Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Análise de variância multivariada nas estimativas dos parâmetros
do modelo log-logístico para susceptibilidade do
capim-pé-de-galinha ao glyphosate

César Augusto Degiato Jotta

Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Estatística e Experimentação Agronômica

Piracicaba
2016

César Augusto Degiatio Jotta
Físico Teórico-Experimental

**Análise de variância multivariada nas estimativas dos parâmetros do modelo
log-logístico para susceptibilidade do capim-pé-de-galinha ao glyphosate**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **CARLOS TADEU DOS SANTOS DIAS**

Dissertação apresentado para obtenção do título de Mestre
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

**Piracicaba
2016**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Jotta, César Augusto Degiato

Análise de variância multivariada nas estimativas dos parâmetros do modelo log-logístico para susceptibilidade do capim-pé-de-galinha ao glyphosate / César Augusto Degiato Jotta. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016.

83 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Modelo não-linear 2. Teste da razão de verossimilhança 3. MANOVA I. Título

CDD 632.58
J85a

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

*Aos meus familiares e às pessoas sem as quais
a arte de viver para mim, se tornaria extenuante, dedico.*

AGRADECIMENTOS

Primeiramente, agradeço a Deus por sua infinda bondade e paciência, à minha família que mesmo com os percalços da vida se manteve unida e me proporcionou uma trajetória tranquila e feliz, em especial pelo papel fundamental em minha educação, à minha mãe , Vera Lúcia Degiato e ao meu pai, Jair Antônio Jotta, os responsáveis por todas as minhas conquistas.

A todos os meus professores e em especial, ao meu professor do ensino fundamental Emerson Silva dos Santos, que me incentivou a nunca parar de estudar e ter o estudo como prática diária e contínua , sem seu trabalho as dificuldades encontradas seriam ainda maiores.

Aos meus velhos e novos amigos que ao longo dos anos vêm se somando, tanto em quantidade como em posicionamento político. Espero que continue sempre assim.

Em especial, à minha namorada e companheira de momentos bons e ruins, Patrícia Bongiovanni Catandi, por sempre ser a melhor companhia.

A toda equipe do departamento de Ciências Exatas da ESALQ, em particular, à prof. Dra. Sônia Maria de Stefano Piedade por deixar as aulas e os dias mais agradáveis com seu espírito acolhedor e gentil, à prof. Dra. Renata Alcarde Sermarini por sempre estar à disposição para tirar dúvidas demonstrando se importar com o desenvolvimento do aluno. Ao meu professor e orientador Dr. Carlos Tadeu dos Santos Dias pelo acolhimento de prontidão e pela disposição sempre presentes com os assuntos de trabalho além da excelente companhia.

Ao Lucas Rosa pelo seu trabalho e concessão do conjunto de dados.

À Capes pelo auxílio financeiro concedido nestes dois anos de mestrado.

A todos, o meu muito obrigado.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Modelos de regressão	19
2.2 Modelos de regressão linear	20
2.3 Modelos de regressão não-linear	21
2.4 Modelo log-logístico com erro normal	22
2.5 Métodos de estimação dos parâmetros não-lineares	24
2.5.1 Método dos mínimos quadrados ordinários	25
2.5.2 Método dos mínimos quadrados ponderados	27
2.5.3 Método dos mínimos quadrados generalizados	28
2.5.4 Método da máxima verossimilhança	29
2.6 Métodos iterativos	31
2.6.1 Método de Gauss-Newton	32
2.7 Igualdade de parâmetros em modelos não-lineares	33
2.8 Análise de dados multivariados	35
2.8.1 Análise de variância multivariada	36
2.8.2 Delineamento aleatório em blocos	36
2.8.3 Decomposição da variabilidade total	38
2.8.4 Análise estatística	40
3 MATERIAL	45
3.1 Obtenção das sementes, produções das mudas e aplicação do tratamento	45
4 RESULTADOS E DISCUSSÃO	47
4.1 Análise exploratória	47
4.2 Análise de regressão não-linear	51
4.3 Análise de variância multivariada	59
5 CONCLUSÃO	63
REFERÊNCIAS	65

APÊNDICE	67
--------------------	----

RESUMO

Análise de variância multivariada nas estimativas dos parâmetros do modelo log-logístico para susceptibilidade do capim-pé-de-galinha ao glyphosate

O cenário agrícola nacional tem se tornado cada vez mais competitivo ao longo dos anos, manter o crescimento da produtividade a um baixo custo operacional e com baixo impacto ambiental tem sido os três ingredientes de maior relevância na área. A produtividade por sua vez, é função de várias variáveis, sendo o controle de plantas daninhas uma dessas variáveis a ser considerada. Nesse trabalho é analisado um conjunto de dados de um experimento realizado no departamento de Produção Vegetal da ESALQ-USP, Piracicaba - SP. Foram avaliadas 4 biótipos de capim-pé-de-galinha provenientes de três estados brasileiros e em três estágios morfológicos com 4 repetições para cada biótipo, a variável resposta utilizada foi massa seca (g) e como variável regressora foi utilizada a dose de glyphosate nas concentrações variando de $1/16 D$ a $16 D$ mais a testemunha, sem aplicação de herbicida, em que D varia de 480 gramas de equivalente ácido de glyphosate por hectare ($g .e a. ha^{-1}$) para o estágio de 2 a 3 perfilhos, 720 ($g .e a. ha^{-1}$) para o estágio de 6 a 8 perfilhos e de 960 para o estágio de 10-12 perfilhos. O trabalho teve como objetivo primário avaliar se, ao longo dos anos, as populações de capim-pé-de-galinha tem se tornado resistentes ao herbicida glyphosate, visando detecção de biótipos resistentes. O experimento foi instalado segundo o delineamento inteiramente aleatorizado, sendo feito em três estágios diferentes. Para a análise dos dados foi utilizado o modelo não-linear log-logístico proposto em Knezevic, S. e Ritz (2007) como método univariado, foi utilizado ainda o método da máxima verossimilhança para verificar a igualdade do parâmetro e . O modelo utilizado convergiu para quase todas as repetições, mas não houve um comportamento sistemático observado que explicasse a não convergência de uma repetição em particular. Num segundo momento, as estimativas dos três parâmetros do modelo foram tomadas como variáveis dependentes em uma análise de variância multivariada. Observando que as três, conjuntamente, foram significativas pelos testes de Pillai, Wilks, Roy e Hotelling-Lawley, foi realizado o teste de Tukey para o mesmo parâmetro e comparado com o primeiro método utilizado. Esse procedimento apresentou, com o mesmo coeficiente de significância, menor capacidade de identificar diferença entre as médias dos parâmetros das variedades de capim do que o método proposto por Regazzi (2015).

Palavras-chave: Modelo não-linear; Teste da razão de verossimilhança; MANOVA

ABSTRACT

Multivariate analysis of variance in the estimates of the log-logistic model parameters for susceptibility of grass chicken feet to glyphosate

The national agricultural scenery has become increasingly competitive over the years, maintaining productivity growth at a low operating cost and low environmental impact has been the three most important ingredients in the area. Productivity in turn is a function of several variables, and the weed control is one of these variables to be considered. In this work it is analyzed a dataset of an experiment conducted in the Plant Production Department of ESALQ-USP, Piracicaba - SP. Were evaluated 4 grass chicken's feet biotypes from three Brazilian states in three morphological stages with 4 repetitions for each biotype, the response variable used was dry mass (g) and as regressor variable were used the dose of glyphosate in concentrations ranging from 1/16 D to 16 D plus the control without herbicide, wherein D ranges from 480 grams of glyphosate acid equivalent per hectare (g .e a. ha⁻¹) for 2 to 3 stage tillers, 720 grams of glyphosate acid equivalent per hectare (g .e a. ha⁻¹) for 6 to 8 tillers and 960 for stage 10-12 tillers. The work had as main objective to evaluate , if over the years, populations of grass chicken's feet has become resistant to glyphosate, aiming detection of resistant biotypes. The experiment was conducted under completely randomized design being done in three stages. For data analysis was used the non-linear log-logistic proposed in Knezevic, S. e Ritz (2007) as univariate method, it was still used the maximum likelihood method to verify the equality of the parameter e. The model converged to almost all repetitions, but there was an observed systematic behavior to explain the non-convergence of a particular repetition. Secondly, estimates of the three model parameters were taken as dependent variables in a multivariate analysis of variance. Noting that all three together, were significant by Pillai, Wilks, Roy and Hotelling-Lawley tests, was performed Tukey test for the same parameter e and compared with the first method. This procedure presented, with the same coefficient of significance, less able to identify differences between the means of the parameters of grass varieties than the method proposed by Regazzi (2015).

Keywords: Nonlinear model; Likelihood ratio test; MANOVA

LISTA DE FIGURAS

Figura 1 - Diferentes valores do parâmetro d , com $b = 0,5$ e $e = 50$	23
Figura 2 - Diferentes valores do parâmetro b , com $d = 4$ e $e = 50$	23
Figura 3 - Diferentes valores do parâmetro e , com $d = 4$ e $b = 0,5$	24
Figura 4 - Foto do estágio de 10 a 12 perfilhos da variedade de suspeita suscetível após 28 dias da aplicação do herbicida	46
Figura 5 - Evolução da massa seca em função da dose de herbicida no estágio em que a planta apresenta de dois a três perfilhos	48
Figura 6 - Evolução da massa seca em função da dose de herbicida no estágio em que a planta apresenta de seis a oito perfilhos	49
Figura 7 - Evolução da massa seca em função da dose de herbicida no estágio em que a planta apresenta de dez a doze perfilhos	50
Figura 8 - Diagrama de caixa da massa-seca em função da dose de herbicida	51
Figura 9 - Análise de resíduo <i>versus</i> observação nos estágios em que a planta apre- sentava entre 2 a 3 perfilhos	52
Figura 10 - Análise de resíduo <i>versus</i> observação nos estágios em que a planta apre- sentava entre 6 a 8 perfilhos	53
Figura 11 - Análise de resíduo <i>versus</i> observação nos estágios em que a planta apre- sentava entre 10 a 12 perfilhos	54

LISTA DE TABELAS

Tabela 1 - Algumas equações não-lineares utilizadas no estudo de curva de crescimento	22
Tabela 2 - Organização dos dados de um delineamento aleatorizado em blocos	37
Tabela 3 - Esquema de análise de variância multivariada de um delineamento alea- torizado em blocos	40
Tabela 4 - Transformações da distribuição Λ de Wilk para a distribuição F	42
Tabela 5 - Espécie, origem, identificação (sigla) e suspeita de resistência ao glyphosate	45
Tabela 6 - Estimativa dos parâmetros do modelo no estágio de 2-3 perfilhos	53
Tabela 7 - Estimativa dos parâmetros do modelo no estágio de 6-8 perfilhos	55
Tabela 8 - Estimativa dos parâmetros do modelo no estágio de 10-12 perfilhos	55
Tabela 9 - Valores previstos pelo modelo log-logístico de dose de herbicida que acar- retam a diminuição de 90% e 95% da matéria seca do capim-pé-de-galinha após de 28 dias da aplicação	56
Tabela 10 -Análise da igualdade dos parâmetros em cada variedade dentro dos três estágios considerados	57
Tabela 11 -Valores p calculados	58
Tabela 12 -Análise de variância multivariada	59
Tabela 13 -Valores-p calculados para os testes de Shapiro-Wilk e M-Box	60
Tabela 14 -Teste de Tukey para o parâmetro e nos os três estágios avaliados	60

1 INTRODUÇÃO

Plantas daninhas têm sido uma fonte de estudo muito importante para aumentar os índices de melhoria contínua nos campos agrícolas do Brasil, tais plantas possuem características como, alta robustez, fácil e eficiente reprodução, competindo em nutrientes, água e luz com as plantas de interesse comercial, soja, milho, feijão, cebola, batata, arroz entre outras. Podendo causar, se não combatidas, prejuízos econômicos ou ainda a completa destruição da plantação dependendo da cultura a ser cultivada.

Uma das formas mais eficientes de controle de população das plantas invasoras tanto em grandes, quanto em pequenas propriedades é o controle químico por meio de aplicação de herbicidas. A eficiência desse método, por sua vez, acaba fazendo com que os agricultores utilizem na maioria das vezes apenas o uso do herbicida para o controle de ervas daninhas, deixando de associar outros métodos de controle e diferentes mecanismos de ação, podendo levar ao longo do tempo o surgimento de biótipos (variedades) resistentes por pressão de seleção (ROSA, 2014).

Neste trabalho, por meio de métodos estatísticos, como o modelo log-logístico com erros normais e análise de variância uni e multivariada, analisa-se o conjunto de dados obtido em um experimento realizado por Rosa (2014). O objetivo geral desse trabalho pode ser dividido em duas partes, sendo: *i*) introduzir de maneira didática os conceitos estatísticos necessários; *ii*) utilizar dois métodos para verificar a resistências das variedades de capim-pé-de-galinha avaliadas.

O conceito de análise de regressão é muito importante para se descrever alguns comportamentos de interesse do pesquisador, nesse contexto, verifica-se dois elementos: a variável resposta ou dependente e a variável regressora ou independente. A finalidade está em se poder prever para um dado valor da variável regressora o valor da variável resposta. A análise de regressão e suas subdivisões são descritas a partir da seção 2.1.

O trabalho realizado por Knezevic, S. e Ritz (2007) foi utilizado como pesquisa e os cálculos necessários foram feitos por meio do software R versão 3.1.3, contudo não foi utilizado o pacote drc, proposto no mesmo artigo, os códigos utilizados estão em anexo.

As estimativas dos parâmetros do modelo log-logístico foram usadas como vetor de variável resposta na análise de variância multivariada, observando que foram significativas optou-se por fazer uma análise univariada da variável considerada como o pa-

parâmetro e do modelo, para essa variável, então foi realizado o teste de Tukey, observando que comparado com o método utilizado por Regazzi (2015), o teste de Tukey conseguiu identificar menos diferença entre as médias do parâmetro e . Não foi observado na variedade de capim-pé-de-galinha obtida em região que não possuía histórico de controle de ervas daninhas por herbicida, um comportamento de susceptibilidade frente as demais variedades analisadas.

2 REVISÃO BIBLIOGRÁFICA

2.1 Modelos de regressão

Políticas que valorizem a captação de informação são cada vez mais importantes para uma posição privilegiada diante das mais diferentes relações possíveis na sociedade. Dessa maneira a transformação de um conjunto de dados em informação prática, com instruções diretas, passa a ser cada vez mais valorizada no contexto atual. São empresas buscando a liderança no mercado por apresentarem um produto com maior qualidade e menor custo, são países buscando a superioridade bélica para manterem seus interesses comerciais e suas soberanias asseguradas, entre inúmeros outros exemplos, a aquisição de informação é um processo contínuo e veloz.

Em grande medida, modelos de regressão são parte fundamentais no processo de aquisição de informação por se preocupar com o estudo da relação entre duas ou mais variáveis tentando responder perguntas de maneira mais objetivas possível utilizando-se de critérios probabilistas, são exemplos, os experimentos em que se verifica um determinado herbicida é eficiente no controle de uma planta invasora e sendo eficiente qual dose é a ideal para o controle, na pecuária tem-se o problema de se estimar a quantidade ideal de ração que maximiza a produção de leite de um certo rebanho, ou a quantidade de carne por animal, entre outros.

Segundo (RATKOWSKY, 1983), as variáveis utilizadas para descrever um fenômeno de interesse do pesquisador são classificadas, em modelos de regressão como, dependentes ou independentes. A primeira (variáveis dependentes), sua medida é verificada como função das demais variáveis, as quais são tomadas como independentes. A variável dependente apresenta uma função de distribuição de probabilidade, por essa razão é chamada de aleatória enquanto que as variáveis independentes podem ou não seguir uma distribuição de probabilidade podendo assim serem chamadas de aleatórias e/ou fixas.

Um modelo de regressão univariado apresenta três componentes sendo: *i*) Y a variável dependente; *ii*) $f(x)$ função linear ou não-linear de x denominada de variável independente e *iii*) ϵ o erro de regressão, não controlado. De uma maneira geral um modelo de regressão univariado pode ser escrito matematicamente como:

$$Y = f(x) + \epsilon \quad (1)$$

Para sintetizar ainda mais, de acordo com Montgomery, P. e Viningy (2015), modelos de regressão são utilizados com a finalidade de: *i*) descrição de dados, uma vez que, em posse de um modelo bem ajustado se espera que esse reproduza de maneira fiel o conjunto de dados, havendo um compromisso entre o conjunto de dados e o modelo utilizado ; *ii*) estimativa de parâmetros, as funções utilizadas para modelar o conjunto de dados possuem alguns parâmetros a serem encontrados, com a utilização de algumas técnicas é possível determiná-los; *iii*) previsão, com algum valor da variável não observada calcula-se o valor da variável resposta, obtendo assim uma estimativa para variável resposta sob um determinado valor da variável independente, que a princípio pode ser de valor diferente do valor da variável observada.

