res Log	AIC	BIC	
A	dias, densidades e interações	1	dias, dias ²	VC	UN	Não Convergiu			
		2	dias, dias ²	UN	UN				
		3	dias, dias ²	AR(1)	UN				
		4	dias, dias ²	CSH	UN				
		5	dias, dias ²	ARH(1)	UN				
		6	dias, dias ²	UN(1)	UN				
B	dias, densidades e interações	1	dias, dias ²	VC	VC	580,7	584,7	588,7	3
		2	dias, dias ²	UN	VC	Não Convergiu			
		3	dias, dias ²	AR(1)	VC	580,7	584,7	585,1	4
		4	dias, dias ²	CSH	VC	Não Convergiu			
		5	dias, dias ²	ARH(1)	VC	Não Convergiu			
		6	dias, dias ²	UN(1)	VC	Não Convergiu			
C	dias, densidades e interações	1	dias, dias ²	VC	AR(1)	Não Convergiu			
		2		UN	AR(1)	Não Convergiu			
		3		AR(1)	AR(1)	Não Convergiu			
		4	dias, dias ²	CSH	AR(1)	Não Convergiu			
		5	dias, dias ²	ARH(1)	AR(1)	Não Convergiu			
		6	dias, dias ²	UN(1)	AR(1)	Não Convergiu			
D	dias, densidades e interações	1	dias, dias ²	VC	CSH	580,7	584,7	585,7	4
		2	dias, dias ²	UN	CSH	Não Convergiu			
		3	dias, dias ²	AR(1)	CSH	580,7	588,7	590,6	5
		4	dias, dias ²	CSH	CSH	Não Convergiu			
		5	dias, dias ²	ARH(1)	CSH	Não Convergiu			
		6	dias, dias ²	UN(1)	CSH	Não Convergiu			
E	dias, densidades e interações	1	dias, dias ²	VC	ARH(1)	580,7	584,7	585,7	4
		2	dias, dias ²	UN	ARH(1)	Não Convergiu			
		3	dias, dias ²	AR(1)	ARH(1)	580,7	588,7	590,6	5
		4	dias, dias ²	CSH	ARH(1)	Não Convergiu			
		5	dias, dias ²	ARH(1)	ARH(1)	Não Convergiu			
		6	dias, dias ²	UN(1)	ARH(1)	Não Convergiu			
F	dias, densidades e interações	1	dias, dias ²	VC	UN(1)	580,7	582,7	584,7	4
		2	dias, dias ²	UN	UN(1)				
		3	dias, dias ²	AR(1)	UN(1)	580,7	586,7	588,1	5
		4	dias, dias ²	CSH	UN(1)	Não Convergiu			
		5	dias, dias ²	ARH(1)	UN(1)	Não Convergiu			
		6	dias, dias ²	UN(1)	UN(1)	Não Convergiu			