



Roney Peterson Pereira

**APLICAÇÃO DE UM MODELO LINEAR MISTO
PARA O ESTUDO DO PESO DE CAMUNDONGOS
INFECTADOS PELO *TRYPANOSOMA CRUZI***

MARINGÁ – PARANÁ
Maio de 2019

Roney Peterson Pereira

**APLICAÇÃO DE UM MODELO LINEAR MISTO PARA O
ESTUDO DO PESO DE CAMUNDONGOS INFECTADOS
PELO *TRYPANOSOMA CRUZI***

Dissertação apresentada ao Programa de Pós-graduação em Bioestatística do centro de ciências exatas da Universidade Estadual de Maringá como requisito parcial para obtenção do título de mestre em Bioestatística.

Orientador: Prof.^a Dr.^a Terezinha Aparecida Guedes

Universidade Estadual de Maringá - UEM

Departamento de Estatística - DES

Programa de Pós-Graduação em Bioestatística

MARINGÁ – PARANÁ

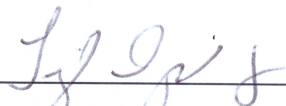
Maio de 2019

RONEY PETERSON PEREIRA

**Aplicação de um modelo linear misto para o estudo do peso de
camundongos infectados pelo *Trypanosoma cruzi***

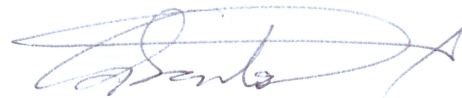
Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

BANCA EXAMINADORA



Prof.^a. Dra. Terezinha Aparecida Guedes

Universidade Estadual de Maringá – PBE/UEM



Prof. Dr. Carlos Aparecido dos Santos

Universidade Estadual de Maringá – PBE/UEM



Prof. Dr. Wyrllen Everson de Souza

Universidade Tecnologia Federal do Paraná – UTFPR/CM

Maringá, 30 de maio de 2019.

Agradecimentos

Sei que estes poucos parágrafos não serão suficientes para agradecer a todos que fizeram parte desta importante fase de minha vida. Então, desde já peço humildes desculpas àqueles que não foram citados, mas com certeza os levo no meu coração, assim como todos os que foram aqui lembrados.

Aos meus pais, Sirlei e Sérgio aos quais dedico tudo, que sempre me conduziram com sabedoria e conhecimento de vida. Minha esposa Viviane, por momentos doces, amor, pela infinita paciência, compreensão e incentivo em momentos difíceis.

Aos amigos de trabalho, Neto, Thelma e Néryla, que sempre estiveram dispostos a uma boa conversa como incentivo de me manter nessa caminhada da Pós-Graduação. Aos amigos de Umuarama, Carlão, Diegão, Leandro, Fabinho e Leandrinho, pela compreensão da minha ausência em churrascos e bate papos por diversos finais de semana.

A todos os amigos e colegas de curso da minha turma, Everton, Vinícius Tosta, Naiara, Jéssica, Patrícia, Talita, Stéfane e Mônica, que sempre estiveram presentes em estudos, conversas e poucos momentos de divertimento. Aos meus amigos da turma anterior, João Pedro e Jean que deram dicas fundamentais sobre o curso. Espero manter essa amizade mesmo com o término do mestrado.

A minha orientadora Prof^a. Dr^a. Terezinha Aparecida Guedes, que me propôs este tema como trabalho e sempre, de forma cordial e atenciosa, me atendeu em nossas reuniões de orientações. Aos professores da banca examinadora, pela presença e contribuições.

Ao Departamento de Ciências e à chefe em exercício Prof^a. Dr^a. Simoni Fiori, por possibilitar meus estudos no mestrado.

A todo corpo docente do Programa de Pós-Graduação em Bioestatística da Universidade Estadual de Maringá, pela formação técnica, profissional e pessoal.

"The greatest mistake you can make in life is continually fearing that you'll make one "
(Elbert Hubbard)

Resumo

O uso de Modelos Lineares Mistos para dados longitudinais com estrutura aninhada, denominados Modelos Lineares Hierárquicos, esses Modelos levam em consideração a dependência dos dados existentes dentro de cada nível e entre os níveis hierárquicos. O processo de modelagem, estimação e análise de diagnósticos foram ilustrados através de dados referentes aos pesos de camundongos experimentalmente infectados pelo *Trypanosoma Cruzi*, divididos em diferentes grupos de tratamento, com objetivo de verificar a evolução do peso corporal dos animais a partir da utilização de diferentes tipos de bioterápicos produzido a partir de soro de *Gallus gallus domesticus* (galinha) no tratamento contra o *Trypanosoma cruzi*. Através das medidas de decisão da ANOVA, um modelo foi selecionado de forma que melhor explicasse os dados, em seguida foi realizado os diagnósticos do Modelo por meio da Análise de Resíduos para os dois níveis e a Análise de Observações Influentes para verificar se alguma observação era sinalizada como influente nos efeitos fixos, componentes de variância e valores ajustados. Após as análises se constatou que as observações que foram sinalizadas como influentes pouco impactavam o Modelo escolhido inicialmente e, assim, o mesmo foi mantido, apontando que não houve diferença entre os tratamentos com os bioterápicos testados e sendo necessário apenas a variável Tempo e Intercepto Aleatório para descrever o peso dos camundongos.

Palavras-chave: Modelos Mistos Lineares; Dados Longitudinais; Níveis Hierárquicos; Trypanosoma cruzi; Análise de Diagnósticos.

Abstract

The use of mixed linear models for nested structure longitudinal data are called hierarchical linear Models. These Models take into account the dependence of existing data within each level and between hierarchical levels. The process of modeling, estimation and analysis of diagnoses were illustrated through data on the weights of mice experimentally infected by *Trypanosoma Cruzi*, divided into different treatment groups, with the purpose of verifying the evolution of the body weight of the animals from the use of different types of biotherapeutics produced from *Gallus gallus domesticus* (chicken) serum in the treatment of *Trypanosoma Cruzi*. Through the ANOVA decision-making measures, a Model was selected to describe correctly the data. The Model diagnoses were then performed through the Residue Analysis for the two levels and the analysis of influential observations to verify if any observations were signaled as influencing the fixed effects, components of variance and adjusted values. After the analysis, it was noticed that the observations that were signaled as influential had little impact on the Model chosen initially and so it was maintained, pointing out that there was no difference between the treatments with the tested biotherapeutics and only the Time variable and the Random Intercept were necessary to describe the weight of the mice.

Keywords: Linear Mixed Models; Longitudinal data; Hierarchical levels; *Trypanosoma cruzi*. Diagnostic Analysis.

Lista de ilustrações

Figura 1.0.1– Níveis hierárquicos	16
Figura 3.1.1– Gráfico de Dispersão do Peso Inicial dos Sujeitos	35
Figura 3.1.2– Gráfico de Dispersão do Peso Inicial dos Sujeitos por Grupo	35
Figura 3.1.3– Gráfico de Dispersão do Peso Inicial dos Sujeitos por Gaiola	36
Figura 3.1.4– Boxplot dos Pesos Iniciais por Grupo	37
Figura 3.1.5– Boxplot dos Pesos Iniciais por Gaiola	37
Figura 3.2.1– Gráfico de Dispersão dos Pesos dos Sujeitos entre a 2ª e 6ª Semana	39
Figura 3.2.2– Gráfico de Perfil de Cada Sujeito no Tempo	39
Figura 3.2.3– Gráfico de Dispersão do Peso dos Sujeitos por Grupo Durante o Tratamento	40
Figura 3.2.4– Peso Médio dos Grupos em Função do Tempo	40
Figura 3.2.5– Gráfico de Dispersão do Peso dos Sujeitos por Gaiola entre a 2ª e 6ª Semana	41
Figura 3.2.6– Histograma, Função de Densidade Ajustadas, Distribuição Acumulada Empírica, Função de Distribuição Ajustadas, Quantis Teóricos, Quantis Empíricos e Probabilidades Teóricas, Probabilidades Empíricas	42
Figura 3.3.1– QQ-Plot dos ajustes nos modelos propostos	44
Figura 3.3.2– Gráfico dos Resíduos MQ de nível 1 pelo Tempo	46
Figura 3.3.3– Gráfico dos Resíduos Semi-padronizados pelo Tempo	46
Figura 3.3.4– Painel dos Vinte Gráficos que Apresentam Resíduos Semi-padronizados de um Modelo Hierárquico em Relação à Variável Preditora Tempo. A Plotagem dos Dados Reais foi Incorporada Aleatoriamente em Dezenove Gráficos Simulados	47
Figura 3.3.5– Gráfico de Quantil Normal dos Resíduos Semi-padronizados de Nível 1	48
Figura 3.3.6– Gráfico de Quantil Normal dos Resíduos do Nível 2 do EB para o Intercepto	48