2.2 Modelos de regressão linear

Segundo Rencher e Schaalje (2008), modelos de regressão linear são muito utilizados em diversas áreas do conhecimento, como parte do processo de aprendizagem do método científico, principalmente nos estágios de planejamento da pesquisa e análise dos dados resultantes.

Quando os níveis de um fator é quantitativo como dose, tempo, quantidade de ração, entre outros, a análise de comparações múltiplas para verificar diferenças entre os níveis não é indicada segundo Regazzi e Silva (2004), sendo necessário incorporar no modelo de alguma forma a natureza quantitativa da variável e assim modelos de regressão linear e não-linear são técnicas que podem ser muito úteis em tal situação.

Um modelo de regressão linear pode ser simples, se apresentar apenas uma variável independente, múltipla se apresentar mais de uma variável independente, ou ainda multivariada se apresentar mais de uma variável dependente, um modelo de regressão linear múltipla apresenta a seguinte forma

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i \quad (2)$$

em que os parâmetros $\beta_0, \beta_1, \dots, \beta_k$ são os coeficientes de regressão; ϵ_i é o erro de regressão do modelo em que $\epsilon_i \sim N(0, \sigma^2)$, Y_i é a variável dependente , $X_{i1}, X_{i2}, \dots, X_{ik}$ são as variáveis independentes e i é uma variável indexadora da observação, variando de 1 até n , em que n é o número total de observações.

Ainda, segundo Rencher e Schaalje (2008), modelos de regressão escritos como em 2 são utilizados para os seguintes fins: *i*) predição, por meio de um modelo bem ajustado, é esperado que sob as mesmas condições do objeto de estudo, novos valores das variáveis independentes observadas possam ser utilizadas para calcular com precisão suficiente a variável resposta; *ii*) descrição ou exploração dos dados, cientistas e/ou engenheiros usam o modelo estimado para resumir ou descrever os dados utilizados; *iii*) estimação dos parâmetros, os valores das estimativas podem ter implicações teóricas para um modelo postulado; *iv*) seleção de variáveis, um dos objetivos do pesquisador pode estar na verificação da importância de uma variável independente utilizada para explicar a variável resposta, em outras palavras, as variáveis independentes (ou preditoras) que estão associadas com uma importante quantidade de variação em Y são mantidas, enquanto que, aquelas que contribuem pouco podem ser deletadas; *v*) controle da saída, se uma relação de causa-efeito entre Y e X é assumida, o modelo estimado deve ser usado para controlar as saídas de um processo variando as entradas. Por experimentação sistemática, pode ser possível conseguir a saída ótima.

As pressuposições mais usuais, segundo Montgomery, P. e Viningy (2015), nos modelos de regressão linear são: *i*) normalidade da variável resposta, independência dos resíduos, homogeneidade de variância, linearidade e aditividade do modelo.

2.3 Modelos de regressão não-linear

De acordo com Seber e Wild (1989) uma importante tarefa em estatística é encontrar as relações de um grupo de variáveis em que pelo menos uma seja aleatória, sendo o conceito de aleatório, definido de maneira prática, como sujeito a uma descrição por meio de uma função de distribuição de probabilidade. Segundo o mesmo autor, particularmente na biologia, em oposição as ciências físicas, os processos que estão por trás do fenômeno estudado são geralmente mais complexos e não bem compreendidos, o que significa ter pouca ou nenhuma noção da forma da relação existente entre variável resposta (dependente) e variável(eis) explicativa (independentes) e o objetivo passa a ser a determinação de alguma função que se aproxime ao máximo da função 'verdadeira' desconhecida.

Seja o modelo de regressão univariado escrito como

$$y_i = f(x_i, \boldsymbol{\theta}) + \epsilon_i \quad (3)$$

em que y_i é a observação da variável dependente, x_i é a observação da variável independente, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ é o vetor de parâmetros multidimensional da função f e ϵ_i é o erro de regressão não controlado, dizemos que o modelo de regressão em estudo é linear, se e somente se, a derivada parcial em relação a θ_j , $j = 1, \dots, p$, ou seja, $\frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \theta_j}$ não depender de maneira funcional de nenhum dos parâmetros $\boldsymbol{\theta}$ da função f utilizada. Nesse momento, vale salientar que a classificação de linear, não-linear é referente ao parâmetro do modelo e não às variáveis regressoras. Na Tabela 1 todas funções são não lineares (MENDES et al., 2009) e sintetizam a maioria das curvas de crescimento.

Tabela 1 - Algumas equações não-lineares utilizadas no estudo de curva de crescimento

Função	Equação	M
Brody	$A(1 - Be^{-kt})$	1
Von Bertalanffy	$A(1 - Be^{-kt})^3$	3
Logística	$A(1 - Be^{-kt})^{-1}$	-1
Gompertz	$A \exp(-e^{-kt})$	∞
Richards	$A(1 - Be^{-kt})^{-M}$	variável

2.4 Modelo log-logístico com erro normal

O modelo de regressão log-logístico proposto por (KNEZEVIC; S.; RITZ, 2007) possui a seguinte parametrização,

$$y = \frac{d}{1 + \exp(b \log(x) - b \log(e))} + \epsilon \quad (4)$$

em que y é a variável avaliada, no presente trabalho é a matéria seca, x é a dose utilizada de herbicida, b está relacionado com a declividade da curva, d é o limite superior da curva, e é a dose responsável por corresponder a uma diminuição de 50 % da variável resposta e $\epsilon \sim N(0, \sigma^2)$ é o erro aleatório .

Nos gráficos 1 a 3 pode-se perceber como cada parâmetro modifica a curva.

Figura 1 - Diferentes valores do parâmetro d , com $b = 0,5$ e $e = 50$

Figura 2 - Diferentes valores do parâmetro b , com $d = 4$ e $e = 50$

Figura 3 - Diferentes valores do parâmetro e , com $d = 4$ e $b = 0,5$

No gráfico 1 pode-se perceber que d muda o valor do ponto de máximo da curva que acontece em $y(0)$, em 2, o parâmetro b explica a declividade da curva, assim para valores maiores de b observa-se-se uma maior velocidade de decrescimento de y , já em 3 tem-se quando $x = e$

$$y(e) = \frac{d}{1 + \exp \{b(\log(e) - \log(e))\}} = \frac{d}{2} \quad (5)$$

assim e é um valor de interpretação prática direta, de tal forma que ao se aplicar a dose e observa-se 50% de diminuição da matéria seca teórica.

2.5 Métodos de estimação dos parâmetros não-lineares

A estimação dos parâmetros de modelos não-lineares podem ser realizadas por diversos métodos, nesse trabalho, será apresetado os dois métodos mais utilizados, o método dos mínimos quadrados e o método da máxima verossimilhança, contudo, em ambos não há solução algébrica, fechada, precisando além do método de estimação um método iterativo para a obtenção dos valores estimados utilizados na equação de regressão, esses por sua vez, também não são únicos sendo um deles descrito mais adiante nesse trabalho.

2.5.1 Método dos mínimos quadrados ordinários

Segundo D. e Smith (1998), sob condições de homogeneidade de variância e independência das observações, o método dos mínimos quadrados, apresentam estimadores não viesados e de variância mínima, sendo por essa maneira um dos métodos mais utilizados nos modelos não lineares. De acordo com Seber e Wild (1989), um modelo não-linear na forma matricial pode ser escrito como,

$$\mathbf{Y} = \mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\epsilon} \quad (6)$$

em que $\mathbf{Y}$ é o vetor da variável resposta de dimensão $n \times 1$ composto por y_i , $\mathbf{f}$ é o vetor de uma função conhecida utilizada no modelo de regressão, composto por $f = f(x_i, \boldsymbol{\theta})$, $\boldsymbol{\theta}$ é o vetor de parâmetros de dimensão $k \times 1$ e $\boldsymbol{\epsilon}$ é o erro aleatório de dimensão $n \times 1$. Se houver independência entre as variáveis dependentes e ainda não existir diferença significativa entre as variâncias das mesmas, então $\epsilon \sim N(0, \Sigma = \sigma^2 I)$, sendo I a matriz identidade de dimensão n . Sob esta condição o nome do método utilizado é método do mínimos quadrados ordinários.

Escrita na forma matricial,

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} f(X_1, \boldsymbol{\theta}) \\ f(X_2, \boldsymbol{\theta}) \\ \vdots \\ f(X_n, \boldsymbol{\theta}) \end{bmatrix}; \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (7)$$

A estimativa de mínimos quadrados de $\boldsymbol{\theta}$, denotada por $\hat{\boldsymbol{\theta}}$, minimiza a seguinte soma de quadrados

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2 \quad (8)$$

Escrita matricialmente, 8 fica,

$$S(\boldsymbol{\theta}) = [\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})]'[\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})] \quad (9)$$

$$S(\boldsymbol{\theta}) = \mathbf{Y}'\mathbf{Y} - \mathbf{f}'(\boldsymbol{\theta})\mathbf{Y} - \mathbf{Y}'\mathbf{f}'(\boldsymbol{\theta}) + \mathbf{f}'(\boldsymbol{\theta})\mathbf{f}'(\boldsymbol{\theta})$$

$$S(\boldsymbol{\theta}) = \mathbf{Y}'\mathbf{Y} - \mathbf{f}'(\boldsymbol{\theta})\mathbf{Y} - \mathbf{Y}'\mathbf{f}(\boldsymbol{\theta}) + \mathbf{f}'(\boldsymbol{\theta})\mathbf{f}(\boldsymbol{\theta})$$

$$S(\boldsymbol{\theta}) = \mathbf{Y}'\mathbf{Y} - 2\mathbf{f}'(\boldsymbol{\theta})\mathbf{Y} + \mathbf{f}'(\boldsymbol{\theta})\mathbf{f}(\boldsymbol{\theta})$$

Para encontrar os valores dos parâmetros $\hat{\boldsymbol{\theta}}$ é necessário que

$$\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0 \quad (10)$$

Desenvolvendo 10,

$$\begin{aligned} \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -2 \left[\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]' \mathbf{Y} + 2 \left[\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]' \mathbf{f}(\boldsymbol{\theta}) \\ \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -2 \left[\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]' [\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})] \end{aligned} \quad (11)$$

A matriz

$$\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial f(x_1, \boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f(x_1, \boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f(x_1, \boldsymbol{\theta})}{\partial \theta_k} \\ \frac{\partial f(x_2, \boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f(x_2, \boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f(x_2, \boldsymbol{\theta})}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(x_n, \boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f(x_n, \boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f(x_n, \boldsymbol{\theta})}{\partial \theta_k} \end{bmatrix} \quad (12)$$

é chamada de matriz jacobiana de X e o lado direito da equação 11 igualado a zero

$$\left[\frac{\partial \mathbf{f}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \right]' [\mathbf{Y} - \mathbf{f}(\hat{\boldsymbol{\theta}})] = 0 \quad (13)$$

é o sistema de equações normais não-lineares, segundo (SEBER; WILD, 1989) não é possível encontrar uma expressão fechada ou explícita de maneira geral para $\hat{\boldsymbol{\theta}}$ tendo que ser utilizado algum algoritmo computacional para encontrá-lo.

Como já mencionado, não é necessária nenhuma pressuposição para a aplicação do método dos mínimos quadrados, contudo as considerações a respeito do erro aleatório $\boldsymbol{\epsilon}$ necessitam ser verificadas para que as hipóteses sobre os parâmetros $\boldsymbol{\theta}$ possam ser utilizadas, ou seja, se $\text{var}(\boldsymbol{\epsilon}) = \sigma^2 I$ é, de fato, verdadeira.

2.5.2 Método dos mínimos quadrados ponderados

Para o método de mínimos quadrados continuar fornecendo estimadores com propriedades ótimas como sendo nao-viesados e de variância mínima quando a suposição de homocedasticidade (igualdade de variância) não for satisfeita é preciso utilizar o método de mínimos quadrados ponderados. Assim, para ilustrar o método considere um modelo de regressão não-linear múltipla como em 6, em que a variável dependente $\mathbf{Y}$ continua ainda apresentando uma estrutura não correlacionada, mas a matriz de variância-covariância não é mais assumida $\sigma^2 I$, o que leva a uma reformulação do problema e que o erro de regressão do modelo passa a ser descrito como $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{V}\sigma^2)$, em que $\mathbf{V}$ é uma matriz diagonal positiva definida

$$\mathbf{V}\sigma^2 = \begin{bmatrix} V_1 & 0 & \dots & 0 \\ 0 & V_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & V_n \end{bmatrix} \sigma^2 \quad (14)$$

em que V_i é a variância da variável X_i .

A soma de quadrados a ser minimizada passa a ser escrita como

$$S(\boldsymbol{\theta}) = [\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})]' \mathbf{V} [\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})] \quad (15)$$

A ideia para encontrar uma nova solução para o modelo que admita heterocedasticidade necessita basicamente de uma transformação linear, para isso seja

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_n \end{bmatrix} \sigma^2 \quad (16)$$

$\lambda_i = \frac{1}{\sqrt{V_i}}$, dessa maneira multiplicando a esquerda a equação 2 na forma matricial por $\boldsymbol{\Lambda}$,

$$\boldsymbol{\Lambda}\mathbf{Y} = \boldsymbol{\Lambda}\mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\Lambda}\boldsymbol{\epsilon} \quad (17)$$

O que é um novo modelo, só que com um novo vetor coluna de erro $\boldsymbol{\varepsilon} = \boldsymbol{\Lambda}\boldsymbol{\epsilon}$ que apresentam as seguintes propriedades:

$$\begin{aligned}
E[\boldsymbol{\varepsilon}] &= E[\boldsymbol{\Lambda}\boldsymbol{\epsilon}] = \boldsymbol{\Lambda}E[\boldsymbol{\epsilon}] \\
&= \boldsymbol{\Lambda}\mathbf{0} = \mathbf{0} \\
Var[\boldsymbol{\varepsilon}] &= Var[\boldsymbol{\Lambda}\boldsymbol{\epsilon}] = \boldsymbol{\Lambda}Var[\boldsymbol{\epsilon}]\boldsymbol{\Lambda}' \\
&= \sigma^2 \boldsymbol{\Lambda} \underbrace{\mathbf{V}}_{\mathbf{V}^{\frac{1}{2}}\mathbf{V}^{\frac{1}{2}}} \boldsymbol{\Lambda}' = \sigma^2 \underbrace{\boldsymbol{\Lambda}\mathbf{V}^{\frac{1}{2}}\mathbf{V}^{\frac{1}{2}}\boldsymbol{\Lambda}'}_{\mathbf{I}} \\
&= \sigma^2 \mathbf{I}
\end{aligned} \tag{18}$$

Dessa maneira com a transformação em 17 recai no método de mínimos quadrados ordinários, em que o vetor coluna $\mathbf{Y}$ passa a ser $\boldsymbol{\Lambda}\mathbf{Y}$, a matriz $\mathbf{X}$ passa a ser $\boldsymbol{\Lambda}\mathbf{X}$ e o vetor de erros $\boldsymbol{\epsilon}$, $\boldsymbol{\Lambda}\boldsymbol{\epsilon}$.

2.5.3 Método dos mínimos quadrados generalizados

Quando além da pressuposição não atendida de homocedasticidade as variáveis apresentam autocorrelação entre si (elementos diferentes de zero fora da diagonal principal na matriz de variância-covariância), a matriz a ser multiplicada afim de recair de maneira análoga no método dos mínimos quadrados ponderados não é mais uma matriz diagonal. Para tanto, considere que o modelo apresente uma matriz de variância-covariância $\boldsymbol{\Omega}$ de modo que $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}\sigma^2)$.