Figura 3.3.7– <i>Dotplot</i> e <i>Dotplot</i> modificado da distância de Cook e do MDFFITS . . .	49
Figura 3.3.8– Diagrama de Pontos Modificado do RVC de Nível 2 para o Intercepto Sujeito	51

Lista de tabelas

Tabela 1 – Desenho do experimento	19
Tabela 2 – Medidas de Resumo Estatístico	34
Tabela 3 – Médias e desvios padrão dos pesos dos grupo e gaiolas em cada semana	38
Tabela 4 – Teste de Kolmogorov-Smirnov	42
Tabela 5 – Modelos Propostos	43
Tabela 6 – Teste de Shapiro-Wilk para Normalidade de Resíduos	43
Tabela 7 – Observações do Sujeito 4	44
Tabela 8 – Teste de Shapiro-Wilk para Normalidade de Resíduos, Sem a Observação 16	44
Tabela 9 – Medidas de Decisão da ANOVA.	45
Tabela 10 – Estimativas dos Parâmetros de Efeitos Fixos	45
Tabela 11 – Matriz de Mudança de Variância Relativa	50
Tabela 12 – Peso dos Sujeitos na Segunda Semana	51
Tabela 13 – Pontos de Alavanca para os Efeitos Fixos, Efeitos Aleatórios e Efeitos Aleatórios Não Confundidos dos Sujeitos	52
Tabela 14 – Estimativas dos Parâmetros sem as Observações Influentes	53
Tabela 15 – Estimativas dos Parâmetros com Todas Observações	53

Sumário

Introdução	12
1 Revisão de Literatura	15
2 Materiais e Métodos	18
2.1 Experimento	18
2.2 Modelos Lineares Mistos	19
2.3 Métodos de Estimação	22
2.3.1 Máxima Verossimilhança (MV)	23
2.3.2 Máxima Verossimilhança Restrita (MVR)	25
2.4 Diagnósticos do Modelo	26
2.4.1 Análise de Resíduos	27
2.4.1.1 Resíduos de nível 1	28
2.4.1.2 Resíduos de Nível 2	28
2.4.2 Análise de Observações Influentes	29
2.4.2.1 Efeitos Fixos	30
2.4.2.2 Componentes de Variância	30
2.4.2.3 Valores Ajustados	31
3 Resultados e Discussões	33
3.1 Análise Descritiva do Peso Inicial dos Camundongos	33
3.2 Análise Descritiva do Peso dos Camundongos Avaliados no Período da 2ª à 6ª Semana	38
3.3 Ajuste	42
3.3.1 Análise dos Resíduos	45
3.3.1.1 Nível 1	45
3.3.2 Análise de Observações Influentes	48
3.3.2.1 Efeitos Fixos	49
3.3.2.2 Componentes de Variância	50
3.3.2.3 Valores Ajustados	52
4 Conclusão	54
5 Trabalhos Futuros	55

Referências	56
------------------------------	-----------

Introdução

A partir do século XX houve um aumento significativo na capacidade de produzir, armazenar e transmitir informações, que perdura até os dias de hoje de forma crescente. Tal aumento está associado ao crescimento acentuado da demanda por informações no mundo globalizado em que vivemos, tal mundo vêm exigindo da estatística cada vez mais avanços no desenvolvimento de metodologias e novos indicadores cada vez mais complexos, que exigem equipamentos modernos, *softwares* estatísticos e técnicos capacitados (IGNÁCIO, 2012).

Devido a estes avanços a estatística revolucionou a ciência no século XX, segundo Salsburg (2009) esta revolução se deu através do fornecimento de modelos úteis que sofisticaram o processo de estudos na direção de melhores parâmetros de investigação, permitindo orientar a tomada de decisões.

Assim sendo, diversas pesquisas em diferentes campos da ciência, tais como agrárias, biológicas e humanas, essas que precisam passar por estudos estatísticos, afim de que o pesquisador possa tomar a melhor decisão em rejeitar ou não sua pesquisa. Como exemplo de pesquisa podemos citar o teste de novos medicamentos, aceitação de produtos por um determinado público consumidor, abertura de um novo estabelecimento, entre diversos outros. No desenvolvimento desses estudos estatísticos pode ocorrer um planejamento para a coleta de dados. Tal planejamento define quais estruturas serão utilizadas para a organização dos dados, que muitas vezes são provenientes de um experimento de campo ou laboratorial.

Em inúmeros experimentos pode constar dados que são provenientes de duas ou mais observações de um mesmo sujeito no tempo, tais experimentos são caracterizados pela sequência temporal e são denominados dados longitudinais. Os dados longitudinais apresentarão estrutura hierárquica se as medidas repetidas são aninhadas dentro do indivíduo, tal estrutura hierárquica faz com que possamos fazer a suposição de que as observações

entre os indivíduos sejam independentes e que as aninhadas no indivíduo, possuam a característica da dependência com erros correlacionados (FAUSTO et al., 2008). Devido à tal natureza, um modelo estatístico que permite analisar dados longitudinais em estrutura hierárquica incorporando a dependência e a estrutura de correlação dos erros é o Modelo Linear Hierárquico (MLH), que nada mais é que um Modelo Misto Linear (MML) com dados hierárquicos ou multiníveis.

Segundo Santos et al. (2000), os Modelos Lineares Hierárquicos possuem importantes propriedades. Sendo que a primeira delas permite que a variabilidade da variável resposta nos diferentes níveis hierárquicos possa ser explicada através de variáveis preditoras incluídas no modelo nesses diferentes níveis e a segunda propriedade permite quantificar quanto da variabilidade da resposta se deve a cada nível, tal que a proporção da variabilidade explicada possa ser comparada diretamente.

Ademais, tais modelos têm como vantagem acomodar as dependências das observações, permitindo uma estrutura de covariância mais geral, na qual as observações do mesmo grupo ou indivíduo possam ser correlacionadas. A estrutura de erro aninhado dos modelos supracitados, torna seu diagnóstico mais complexo, cujos diagnósticos precisam ser ajustados para refletir apropriadamente as dependências dentro dos grupos e entre vários níveis (LOY, 2013).

As aplicações dos Modelos Lineares Hierárquicos vêm crescendo, devido a grande extensão de problemas que possuem dados estruturados de forma hierárquica, em praticamente todas as áreas de conhecimento, como no presente trabalho, que trata de dados do experimento realizado por Ferreira et al. (2018), que como objetivo buscou encontrar medidas que venham a contribuir para o seguimento do tratamento medicamentoso e consequentemente numa possível melhoria de qualidade de vida dos indivíduos infectados por *Trypanosoma cruzi*, que é o agente da doença de Chagas, principalmente na fase crônica. Através da avaliação do efeito do bioterápico de soro de galinha em roedores experimentalmente infectados por *Trypanosoma cruzi*, considerando parâmetros clínicos, imunológicos e parasitológicos, considerando que tal avaliação se deu pela relação de uma ou mais variáveis dependentes a uma ou mais variáveis preditoras.

A doença de Chagas é uma das patologias de mais larga distribuição no continente americano. É conhecida a existência de vetores da doença desde o sul dos Estados Unidos à Argentina. São mais de cem espécies responsáveis pela transmissão natural da infecção pelo *Trypanosoma Cruzi*, intervindo diretamente na sua veiculação no ambiente domiciliar ou participando na manutenção da enzootia chagásica. Estima-se que seja de 16 a 18 milhões o número de indivíduos infectados e de aproximadamente 80 milhões a parte da população em risco de contaminação na América Latina (SCHMUNIS; ORGANIZATION et

al., 1997, 1991 apud VINHAES; DIAS, 2000). O que justifica a importância dos diversos estudos realizados nessa área.

O objetivo geral deste trabalho foi descrever por meio de um Modelo Linear Hierárquico o comportamento do peso corporal dos sujeitos no experimento realizado por Ferreira et al. (2018) em diferentes grupos de tratamentos, divididos em números distintos de gaiolas e em um tempo determinado de semanas. Para isso foram propostos os seguintes objetivos específicos:

- Análise descritiva dos pesos iniciais e no decorrer do tratamento;
- Ajustar um modelo que explique o comportamento dos dados;
- Validar o modelo através dos diagnósticos;
- Avaliar se há pontos de *outliers*;
- Se os pontos de *outliers* são observações influentes;
- E como essas observações influentes afetam o modelo proposto.

O capítulo 1 apresenta uma breve revisão de literatura acerca dos Modelos Mistos Lineares e demais nomenclaturas que o mesmo recebe. O capítulo 2, intitulado materiais e métodos, descreve o desenho do experimento e os métodos apropriados no trabalho (Modelos Lineares Hierárquicos, processo de estimação e diagnósticos). O capítulo 3 apresenta os resultados e discussões da análise descritiva, a escolha do Modelo e os diagnósticos. O capítulo 4 elenca as conclusões finais. E, por fim, o capítulo 5 apresenta as propostas futuras de trabalho.

Capítulo 1

Revisão de Literatura

Os Modelos Mistos foram amplamente estudado por Fisher em 1918, com grande repercussão nos estudos de genética quantitativa, cujos estudos foram denominados pelo autor como modelos de componentes de variância ([SCHEFFE, 1959](#)). O termo Modelo Linear Hierárquico, foi inserido pela primeira vez por [Lindley e Smith \(1972\)](#) como parte de suas contribuições em estimação bayesiana de modelos lineares; neste âmbito os autores elaboraram uma modelagem para dados hierárquicos englobando erros com estruturas complexas. Já o desenvolvimento dos Modelos Lineares Mistos na forma combinada em uma única equação, se deve aos trabalhos primordiais [Harville \(1976\)](#), [Harville \(1977\)](#) que facilitaram tal feito, que posteriormente seu uso foi discutido em ([LAIRD; WARE, 1982](#)) para dados longitudinais, que são dados caracterizados pela sequência temporal de duas ou mais observações em cada indivíduo.

Os dados longitudinais apresentam estrutura hierárquica, uma vez que as medidas repetidas são aninhadas dentro do indivíduo, tal estrutura hierárquica faz com que possamos fazer a suposição de que as observações entre os indivíduos sejam independentes e que as aninhadas no indivíduo possuam a característica da dependência com erros correlacionados ([FAUSTO et al., 2008](#)).

Dados que possuem estrutura aninhada, com unidades que possuem diferentes níveis de hierarquia devem ser analisados usando um modelo multinível. Os Modelos Multiníveis levam em conta a variabilidade dos dados existentes dentro de cada nível hierárquico e entre os níveis hierárquicos ([PINHEIRO, 2005](#)). A figura abaixo apresenta esquematicamente os níveis hierárquicos provenientes da estrutura aninhada dos dados.

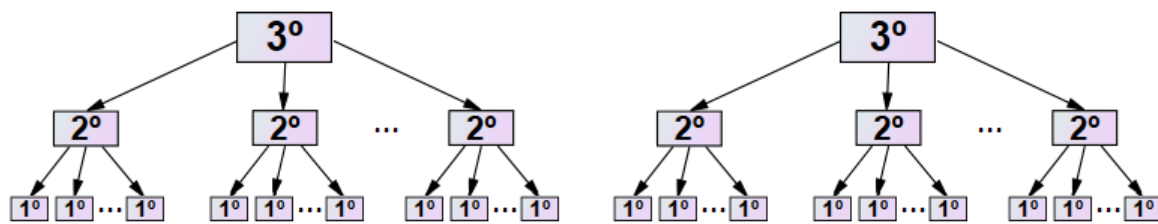


Figura 1.0.1 – Níveis hierárquicos

Até o final da década de 80, a grande maioria da modelagem realizada não levava em consideração como os dados estavam estruturados e, portanto, as análises para os dados não era realizada de maneira apropriada. Em parte, isto se devia à falta de métodos e *softwares* que viabilizassem a generalização da abordagem multinível ou hierárquica. Assim, o pesquisador deveria definir uma unidade sobre a qual o seu estudo iria ser aplicado.

Como exemplo da desconsideração da estrutura dos dados, temos o ocorrido na Inglaterra onde [Bennett \(1977\)](#) demonstrou que crianças tinham um aprendizado maior quando aprendiam a ler de uma maneira tradicional, em relação as demais crianças do ensino fundamental. A análise dos dados era baseada na regressão múltipla comum, cujas crianças eram consideradas unidades analisadas e os grupos de classes escolares e professores eram inexplorados. Posteriormente e [AITKIN; ANDERSON; HINDE, 1981](#)) foi comprovado que as diferenças significantes observadas em [Bennett \(1977\)](#) desapareciam, ao nortear a análise agrupando as crianças em classes escolares; sendo assim, as crianças que foram sujeitas ao ensino tradicional não apresentaram qualquer disparidade em referência às outras crianças.

Para [Bergamo \(2002\)](#), quando a finalidade é compreender tudo o que é abrangido em um determinado evento, é preciso ajustar um modelo que leve em conta toda a variabilidade entre os experimentos e consequentemente englobe os diferentes aspectos de cada um deles. Considerar os dados de acordo com uma estrutura aninhada leva em conta tal tipo de análise. O termo análise multinível, de acordo com [Diez-Roux \(2000\)](#), tem sido utilizado em diversos campos do conhecimento, além da educação, como demografia e sociologia entre outros, com intuito de descrever uma abordagem analítica que permita averiguar conjuntamente os efeitos exercidos sobre a variável resposta, onde esses efeitos são oriundos das variáveis do nível do grupo e do nível do indivíduo.

Com o passar dos anos, muitos *softwares* estatísticos foram desenvolvidos com capacidade de se adequarem aos modelos lineares hierárquicos, contornando, assim, a escassez sofrida na década de 80. Podemos destacar alguns, que são especializados e dedicados somente a essa classe de modelo, como **HLM** e **MLwiN**, enquanto **PROC MIXED**

no **SAS**, **xtmixed** e **gllamm** em **Stata** e **nlmn**, **lme4** em **R** são pacotes complementares de *softwares* de aspectos mais amplos (WEST; GALECKI, 2011). Em consonância com o desenvolvimento dos *softwares* estatísticos, segundo Loy (2013) houveram inúmeros progressos nos procedimentos de diagnóstico para os modelos lineares hierárquicos, que se concentraram, principalmente, na busca de formulação para detecção de *outliers* e pontos influentes em cada nível desses modelos.

Um exemplo da utilização de análise de diagnóstico em modelos mistos lineares, foi dada em problemas ¹atuariais práticos em (PINHO; NOBRE; FREITAS, 2012), que utilizaram o conjunto de dados descritos por Hachemeister no ano de 1975, em que justificaram que métodos diagnósticos podem ajudar a melhorar a escolha do modelo e identificar *outliers* ou assuntos influentes que mereçam melhor atenção das seguradoras.

Embora esses desenvolvimentos em procedimentos de diagnósticos melhorem muito a capacidade de um analista de verificar um modelo ajustado, sua incorporação em *softwares* estatísticos ficaram para trás. Essa defasagem para com os *softwares* foi resolvida em (LOY, 2013), que apresenta um pacote para o *software R*, chamado **HLMdiag**, que incorpora tanto as ferramentas necessárias para a análise de resíduos, quanto as necessárias para a análise de influentes, para modelos lineares hierárquicos de todos os aspectos e níveis com estruturas de dependência.

Por questão de padronização utilizaremos o termo Modelos Lineares Hierárquicos no decorrer deste trabalho para refletir a organização hierárquica dos dados, já que são comumente referidos por Goldstein (1995), Raudenbush e Bryk (2002). Os Modelos Lineares Hierárquicos podem ainda ser encontrados na literatura sob outros títulos como Modelos Lineares Multinível, Modelos Lineares de Efeitos Mistos, Modelos de Efeitos Aleatórios, Modelos de Regressão com Coeficientes Aleatórios e Modelos de Componentes de Variância (NATIS, 2001).

¹ A ciência atuarial é a ciência das técnicas específicas de análise de riscos e expectativas, principalmente na administração de seguros e fundos de pensão.

Capítulo 2

Materiais e Métodos

2.1 Experimento

O experimento descrito neste trabalho foi realizado por (FERREIRA et al., 2018), os quais foram conduzidos por ensaios cegos, controlados e aleatorizados. Como objetivo buscou-se verificar a eficiência da utilização de variações de um bioterápico produzido à partir de soro de *Gallus gallus domesticus* (galinha) sob parâmetro parasitológicos, clínicos e imunológicos em camundongos experimentalmente infectados por *Trypanosoma cruzi*. A eficiência esperada de cada bioterápico seria observado no peso corporal dos sujeitos.