Utilizando mais uma vez a ideia abordada no método dos mínimos quadrados ponderados de tal forma que se precisa fazer uma transformação linear afim de recair no modelo com $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ e $Var[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I}$. Seja uma matriz $\mathbf{T}$ não singular de modo que $\mathbf{T}\mathbf{T}' = \boldsymbol{\Omega}$, a transformação linear

$$\mathbf{T}^{-1}\mathbf{Y} = \mathbf{T}^{-1}\mathbf{f}(\boldsymbol{\theta}) + \mathbf{T}^{-1}\mathbf{u} \tag{19}$$

transforma o modelo original em um novo modelo apresentando média e variância do erro, respectivamente,

$$\begin{aligned}
E[\mathbf{T}^{-1}\mathbf{u}] &= \mathbf{T}^{-1}E[\mathbf{u}] \\
&= \mathbf{0} \\
var[\mathbf{T}^{-1}\mathbf{u}] &= \mathbf{T}^{-1}var[\mathbf{u}](\mathbf{T}^{-1})' \\
var[\mathbf{T}^{-1}\mathbf{u}] &= \mathbf{T}^{-1}\boldsymbol{\Omega}(\mathbf{T}^{-1})' = \sigma^2 \mathbf{T}^{-1}\mathbf{T}\mathbf{T}'(\mathbf{T}^{-1})' \\
&= \sigma^2 \mathbf{I}
\end{aligned} \tag{20}$$

Seja o erro autocorrelacionado de primeira ordem escrito na seguinte forma

$$u_i = \phi_1 u_{i-1} + \epsilon_i \quad (21)$$

em que u_i é o erro no instante i , ϕ_1 é o parâmetro autorregressivo de ordem 1, η_{i-1} é o erro no instante $i - 1$ e ϵ_i é o ruído branco Morettin e Toloi (2004), com as seguintes condições:

$$\begin{aligned} E[\epsilon_i] &= 0 \\ E[\epsilon_i^2] &= \sigma_\epsilon^2 \\ E[\epsilon_i, \epsilon_{i-j}] &= 0 \end{aligned} \quad (22)$$

se $j \neq 0$

Em que

$$\Omega = \frac{\sigma_\epsilon^2}{1 - \phi_1^2} \begin{bmatrix} 1 & \phi_1 & \phi_1^2 & \dots & \phi_1^{n-1} \\ \phi_1 & 1 & \phi_1 & \dots & \phi_1^{n-2} \\ \phi_1^2 & \phi_1 & 1 & \dots & \phi_1^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1^{n-1} & \phi_1^{n-2} & \phi_1^{n-3} & \dots & 1 \end{bmatrix} \quad (23)$$

É importante notar que esse modelo contempla, para observações mais próximas correlações maiores, uma vez que $\phi_1 < 0$ e que se $\phi_1 = 0$ o método volta a ser o método dos mínimos quadrados ordinários. Outros processos autorregressivos além de 21 para a determinação da matriz de variância-covariância podem ser considerados (MORETTIN; TOLOI, 2004).

2.5.4 Método da máxima verossimilhança

O método da máxima verossimilhança consiste em considerar as observações da variável dependente como provindas de uma amostra aleatória com função de distribuição de probabilidade pré-determinada. Para exemplificar o método, considere $Y_1, Y_2, \dots, Y_n$ uma amostra aleatória de tamanho n , provinda da variável aleatória Y , com função de distribuição de probabilidade $f(y_i|\theta)$, com $\theta \in \Theta$, sendo Θ o espaço paramétrico (H.BOLFARINE; C.SANDOVAL, 2001). Define-se a função verossimilhança de θ , como

$$L(\theta; y_1, \dots, y_n) = \prod_{i=1}^n f(y_i, \theta). \quad (24)$$

Na função de verossimilhança, observa-se que ela não é mais função de Y , uma vez que, no momento de se calcular tal função, a amostra já foi obtida, sendo função apenas de θ , observando uma outra amostra de tamanho n da variável Y a função de verossimilhança poderia ser outra. Assim, procura-se um por um vetor $\hat{\theta}$ que maximize-a, dizemos que $\hat{\theta}$ é o estimador de máxima verossimilhança de θ .

Considere para ilustração do método o modelo de crescimento log-logístico, adaptado de Regazzi e Silva (2004) :

$$y_{ij} = \frac{d_i}{1 + \exp(b_i(\log(x_{ij}) - \log(e_i)))} + \epsilon_{ij} \quad (25)$$

em que, y_{ij} é o valor observado na j -ésima unidade experimental do i -ésimo grupo que corresponde a porcentagem de controle do herbicida, $i = 1, \dots, g$, $j = 1, \dots, n_i$ (podendo ser desbalanceado); x_{ij} é o valor da dose do herbicida (g e.a. ha⁻¹) associado ao valor observado y_{ij} ; n_i é o número de observações da unidade experimental i sendo $n = n_1 + n_2 + \dots + n_g$ o número total de observações; d_i, b_i e e_i são os parâmetros do modelo em cada grupo i , d é o limite superior da curva, b é a declividade da curva e e é a dose que corresponde a diminuição de 50% da variável y ; $\epsilon_{ij} \sim N(0, \sigma^2)$ é o erro de regressão.

Calculando a esperança da variável dependente,

$$\begin{aligned} E[y_{ij}] &= \frac{d_i}{1 + \exp(b_i(\log(x_{ij}) - \log(e_i)))} + \underbrace{E[\epsilon_{ij}]}_{=0} \\ &= \mu(d_i, b_i, e_i, x_{ij}) \end{aligned}$$

A função de verossimilhança do modelo pode ser escrita como

$$\begin{aligned} L(\theta; y_1, \dots, y_n) &= \prod_{i=1}^g \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_{ij} - \mu(d_i, b_i, e_i, x_{ij}))^2}{2\sigma^2}\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \mu(d_i, b_i, e_i, x_{ij}))^2\right\} \end{aligned} \quad (26)$$

em que $\theta = (d_1, \dots, d_g, b_1, \dots, b_g, e_1, \dots, e_g, \sigma^2)$. Para se tornar mais simples a busca pelos parâmetros que maximizem a função de verossimilhança é usual nos casos em que se admite distribuição normal, trabalhar com o logaritmo natural da função de verossimilhança por ser uma função monótona e crescente, de modo que os valores dos parâmetros que maximizam a função do logaritmo da verossimilhança também irão maximizar a função de verossimilhança.

$$\begin{aligned}
l(\boldsymbol{\theta}, y_1, \dots, y_n) &= \log(L(\boldsymbol{\theta}; y_1, \dots, y_n)) \\
&= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \mu(d_i, b_i, e_i, x_{ij}))^2
\end{aligned} \tag{27}$$

Em posse de 27, o problema recai em encontrar um máximo global $\hat{\boldsymbol{\theta}}$ no domínio de $l(\boldsymbol{\theta}, y_1, \dots, y_n)$, ou seja, em $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. No caso, de maneira análoga ao método dos mínimos quadrados não é garantida uma expressão fechada para $\hat{\boldsymbol{\theta}}$, sendo necessário recorrer a algum método iterativo. Vale ressaltar, com referência a 27 que tomando σ^2 como mais um parâmetro a ser estimado,

$$\frac{\partial l(\boldsymbol{\theta}, y_1, \dots, y_n)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{(\sigma^2)^2} \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \mu(d_i, b_i, e_i, x_{ij}))^2$$

Igualando a zero e observando que de fato é um ponto de máximo (pelo teste da derivada segunda), o estimador de σ^2 de máxima verossimilhança é

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \mu(\hat{d}_i, \hat{b}_i, \hat{e}_i, x_{ij}))^2 \tag{28}$$

Ao se derivar a verossimilhança com relação a $\boldsymbol{\theta}$ a equação a ser resolvida é a mesma encontrada no método dos mínimos quadrados, nessa situação o método de máxima verossimilhança para $\boldsymbol{\theta}$ é igual ao método dos mínimos quadrados ordinários (SEBER; WILD, 1989).

2.6 Métodos iterativos

Os métodos numéricos são ferramentas de fundamental importância para a obtenção das estimativas dos parâmetros dos modelos estatísticos, sem eles não seria possível avançar nos estudos e a maioria dos problemas ficaria sem solução por não apresentarem soluções algébricas explícitas. Existem muitos métodos numéricos, cada um com sua particularidade, ordem de convergência e simplicidade algorítmica. Os métodos mais utilizados são o método de Gauss-Newton, o método de Newton-Raphison, o método do Gradiente e o método de Marquardt (BATES; WATTS, 1988).

Mesmo com uma grande quantidade de métodos numéricos, a convergência não é garantida em nenhum deles, bem como, a possibilidade de convergência da estimativa

do parâmetro para valores fora do espaço paramétrico, o que não faria sentido, sendo de extrema importância, segundo Seber e Wild (1989), a escolha de uma função apropriada para explicar o conjunto de dados e um bom conjunto de valores iniciais para os parâmetros. Neste trabalho será apresentado o método de Gauss-Newton, utilizado na obtenção dos parâmetros do modelo.

2.6.1 Método de Gauss-Newton

O método de Gauss-Newton se utiliza de uma aproximação da função $f(\boldsymbol{\theta})$ em 6 em uma série de Taylor até primeira ordem em torno de um valor de parâmetros iniciais $\boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20}, \dots, \theta_{p0})$, em que p é a dimensão do espaço paramétrico, assim

$$f(\boldsymbol{\theta}) = f(\boldsymbol{\theta}_0) + \mathbf{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

em que

$$\mathbf{D}_0 = \left(\frac{\partial f(\boldsymbol{\theta}_0)}{\partial \theta_1}, \frac{\partial f(\boldsymbol{\theta}_0)}{\partial \theta_2}, \dots, \frac{\partial f(\boldsymbol{\theta}_0)}{\partial \theta_p} \right)$$

Escrevendo o resíduo $\mathbf{r}$ como função de $\boldsymbol{\theta}$, pode-se escrever

$$\begin{aligned} \mathbf{r}(\boldsymbol{\theta}) &= \mathbf{Y} - \mathbf{f}(\boldsymbol{\theta}) \\ &\approx \mathbf{r}(\boldsymbol{\theta}_0) - \mathbf{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \end{aligned} \tag{29}$$

A soma de quadrado do resíduo, $S(\boldsymbol{\theta})$, escrita com a aproximação 29

$$\begin{aligned} S(\boldsymbol{\theta}) &= [\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})]'[\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})] \\ &= \mathbf{r}(\boldsymbol{\theta}_0)' \mathbf{r}(\boldsymbol{\theta}_0) - 2\mathbf{r}(\boldsymbol{\theta}_0)' \mathbf{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{D}_0' \mathbf{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \end{aligned} \tag{30}$$

Dessa maneira, tomando a derivada parcial com relação a $\boldsymbol{\theta}$ para minimizar a soma de quadrado de resíduo

$$\boldsymbol{\theta} - \boldsymbol{\theta}_0 = [\mathbf{D}_0' \mathbf{D}_0]^{-1} \mathbf{D}_0' \mathbf{r}(\boldsymbol{\theta}_0) \tag{31}$$

Num processo de iteração, as primeiras q estimativas ficam

$$\begin{aligned}
\theta_1 &= \theta_0 + [D'_0 D_0]^{-1} D'_0 r(\theta_0) \\
\theta_2 &= \theta_1 + [D'_1 D_1]^{-1} D'_1 r(\theta_1) \\
&\vdots \\
\theta_q &= \theta_{q-1} + [D'_{q-1} D_{q-1}]^{-1} D'_{q-1} r(\theta_{q-1})
\end{aligned} \tag{32}$$

O processo iterativo acaba até se obter a convergência do modelo, na prática quando se observa uma diferença pré-determinada muito pequena entre θ_q e θ_{q-1} .

2.7 Igualdade de parâmetros em modelos não-lineares

Em posse dos parâmetros ajustados para cada unidade experimental, ou para um certo conjunto das mesmas (incluindo a informação das repetições nas mesmas condições experimentais), faz-se necessário a conjectura a respeito da igualdade entre as condições experimentais de interesse do pesquisador. Regazzi (2015) apresenta uma metodologia geral para o ajustamento de g equações de regressão não-linear, bem como uma metodologia para testar as seguintes hipóteses: *i*) H_0 : as g equações são idênticas, podendo uma única equação ser utilizada como estimativa nos g grupos; *ii*) H_0 : um determinado subconjunto é igual nos g grupos.

Considere as variáveis 'dummy' definidas como,

$$D_i = \begin{cases} 1, & \text{se a observação } y_{ij} \text{ pertence ao grupo } i \\ 0, & \text{caso contrário.} \end{cases} \tag{33}$$

Pode-se reescrever o modelo em 25, como

$$y_{ij} = \sum_{u=1}^g D_u \left[\frac{d_u}{1 + \exp(b_u(\log(x_{ij}) - \log(e_u)))} \right] + \epsilon_{ij} \tag{34}$$

em que y_{ij} é a massa de matéria seca observada após os 28 dias de aplicação de uma dose x_{ij} de herbicida glyphosate, e ϵ_{ij} é o erro de regressão e $\theta = (d_u, b_u, e_u)$ são os parâmetros do modelo log-logístico, D_u é a variável 'dummy' definida como em 33.

Lembrando que 28 pode ser reescrita como,

$$\hat{\sigma}^2 = \frac{1}{n} S(\hat{\theta}) \tag{35}$$

Utilizando o teste da razão de verossimilhança para testar as hipóteses for-

muladas tem-se que, de maneira geral o problema pode ser exposto da seguinte maneira:

Seja H_0 a hipótese nula em que $H_0 : \boldsymbol{\theta} \in \boldsymbol{w}$ e a hipótese alternativa $H_a : \boldsymbol{\theta}^c \in \boldsymbol{w}^c$ de tal forma que $\boldsymbol{\Theta} = \boldsymbol{w} \cup \boldsymbol{w}^c$ é o espaço paramétrico. A estatística do teste K de razão de verossimilhança é

$$K = \left(\frac{\hat{\sigma}_{\boldsymbol{\Theta}}^2}{\hat{\sigma}_{\boldsymbol{w}}^2} \right)^{n/2} \quad (36)$$

sendo $\hat{\sigma}_{\boldsymbol{\Theta}}^2$ a estimativa de máxima verossimilhança de σ^2 quando nenhuma restrição no espaço paramétrico é feita e $\hat{\sigma}_{\boldsymbol{w}}^2$ é a estimativa de σ^2 quando as restrições em H_0 são impostas no espaço paramétrico.

Nessas condições, pode-se utilizar o resultado assintótico para grandes amostras que diz que a distribuição $-2 \log(k)$ é aproximadamente qui-quadrado com ν graus de liberdade, sendo ν a diferença entre o número de parâmetros estimados em $\boldsymbol{\Theta}$ e o número de parâmetros estimados em $\boldsymbol{w}$, em símbolos

$$-2 \log(K) = -n \log \left(\frac{\hat{\sigma}_{\boldsymbol{\Theta}}^2}{\hat{\sigma}_{\boldsymbol{w}}^2} \right) \xrightarrow[n \rightarrow \infty]{d} \chi_{\nu}^2 \quad (37)$$

Um procedimento prático para a aplicação do método consiste em

- i) Ajustar o modelo completo $\boldsymbol{\Theta}$, obtendo as estimativas dos parâmetros e calcular $\hat{\sigma}_{\boldsymbol{\Theta}}^2$;
- ii) Ajustar o modelo restrito $\boldsymbol{w}$, obtendo as estimativas dos parâmetros e calcular $\hat{\sigma}_{\boldsymbol{w}}^2$;
- iii) Obter o número ν , dado por $p_{\boldsymbol{\Theta}} - p_{\boldsymbol{w}}$, em que $p_{\boldsymbol{\Theta}}$ é o número de parâmetros do modelo completo e $p_{\boldsymbol{w}}$ é o número de parâmetros do modelo restrito;
- iii) Obter a estatística do teste

$$K = -n \log \left(\frac{\hat{\sigma}_{\boldsymbol{\Theta}}^2}{\hat{\sigma}_{\boldsymbol{w}}^2} \right);$$

iv) Regra de decisão, tendo a estatística calculada e sabendo que a distribuição assintótica é a χ_{ν}^2 , basta prosseguir como, se

$$K \geq \chi_{tabela}^2$$

rejeita-se H_0 , caso contrário não se rejeita H_0 .

2.8 Análise de dados multivariados

A análise de dados multivariados, mais do que uma extensão da análise univariada, tem por objetivo extrair informação de maneira global não desprezando *a priori* nenhuma variável em estudo. Podendo-se observar, dentre outras, as relações entre as variáveis por meio de alguma estrutura de correlação, diminuição quando possível o número de variáveis, a quantificação da informação de modo a definir o que é importante e o que pode ser desconsiderado, o agrupamento por meio de algum critério definido, podendo-se classificar indivíduos sob estudo (MINGOTI, 2005).

Um conjunto de dados multivariados é representado por uma matriz $\mathbf{X}$ de dimensão $(n \times k)$ de modo que a convenção mais usual é a representação dos indivíduos ou objetos no espaço das linhas, assim $(X_{i1}, X_{i2}, \dots, X_{ik})$ são as medidas tomadas do indivíduo i .

De maneira explícita, considere a matriz de dados multivariados observada

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix} \quad (38)$$

Seja $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ ¹ um vetor de variáveis aleatórias p -dimensional, defini-se o vetor de média $\bar{X}$ amostral e a matriz de variância-covariância amostral $\mathbf{S}$ respectivamente, da seguinte forma

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n [X_{i1}, X_{i2}, \dots, X_{ip}]' = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{bmatrix} \quad (39)$$

¹Nota: não confundir com a matriz $\mathbf{X}$ que está em negrito

em que $s_{ij} = \frac{1}{n-1} \sum_{k=1}^p (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$ é a covariância entre as variáveis X_i e X_j .

A matriz de variância-covariância $\mathbf{S}$ é uma matriz simétrica ($s_{ij} = s_{ji}$) e que os elementos da diagonal são as variâncias amostrais usuais das variáveis consideradas, além do mais, ao menos que a matriz $\mathbf{X}$ não apresente multicolinearidade, ou seja, alguma coluna seja alguma combinação linear das demais, a matriz $\mathbf{S}$ será definida positiva, podendo dessa forma ser decomposta (RENCHEER; SCHAALJE, 2008).

2.8.1 Análise de variância multivariada

Segundo Rencher (2002) existem várias razões para se utilizar, quando possível, a análise multivariada, dentre elas pode-se citar: (i) a não alteração do nível de significância α do teste multivariado, enquanto que para cada teste univariado para as variáveis em questão há o aumento do erro tipo I (rejeição da hipótese nula dado que a hipótese nula é verdadeira); (ii) os testes multivariados levam em conta a matriz de covariância estimada das variáveis, enquanto que os testes univariados não incorporam essa informação no modelo; (iii) a significância no teste multivariado pode ser significativa enquanto que os testes univariados individuais podem não ser, aumentando dessa forma o poder do teste multivariado frente aos univariados.

2.8.2 Delineamento aleatório em blocos

A análise de variância multivariada (MANOVA) tem como objetivo comparar vetores de médias da mesma maneira que a análise de variância (ANOVA) visa comparar médias de tratamentos. Os princípios do planejamento de experimento na MANOVA são os mesmos da ANOVA, a saberem, (i) repetição; (ii) aleatorização; e (iii) controle local.

Com as repetições é possível obter uma estimativa para erro experimental, necessário para verificar se as diferenças observadas na variável resposta (ou vetor de variável resposta) podem ser consideradas devido ao acaso ou não.

O segundo princípio se faz importante para que na hipótese de haver algum tipo de fonte de variabilidade não controlada no experimento, a mesma apareça de maneira aleatória nas observações, diminuindo a probabilidade de, indevidamente, identificar uma mudança na variável resposta como sendo algo devido aos fatores considerados no estudo.

O controle local por sua vez, é importante quando há a hipótese de que as observações estão sendo influenciadas por algum fator que não seja de interesse primário do

pesquisador, nessa situação há a necessidade de retirar do resíduo essa variabilidade que não é devida aos fatores de interesse, para tanto a blocagem deve ser realizada afim de tornar as condições experimentais o mais homogênea possível dentro dos blocos e o mais heterogênea entre eles. Uma possível organização do conjunto de dados proveniente de um experimento multivariado no delineamento aleatorizados em blocos é apresentado na Tabela 2 .

Tabela 2 - Organização dos dados de um delineamento aleatorizado em blocos

		Bloco			
		1	2	...	b
	1	$\begin{pmatrix} Y_{111} \\ Y_{112} \\ \vdots \\ Y_{11p} \end{pmatrix}$	$\begin{pmatrix} Y_{121} \\ Y_{122} \\ \vdots \\ Y_{12p} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1b1} \\ Y_{1b2} \\ \vdots \\ Y_{1bp} \end{pmatrix}$
		$\begin{pmatrix} Y_{211} \\ Y_{212} \\ \vdots \\ Y_{21p} \end{pmatrix}$	$\begin{pmatrix} Y_{221} \\ Y_{222} \\ \vdots \\ Y_{22p} \end{pmatrix}$		$\begin{pmatrix} Y_{2b1} \\ Y_{2b2} \\ \vdots \\ Y_{2bp} \end{pmatrix}$
		$\vdots$	$\vdots$		$\vdots$
		$\begin{pmatrix} Y_{k11} \\ Y_{k12} \\ \vdots \\ Y_{k1p} \end{pmatrix}$	$\begin{pmatrix} Y_{k21} \\ Y_{k22} \\ \vdots \\ Y_{k2p} \end{pmatrix}$		$\begin{pmatrix} Y_{kb1} \\ Y_{kb2} \\ \vdots \\ Y_{kbp} \end{pmatrix}$
Tratamento	2				

Em busca de generalizar a ideia univariada de um delineamento aleatorizado em blocos considere o seguinte modelo multivariado :

$$Y_{ijr} = \mu_r + t_{ir} + b_{jr} + \epsilon_{ijr} \quad (40)$$

em que $i = 1, \dots, k$ sendo k o número de tratamentos; $j = 1, \dots, b$, b é o número de blocos; $r = 1, \dots, p$, p é o número de variáveis que compõem $\mathbf{Y}$; μ_r é a média geral da r -ésima variável; t_{ir} é o efeito do i -ésimo tratamento na r -ésima variável; b_{jr} é o efeito do j -ésimo bloco na r -ésima variável; e ϵ_{ijr} é o efeito aleatório associado à observação Y_{ijr} , a função de distribuição que se assume para ϵ_{ij} é a normal p -variada $N_p(0, \Sigma)$, em que Σ é a matriz de variância-covariância. A função de distribuição de probabilidade para ϵ_{ij} pode ser escrita como

$$f_{\epsilon_{ij}}(\epsilon_{ij}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \epsilon'_{ij} \Sigma^{-1} \epsilon_{ij}\right) \quad (41)$$

em que $\epsilon'_{ij} = (\epsilon_{ij1}, \epsilon_{ij2}, \dots, \epsilon_{ijp})$.

Escrito de maneira matricial o modelo 40 fica

$$\underbrace{\begin{pmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ijp} \end{pmatrix}}_{\mathbf{Y}_{ij}} = \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{pmatrix} t_{ij1} \\ t_{ij2} \\ \vdots \\ t_{ijp} \end{pmatrix}}_{\mathbf{t}_{ij}} + \underbrace{\begin{pmatrix} b_{ij1} \\ b_{ij2} \\ \vdots \\ b_{ijp} \end{pmatrix}}_{\mathbf{b}_{ij}} + \underbrace{\begin{pmatrix} \epsilon_{ij1} \\ \epsilon_{ij2} \\ \vdots \\ \epsilon_{ijp} \end{pmatrix}}_{\boldsymbol{\epsilon}_{ij}} \quad (42)$$

2.8.3 Decomposição da variabilidade total

De acordo com Rencher (2002), utiliza-se a seguinte nomenclatura para os objetos de interesse da análise estatística multivariada, o vetor de médias geral ($\bar{\mathbf{Y}}_{..}$)

$$\bar{\mathbf{Y}}_{..} = \frac{1}{kb} \sum_{i=1}^k \sum_{j=1}^b \bar{\mathbf{Y}}_{ij} = \begin{pmatrix} \bar{Y}_{..1} \\ \bar{Y}_{..2} \\ \vdots \\ \bar{Y}_{..p} \end{pmatrix}$$

o vetor de médias de tratamento ($\bar{\mathbf{Y}}_{i.}$)

$$\bar{\mathbf{Y}}_{i.} = \frac{1}{b} \sum_{j=1}^b \bar{\mathbf{Y}}_{ij} = \begin{pmatrix} \bar{Y}_{i.1} \\ \bar{Y}_{i.j2} \\ \vdots \\ \bar{Y}_{i.jp} \end{pmatrix}$$

o vetor de médias de bloco ($\bar{\mathbf{Y}}_{.j}$)

$$\bar{\mathbf{Y}}_{.j} = \frac{1}{k} \sum_{i=1}^k \bar{\mathbf{Y}}_{ij} = \begin{pmatrix} \bar{Y}_{.j1} \\ \bar{Y}_{.j2} \\ \vdots \\ \bar{Y}_{.jp} \end{pmatrix}$$

De maneira geral a variabilidade total do conjunto de dados pode ser escrita como

$$\mathbf{T} = \sum_{i=1}^k \sum_{j=1}^b (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})' \quad (43)$$

Somando e subtraindo os termos $\bar{\mathbf{Y}}_{.j}$, $\bar{\mathbf{Y}}_{i.}$ e $\bar{\mathbf{Y}}_{..}$ em 43

$$\begin{aligned} & \sum_{i=1}^k \sum_{j=1}^b (\mathbf{Y}_{ij} + \bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{i.} + \bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{..} + 2\bar{\mathbf{Y}}_{..}) \\ & (\mathbf{Y}_{ij} + \bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{i.} + \bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{.j} + \bar{\mathbf{Y}}_{..} - 2\bar{\mathbf{Y}}_{..})'. \end{aligned} \quad (44)$$

Rearranjando os termos de maneira a construir uma partição ortogonal da soma de quadrado e produto cruzado total

$$\begin{aligned} & \sum_{i=1}^k \sum_{j=1}^b (\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..} + \bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{..} + \bar{\mathbf{Y}}_{ij} - \bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{.j} + \mathbf{Y}_{ij}) \\ & (\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..} + \bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{..} + \bar{\mathbf{Y}}_{ij} - \bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{.j} + \mathbf{Y}_{ij})'. \end{aligned}$$

$$\begin{aligned} & = \underbrace{\sum_{i=1}^k \sum_{j=1}^b (\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})'}_H + \underbrace{\sum_{i=1}^k \sum_{j=1}^b (\bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{..})(\bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{..})'}_B + \\ & \underbrace{\sum_{i=1}^k \sum_{j=1}^b (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{.j} + \bar{\mathbf{Y}}_{ij})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{.j} + \mathbf{Y}_{ij})'}_E. \end{aligned} \quad (45)$$

Uma vez que a somatória dos produtos cruzados entre os três termos $(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})$, $(\bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{..})$ e $(\bar{\mathbf{Y}}_{ij} - \bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{.j} + \bar{\mathbf{Y}}_{..})$ é nula, por exemplo,

$$\begin{aligned} & \sum_{i=1}^k \sum_{j=1}^b (\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})(\bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{..})' \\ &= \sum_{i=1}^k (\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..}) \sum_{j=1}^b (\bar{\mathbf{Y}}_{.j} - \bar{\mathbf{Y}}_{..})' \\ &= \mathbf{0}_{(p \times p)} \end{aligned} \quad (46)$$

pois a somatória em i é o vetor nulo coluna p -variado e a somatória em j é o vetor nulo linha p -variado.

Dessa maneira pode-se escrever

$$\mathbf{T} = \mathbf{H} + \mathbf{B} + \mathbf{E} \quad (47)$$

em que $\mathbf{T}$ é a matriz da soma de quadrados e produtos cruzados total; $\mathbf{H}$ é a matriz da soma de quadrados e produtos cruzados de tratamento; $\mathbf{B}$ é a matriz da soma de quadrados e produtos cruzados de bloco e $\mathbf{E}$ é a matriz da soma de quadrados e produtos cruzados de resíduo.

2.8.4 Análise estatística

A ideia em 47 está em separar a causa de variação por parte do que pode e não pode ser controlado, o objetivo está em se verificar se tais variações podem ser consideradas devidas ao tratamento ou não. Os graus de liberdade de cada elemento da partição em 47 é bastante similar à decomposição univariada, a Tabela 3 apresenta os graus de liberdade dessa partição e MSQPC é a matriz de soma de quadrados e produtos cruzados.

Tabela 3 - Esquema de análise de variância multivariada de um delineamento aleatorizado em blocos

Fonte variação	GL	MSQPC
blocos	b-1	$\mathbf{B}$
tratamento	t-1	$\mathbf{H}$
resíduo	(b-1)(t-1)	$\mathbf{E}$
total	bt-1	$\mathbf{T}$

Tomando

$$\boldsymbol{\mu}_i = +\boldsymbol{\mu} + \boldsymbol{t}_i \quad (48)$$

pode-se fazer a seguinte hipótese

$$\begin{aligned} H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_k \\ H_1 : \text{pelo menos um contraste é significativo} \end{aligned} \quad (49)$$

A hipótese em 49 pode ser testada por várias estatísticas, a saber as mais utilizadas são: (i) Teste lambda de Wilk; (ii) Teste da maior raiz característica de Roy; (iii) Teste de Pillai; e (iv) Teste de Lawley - Hotelling.

No teste de Wilk a estatística é

$$\Lambda = \frac{|\boldsymbol{E}|}{|\boldsymbol{H} + \boldsymbol{E}|} \quad (50)$$

em que o numerador é o determinante da matriz de quadrados e produtos cruzados de resíduo e o denominador é o determinante da soma da matriz de quadrados e produtos cruzados de tratamento e de resíduo. Rejeita-se H_0 se $\Lambda < \Lambda_{(\alpha, p, v_{\boldsymbol{H}}, v_{\boldsymbol{E}})}$, rejeitando H_0 para valores pequenos de Λ , sendo p o número de variáveis, $v_{\boldsymbol{H}}$ o número de graus de liberdade para o tratamento, $v_{\boldsymbol{E}}$ o número de graus de liberdade para o resíduo. A estatística Λ de Wilks possui as seguintes propriedades (RENCHE, 2002):

- (i) Para os determinantes em 50 serem positivos, é necessário que $v_{\boldsymbol{E}} \geq p$;
- (ii) Para qualquer modelo MANOVA, os graus de liberdade $v_{\boldsymbol{H}}$ e $v_{\boldsymbol{E}}$ serão sempre os mesmos em analogia com o modelo univariado;
- (iii) Os parâmetros p e $v_{\boldsymbol{H}}$ podem ser trocados de modo que a distribuição de $\Lambda_{(p, v_{\boldsymbol{H}}, v_{\boldsymbol{E}})}$ é a mesma que $\Lambda_{(v_{\boldsymbol{H}}, p, v_{\boldsymbol{E}} + v_{\boldsymbol{H}} - p)}$;
- (iv) A expressão em 50 pode ser escrita em termos dos autovalores $\lambda_1, \lambda_2, \dots, \lambda_s$ de $\boldsymbol{E}^{-1}\boldsymbol{H}$ como

$$\Lambda = \prod_{i=1}^s \frac{1}{1 + \lambda_i} \quad (51)$$

em que $s = \min(p, v_{\boldsymbol{H}})$

- (v) Λ varia dentro do intervalo $[0, 1]$ e o teste de Wilk é um teste invertido no sentido em que rejeita-se H_0 para valores pequenos de Λ , assim se o vetor $(\bar{\boldsymbol{Y}}_{i.} - \bar{\boldsymbol{Y}}_{..})$ fosse nulo. $\boldsymbol{H} = \mathbf{0}$, e $\Lambda = |\boldsymbol{E}|/|\boldsymbol{E} + \mathbf{0}| = 1$. Por outro lado, se houvesse muita diferença entre

$(\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}}_{..})$, $|\mathbf{E} + \mathbf{H}|$ ficaria muito grande e Λ tenderia a 0.

(vi) Os valores tabelados de Λ diminuem com o aumento do número de variáveis p , assim o aumento de variáveis reduz a capacidade de encontrar diferença entre os tratamentos a não ser que a variável contribua para a rejeição da hipótese nula produzindo diminuição significativa em Λ .