Ainda segundo os autores, foram utilizados no experimento 57 camundongos suíços, machos com 56 dias de idade, provenientes do Biotério Central da Universidade Estadual de Maringá. Os animais foram distribuídos em grupos de tratamentos e mantidos em gaiolas com no máximo 5 animais. As gaiolas são microambientadas de polisulfona (ALESCO®), com dimensão de $20 \times 32 \times 21 \text{ cm}^3$, com temperatura controlada ($22 \pm 2^\circ \text{C}$), com ciclo de claro/escuro de 12 horas, recebendo água e ração *ad libitum*. Todos os grupos foram mantidos nas mesmas condições experimentais.

A Tabela 1 apresenta a quantidade de sujeitos em cada grupo e como foram alocados nas gaiolas. Os animais infectados foram inoculados por via intraperitoneal, com 1.400 tripomastigotas sanguíneos (formas infectantes) de *Trypanosoma cruzi*-cepa Y (NUSSENZWEIG et al., 1953).

Tabela 1 – Desenho do experimento

Grupo	n ° de sujeitos	Gaiola
1	5	1
2	13	2, 3, 4
3	13	5, 6, 7
4	13	8, 9, 10
5	13	11, 12, 13, 14

Os grupos experimentais (tratamentos) foram os seguintes:

- 1 = CNI - (Controle não infectado) Animais não infectados e não tratados ($n = 5$);
- 2 = G13cH - Animais tratados com bioterápico de soro de galinha 13cH ($n = 13$);
- 3 = G6cH - Animais tratados com bioterápico de soro de galinha 6cH ($n = 13$);
- 4 = GCI - (Controle infecção) Os animais foram infectados e não receberam tratamento ($n = 13$);
- 5 = G3cH - Animais tratados com bioterápico de soro de galinha 3cH ($n = 13$).

O medicamento foi diluído em água (1mL/10mL) e fornecido *ad libitum* em garrafa âmbar estéril de acordo com [Aleixo et al. \(2013\)](#), durante 16 horas consecutivas (medicamento disponível para os animais das 16:00 às 08:00), no 4º, 7º e 10º dia após a infecção (totalizando 3 doses). O esquema de tratamento se baseia na ação do medicamento que está ligada aos seus efeitos imunológicos e à evolução específica da cepa Y de *Trypanosoma cruzi* em camundongos Swiss ([ALEIXO et al., 2013](#); [FERRAZ et al., 2016](#)). O projeto para a execução dos experimentos foi aprovado pelo Comitê de Ética em Pesquisa com Animais da Universidade Estadual de Maringá, Paraná, parecer CEUA 2401220716/2016.

A evolução do peso corporal foi acompanhada durante 12 semanas e este foi mensurado em balança semianalítica (Balance BEL®).

Neste estudo, a avaliação realizada no início do tratamento foi utilizada para analisar o peso inicial dos camundongos. Devido à grande perda de informação a partir da 7ª semana de estudo, a modelagem foi realizada considerando as medidas longitudinais observadas até a 6ª semana.

2.2 Modelos Lineares Mistos

Para a modelagem do peso dos camundongos (y) em função das variáveis preditoras (Z) ao nível do indivíduo e/ou a um nível superior (W), assume-se que o conjunto de da-

dos seja multinível, conceitualmente o modelo pode ser visto como um sistema hierárquico de equações de regressão (PINHEIRO, 2005).

Segundo Raudenbush e Bryk (2002), o Modelo Linear Hierárquico mais simples possível é o modelo com apenas o intercepto aleatório ou análise de variância (ANOVA) com efeitos aleatórios. Assume-se que não há variável explicativa em nenhum dos dois níveis, sendo assim temos que $Z_i = 1$ e $W_i = 1$. Este modelo é o primeiro passo na análise de dados hierárquicos e sua equação é dada por:

$$y_{ij} = \beta_i + b_i + e_{ij} \quad (2.2.1)$$

Com $b_i \sim N(0, \sigma_b^2)$, $e_{ij} \sim N(0, \sigma^2)$ e j denotando o j -ésimo indivíduo ($j = 1, \dots, n_i$) de cada grupo i ($i = 1, \dots, M$).

A equação (2.2.1) é um Modelo de Efeitos Aleatórios porque os efeitos dos grupos (b_i) são interpretados como aleatórios. Este Modelo decompõe a variância em duas componentes independentes: σ^2 , que é a variância dos erros do nível 1 (do indivíduo), aqui denominado e_{ij} ; e σ_b^2 , que é a variância dos erros do nível 2 (do grupo), definidos por b_i .

$$Var(y_i) = Var(b_i + e_{ij}) = \sigma_b^2 + \sigma^2 \quad (2.2.2)$$

O Modelo Linear Hierárquico na sua forma geral, pode ser formulado através de duas equações:

$$y_i = Z_i \beta_i + e_i \quad (2.2.3)$$

$$\beta_i = W_i \gamma + b_i \quad (2.2.4)$$

em que (2.2.3) representa a equação do Modelo dentro do grupo, dito nível 1 e (2.2.4) representa o Modelo entre grupos, dito nível 2. As variáveis explicativas à nível do sujeito são representadas pelo índice (q), onde ($q = 1, \dots, Q$) e as variáveis explicativas à nível de grupo são representadas pelo índice (p), onde ($p=1, \dots, P$), logo (q) (p) representam o número de variáveis. Assim sendo temos:

- y_i é um vetor de resultado com dimensão ($n \times 1$);
- Z_i é uma matriz de variáveis explicativas com dimensão ($j \times q$);
- β_i é um dos componentes do vetor de parâmetros fixos desconhecidos com dimensão ($q \times 1$);
- e_i é o erro do nível 1;
- W_i é uma matriz de variáveis explicativas de nível 2 com dimensão de ($q \times p$);

- γ é um vetor de efeitos fixos com dimensão $(p \times 1)$;
- b_i é um vetor de efeitos aleatórios com dimensão $(q \times 1)$.

Tal modelo pode ser combinado, assim obtemos um único Modelo Linear Misto, sendo $X_i = Z_i W_i$ e $\beta = \gamma$

$$y_i = X_i \beta + Z_i b_i + e_i \quad (2.2.5)$$

Comumente referido como Modelo Hierárquico de (VERBEKE; MOLENBERGHS, 2000). $X_i \beta$ corresponde a parte fixa, com β representando o vetor dos componentes β_i e $Z_i b_i + e_i$ corresponde a parte aleatória do modelo.

Desta forma, a partir do modelo dado por (2.2.5), os erros serão assumidos como independentes entre grupos e diferentes entre níveis, isto é, se assume que os efeitos aleatórios e os erros seguem uma distribuição normal com média zero e possuem resíduos correlacionados, com as matrizes de covariâncias (D) e (R), onde ambas matrizes são positivas definidas por hipótese, logo não singulares, dados como:

$$Var(Y) = \begin{bmatrix} b_i \\ e_i \end{bmatrix} = \begin{bmatrix} D & \phi \\ \phi & R_i \end{bmatrix} \quad (2.2.6)$$

onde

$$Var(b_i) = E[b_i b_i^T] = D \quad (2.2.7)$$

e

$$Var(e_i) = E[e_i e_i^T] = R_i \quad (2.2.8)$$

Com as seguintes pressuposições

$$e_i \sim N(0, R_i) \quad \text{e} \quad b_i \sim N(0, D)$$

e

$$Cov(e_i, b_j) = 0 \quad \forall \quad i; j$$

e $Cov(b_i, b_j)$ podendo ser zero para $i \neq j$.

Frequentemente, assume-se, que $R_i = \sigma^2 I_{ni}$. Assim a variância total do modelo (2.2.5) para o vetor de resposta y_i é dado como

$$V(y_i) = Z_i D Z_i^T + R_i \quad (2.2.9)$$

onde tais considerações implicam que

$$y_i \sim N(X_i \beta, V_i) \quad (2.2.10)$$

Na maioria das vezes segundo [Gurka e Edwards \(2007\)](#), os analistas estão interessados em inferência sobre os parâmetros de efeitos fixos, β e possivelmente nos "componentes de variância", de b_i . Nesse cenário, o modelo (2.2.5) com

$$\begin{aligned}\varepsilon_i &= Z_i b_i + e_i \\ y_i &= X_i \beta + \varepsilon_i\end{aligned}\tag{2.2.11}$$

é muitas vezes referido como o modelo marginal de [Verbeke e Molenberghs \(2000\)](#).

Ainda sob a perspectiva de [Gurka e Edwards \(2007\)](#), o uso do Modelo Marginal não implica que os efeitos aleatórios sejam dispensáveis para tal análise. Longe disso, a modelagem apropriada dos efeitos aleatórios proporciona uma maneira tipicamente intuitiva de modelar a variação de dados complexos que permite estimativas precisas e inferências sobre os parâmetros de interesse, β .

De forma recorrente, os efeitos aleatórios também merecem sua devida atenção, quando o modelo (2.2.5) é visto como hierárquico, podemos definir os efeitos aleatórios com pressupostos distributivos do Modelo Condicional em b_i , ao invés de simplesmente os desconsiderar do modelo. Assim sendo temos

$$\begin{aligned}y_i | b_i &\sim N(X_i \beta + Z_i b_i, R_i) \\ b_i &\sim N(0, D)\end{aligned}\tag{2.2.12}$$

com $y_i \sim N(X_i \beta, V_i)$ que é a mesma suposição distributiva do Modelo Marginal. No entanto, o Modelo Marginal e o Modelo Hierárquico não são equivalentes, pelo menos em termos de interpretação e utilidade ([GURKA; EDWARDS, 2007](#)).

A metodologia recomendada para a estimação dos efeitos fixos, juntamente com a estimação dos componentes de variância e covariâncias dos efeitos aleatórios, tem sido um dos problemas centrais do ajuste dos Modelos Mistos a um conjunto de observações. Assim, no uso de Modelos Fixos, deve-se estimar os próprios efeitos fixos, enquanto que os Modelos Aleatórios prestam-se para estimar os componentes de variância (das variáveis aleatórias), bem como para a predição das próprias variáveis aleatórias ([FILHO, 2002](#)).

2.3 Métodos de Estimação

Quando há um Modelo Linear Hierárquico, na forma dada em (2.2.5) com matriz de variâncias e covariâncias dadas como em (2.2.9), há geralmente interesse em estimar os efeitos fixos e efeitos aleatórios e na estimação dos componentes de variância. Geralmente como procedimento de estimação de Modelos Lineares Hierárquicos, os *softwares* estatísticos utilizam dois procedimentos: a Máxima Verossimilhança (MV), que inclui na função de

verossimilhança os efeitos fixos e as componentes de variância e o de Máxima Verossimilhança Restrita (MVR), que inclui apenas as componentes de variância na função de verossimilhança (PINHEIRO, 2005).

Segundo Hox (1995) a diferença entre os métodos de estimação é que o procedimento de Máxima Verossimilhança trata as estimativas dos coeficientes de regressão como quantidades que são conhecidas quando estimam as componentes de variância, e o procedimento de Máxima Verossimilhança Restrito trata estas estimativas como quantidades que carregam uma certa incerteza. Na prática, a diferença entre os dois métodos de estimação não é muito grande.

2.3.1 Máxima Verossimilhança (MV)

O método da Máxima Verossimilhança consiste em maximizar a função de verossimilhança das observações em relação aos efeitos fixos e aos componentes de variância, exigindo a suposição de normalidade dos dados. Assumindo que o vetor \mathbf{y}_i das observações tem média $\mathbf{X}_i\beta_i$ e matriz de variâncias e covariâncias \mathbf{V}_i , temos que a função de verossimilhança (L) de \mathbf{y}_i é dada por:

$$L_{MV}(y_n; \beta_q) = \prod_{i=1}^N \left\{ (2\pi)^{(-n_i/2)} |\mathbf{V}_i|^{(1/2)} \times \exp \left[-\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i\beta)^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\beta) \right] \right\} \quad (2.3.1)$$

onde $|\mathbf{V}_i|$ representa o determinante da matriz \mathbf{V}_i .

Fazendo $\ell_{MV}(y_n; \beta_q) = \log(L_{MV}(y_n; \beta_q))$, obtemos a log-verossimilhança dado como:

$$\ell_{MV}(y_n; \beta_q) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^M \log|\mathbf{V}_i| - \frac{1}{2} \sum_{i=1}^M (\mathbf{y}_i - \mathbf{X}_i\beta)^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\beta) \quad (2.3.2)$$

Aplicando as derivadas parciais em relação a β e em relação aos componentes σ_{ij}^2 de \mathbf{V}_i e igualando a zero ambas, a fim de maximizar ℓ obtemos um sistema de equações

$$\begin{cases} \mathbf{X}_i^T \widehat{\mathbf{V}}^{-1} \mathbf{X}_i \hat{\beta} = \mathbf{X}_i^T \widehat{\mathbf{V}}^{-1} \mathbf{y} \\ \text{tr} \left(\widehat{\mathbf{V}}^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \widehat{\mathbf{V}}^{-1} \mathbf{Z}_j \mathbf{Z}_j^T \right) \hat{\sigma}_i^2 = \mathbf{y}^T \widehat{\mathbf{P}} \mathbf{Z}_i \mathbf{Z}_i^T \widehat{\mathbf{P}} \mathbf{y} \end{cases} \quad (2.3.3)$$

que, resolvido de forma iterativa para $\hat{\beta}$ e $\hat{\sigma}_i^2$, fornece os estimadores de máxima verossimilhança para os efeitos fixos β e para os componentes de variância σ_i^2 .

Sendo que:

- tr denota o traço da matriz;

- e P o projetor, que é definido como:

$$P = V^{-1} - V^{-1}X(X^T V^{-1}X)^{-1}X^T V^{-1} \quad (2.3.4)$$

onde V é a variância total do modelo.

Podemos ver algumas propriedades desses estimadores em [Searle \(1987\)](#), destacando a invariância e que tais estimadores fornecem sempre estimativas não negativas de componentes de variância, mas essas são viesadas porque o método não considera a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo.

A fim de estimar o parâmetro de b_i podemos utilizar, segundo [Filho \(2002\)](#), a função densidade de probabilidade conjunta $f(y_i, b_i)$, que pode ser obtida através do produto entre a função densidade condicional $f(y_i|b_i)$, e a função densidade de probabilidade de $f(b_i)$. Sendo assim temos:

$$f(y_i, b_i) = f(y_i|b_i) \cdot f(b_i) \quad (2.3.5)$$

como

$$f(y, b_i) = \left[\frac{\exp[-\frac{1}{2}(y_i - X_i\beta - Z_i b_i)^T (R_i)^{-1}(y_i - X_i\beta - Z_i b_i)]}{(2\pi)^{\frac{N}{2}} |R|^{\frac{1}{2}}} \right] \cdot \left[\frac{\exp[-\frac{1}{2}(b_i - 0)^T (D)^{-1}(b_i - 0)]}{(2\pi)^{\frac{1}{2}} |D|^{\frac{1}{2}}} \right] \quad (2.3.6)$$

com $|R_i|$ e $|D|$ sendo os determinantes das matrizes de covariâncias. E a função de log-verossimilhança ($\ell_{MV}(\beta, b_i)$) é dada como:

$$\begin{aligned} \ell_{MV}(\beta, b_i) = & \frac{1}{2} \left[2N \log(2\pi) - (\log(R_i) + \log(D)) \right] - \frac{1}{2} \left[y_i^T R_i^{-1} y_i - 2y_i^T R_i^{-1} X_i \beta \right. \\ & - 2y_i^T R_i^{-1} Z_i b_i + 2\beta^T X_i^T R_i^{-1} Z_i b_i + \beta^T X_i^T R_i^{-1} X_i \beta \\ & \left. + b_i^T Z_i^T R_i^{-1} Z_i b_i + b_i^T D^{-1} b_i \right] \end{aligned} \quad (2.3.7)$$

aplicando as derivadas parciais em relação a β e b_i , e igualando a zero, obtêm-se:

$$\begin{bmatrix} \frac{\partial \ell}{\partial \beta} \\ \frac{\partial \ell}{\partial b_i} \end{bmatrix} = \begin{bmatrix} -X_i^T R_i^{-1} y_i + X_i^T R_i^{-1} X_i \hat{\beta} + X_i^T R_i^{-1} Z_i \hat{b}_i \\ -Z_i^T R_i^{-1} y_i + Z_i^T R_i^{-1} X_i \hat{\beta} + Z_i^T R_i^{-1} Z_i \hat{b}_i + D^{-1} \hat{b}_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

de forma equivalente

$$\begin{bmatrix} X_i^T R_i^{-1} X_i \hat{\beta} + X_i^T R_i^{-1} Z_i \hat{b}_i \\ Z_i^T R_i^{-1} X_i \hat{\beta} + Z_i^T R_i^{-1} Z_i \hat{b}_i + D^{-1} \hat{b}_i \end{bmatrix} = \begin{bmatrix} X_i^T R_i^{-1} y_i \\ Z_i^T R_i^{-1} y_i \end{bmatrix}$$

simplificando

$$\begin{bmatrix} \mathbf{X}_i^T \mathbf{R}_i^{-1} \mathbf{X}_i + \mathbf{X}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i \\ \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{X}_i + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{b}_i \end{bmatrix} = \begin{bmatrix} \mathbf{X}_i^T \mathbf{R}_i^{-1} \mathbf{y}_i \\ \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{y}_i \end{bmatrix} \quad (2.3.8)$$