(vii) O teste Λ de Wilk se transforma num teste F exato para alguns valores de $v_{\mathbf{H}}$ e p , tais valores são apresentados na tabela 4

Tabela 4 - Transformações da distribuição Λ de Wilk para a distribuição F

parâmetros p e $v_{\mathbf{H}}$	estatística tendo distribuição F	graus de liberdade
qualquer p e $v_{\mathbf{H}} = 1$	$\frac{1-\Lambda}{\Lambda} \frac{v_{\mathbf{E}}-p+1}{p}$	$p, v_{\mathbf{E}} - p + 1$
qualquer p e $v_{\mathbf{H}} = 2$	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{v_{\mathbf{E}}-p+1}{p}$	$2p, 2(v_{\mathbf{E}} - p + 1)$
$p = 1$ e qualquer $v_{\mathbf{H}}$	$\frac{1-\Lambda}{\Lambda} \frac{v_{\mathbf{E}}}{v_{\mathbf{H}}}$	$v_{\mathbf{H}}, v_{\mathbf{E}}$
$p = 2$ e qualquer $v_{\mathbf{H}}$	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{v_{\mathbf{E}}-1}{v_{\mathbf{H}}}$	$2v_{\mathbf{H}}, 2(v_{\mathbf{E}} - 1)$

(viii) Para outros valores de p e $v_{\mathbf{H}}$ que não estão na tabela 4 uma aproximação para a estatística F é dada por:

$$F = \frac{1-\Lambda^{1/t}}{\Lambda^{1/t}} \frac{df_2}{df_1} \quad (52)$$

em que df_1 e df_2 são

$$\begin{aligned} df_1 &= pv_{\mathbf{H}}, & df_2 &= wt - \frac{1}{2}(pv_{\mathbf{H}} - 2), \\ w &= v_{\mathbf{E}} - v_{\mathbf{H}} + \frac{1}{2}(p + v_{\mathbf{H}} + 1), & t &= \sqrt{\frac{p^2 v_{\mathbf{H}}^2 - 4}{p^2 + v_{\mathbf{H}}^2 - 5}} \end{aligned} \quad (53)$$

O Teste da maior raiz característica de Roy considera a seguinte estatística

$$\theta = \frac{\lambda_1}{1+\lambda_1} \quad (54)$$

em que λ_1 é o maior autovalor da matriz $\mathbf{E}^{-1}\mathbf{H}$. Se $\theta \geq \theta_{\alpha,s,m,N}$ rejeita-se H_0 sendo α o nível de significância do teste $s = \min(v_{\mathbf{H}}, p)$, $p = \frac{1}{2}(|v_{\mathbf{H}} - p| - 1)$ e $N = \frac{1}{2}(v_{\mathbf{E}} - p - 1)$, o valor de $\theta_{\alpha,s,m,N}$ encontra-se tabelado.

O teste de Pillai é uma generalização do teste de Roy e considera a seguinte estatística

$$V^s = tr[(\mathbf{E} + \mathbf{H})^{-1}\mathbf{H}] = \sum_{i=1}^s \frac{\lambda_i}{1+\lambda_i} \quad (55)$$

que leva em conta o número total de autovalores (s) da matriz $\mathbf{E}^{-1}\mathbf{H}$, e de maneira análoga ao teste de Roy, se $V^s \geq V_\alpha$ rejeita-se H_0 .

O teste de Lawley-Hotelling baseia-se na estatística

$$U^s = tr[(\mathbf{E}^{-1}\mathbf{H})] = \sum_{i=1}^s \lambda_i \quad (56)$$

rejeita-se H_0 se a estatística U_s for maior que valor tabelado Rencher (2002).

As pressuposições do modelo para o delineamento aleatorizado em blocos são:

(i) O vetor de erros $\boldsymbol{\epsilon}_{ij}$ aleatórios possui vetor de médias igual a $\mathbf{0}$; (ii) apresentam uma mesma matriz $\boldsymbol{\Sigma}$ de variância-covariância; (iii) $\boldsymbol{\epsilon}_{ij}$ são independentes (iv) provém de uma distribuição normal p-variada; (v) não há interação entre blocos e os tratamentos, o que significa dizer que o efeito de tratamento é considerado o mesmo para todos os blocos.

3 MATERIAL

3.1 Obtenção das sementes, produções das mudas e aplicação do tratamento

Os dados utilizados neste trabalho são provenientes de um experimento realizado em Piracicaba-SP, na casa de vegetação do departamento de Produção Vegetal da ESALQ-USP, as sementes de capim-pé-de-galinha *Eleusine indica* utilizadas nesse experimento provem de quatro localidades do território nacional e estão descritas na tabela 5.

A princípio, foram dadas às populações os nomes de resistente ou suscetível de acordo com a suspeita devido ao histórico de ter havido método de controle por meio de glyphosate em suas respectivas localidades. As quatro amostras de sementes foram então, designadas como EIR1, EIR2, EIR3 e EIS, havendo apenas uma amostra com suspeita de susceptibilidade, as três primeiras amostras de sementes (EIR1, EIR2 e EIR3) foram coletadas em regiões com histórico de controle sistemático da planta pé-de-galinha por meio de aplicação de herbicida glyphosate, sendo as duas primeiras em região agrícola e a terceira de região urbana, houve o cuidado de se coletar as sementes das plantas isentas de aplicação de herbicida. Na Tabela 5 descreve-se as cidades e os estados das quatro amostras, feita a coleta, as sementes foram secas, guardadas em sacos de papel etiquetados e armazenados em câmaras frias (ROSA, 2014)

Tabela 5 - Espécie, origem, identificação (sigla) e suspeita de resistência ao glyphosate

Espécie	Origem (Cidade-Estado)	Sigla	Suspeita
<i>Eleusine indica</i>	Guarapuava-PR	EIR1	Resistente
<i>Eleusine indica</i>	Primavera do Leste-MT	EIR2	Resistente
<i>Eleusine indica</i>	Piracicaba-SP	EIR3	Resistente
<i>Eleusine indica</i>	Pedra Preta-MT	EIS	Suscetível

Para a produção das mudas foram utilizadas as amostras das sementes coletadas, essas sementes foram plantadas em bandejas plásticas, preenchidas com substrato comercial e mantidas sob condição de casa de vegetação. Atingindo de 1-2 folhas definiti-

vas, as mudas foram transferidas para novas bandejas plásticas, sendo deixado apenas uma planta por vaso.

O delineamento experimental utilizado foi o inteiramente aleatorizado, sendo o estágio de desenvolvimento da planta a variável de blocagem, foi considerado apenas um fator para explicar a variável resposta y , o herbicida, e os níveis foram as doses do mesmo que variaram igualmente de 1/16 D a 16 D mais a testemunha, sem aplicação de herbicidas, totalizando dez níveis, houve quatro repetições para cada nível. No experimento, D foi diferente para cada estágio, sendo D igual a 480 gramas de equivalente ácido de glyphosate por hectare (g .e a. ha⁻¹) para o estágio de 2 a 3 perfilhos, 720 g .e a. ha⁻¹ para o estágio de 6 a 8 perfilhos e de 960 g para o estágio de 10-12 perfilhos.

Chegando nos estágios pré-determinados (de 2 a 3 perfilhos, de 6 a 8 perfilhos e de 10-12 perfilhos), foram feitas pulverizações com as doses específicas a uma pressão de 23 lb pol⁻² gerando um volume de aplicação proporcional a 150 L ha⁻¹. Passado 28 dias após aplicação, foi coletado a parte aérea da planta e posteriormente secas em estufa de circulação forçada de ar a 72°C por 72 horas.

Na Figura 4 podemos observar todos os níveis do fator sendo aplicados nas unidades experimentais (vasos) com 4 repetições por nível, na variedade com suspeita suscetível no estágio de 10 a 12 perfilhos.

Figura 4 - Foto do estágio de 10 a 12 perfilhos da variedade de suspeita suscetível após 28 dias da aplicação do herbicida

4 RESULTADOS E DISCUSSÃO

4.1 Análise exploratória

O conjunto de dados é constituído de 480 observações, essas observações são provenientes de três experimentos realizados no delineamento inteiramente aleatorizado, em cada delineamento foi considerado como tratamento a variedade provinda das localidades listadas na Tabela 5, para cada variedade foi realizado 4 repetições. Nas Figuras 5, 6, 7 são apresentadas todas as medidas realizadas no experimento em função da dose administrada.

Figura 5 - Evolução da massa seca em função da dose de herbicida no estágio em que a planta apresenta de dois a três perfilhos

Figura 6 - Evolução da massa seca em função da dose de herbicida no estágio em que a planta apresenta de seis a oito perfilhos

Figura 7 - Evolução da massa seca em função da dose de herbicida no estágio em que a planta apresenta de dez a doze perfilhos

Em todas as observações constata-se um comportamento geral de diminuição da matéria seca com o aumento da concentração de herbicida, contudo em algumas medidas observa-se que, não raro, esse padrão não é verificado quando olhado de maneira mais pontual, ou seja, por algum motivo, pode-se dizer que algum fator não controlado no experimento está influenciando a variável resposta, podendo ser observado o comportamento descrito por Knezevic, S. e Ritz (2007).

Na Figura 8 o gráfico boxplot ou diagrama de caixa indica, como nos gráficos de perfil, que alguma fonte de variação no experimento foi não controlada, uma vez que as caldas no diagrama de caixa em todos os perfis estão grandes demais em relação à média, outro fator importante é que tais caldas, aparentemente variam de acordo com a dose de herbicida, indicando dessa maneira uma possível heterogeneidade de variância. Destaca-se ainda a presença em todos os perfis analisados de pontos atípicos detectados pelo diagrama de caixa.

Figura 8 - Diagrama de caixa da massa-seca em função da dose de herbicida

4.2 Análise de regressão não-linear

O modelo para explicar o fenômeno abordado no experimento descrito no capítulo 3 foi proposto por Knezevic, S. e Ritz (2007) e utilizado em inúmeros trabalhos na área de controle de ervas daninhas por meio de controle químico como Brunharo (2014) e Rosa (2014).

Para a análise de regressão não-linear os números de perfis foram considerados independentemente, sendo realizado para cada grupo de perfis uma regressão não-linear. No estágio das plantas de 2 a 3 perfis, as observações 47, 48, 57, 78 apareceram como outliers na Figura 8 contudo na Figura 9 as observações 32, 33, 62 e 71 se destacam com resíduo estudatizado internamente maiores que 3, sendo todas essas observações substituídas pela média das respectivas observações restantes.

Uma justificativa para tal atitude está no fato de que a análise de resíduo realizada nos modelos lineares podem não ser apropriada para modelos não-lineares uma vez que a métrica muda de acordo com a vizinhança em que se está verificando o resíduo (COOK; TSAI, 1985), dessa forma pode-se penalizar as observações com baixa magnitude, não identificando dessa maneira observações pequenas como outlier ou identificando indevidamente como outlier observações de valores grandes, dessa maneira optou-se por utilizar um método gráfico e outro como limite de resíduo estudentizado igual a 3.

Figura 9 - Análise de resíduo *versus* observação nos estágios em que a planta apresentava entre 2 a 3 perfilhos

As estimativas dos parâmetros para o estágio das plantas de 2 a 3 perfilhos são apresentadas na Tabela 6, observa-se que para a maioria dos parâmetros o método utilizado apresentou estimativas consistentes, apresentando o erro padronizado das estimativas (EPR) muito menores que os valores das estimativas e valor-p significativos para todas as estimativas, com exceção da estimativa do parâmetro b para EIS que apresentou valor-p igual a 0,4.

Tabela 6 - Estimativa dos parâmetros do modelo no estágio de 2-3 perfilhos

Variedade	2-3 perfilhos					
	$\hat{d}$	EPE	$\hat{b}$	EPE	$\hat{e}$	EPE
EIR1	5,20	0,21	1,61	0,21	168,08	17,95
EIR2	5,91	0,20	1,49	0,17	219,19	21,42
EIR3	3,83	0,25	0,90	0,15	62,63	14,62
EIS	2,45	0,13	5,96	7,18	242,71	17,71

De maneira análoga para os demais estágios, nas figuras 10 e 11 são apresentados os gráficos de resíduo padronizado internamente *versus* observação e valor ajustado.

Figura 10 - Análise de resíduo *versus* observação nos estágios em que a planta apresentava entre 6 a 8 perfilhos

Figura 11 - Análise de resíduo *versus* observação nos estágios em que a planta apresentava entre 10 a 12 perfilhos

Utilizando o mesmo critério para as observações no estágio de 6 a 8 perfilhos, foram detectadas apenas uma observações com resíduo maior que 3, a observação 72, a mesma aparece no gráfico na Figura 8 e é substituídas pela média das observações restantes. No estágio de de 10 a 12 perfilhos, a observação 130 foi identificada como outlier em 8 e as observações com resíduo estudentizado internamente maior que 3 foram as observações 41 e 125.

Na Tabela 7 encontra-se as estimativas dos parâmetros para o estágio de 6 a 8 perfilhos. Por meio do teste t, a um nível de significância de 5% apenas a os parâmetros e_1 e e_4 não foram significativos para explicar a variável resposta. valor-p respectivamente, 0,13 e 0,11. Na Tabela 8 encontra-se as estimativas do modelo para o estágio de 10 a 12 perfilhos, todos os parâmetros foram significativos.

Tabela 7 - Estimativa dos parâmetros do modelo no estágio de 6-8 perfilhos

Variedade	6-8 perfilhos					
	$\hat{d}$	EPE	$\hat{b}$	EPE	$\hat{e}$	EPE
EIR1	4,66	0,40	0,64	0,20	19,52	12,97
EIR2	5,65	0,27	1,82	0,33	318,64	39,30
EIR3	4,00	0,34	0,69	0,15	108,98	42,47
EIS	3,95	0,36	0,49	0,12	50,18	31,04

Tabela 8 - Estimativa dos parâmetros do modelo no estágio de 10-12 perfilhos

Variedade	10-12 perfilhos					
	$\hat{d}$	EPE	$\hat{b}$	EPE	$\hat{e}$	EPE
EIR1	24,50	0,79	3,30	0,475	119,5	5,6
EIR2	17,90	0,51	3,00	0,52	494,39	30,85
EIR3	15,77	0,61	5,67	2,57	213,67	14,85
EIS	14,62	0,84	0,64	0,08	416,73	110,16

Em posse das estimativas dos parâmetros e se utilizando da interpretação do parâmetro e , pode-se calcular o valor da dose de herbicida que reduz em dada porcentagem a massa seca, lembrando que $y(0) = d$ sendo a quantidade matéria seca c e a dose correspondente à matéria seca de x tem-se que,

$$c = y(x) = \frac{d}{1 + \exp(b \ln(x) + b \ln(e))}$$

Isolando x

$$c = \frac{d}{1 + \exp(b \ln(\frac{x}{e}))} = \frac{d}{1 + \exp(b \ln(\frac{x}{e}))} = \frac{d}{1 + (\frac{x}{e})^b} \rightarrow$$

$$\frac{c}{d} + \frac{c}{d} \left(\frac{x}{e}\right)^b = 1 \rightarrow \left(\frac{x}{e}\right)^b = \frac{d-c}{c} \rightarrow \quad (57)$$

$$\left(\frac{x}{e}\right) = \left(\frac{d-c}{c}\right)^{1/b} \rightarrow x = e \left(\frac{d-c}{c}\right)^{1/b}.$$

Vale notar que, como dito anteriormente, a equação 57 calculada para $c = d/2$,

Tabela 9 - Valores previstos pelo modelo log-logístico de dose de herbicida que acarretam a diminuição de 90% e 95% da matéria seca do capim-pé-de-galinha após de 28 dias da aplicação

Variedade	2-3 perfilhos		6-8 perfilhos		10-12 perfilhos	
	90%	95%	90%	95%	90%	95%
EIR1	657,29	1045,11	422,14	1393,57	232,87	292,11
EIR2	951,37	1567,31	1156,18	1775,35	1029,91	1321,87
EIR3	722,30	1658,99	3836,66	11489,44	314,84	359,20
EIS	350,80	397,61	7754,61	37537,58	12679,63	40505,84

ou seja, o valor da dose que diminui em 50% a matéria seca, x fica sendo igual ao próprio parâmetro e . Os valores que reduzem 90% e 95% são apresentados na Tabela 9

As doses recomendadas de herbicida glyphosate para o controle de *Eleusine indica* variam entre 360 a 1440 g e.a. ha⁻¹ no Brasil (MUELLER et al., 2011), ao analisar as estimativas da Tabela 9, pode-se afirmar que para o estágio de 2 a 3 perfilhos há uma boa concordância com o que é feito no que se refere ao controle com as dose mínima e dose máxima praticadas, considerando, que as amostras coletadas representem a população de *Eleusine indica* no Brasil. Observa-se ainda que para os estágios de 6 a 8 e de 10 a 12 valores mais extremos, segundo Rosa (2014) o método que avalia a matéria seca como função da dose passa a superestimar a dose necessária para matar a planta.

Continuando a abordagem, uma das hipóteses a ser analisada, foi se a variável considerada suscetível é de fato suscetível ou não para todos os estágios da planta considerados na aplicação do herbicida. O modelo utilizado foi

$$y_{ij} = \sum_{u=1}^g D_u \left[\frac{d_u}{1 + \exp(b_u(\log(x_{ij}) - \log(e_u)))} \right] + \epsilon_{ij} \quad (58)$$

em que D_u é a variável 'dummy', $i = 1, 2, 3, 4$, $j = 1, \dots, 40$, com número total de observações igual a 160 por número de perfilhos.