Essas são as Equações de Modelos Mistos (EMM), que permitem obter as estimativas para os efeitos fixos $\hat{\beta}$ e predições para os efeitos aleatórios \hat{b}_i . Tais equações podem ser escritas como:

$$\begin{bmatrix} \hat{\beta} \\ \hat{b}_i \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{y}_i \\ \hat{\mathbf{D}} \mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}) \end{bmatrix} = \hat{\mathbf{C}} \begin{bmatrix} \mathbf{X}_i^T \hat{\mathbf{R}}_i^{-1} \mathbf{y}_i \\ \mathbf{Z}_i^T \hat{\mathbf{R}}_i^{-1} \mathbf{y}_i \end{bmatrix} \quad (2.3.9)$$

onde

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}_i^T \hat{\mathbf{R}}_i^{-1} \mathbf{X}_i & \mathbf{X}_i^T \hat{\mathbf{R}}_i^{-1} \mathbf{Z}_i \\ \mathbf{Z}_i^T \hat{\mathbf{R}}_i^{-1} \mathbf{X}_i & \mathbf{Z}_i^T \hat{\mathbf{R}}_i^{-1} \mathbf{Z}_i \hat{\mathbf{D}}^{-1} \end{bmatrix}^{-1}$$

Segundo [Henderson et al. \(1959\)](#), \mathbf{V} é dado como,

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{Z} (\mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{D}^{-1})^{-1} \mathbf{Z}^T \mathbf{R}^{-1} \quad (2.3.10)$$

Qualquer inversa generalizada da matriz de coeficientes nas equações do Modelo Linear Hierárquico, produz a matriz de covariâncias \mathbf{C} . Se as matrizes \mathbf{D} e \mathbf{R}_i são conhecidas, $\hat{\beta}$ é o melhor estimador linear não viesado (BLUE) de β e \hat{b}_i é o melhor preditor linear não viesado (BLUP) de b_i . No entanto, $\hat{\mathbf{C}}$ pode ser obtida através da substituição das estimativas de \mathbf{D} e \mathbf{R}_i em \mathbf{C} , que possibilita utilizar o teste da razão de verossimilhança que é aconselhado, embora seja possível construir estatísticas t e F aproximadas ([WOLFINGER, 1993](#) apud [FILHO, 2002](#)).

Outra propriedade vantajosa dos estimadores de Máxima Verossimilhança é que, à medida que o tamanho das amostras aumenta, eles podem ser facilmente estimados. Assim, mesmo que o método para obter o estimador MV seja iterativo, a distribuição para grandes amostras do estimador é bem definida.

2.3.2 Máxima Verossimilhança Restrita (MVR)

Foi [Harville \(1974\)](#) que introduziu o uso da Verossimilhança Restrita, desenvolvido por [Patterson e Thompson \(1971\)](#) para estimar os parâmetros de covariância do Modelo Linear Misto. As estimativas dos parâmetros no MV têm um viés decrescente, uma vez que a perda de graus de liberdade resultante da estimativa dos efeitos fixos não é levada em

consideração. A estimativa de MVR reconhece essa perda de graus de liberdade e, portanto, leva à estimativas menos tendenciosas (GURKA; EDWARDS, 2007).

O estimador MVR é calculado maximizando a função de verossimilhança de um conjunto de contrastes de erro de \mathbf{y}_i , dado como $\mathbf{U} = \mathbf{A}^T \mathbf{y}_i$, onde \mathbf{A} é qualquer matriz posto completo ($n \times (n - p)$) com colunas ortogonais às colunas da matriz \mathbf{X} e escolhido de modo que $E(\mathbf{U}) = 0$. O vetor \mathbf{U} então segue uma distribuição normal com média zero e covariância $\mathbf{A}^T \mathbf{V} \mathbf{A}$ (VERBEKE; MOLENBERGHS, 2000).

Segundo Gurka e Edwards (2007), a função resultante da maximização da função de verossimilhança de um conjunto de contrastes de erro, não dependente de β , é baseada em uma transformação das observações originais que levam a um novo conjunto de observações ($N - p$).

Harville (1974) também mostrou que a função de log-verossimilhança restrita pode ser escrita da seguinte forma com base nas observações originais:

$$\begin{aligned} \ell_{MVR}(\beta; b_i) = & -\frac{N-p}{2} \log(2\pi) + \frac{1}{2} \log \left| \sum_{i=1}^M X_i^T X_i \right| - \frac{1}{2} \sum_{i=1}^M \log |V_i| \\ & - \frac{1}{2} \log \left| \sum_{i=1}^M X_i^T V_i^{-1} X_i \right| - \frac{1}{2} \sum_{i=1}^M (y_i - X_i \hat{\beta})^T V_i^{-1} (y_i - X_i \hat{\beta}) \end{aligned} \quad (2.3.11)$$

$\hat{\beta}$ é dado pela equação 2.3.9.

Os estimadores MVR para os componentes de variâncias segundo Marcelino e lemma (2000), também podem ser obtidos pela resolução das equações:

$$tr \left[\hat{\mathbf{P}} \mathbf{Z}_i \mathbf{Z}_i^T \hat{\mathbf{P}} \mathbf{Z}_j \mathbf{Z}_j^T \right] \sigma_i^2 = \left[\mathbf{y}_i^T \hat{\mathbf{P}} \mathbf{Z}_i \mathbf{Z}_i^T \hat{\mathbf{P}} \mathbf{y} \right] \quad (2.3.12)$$

Veja que essas equações são similares às equações de Máxima Verossimilhança, exceto pelo projetor $\hat{\mathbf{P}}$ em vez de $\hat{\mathbf{V}}^{-1}$, para o operador traço.

À princípio, o MVR era pouco empregado na prática, devido a suas exigências de cálculo. Entretanto, com o avanço da informática, vários algoritmos iterativos têm sido desenvolvidos para sanar tal dificuldade (SEARLE; CASELLA; MCCULLOCH, 2009). Mesmo com tal entrave, o MVR apresenta propriedades muito desejáveis, como exemplo podemos citar a geração de estimadores únicos.

2.4 Diagnósticos do Modelo

O Modelo Linear Hierárquico, assim como na Regressão Linear Ordinária, tem suposições distributivas que podem ou não ser válidas quando utilizados na prática. Entretanto,

os diagnósticos para avaliar essas suposições e as consequentes alternativas quando suspeitas de violações das suposições, não foram desenvolvidas totalmente para tal modelo, principalmente devido à ferramenta de análise ser relativamente recente.

2.4.1 Análise de Resíduos

Com o intuito de verificar a validade das suposições do Modelo Estatístico, o uso dos resíduos se faz necessário. Como no Modelo Linear Misto, existem mais de uma fonte de variação, e, conseqüentemente, mais de um tipo de resíduo, a forma geral de Cox e Snell (1968) para modelos com uma única fonte de variação pode ser estendida para definir resíduos do Modelo Linear Misto (NOBRE; SINGER, 2004).

De acordo com Loy (2013), um termo residual deve possibilitar a apreciação das premissas distribucionais do modelo, assim consideramos como essenciais. Conseqüentemente, independente da forma da matriz de covariância esses resíduos são de interesse.

Para a análise de resíduos o modelo dado em 2.2.5, engloba a incerteza tanto no nível individual e_i , como no nível do grupo b_i .

- Resíduos de nível 1 (individual) \hat{e}_i
- Resíduos de nível 2 (grupo) \hat{b}_i

Embora os resíduos de nível 1 e 2 sejam fundamentais para a modelagem é interessante verificar se suas definições estão interrelacionadas, pois uma deficiência de um nível do modelo pode ser percebida na análise residual em outro nível. Recomenda-se uma análise residual crescente, de tal forma que seja verificado a validação dos resíduos de nível 1, assim sendo, concluído o modelo adequado para esse nível, mude para o nível 2. Com isso, os impactos na análise residual, provenientes de um possível confundimento na hora de modelar, serão minimizados (LOY, 2013).

Conforme Loy (2013) sugere, uma forma de indicar se a variação residual dentro do grupo é constante entre os grupos é utilizar a inferência visual por meio do uso do protocolo de programação apresentado por Buja et al. (2009), em tal protocolo, introduz de forma aleatória o gráfico construído a partir dos dados observados em uma grade com outros gráficos nulos, em que esses gráficos nulos são construídos a partir de um número de dados conjuntos simulados.

Se for possível para o pesquisador identificar visualmente o gráfico real entre os demais gráficos nulos, tem-se evidência contra a hipótese nula (que a variação residual dentro do grupo é constante entre os grupos) e podemos assumir que os dados possivelmente se desviam de pelo menos uma suposição sobre os pressupostos do Modelo.

2.4.1.1 Resíduos de nível 1

Dependendo do método que se utiliza para a estimação dos parâmetros de um Modelo, pode-se causar variação nos valores dos diferentes tipos de resíduos deste Modelo. Assim, a estimativa dos parâmetros de efeitos fixos causa dependência nos resíduos de nível 1, logo, através do Modelo dado em (2.2.5) podemos defini-los de modo direto como:

$$\hat{e}_i = \hat{y}_i - X_i\hat{\beta} - Z\hat{b}_i \quad (2.4.1)$$

Os resíduos de nível 1 encontrados usando a estimativa de \hat{V}_i são frequentemente chamados de Resíduos Empíricos de Bayes (EB) nível 1, que são estabelecidos como os modos condicionais dos b_i , dados os valores estimados dos parâmetros. Os resíduos do EB em cada nível estão inter-relacionados, causando confundimento entre os níveis, sendo assim, os resíduos de mínimos quadrados (MQ) se tornam preferíveis para o nível 1 (LOY, 2013).

O primeiro conjunto de resíduos provem do ajuste dos Modelos Lineares de forma separada para cada grupo, tais ajustes se dão afim de estimar β e b_i através dos modelos de regressão de mínimos quadrados.

Ajustando Modelos Lineares de forma separada para cada grupo, através dos Modelos de Regressão de Mínimos Quadrados, afim de estimar β e b_i gera o primeiro conjunto de resíduos.

A vantagem desta forma de estimação é que os resíduos dependem apenas do nível mais baixo da hierarquia (nível 1). Assim, os resíduos resultantes do nível 1 do MQ, não são confundidos pelos resíduos no nível 2 (HILDEN-MINTON, 1995). Mesmo não havendo tal confundimento entre os níveis, a utilização dos resíduos EB é recomendada se tivermos tamanhos pequenos de amostras dentro do grupo (LOY, 2013).

2.4.1.2 Resíduos de Nível 2

Visto que um Modelo apropriado de nível 1 foi encontrado, inicia-se a avaliação das premissas do Modelo ajustado nos efeitos aleatórios de nível 2. Neste nível, inclinação e intercepto podem surgir como efeitos aleatórios, dentre outros, onde cada um deve ser avaliado. As estimativas do efeito aleatório b_i provenientes do HLM apresentada na equação 2.3.9 são ditas resíduos empíricos de Bayes de nível 2.

Os resíduos de MQ são obtidos como diferenças entre os coeficientes de regressão média individual e populacional encontrados por MQ, que devem ser usados apenas no caso de tamanho de amostra suficiente dentro de grupos (LOY, 2013). Os resíduos de EB são BLUPs se o modelo for apropriado em todos os níveis (HILDEN-MINTON, 1995).

Com o intuito de avaliar a acomodação e plenitude da estrutura de efeitos aleatórios (Z_i), investiga-se seus elementos e suas relações com o elemento sistemático (X_i). Diante disto, para cada componente na equação 2.3.9 de b_i em todos os grupos i serão utilizadas parcelas residuais.

Se deve ter cuidado ao interpretar a heterogeneidade observada entre os grupos, a variância EB dos resíduos do nível 2 pode depender dos efeitos fixos estimados (GOLDSTEIN, 1995). Assim sendo, a heterogeneidade de variância pode ser um artifício disso. Resíduos padronizados são utilizados para uma possível avaliação de homogeneidade de parcela única. Dentre as diversas sugestões presentes na literatura de como estimar a variância dos resíduos EB de nível 2, ($Var(\hat{b}_i)$) destacamos a estimativa proveniente da equação (2.2.5) que é dada por:

$$Var(\hat{b}_i) = \widehat{D}Z_i \left(\widehat{V}_i^{-1} - \widehat{V}_i^{-1}X_i \left(\sum_{i=1}^m X_i' \widehat{V}_i^{-1} X_i \right)^{-1} X_i' \widehat{V}_i^{-1} \right) Z_i \widehat{D} \quad (2.4.2)$$

Como observado por Laird e Ware (1982), o Modelo (2.4.2) subestima a variabilidade ($\hat{b}_i - b_i$) porque ignora a variação de (b_i). Consequentemente, a inferência sobre (b_i) é tipicamente baseado em:

$$Var(\hat{b}_i - b_i) = D - Var(\hat{b}_i) \quad (2.4.3)$$

que é denominada como covariância condicional (GURKA; EDWARDS, 2007).

Uma proposta alternativa é a variância marginal

$$Var(\hat{b}_i) = \widehat{D}Z_i^T \widehat{V}_i^{-1} Z_i \widehat{D} \quad (2.4.4)$$

onde ambas são denominadas e propostas respectivamente por Goldstein (1995).

Como verificação do Modelo, recomenda-se a padronização dos resíduos, com base no pressuposto de que a variação contribuída pelos efeitos fixos é pequena, considerando-se grandes amostras. Já para inferência de efeitos aleatórios, aconselha-se a utilização da covariância condicional. Gráficos residuais como exemplo (resíduos, valores ajustados), quando não indicam nenhuma tendência no comportamento dos pontos, podem então ser usados para detectar heterogeneidade, se forem utilizados os resíduos padronizados EB de nível 2.

2.4.2 Análise de Observações Influentes

No ajuste dos Modelos e nas estimativas de Parâmetros não são todas as observações, ou grupos, que têm o mesmo efeito. Algumas observações ou grupos podem ter efeitos exagerados no decorrer do processo de ajuste de Modelo. Essas observações ou

grupos são ditos influentes ou pontos de alavanca ("leverage"). Especialmente, estamos interessados nas observações influentes sobre os valores ajustados, os efeitos fixos (estimativas) e os componentes de variância.

Nos últimos anos, a análise de influentes dos Modelos Lineares Hierárquicos tem sido estudada, comumente sob a ótica dos Modelos Lineares Mistos, resultando em generalizações das medidas de influentes para a Regressão Linear.

2.4.2.1 Efeitos Fixos

Para o modelo de Regressão Linear, a distância de Cook dado por Cook (1977) é frequentemente utilizada, a fim de medir a influência de uma observação, ou grupo, nas estimativas de efeitos fixos. Onde se é observado a mudança nessas estimativas após a eliminação da i -ésima unidade desse nível. Podemos estender tal técnica para os Modelos Lineares Hierárquicos.

Além da estatística da distância de Cook, podemos também utilizar a estatística MDFITS, que é uma versão multivariada da estatística DFFITS (BELSLEY, 1980). As duas estatísticas determinam a distância entre as estimativas de efeitos fixos provenientes dos dados completos e aquelas provenientes dos dados reduzidos e são generalizadas por (CHRISTENSEN; PEARSON; JOHNSON, 1992) e (SCHABENBERGER, 2005) para o Modelo Linear Hierárquico da seguinte forma:

$$C_i(\hat{\beta}) = (\hat{\beta} - \hat{\beta}_i)^T \widehat{\text{Var}}(\hat{\beta})^{-1} (\hat{\beta} - \hat{\beta}_i)/p \quad (2.4.5)$$

$$\text{MDFITS}_i(\hat{\beta}) = (\hat{\beta} - \hat{\beta}_i)^T \widehat{\text{Var}}(\hat{\beta})^{-1} (\hat{\beta} - \hat{\beta}_i)/p \quad (2.4.6)$$

Tais estatísticas apresentam grandes valores para observação influentes e como utilizamos $\widehat{\mathbf{V}}$, não há distribuição de referência exata para a mesma.