Realizando o teste da razão de verossimilhança para verificar a igualdade do parâmetro e para os três estágios o teste encontrou diferença significativa dentro de todos os estágios (os valores-p correspondentes foram todos menores que 0,01). Com esse resultado a análise prosseguiu testando as combinações dois a dois obtendo-se as Tabelas 10 e 11.

Tabela 10 - Análise da igualdade dos parâmetros em cada variedade dentro dos três estágios considerados

2-3 perfilhos			
EIS	a		
EIR2	a	b	
EIR1		b	c
EIR3			d
6-8 perfilhos			
EIR2	a		
EIR3		b	
EIS		b	c
EIR1			c
10-12 perfilhos			
EIR2	a		
EIS		b	
EIR3		b	c
EIR1			c

Tabela 11 - Valores p calculados

2-3 perfilhos				
	EIS	EIR2	EIR1	EIR3
EIS	-	0,35	0,01	-
EIR2		-	0,14	4e-3
EIR1			-	1,8e-4
EIR3				-
6-8 perfilhos				
	EIR2	EIR3	EIS	EIR1
EIR2	-	8,8e-3	-	-
EIR3		-	0,18	0,01
EIS			-	0,31
EIR1				-
10-12 perfilhos				
	EIR2	EIS	EIR3	EIR1
EIR2	-	0,49	4,4e-12	-
EIS		-	0,07	2,94e-6
EIR3			-	9,65e-10
EIR1				-

Na Tabela 10 as letras iguais indicam que, ao nível de significância de 5%, o parâmetro e do modelo é considerado igual para as variedades analisadas, por outro lado, para letras diferentes, as variedades não apresentam o mesmo valor do parâmetro.

A variedade considerada suscetível, ao contrário do esperado, demonstrou-se bastante resistente, principalmente no estágio avaliado de 2 a 3 perfilhos, mas a variedade mais resistente, de maneira geral, que esteve nos três estágios os maiores valores do parâmetro relacionado com a diminuição da matéria seca foi a variedade coletada na cidade de Primavera do Leste-MT.

Na Tabela 11 para os valores-p maiores que o nível de significância adotado ($\alpha = 0,05$), os parâmetros são considerados iguais, e para os valores-p menores, os parâmetros são considerados diferentes, com esses valores foi possível fazer a classificação da Tabela 10 de maneira a facilitar a visualização.

O coeficiente de determinação não corrigido pela média foi utilizado como ferramenta para se avaliar o ajuste do modelo (SOUZA, 1998), os valores calculados foram, para o estágio de 2 a 3 perfilhos, 0,93, para o estágio de 6 a 8, 0,82 e para o estágio de 10 a 12, 0,94, mostrando um ajuste satisfatório. A implementação utilizando o software R foi realizada como descrito em (ZEVIANE, 2012)

4.3 Análise de variância multivariada

A partir dos parâmetros estimados e se utilizando de suas interpretações, optou-se por utilizá-los como variável resposta em uma análise de variância multivariada, com o objetivo de verificar se as novas variáveis criadas são, conjuntamente explicadas pelas variedades de capim-pé-de-galinha. A Tabela 12 refere-se a análise variância multivariada nos três estágios.

Tabela 12 - Análise de variância multivariada

2-3 perfilhos			
	GL	Estatística Λ de Wilk	Valor-p
Variedade	3	2,4 e-2	3,1 e-6
Resíduo	12		
6-8 perfilhos			
	GL	Estatística Λ de Wilk	Valor-p
Variedade	3	6,5 e-2	3,2 e-4
Resíduo	12		
10-12 perfilhos			
	GL	Estatística Λ de Wilk	Valor-p
Variedade	3	3,0 e-3	1,6 e-10
Resíduo	12		

Observa-se na Tabela 12 que em todos os estágios a variedade considerada como fator e com quatro níveis foi significativa para explicação conjunta do vetor de variáveis respostas consideradas como os parâmetros estimados $(\hat{d}, \hat{b}, \hat{e})$. Os testes de Pillai, Roy e Lawley-Hotelling também tiveram valor-p menores que 0,01.

Os testes para igualdade de covariância e para normalidade multivariada são

importantes para se verificar o ajuste do modelo, na Tabela 13 são colocados os valores obtidos para o teste de normalidade multivariada de Shapiro-Wilk e o teste para igualdade de matriz M-Box.

Tabela 13 - Valores-p calculados para os testes de Shapiro-Wilk e M-Box

Teste	Valor-p		
	2-3 perfilhos	6-8 perfilhos	10-12 perfilhos
Shapiro-Wilk	0,07	0,08	5,7 e-3
M-Box	0,06	0,05	1,4 e-4

Observa-se que apenas os estágios de 10 a 12 perfilhos foram rejeitados (valor- $p < 0,05$), tanto para normalidade multivariada quanto para a igualdade da matriz de variância-covariância, dessa maneira para o estágios restantes o ajuste do modelo foi considerado satisfatório, de modo a aumentar a veracidade das decisões tomadas nos testes de hipóteses. O último teste realizado foi o teste de Tukey para igualdade de médias para os parâmetros e . Os resultados estão na Tabela 14

Tabela 14 - Teste de Tukey para o parâmetro e nos os três estágios avaliados

2-3 perfilhos	
EIS	a
EIR2	a b
EIR1	a b
EIR3	b
6-8 perfilhos	
EIR2	a
EIR3	b
EIS	b
EIR1	b
10-12 perfilhos	
EIR2	a
EIS	a
EIR3	b
EIR1	b

De maneira análoga à Tabela 10, a Tabela 14 indica quais os valores foram considerados iguais pelo teste de Tukey, observa-se que por esse teste em oposição ao teste de máxima verossimilhança foi detectado menos diferença entre as variedades analisadas. Dessa maneira, nos estágios de 2 a 3 perfilhos, por exemplo, as 3 variedades EIS, EIR2 e EIR1 foram consideradas iguais com relação as variedades mais resistente ao glyphosate, enquanto que pelo teste de máxima verossimilhança as variedades mais resistentes e consideradas iguais foram apenas EIS e EIR2. Embora os resultados quanto a identificação de igualdade dos parâmetros não foram idênticos, Tabelas 4.5 e 4.9, não há *a priori*, preferência sobre algum desses, sendo importante verificar as pressuposições exigidas para cada método. 5

5 CONCLUSÃO

Os dois métodos apresentados para verificar diferença entre qualquer subconjunto de parâmetros do modelo são gerais e podem ser utilizados em qualquer modelo não-linear com erros normais, contudo a convergência das estimativas dos parâmetros não é garantida. Para tal problema não acontecer e impossibilitar a análise estatística, deve-se trabalhar com um modelo que consiga explicar o conjunto de dados necessitando ter algum conhecimento prévio sobre o modelo e o fenômeno em estudo.

A ocorrência de dados atípicos pode afetar as estimativas dos parâmetros, ou também, implicar a não convergência da mesmas. A escolha dos valores iniciais para o processo iterativo de obtenção das estimativas também podem acarretar a não convergência das estimativas.

A hipótese levantada de que ao longo dos anos o uso indiscriminado de herbicida estaria selecionando variedades resistentes não se verificou com o conjunto de dados analisado. A variedade, *a priori* denominada suscetível apresentou o maior valor para o parâmetro e no estágio de 2 a 3 perfilhos, contudo, apesar do resultado, a variedade EIR2 sempre apresentou os maiores valores do mesmo parâmetro, independentemente do estágio analisado, o que de fato sugere que a variedade EIR2 e a variedade EIS são mais resistente quando comparada às variedades ERI1 e ERI3.

Por meio do modelo log-logístico foi possível calcular a dose que reduz a quantidade de matéria seca em 50%, 90% e 95% para as variedades de capim avaliadas e para os três estágios em que a planta se encontrava.

Quando se questiona por qual método utilizar contudo, não há uma resposta geral para todos os conjunto de dados e irá depender do ajuste do modelo nas duas metodologias. No caso do conjunto de dados tratados no trabalho, o método da máxima verossimilhança apresentado é mais aconselhado pois a adequação às pressuposições do modelo foram melhores, pois para os três estágios o coeficiente de correlação foi alto, acima de 80%, já para o método em que se utiliza a análise multivariada, o teste de normalidade multivariada e o teste de igualdade de matriz de variância-covariância deram significativos no estágio de 10 a 12 perfilhos, o que compromete as decisões tomadas sobre o parâmetro *e*.

REFERÊNCIAS

- BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its application**. New York: John Wiley & Sons, 1988. 365 p.
- BRUNHARO, Caio A. d C. G. **Resistência de planta daninha capim-branco *Chloris polydactyloa* herbicida glyphosate**. 2014. 153 p. Tese (Doutorado) — Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2014.
- COOK, R.; TSAI, C. Bias in nonlinear regression. **Biometrika**, Biometrika Trust, Cambridge, v. 72, n. 1, p. 23–29, 1985.
- D., N. R.; SMITH, H. **Applied regression analysis**. New York: John Wiley & Sons, 1998. 706 p.
- H.BOLFARINE; C.SANDOVAL, M. **Introdução à inferência estatística**. São Paulo: SBM, 2001. 125 p.
- KNEZEVIC, S. Z; S., J. C.; RITZ, C. Utilizing r software package for dose-response studies: the concept and data analysis. **Weed Technology**, New York, v. 21, n. 3, p. 840–848, 2007.
- MENDES, P. N.; MUNIZ, J. A.; FONSECA, F.; MAZZINI, A. R. de A.; SILVA, N. A. M. da et al. Análise da curva de crescimento difásica de fêmeas hereford por meio da função não linear de gompertz. **Ciência Animal Brasileira**, Santa Maria, v. 10, n. 2, p. 454–461, 2009.
- MINGOTI, D C. **Análise de dados através de métodos de estatística multivariada**. Minas Gerais: UFMG, 2005. 297 p.
- MONTGOMERY, D. C.; P., E. A.; VININGY, G. G. **Introduction to linear regression analysis**. New York: John Wiley & Sons, 2015.
- MORETTIN, P. A.; TOLOI, C.M.C. **Nonlinear regression modeling: a unified practical approach**. São Paulo: Edgard Blucher, 2004. 556 p.
- MUELLER, T.C.; BARNETT, K.A.; BROSNAN, J.T.; STECKEL, L.E. Glyphosate-resistant goosegrass *Eleusine indica* confirmed in tennessee. **Weed Science**, Nebraska, v. 59, p. 562–566, 2011.
- RATKOWSKY, D. A. **Nonlinear regression modeling: a unified practical approach**. New York: Dekker, 1983. 241 p.
- REGAZZI, A. J. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear / test for parameter equality in nonlinear regression. **Ceres**, Viçosa, v. 50, n. 287, p. 9–26, 2015.
- REGAZZI, A. J.; SILVA, C. H. O. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear. i. dados no delineamento inteiramente casualizado. **Revista de Matemática e Estatística**, São Paulo, v. 22, n. 3, p. 33–45, 2004.

RENCHEER, A. C. **Methods of multivariate analysis**. New York: John Wiley & Sons, 2002. 727 p.

RENCHEER, A. C.; SCHAALJE, G. B. **Linear models in statistics**. New Jersey: John Wiley & Sons, 2008. 678 p.

ROSA, L. E. **Aspectos da biologia, suscetibilidade diferencial e eficácia de herbicidas alternativos ao glyphosate no manejo de populações de capim-pé-de-galinha *Eleusine indica* L. Gaertn.** 2014. 94 p. Tese (Doutorado) — Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2014.

SEBER, G. A. F.; WILD, C. J. **Nonlinear regression**. New Jersey: John Wiley & Sons, 1989. 768 p.

SOUZA, G. da S. e. **Introdução aos modelos de regressão linear e não-linear**. Brasília: EMBRAPA-SPI/EMBRAPA-SEA, 1998. 489 p.

ZEVIANE, W. **Interface gráfica para ajuste de modelos de regressão não linear**. 2012. Disponível em: <www.dm.ufscar.br/sadao>. Acesso em: 30 jul. 2015.

APÊNDICE

Programação utilizada nessa dissertação

```
#####modelo log logistico#####
##### Diferentes valores para os parâmetros
rm(list=ls(all=TRUE))
library(lattice)
Dose = seq(0,10000,.1)

#d diferentes
y=1/(1+exp(.5*(log(Dose)-log(50))))
plot(y~Dose,type='s',col="black", ylim=c(0,6))
y=4/(1+exp(.5*(log(Dose)-log(50))))
lines(y~Dose,type='s',col="blue")
y=10/(1+exp(.5*(log(Dose)-log(50))))
lines(y~Dose,type='s',col="red")
legend("bottomright",c('d=1','d=4','d=10'),
col=c('black','blue','red'),inset = 0.01,bty='n',lwd=2)
#b diferentes

y=4/(1+exp(.1*(log(Dose)-log(50))))
plot(y~Dose,type='s', col="black", ylim=c(0,6))
y=4/(1+exp(.5*(log(Dose)-log(50))))
lines(y~Dose,type='s', col="blue")
y=4/(1+exp(2*(log(Dose)-log(50))))
lines(y~Dose,type='s', col="red")
legend("bottomright",c('b=0.1','b=0.5','b=2'),col=
c('black','blue','red'),inset = 0.01,bty='n',lwd=2)
#e diferentes

y=4/(1+exp(.5*(log(Dose)-log(10))))
plot(y~Dose,type='s', col="black" )
y=4/(1+exp(.5*(log(Dose)-log(50))))
lines(y~Dose,type='s', col="blue" )
y=4/(1+exp(.5*(log(Dose)-log(100))))
lines(y~Dose,type='s', col="red")
legend("bottomright",c('e=10','e=50','e=100'),
col=c('black','blue','red'),inset = 0.01,bty='n',lwd=2)

#####
#Análise exploratória
dados <- read.csv2("C:/Users/estatistica/dadosR/glifosato/anExdesenInd.csv")
require(moments)
library(lattice)
xyplot(Massa~Dose|Biotipo,type='b',cex=.8,ylab =
'Massa-seca',xlab = 'Dose (g. e. a. por ha)',dados)
xyplot(Massa1~Dose1|Biotipo,type='b',cex=.8,ylab =
'Massa-seca',xlab = 'Dose (g. e. a. por ha)',dados)
xyplot(Massa2~Dose2|Biotipo,type='b',cex=.8,ylab =
'Massa-seca',xlab = 'Dose (g. e. a. por ha)',dados)

par(mfrow=c(1,3))
boxplot(Massa~Dose,ylab = 'Massa-seca',xlab =
'Dose (g. e. a. por ha)',main="2-3 perfilhos",dados)
boxplot(Massa1~Dose1,ylab = 'Massa-seca',xlab =
'Dose (g. e. a. por ha)',main="6-8 perfilhos",dados)
boxplot(Massa2~Dose2,range = 1.5,ylab = 'Massa-seca',xlab = 'Dose (g. e. a. por ha)',main="10-12 perfilhos",dados)
```

```

#Regressão não-linear dentro de cada estágio da planta (Número de perfilhos)
#Estágio 2p
dados2p <- read.csv2("C:/Users/estatistica/dadosR/glifosato/estagio2p.csv");dados2p

ajuste2p <- nls(massa~um*d1/(1+exp(b1*(log(dose)-log(e1))))+
dois*d2/(1+exp(b2*(log(dose)-log(e2))))+
tres*d3/(1+exp(b3*(log(dose)-log(e3))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dados2p, start=list(d1=24, b1=2.29,e1=119,
d2=6, b2=1.4,e2=228,
d3=5.7, b3=0.89,e3=62,
d4=14, b4=0.64,e4=480
))

summary(ajuste2p)
residuo2p = summary(ajuste2p)
res_ord2 = residuo2p$residuals
shapiro.test(residuals(residuo2p))
respad_inter2 = res_ord2/sqrt(var(res_ord2))
dados2p$massa[47] = 0.645
dados2p$massa[57] = 0.645
dados2p$massa[48] = 0.35
dados2p$massa[78] = 0.35
dados2p$massa[32] = 4.313
dados2p$massa[33] = 3.745
dados2p$massa[62] = 6.1766
dados2p$massa[71] = 5.37
plot(dados2p$dose,respad_inter2)
plot(fitted(ajuste2p),res_ord2)
plot(fitted(ajuste2p),respad_inter2)

#Estágio 6p
dados6p <- read.csv2("C:/Users/estatistica/dadosR/glifosato/estagio6p.csv");dados6p
ajuste6p <- nls(massa~um*d1/(1+exp(b1*(log(dose)-log(e1))))+
dois*d2/(1+exp(b2*(log(dose)-log(e2))))+
tres*d3/(1+exp(b3*(log(dose)-log(e3))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dados6p, start=list(d1=4.66, b1=.63,e1=19,
d2=6, b2=1.8,e2=302,
d3=4, b3=0.69,e3=108,
d4=4, b4=0.49,e4=50
))
dados6p$massa[72] = 5.94
summary(ajuste6p)
residuo6p = summary(ajuste6p)
res_ord6 = residuo6p$residuals
shapiro.test(residuals(residuo6p))
respad_inter6 = res_ord6/sqrt(var(res_ord6))

#Estágio 12p
dados12p <- read.csv2("C:/Users/estatistica/dadosR/glifosato/estagio12p.csv");dados12p

ajuste12p <- nls(massa~um*d1/(1+exp(b1*(log(dose)-log(e1))))+
dois*d2/(1+exp(b2*(log(dose)-log(e2))))+
tres*d3/(1+exp(b3*(log(dose)-log(e3))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dados12p, start=list(d1=24, b1=2.29,e1=119,

```

```

d2=6, b2=1.4,e2=228,
d3=5.7, b3=0.89,e3=62,
d4=14, b4=0.64,e4=50
))
dados12p$massa[41] = 17.68
dados12p$massa[125] = 7.18
dados12p$massa[130] = 1.97
summary(ajuste12p)
residuo12p = summary(ajuste12p)
res_ord12 = residuo12p$residuals
shapiro.test(residuals(residuo12p))
respad_inter12 = res_ord12/sqrt(var(res_ord12))

#####
###Valores da dose que reduzem 90% e 95% a matéria seca#####
###Para o estágio de 2-3 perfilhos#####
c1 =.10*residuo2p$parameters[1,1]
c12 =.05*residuo2p$parameters[1,1]
c2 =.10*residuo2p$parameters[4,1]
c22 =.05*residuo2p$parameters[4,1]
c3 =.10*residuo2p$parameters[7,1]
c32 =.05*residuo2p$parameters[7,1]
c4 =.10*residuo2p$parameters[10,1]
c42 =.05*residuo2p$parameters[10,1]

##### Valores da dose#####
x1= residuo2p$parameters[3,1]*((residuo2p$parameters[1,1]-c1)/c1)
^1/{residuo2p$parameters[2,1]}};x1
x12= residuo2p$parameters[3,1]*((residuo2p$parameters[1,1]-c12)/c12)
^1/{residuo2p$parameters[2,1]}};x12
x2= residuo2p$parameters[6,1]*((residuo2p$parameters[4,1]-c2)/c2)
^1/{residuo2p$parameters[5,1]}};x2
x22= residuo2p$parameters[6,1]*((residuo2p$parameters[4,1]-c22)
/c22)^1/{residuo2p$parameters[5,1]}};x22
x3= residuo2p$parameters[9,1]*((residuo2p$parameters[7,1]-c3)
/c3)^1/{residuo2p$parameters[8,1]}};x3
x32= residuo2p$parameters[9,1]*((residuo2p$parameters[7,1]-c32)
/c32)^1/{residuo2p$parameters[8,1]}};x32
x4= residuo2p$parameters[12,1]*((residuo2p$parameters[10,1]-c4)
/c4)^1/{residuo2p$parameters[11,1]}};x4
x42= residuo2p$parameters[12,1]*((residuo2p$parameters[10,1]-c42)
/c42)^1/{residuo2p$parameters[11,1]}};x42

doisp9095 = c(x1,x12,x2,x22,x3,x32,x4,x42)

##### De maneira igual para os demais perfilhos #####

###Cálculo do coeficiente de correlação#####

F <- attr(ajuste6p$m$fitted(), "gradient")
F

#-----
# passando a matriz gradiente para a lm(), importante remover intercepto

m0 <- lm(massa~-1+F, data=ajuste6p)

```



```
anova(n0, lm(massa~1, ajuste6p))
```

```
R2 <- function(nls.obj){
da <- eval(nls.obj$data)
resp.name <- all.vars(summary(nls.obj)$formula)[1]
form <- paste(resp.name, "~1", sep="")
m0 <- lm(form, da)
an <- anova(nls.obj, m0)
sqn <- deviance(nls.obj)
sqe <- deviance(m0)
r2 <- 1-(sqn/sqe)
aov <- data.frame(fv=c("regression","residuals"),
gl=c(-an$Df[2],an$Res.Df[1]),
sq=c(-an$Sum[2],an$Res.Sum[1]))
aov$qm <- aov$sq/aov$gl
aov$F <- c(aov$qm[1]/aov$qm[2], NA)
aov$"Pr(>F)" <- c(1-pf(aov$F[1], df1=aov$gl[1], df2=aov$gl[2]), NA)
names(aov) <- c(" ", "Df", "Sum Sq", "Mean Sq", "F value", "Pr(>F)")
return(list(anova=aov, R2=r2))
}
```

```
R2(ajuste6p)
```

```
#Análise da igualdade do parâmetro "e" da curva log-logística
```

```
#modelo sem nenhuma restrição
```

```
#Já obtido anteriormente.
```

```
ajuste2pRest <- nls(massa~um*d1/(1+exp(b1*(log(dose)-log(e))))+
dois*d2/(1+exp(b2*(log(dose)-log(e))))+
tres*d3/(1+exp(b3*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dados2p, start=list(d1=4.62, b1=1.85,e=211,
d2=6, b2=1.4,
d3=2.82, b3=1.33,
d4=2.52, b4=4.39
))
```

```
residuo_rest = summary(ajuste2pRest)
```

```
residuo_rest$parameters
```

```
sigma2_mvsComp=((residuo2p$sigma)^2)*residuo2p$df[2]/(residuo2p$df[1]+residuo2p$df[2]);sigma2_mvsComp
```

```
sigma2_mvsRest=((residuo_rest$sigma)^2)*residuo_rest$df[2]/(residuo_rest$df[1]+residuo_rest$df[2]);sigma2_mvsRest
```

```
# Estatística do Teste
```

```
estat_teste = -(160)*log(sigma2_mvsComp/sigma2_mvsRest)
```

```
graus_lib = residuo2p$df[1] - residuo_rest$df[1]
```

```
pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE)
```

```
#Como houve diferença significativa entre pelo menos uma das estimativas
```

```
#calculadas, ordena-se todos os valores do parâmetro "e" e procura-se fazer uma
```

#comparação duas a duas para isso se constroi os conjuntos de dados:

```
# Comparação entre os valores e2 e e4
newdata <- subset(dados2p, index > 40 & index <=80)
dadosE2 = newdata;dadosE2
newdata <- subset(dados2p, index > 120 & index <=160 )
dadosE4 = newdata;dadosE4
dadosE2E4 = rbind(dadosE2,dadosE4);dadosE2E4

ajusteE2E4 <- nls(massa~ dois*d2/(1+exp(b2*(log(dose)-log(e2))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dadosE2E4, start=list(d2=6, b2=1.4,e2=228,
d4=14, b4=0.64,e4=480
))
residuo_E2E4 = summary(ajusteE2E4)
ajusteE24 <- nls(massa~ dois*d2/(1+exp(b2*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dadosE2E4, start=list(d2=6, b2=1.4,e=228,
d4=14, b4=0.64))
residuo_E24 = summary(ajusteE24)

sigma2E2E4=((residuo_E2E4$sigma)^2)*residuo_E2E4$df[2]/
(residuo_E2E4$df[1]+residuo_E2E4$df[2]);sigma2E2E4
sigma2E24=((residuo_E24$sigma)^2)*residuo_E24$df[2]/
(residuo_E24$df[1]+residuo_E24$df[2]);sigma2E24
estat_teste = -(80)*log(sigma2E2E4/sigma2E24)
graus_lib = residuo_E2E4$df[1] - residuo_E24$df[1]
pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

#Comparação entre os valores e4 e e1

newdata <- subset(dados2p, index >= 1 & index <=40)
dadosE1 = newdata;dadosE1
newdata <- subset(dados2p, index > 120 & index <=160 )
dadosE4 = newdata;dadosE4
dadosE1E4 = rbind(dadosE1,dadosE4);dadosE1E4
ajusteE1E4 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e1))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dadosE1E4, start=list(d1=6, b1=1.4,e1=228,
d4=14, b4=0.64,e4=480
))
residuo_E1E4 = summary(ajusteE1E4)
ajusteE14 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dadosE1E4, start=list(d1=6, b1=1.4,e=228,
d4=14, b4=0.64))
residuo_E14 = summary(ajusteE14)

sigma2E1E4=((residuo_E1E4$sigma)^2)*residuo_E1E4$df[2]/
(residuo_E1E4$df[1]+residuo_E1E4$df[2]);sigma2E1E4
sigma2E14=((residuo_E14$sigma)^2)*residuo_E14$df[2]/
(residuo_E14$df[1]+residuo_E14$df[2]);sigma2E14
estat_teste = -(80)*log(sigma2E1E4/sigma2E14)
graus_lib = residuo_E1E4$df[1] - residuo_E14$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

#Comparação entre os valores e1 e e2
```

```

newdata <- subset(dados2p, index >= 1 & index <=40)
dadosE1 = newdata;dadosE1
newdata <- subset(dados2p, index > 40 & index <=80 )
dadosE2 = newdata;dadosE2
dadosE1E2 = rbind(dadosE1,dadosE2);dadosE1E2
ajusteE1E2 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e1))))+
dois*d2/(1+exp(b2*(log(dose)-log(e2))))
, data=dadosE1E2, start=list(d1=6, b1=1.4,e1=228,
d2=14, b2=0.64,e2=480
))
residuo_E1E2 = summary(ajusteE1E2)
ajusteE12 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e))))+
dois*d2/(1+exp(b2*(log(dose)-log(e))))
, data=dadosE1E2, start=list(d1=6, b1=1.4,e=228,
d2=14, b2=0.64))
residuo_E12 = summary(ajusteE12)

sigma2E1E2=((residuo_E1E2$sigma)^2)*residuo_E1E2$df[2]/
(residuo_E1E2$df[1]+residuo_E1E2$df[2]);sigma2E1E2
sigma2E12=((residuo_E12$sigma)^2)*residuo_E12$df[2]/
(residuo_E12$df[1]+residuo_E12$df[2]);sigma2E12
estat_teste = -(80)*log(sigma2E1E2/sigma2E12)
graus_lib = residuo_E1E2$df[1] - residuo_E12$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

#Comparação entre os valores e2 e e3
newdata <- subset(dados2p, index > 40 & index <=80 )
dadosE2 = newdata;dadosE2
newdata <- subset(dados2p, index > 80 & index <=120 )
dadosE3 = newdata;dadosE3

dadosE2E3 = rbind(dadosE2,dadosE3);dadosE2E3
ajusteE2E3 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e3))))+
dois*d2/(1+exp(b2*(log(dose)-log(e2))))
, data=dadosE2E3, start=list(d3=6, b3=1.4,e3=62,
d2=14, b2=0.64,e2=480
))
residuo_E2E3 = summary(ajusteE2E3)
ajusteE23 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e))))+
dois*d2/(1+exp(b2*(log(dose)-log(e))))
, data=dadosE2E3, start=list(d3=6, b3=1.4,e=100,
d2=14, b2=0.64))
residuo_E23 = summary(ajusteE23)

sigma2E2E3=((residuo_E2E3$sigma)^2)*residuo_E2E3$df[2]/
(residuo_E2E3$df[1]+residuo_E2E3$df[2]);sigma2E2E3
sigma2E23=((residuo_E23$sigma)^2)*residuo_E23$df[2]/
(residuo_E23$df[1]+residuo_E23$df[2]);sigma2E23
estat_teste = -(80)*log(sigma2E2E3/sigma2E23)
graus_lib = residuo_E2E3$df[1] - residuo_E23$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

#Comparação entre os valores e1 e e3

newdata <- subset(dados2p, index >= 1 & index <=40)
dadosE1 = newdata;dadosE1

```

```

newdata <- subset(dados2p, index > 80 & index <=120 )
dadosE3 = newdata;dadosE3
dadosE1E3 = rbind(dadosE1,dadosE3);dadosE1E3
ajusteE1E3 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e))))+
tres*d3/(1+exp(b3*(log(dose)-log(e3))))
, data=dadosE1E3, start=list(d1=6, b1=1.4,e1=160,
d3=14, b3=0.64,e3=62
))
residuo_E1E3 = summary(ajusteE1E3)
ajusteE13 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e))))+
tres*d3/(1+exp(b3*(log(dose)-log(e))))
, data=dadosE1E3, start=list(d1=6, b1=1.4,e=160,
d3=14, b3=0.64))
residuo_E13 = summary(ajusteE13)

sigma2E1E3=((residuo_E1E3$sigma)^2)*residuo_E1E3$df[2]/
(residuo_E1E3$df[1]+residuo_E1E3$df[2]);sigma2E1E3
sigma2E13=((residuo_E13$sigma)^2)*residuo_E13$df[2]/
(residuo_E13$df[1]+residuo_E13$df[2]);sigma2E13
estat_teste = -(80)*log(sigma2E1E3/sigma2E13)
graus_lib = residuo_E1E3$df[1] - residuo_E13$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

##
#Ajuste do modelo no estágio 6p
##
ajuste6pRest <- nls(massa~um*d1/(1+exp(b1*(log(dose)-log(e))))+
dois*d2/(1+exp(b2*(log(dose)-log(e))))+
tres*d3/(1+exp(b3*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dados6p, start=list(d1=2, b1=0.56,e=291,
d2=6, b2=1.4,
d3=2.82, b3=1.33,
d4=2.52, b4=4.39
))

summary(ajuste6pRest)
residuo_rest = summary(ajuste6pRest)
residuo_rest$parameters
sigma2_mvsComp=((residuo6p$sigma)^2)*residuo6p$df[2]/
(residuo6p$df[1]+residuo6p$df[2]);sigma2_mvsComp
sigma2_mvsRest=((residuo_rest$sigma)^2)*residuo_rest$df[2]/
(residuo_rest$df[1]+residuo_rest$df[2]);sigma2_mvsRest

# Estatística do Teste

estat_teste = -(160)*log(sigma2_mvsComp/sigma2_mvsRest)
graus_lib = residuo6p$df[1] - residuo_rest$df[1]
pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE)

#houve diferença,portanto ordena-se os valores das estimativas
#e se procede da mesma maneira no estágio 2p

```

#Comparação entre os valores e2 e e3

```
newdata <- subset(dados6p, index > 40 & index <=80 )
dadosE2 = newdata;dadosE2
newdata <- subset(dados6p, index > 80 & index <=120 )
dadosE3 = newdata;dadosE3

dadosE2E3 = rbind(dadosE2,dadosE3);dadosE2E3
ajusteE2E3 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e3))))+
dois*d2/(1+exp(b2*(log(dose)-log(e2))))
, data=dadosE2E3, start=list(d3=6, b3=1.4,e3=62,
d2=14, b2=0.64,e2=480
))
residuo_E2E3 = summary(ajusteE2E3)
ajusteE23 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e))))+
dois*d2/(1+exp(b2*(log(dose)-log(e))))
, data=dadosE2E3, start=list(d3=6, b3=1.4,e=100,
d2=14, b2=0.64))
residuo_E23 = summary(ajusteE23)

sigma2E2E3=((residuo_E2E3$sigma)^2)*residuo_E2E3$df[2]/
(residuo_E2E3$df[1]+residuo_E2E3$df[2]);sigma2E2E3
sigma2E23=((residuo_E23$sigma)^2)*residuo_E23$df[2]/
(residuo_E23$df[1]+residuo_E23$df[2]);sigma2E23
estat_teste = -(80)*log(sigma2E2E3/sigma2E23)
graus_lib = residuo_E2E3$df[1] - residuo_E23$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor
```

#Comparação entre os valores e3 e e4

```
newdata <- subset(dados6p, index > 80 & index <=120 )
dadosE3 = newdata;dadosE3
newdata <- subset(dados6p, index > 120 & index <=160 )
dadosE4 = newdata;dadosE4
dadosE3E4 = rbind(dadosE3,dadosE4);dadosE3E4

ajusteE3E4 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e3))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dadosE3E4, start=list(d3=4, b3=0.5,e3=100,
d4=4, b4=0.64,e4=60
))
residuo_E3E4 = summary(ajusteE3E4)
ajusteE34 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dadosE3E4, start=list(d3=6, b3=1.4,e=100,
d4=14, b4=0.64))
residuo_E34 = summary(ajusteE34)
sigma2E3E4=((residuo_E3E4$sigma)^2)*residuo_E3E4$df[2]/
(residuo_E3E4$df[1]+residuo_E3E4$df[2]);sigma2E3E4
sigma2E34=((residuo_E34$sigma)^2)*residuo_E34$df[2]/
(residuo_E34$df[1]+residuo_E34$df[2]);sigma2E34
estat_teste = -(80)*log(sigma2E3E4/sigma2E34)
graus_lib = residuo_E3E4$df[1] - residuo_E34$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor
```

#Comparação entre os valores e3 e e1

```
newdata <- subset(dados6p, index >= 1 & index <=40)
dadosE1 = newdata;dadosE1
newdata <- subset(dados6p, index > 80 & index <=120 )
dadosE3 = newdata;dadosE3
dadosE1E3 = rbind(dadosE1,dadosE3);dadosE1E3
ajusteE1E3 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e1))))+
tres*d3/(1+exp(b3*(log(dose)-log(e3))))
, data=dadosE1E3, start=list(d1=6, b1=1.4,e1=160,
d3=14, b3=0.64,e3=62
))
residuo_E1E3 = summary(ajusteE1E3)
ajusteE13 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e))))+
tres*d3/(1+exp(b3*(log(dose)-log(e))))
, data=dadosE1E3, start=list(d1=6, b1=1.4,e=160,
d3=14, b3=0.64))
residuo_E13 = summary(ajusteE13)

sigma2E1E3=((residuo_E1E3$sigma)^2)*residuo_E1E3$df[2]/
(residuo_E1E3$df[1]+residuo_E1E3$df[2]);sigma2E1E3
sigma2E13=((residuo_E13$sigma)^2)*residuo_E13$df[2]/
(residuo_E13$df[1]+residuo_E13$df[2]);sigma2E13
estat_teste = -(80)*log(sigma2E1E3/sigma2E13)
graus_lib = residuo_E1E3$df[1] - residuo_E13$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor
```

#Comparação entre os valores e1 e e4

```
newdata <- subset(dados6p, index >= 1 & index <=40)
dadosE1 = newdata;dadosE1
newdata <- subset(dados6p, index > 120 & index <=160 )
dadosE4 = newdata;dadosE4
dadosE1E4 = rbind(dadosE1,dadosE4);dadosE1E4
ajusteE1E4 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e1))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dadosE1E4, start=list(d1=6, b1=1.4,e1=19,
d4=14, b4=0.64,e4=50
))
residuo_E1E4 = summary(ajusteE1E4)
ajusteE14 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dadosE1E4, start=list(d1=6, b1=1.4,e=30,
d4=14, b4=0.64))
residuo_E14 = summary(ajusteE14)

sigma2E1E4=((residuo_E1E4$sigma)^2)*residuo_E1E4$df[2]/
(residuo_E1E4$df[1]+residuo_E1E4$df[2]);sigma2E1E4
sigma2E14=((residuo_E14$sigma)^2)*residuo_E14$df[2]/
(residuo_E14$df[1]+residuo_E14$df[2]);sigma2E14
estat_teste = -(80)*log(sigma2E1E4/sigma2E14)
graus_lib = residuo_E1E4$df[1] - residuo_E14$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

##
```

```

#Ajuste do modelo no estágio 12p
##
ajuste12pRest <- nls(massa~um*d1/(1+exp(b1*(log(dose)-log(e))))+
dois*d2/(1+exp(b2*(log(dose)-log(e))))+
tres*d3/(1+exp(b3*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dados12p, start=list(d1=24, b1=3,e=291,
d2=18, b2=3,
d3=15, b3=5,
d4=15, b4=.65
))
residuo_rest = summary(ajuste12pRest)
residuo_rest$parameters
sigma2_mvsComp=((residuo12p$sigma)^2)*residuo12p$df[2]/
(residuo12p$df[1]+residuo12p$df[2]);sigma2_mvsComp

sigma2_mvsRest=((residuo_rest$sigma)^2)*residuo_rest$df[2]/
(residuo_rest$df[1]+residuo_rest$df[2]);sigma2_mvsRest

# Estatística do Teste

estat_teste = -(160)*log(sigma2_mvsComp/sigma2_mvsRest)
graus_lib = residuo12p$df[1] - residuo_rest$df[1]
pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE);pvalor

# Comparação entre os valores e2 e e4

newdata <- subset(dados12p, index > 40 & index <=80)
dadosE2 = newdata;dadosE2
newdata <- subset(dados12p, index > 120 & index <=160 )
dadosE4 = newdata;dadosE4
dadosE2E4 = rbind(dadosE2,dadosE4);dadosE2E4

ajusteE2E4 <- nls(massa~ dois*d2/(1+exp(b2*(log(dose)-log(e2))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dadosE2E4, start=list(d2=17, b2=3,e2=494,
d4=14, b4=0.64,e4=416
))
residuo_E2E4 = summary(ajusteE2E4)
ajusteE24 <- nls(massa~ dois*d2/(1+exp(b2*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dadosE2E4, start=list(d2=6, b2=1.4,e=228,
d4=14, b4=0.64))
residuo_E24 = summary(ajusteE24)

sigma2E2E4=((residuo_E2E4$sigma)^2)*residuo_E2E4$df[2]/
(residuo_E2E4$df[1]+residuo_E2E4$df[2]);sigma2E2E4
sigma2E24=((residuo_E24$sigma)^2)*residuo_E24$df[2]/
(residuo_E24$df[1]+residuo_E24$df[2]);sigma2E24
estat_teste = -(80)*log(sigma2E2E4/sigma2E24)
graus_lib = residuo_E2E4$df[1] - residuo_E24$df[1]
pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

#Comparação entre os valores e2 e e3

newdata <- subset(dados12p, index > 40 & index <=80 )
dadosE2 = newdata;dadosE2

```

```

newdata <- subset(dados12p, index > 80 & index <=120 )
dadosE3 = newdata;dadosE3

dadosE2E3 = rbind(dadosE2,dadosE3);dadosE2E3
ajusteE2E3 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e3))))+
dois*d2/(1+exp(b2*(log(dose)-log(e2))))
, data=dadosE2E3, start=list(d3=6, b3=1.4,e3=62,
d2=14, b2=0.64,e2=480
))
residuo_E2E3 = summary(ajusteE2E3)
ajusteE23 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e))))+
dois*d2/(1+exp(b2*(log(dose)-log(e))))
, data=dadosE2E3, start=list(d3=6, b3=1.4,e=100,
d2=14, b2=0.64))
residuo_E23 = summary(ajusteE23)

sigma2E2E3=((residuo_E2E3$sigma)^2)*residuo_E2E3$df[2]/
(residuo_E2E3$df[1]+residuo_E2E3$df[2]);sigma2E2E3
sigma2E23=((residuo_E23$sigma)^2)*residuo_E23$df[2]/
(residuo_E23$df[1]+residuo_E23$df[2]);sigma2E23
estat_teste = -(80)*log(sigma2E2E3/sigma2E23)
graus_lib = residuo_E2E3$df[1] - residuo_E23$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

#Comparação entre os valores e3 e e4

newdata <- subset(dados12p, index > 80 & index <=120 )
dadosE3 = newdata;dadosE3
newdata <- subset(dados12p, index > 120 & index <=160 )
dadosE4 = newdata;dadosE4
dadosE3E4 = rbind(dadosE3,dadosE4);dadosE3E4

ajusteE3E4 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e3))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dadosE3E4, start=list(d3=15.76, b3=5.6,e3=213,
d4=15, b4=0.64,e4=416
))
residuo_E3E4 = summary(ajusteE3E4)
ajusteE34 <- nls(massa~ tres*d3/(1+exp(b3*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dadosE3E4, start=list(d3=15.76, b3=5.6,e=400,
d4=14.68, b4=0.643))
residuo_E34 = summary(ajusteE34)
sigma2E3E4=((residuo_E3E4$sigma)^2)*residuo_E3E4$df[2]/
(residuo_E3E4$df[1]+residuo_E3E4$df[2]);sigma2E3E4
sigma2E34=((residuo_E34$sigma)^2)*residuo_E34$df[2]/
(residuo_E34$df[1]+residuo_E34$df[2]);sigma2E34
estat_teste = -(80)*log(sigma2E3E4/sigma2E34)
graus_lib = residuo_E3E4$df[1] - residuo_E34$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

#Comparação entre os valores e1 e e4

newdata <- subset(dados12p, index >= 1 & index <=40)
dadosE1 = newdata;dadosE1

```



```

newdata <- subset(dados12p, index > 120 & index <=160 )
dadosE4 = newdata;dadosE4
dadosE1E4 = rbind(dadosE1,dadosE4);dadosE1E4
ajusteE1E4 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e1))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e4))))
, data=dadosE1E4, start=list(d1=6, b1=1.4,e1=119,
d4=14, b4=0.64,e4=400
))
residuo_E1E4 = summary(ajusteE1E4)
ajusteE14 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e))))+
quatro*d4/(1+exp(b4*(log(dose)-log(e))))
, data=dadosE1E4, start=list(d1=6, b1=1.4,e=100,
d4=14, b4=0.64))
residuo_E14 = summary(ajusteE14)

sigma2E1E4=((residuo_E1E4$sigma)^2)*residuo_E1E4$df[2]/
(residuo_E1E4$df[1]+residuo_E1E4$df[2]);sigma2E1E4
sigma2E14=((residuo_E14$sigma)^2)*residuo_E14$df[2]/
(residuo_E14$df[1]+residuo_E14$df[2]);sigma2E14
estat_teste = -(80)*log(sigma2E1E4/sigma2E14)
graus_lib = residuo_E1E4$df[1] - residuo_E14$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

#Comparação entre os valores e1 e e3

newdata <- subset(dados12p, index >= 1 & index <=40)
dadosE1 = newdata;dadosE1
newdata <- subset(dados12p, index > 80 & index <=120 )
dadosE3 = newdata;dadosE3
dadosE1E3 = rbind(dadosE1,dadosE3);dadosE1E3
ajusteE1E3 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e1))))+
tres*d3/(1+exp(b3*(log(dose)-log(e3))))
, data=dadosE1E3, start=list(d1=24, b1=3,e1=119,
d3=14, b3=5,e3=213
))
residuo_E1E3 = summary(ajusteE1E3)
ajusteE13 <- nls(massa~ um*d1/(1+exp(b1*(log(dose)-log(e))))+
tres*d3/(1+exp(b3*(log(dose)-log(e))))
, data=dadosE1E3, start=list(d1=23, b1=3,e=160,
d3=18, b3=1.72))
residuo_E13 = summary(ajusteE13)

sigma2E1E3=((residuo_E1E3$sigma)^2)*residuo_E1E3$df[2]/
(residuo_E1E3$df[1]+residuo_E1E3$df[2]);sigma2E1E3
sigma2E13=((residuo_E13$sigma)^2)*residuo_E13$df[2]/
(residuo_E13$df[1]+residuo_E13$df[2]);sigma2E13
estat_teste = -(80)*log(sigma2E1E3/sigma2E13)
graus_lib = residuo_E1E3$df[1] - residuo_E13$df[1]

pvalor = pchisq(estat_teste, df=graus_lib, lower.tail=FALSE) ;pvalor

#Comparação dois a dois dos parâmetros dentro de cada estágio.
#Para poder comparar os parâmetros do modelo foi preciso fazer o ajuste
#para cada repetição

newdata <- subset(dados6p, index <=40 )
dados6p1 = newdata;dados6p1

```

```

newdata <- subset(dados6p, index > 40 & index <=80)
dados6p2 = newdata;dados6p2
newdata <- subset(dados6p, index > 80 & index <=120 )
dados6p3 = newdata;dados6p3
newdata <- subset(dados6p, index > 120 & index <=160 )
dados6p4 = newdata;dados6p4

#Para a variedade EIR1
newdata <- subset(dados6p1, index <= 10 )
dados6p11 = newdata;dados6p11
newdata <- subset(dados6p1, index > 10 & index <=20 )
dados6p12 = newdata;dados6p12
newdata <- subset(dados6p1, index > 20 & index <=30 )
dados6p13 = newdata;dados6p13
newdata <- subset(dados6p1, index > 30 & index <=40 )
dados6p14 = newdata;dados6p14

n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p11, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n1=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p12, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n2=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p13, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n3=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p1, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n4=summary(n)
p61 =rbind(n1$parameters[,1],n2$parameters[,1],n3$parameters[,1],
n4$parameters[,1])

#Para a variedade EIR2
newdata <- subset(dados6p2, index >= 40 & index <=50 )
dados6p21 = newdata;dados6p21
newdata <- subset(dados6p2, index > 50 & index <=60 )
dados6p22 = newdata;dados6p22
newdata <- subset(dados6p2, index > 60 & index <=70 )
dados6p23 = newdata;dados6p23
newdata <- subset(dados6p2, index > 70 & index <=80 )
dados6p24 = newdata;dados6p24
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p21, start=list(D=3.65, B=1.56, E=311),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n1=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p22, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n2=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p23, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n3=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),

```

```

data=dados6p24, start=list(D=3.65, B=1.56, E=289),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n4=summary(n)
p62 =rbind(n1$parameters[,1],n2$parameters[,1],n3$parameters[,1],
n4$parameters[,1])

#Para a variedade EIR3
newdata <- subset(dados6p3, index > 80 & index <=90)
dados6p31 = newdata;dados6p31
newdata <- subset(dados6p3, index > 90 & index <=100 )
dados6p32 = newdata;dados6p32
newdata <- subset(dados6p3, index > 100 & index <=110 )
dados6p33 = newdata;dados6p33
newdata <- subset(dados6p3, index > 110 & index <=120 )
dados6p34 = newdata;dados6p34

n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p3, start=list(D=3.65, B=1.56, E=300),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n1=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p32, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n2=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p33, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n3=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p34, start=list(D=3.65, B=1.56, E=289),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n4=summary(n)
p63 =rbind(n1$parameters[,1],n2$parameters[,1],n3$parameters[,1],
n4$parameters[,1])

#Para a variedade EIRS
newdata <- subset(dados6p4, index > 120 & index <=130)
dados6p41 = newdata;dados6p41
newdata <- subset(dados6p4, index > 130 & index <=140 )
dados6p42 = newdata;dados6p42
newdata <- subset(dados6p4, index > 140 & index <=150 )
dados6p43 = newdata;dados6p43
newdata <- subset(dados6p4, index > 150 & index <=160 )
dados6p44 = newdata;dados6p44

n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p41, start=list(D=3.65, B=0.3, E=22),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n1=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p4, start=list(D=3.65, B=0.35, E=22),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n2=summary(n)
n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p43, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n3=summary(n)

```

```

n <- nls(massa~D/(1+exp(B*(log(dose)-log(E)))),
data=dados6p44, start=list(D=3.65, B=1.56, E=38),
nls.control(maxiter=1000,tol = 1e-5,minFactor = 1/16384))
n4=summary(n)
p64 =rbind(n1$parameters[,1],n2$parameters[,1],n3$parameters[,1],
n4$parameters[,1])

p6 = rbind(p61,p62,p63,p64)
#Criação do conjunto de dados

write.table(p6, file='arquivop616.csv', sep=';', dec=',', row.names=FALSE)

## Feito de maneira análoga para os demais estágios

#####
##### Análise Multivariada #####
#####

#leitura dos dados
dados2 <- read.csv2("C:/Users/estatistica/dadosR/glifosato/arquivop2161.csv")
dados6 <- read.csv2("C:/Users/estatistica/dadosR/glifosato/arquivop616.csv")
dados12 <- read.csv2("C:/Users/estatistica/dadosR/glifosato/arquivop12161.csv")

#Análise de Variância Multivariada
Y=cbind(dados2$D,dados2$B,dados2$E);Y
modeloMult=manova(Y~Trat,dados2)
#Principais estatísticas Multivariadas
modeloMultW=summary.manova(modeloMult,test="Wilks")
modeloMultP=summary.manova(modeloMult,test="Pillai")
modeloMultR=summary.manova(modeloMult,test="Roy")
modeloMultHL=summary.manova(modeloMult,test="Hotelling-Lawley")

#Teste de Shapiro Wilk para verificar normalidade Multivariada
library(mvnormttest)
mshapiro.test(t(dados2[,,-1]))

#Teste M-box para verificar igualdade de matriz de variância-covariância
z2 = boxM(dados6[,,-1],dados2$Trat)

#Teste de Tukey para comparação múltipla da variável e
library(ExpDes.pt)
dic(Trat, dados2$E, mcomp = "tukey", sigT = 0.05,
sigF = 0.05)

#Repete-se, de maneira análoga, para os demais estágios analisados

```