2.4.2.2 Componentes de Variância

Segundo Christensen, Pearson e Johnson (1992), a distância de Cook e MDFFITS podem, novamente, ser aplicada para medir a mudança no vetor de componentes de variância. Basta remover $1/p$ e substituir nas equações (2.4.5) (2.4.6) $\hat{\beta}$ e $\hat{\beta}_i$ por $\hat{\theta}$ e $\hat{\theta}_i$. Onde θ representa o vetor de componentes de variância, formado pelo vetor da variância residual σ^2 e as entradas únicas de D .

Assim sendo, as equações para medir a mudança no vetor de componentes de variância são:

$$D_i(\theta) = (\hat{\theta} - \hat{\theta}_i)^T \widehat{\text{Var}}(\hat{\theta})^{-1} (\hat{\theta} - \hat{\theta}_i) \quad (2.4.7)$$

$$MDFFITs_i(\theta) = (\hat{\theta} - \hat{\theta}_i)^T \widehat{Var(\hat{\theta})}^{-1} (\hat{\theta} - \hat{\theta}_i) \quad (2.4.8)$$

Nota-se que ambas equações dependem de estimativas da matriz de covariância, tais estimativas podem ser obtidas, mas isso aumentaria significativamente a complexidade computacional.

Segundo Dillane (2006), como opção para contornar a dependência de estimativas para a matriz de covariância, podemos comparar diretamente a mudança relativa para cada componente de variância, θ_ℓ

$$RVC_i(\theta_\ell) = \frac{\hat{\theta}_{\ell i}}{\hat{\theta}_\ell} - 1 \quad (2.4.9)$$

Note que a mudança de Variância Relativa (RVC) será próxima de zero quando a i -ésima unidade não é influente na componente de variância em questão. O RVC não requer uma estimativa da variância do vetor de componentes de variância, exigindo portanto, menos computação e parece adequado explorar a influência sobre os componentes de variância (LOY, 2013).

2.4.2.3 Valores Ajustados

Observações ou grupos que influenciam excessivamente os Valores Ajustados são ditos pontos de alavanca ("leverage"). No Modelo de Regressão Comum, uma observação é dita ser alavanca se, h_{ii} que é o i -ésimo elemento da diagonal principal da matriz $H = X(X^T X)^{-1} X^T$, apresentar um grande valor, onde esse grande valor indica uma observação que representa um ponto discrepante, comumente referido como "outlier" no subespaço gerado pelas colunas da matriz X (NOBRE; SINGER, 2004).

Já nos Modelos Lineares Hierárquicos, dispomos de efeitos fixos e aleatórios que podem influenciar os Valores Ajustados. Assim, devemos ajustar uma generalização da Regressão Linear descrita acima, para refletir tais efeitos. Tal generalização é dada por supor que a estrutura de covariância, V_i , seja fixa, e com $H_i = \partial \hat{y}_i / \partial y_i$ denotando a alavancagem no nível i .

Seguindo a definição dada por Demidenko e Stukel (2005), a alavancagem do grupo i é a soma das alavancas para os efeitos fixos e aleatórios, $H_{1i} + H_{2i}$, onde

$$H_{1i} = X_i(X_i^T V_i^{-1} X_i)^{-1} X_i^T V_i^{-1} \quad (2.4.10)$$

$$H_{2i} = Z_i D Z_i^T V_i^{-1} (I - H_{1i}) \quad (2.4.11)$$

Assim, como foi o caso dos resíduos, observamos confusão para o diagnóstico de pontos influentes, que se dá entre os níveis. Pois, a alavancagem dos efeitos aleatórios H_{2i} em (2.4.11) decorre da alavancagem dos efeitos fixos H_{1i} em (2.4.10). Opcionalmente, podemos definir segundo Nobre e Singer (2011), a alavancagem para os efeitos aleatórios sendo:

$$H_{2i}^* = Z_i D Z_i^T \quad (2.4.12)$$

que soluciona tal confundimento.

Da mesma forma, como foi dito na regressão comum, o efeito de alavanca para a observação j é obtido a partir do j -ésimo elemento diagonal da matriz correspondente, H_{1i} ou H_{2i}^* . Um grupo ou observação é considerado como tendo alta alavancagem se qualquer uma dessas quantidades for "grande". Para uma observação, o valor bruto é considerado, enquanto para avaliar a alavancagem de um grupo i , é necessário considerar o traço da quantidade dividido pelo tamanho do grupo (LOY, 2013). Em prática substituímos as estimativas \widehat{V} e \widehat{D} nas equações anteriores.

Capítulo 3

Resultados e Discussões

Para ilustrar a utilização da metodologia estudada, uma análise descritiva foi realizada a fim de averiguar o comportamento do peso dos sujeitos durante experimento, onde tal análise se deu em duas partes, peso inicial dos sujeitos (1ª semana) e peso dos sujeitos nas demais semanas de experimento (2ª à 6ª semana). O conjunto de dados utilizados tanto para análise descritiva, quanto para os ajustes é composto de 342 observações, sendo 16 observações perdidas (não registradas).

Na semana inicial, investigou-se como estavam o peso dos sujeitos e apresentou-se medidas de resumo, por meio de tabelas, gráficos de dispersão e *boxplot*.

Para as observações das semanas decorrentes, repetimos os procedimentos utilizados na semana inicial com acréscimo de gráficos de perfis para os sujeitos. E foi verificado se as observações seguiam uma distribuição normal através do teste de Kolmogorov-Smirnov.

Todas as medidas de decisão para a escolha do melhor ajuste proposto, estimação dos parâmetros e diagnósticos do modelo, foram obtidas através do *software* R.

3.1 Análise Descritiva do Peso Inicial dos Camundongos

Para o estudo do peso inicial, um resumo estatístico foi produzido e apresentado na Tabela 2, cuja diferença da quantidade de camundongos nas gaiolas se deve ao fato de que alguns camundongos acabavam brigando, assim os mesmo eram trocados de gaiola, respeitando o grupo ao qual pertenciam.

Tabela 2 – Medidas de Resumo Estatístico

		n	Mediana	Média	Desvio Padrão	Min.	Max.	CV
GRUPO	1	5	41,00	41,18	1,70	39,20	43,40	4,12
	2	13	41,20	42,19	4,04	36,20	50,00	9,57
	3	13	39,50	40,25	2,96	36,20	46,70	7,36
	4	13	38,60	39,24	3,68	33,80	47,80	9,37
	5	13	38,20	39,94	3,77	34,20	45,40	9,43
GAIOLA	1	4	41,00	41,18	1,70	39,20	43,40	4,12
	2	4	43,75	44,23	4,83	39,40	50,00	10,93
	3	5	40,30	41,24	4,00	36,20	47,20	9,70
	4	4	41,70	41,35	3,54	37,50	44,50	8,55
	5	5	38,30	39,74	3,03	37,50	44,80	7,63
	6	4	39,35	38,65	1,65	36,20	39,70	4,26
	7	4	42,00	42,48	3,11	39,20	46,70	7,33
	8	5	38,50	38,22	2,20	34,80	40,40	5,76
	9	4	40,70	40,75	6,22	33,80	47,80	15,26
	10	4	39,20	39,00	2,03	36,40	41,20	5,21
	11	5	41,20	40,36	4,43	34,20	45,40	10,99
	12	4	38,80	40,00	3,64	37,20	45,20	9,10
	13	2	37,00	37,00	1,70	35,80	38,20	4,59
	14	2	41,70	41,70	4,95	38,20	45,20	11,87

Observa-se que as médias dos grupos variam de 39,24g para o grupo 4 até 42,19g para o grupo 2, com desvios padrões iguais a 3,68g e 4,04g respectivamente, enquanto as médias das gaiolas variam de 38,22g para a gaiola 8 até 44,23g para a gaiola 2, com desvios padrões iguais a 2,20g e 4,83g respectivamente. Além disso, se percebe uma grande diferença entre os valores de mínimo e máximo para os grupo e gaiolas. Nota-se, também, que os valores dos Coeficientes de Variação (CV) que representando o quanto, em média, os desvios atingem o valor da média amostral, variam de 0,04 até 0,09 para os grupos e 0,041 até 0,152 para as gaiolas.

Em seguida foram construídos os gráficos exploratórios de dispersão do peso inicial dos sujeitos, dispersão do peso inicial dos sujeitos por grupo, dispersão do peso inicial dos sujeitos por gaiola, *Boxplot* do peso inicial dos sujeitos por grupo e *Boxplot* do peso inicial dos sujeitos por gaiola, que são apresentados na sequência.

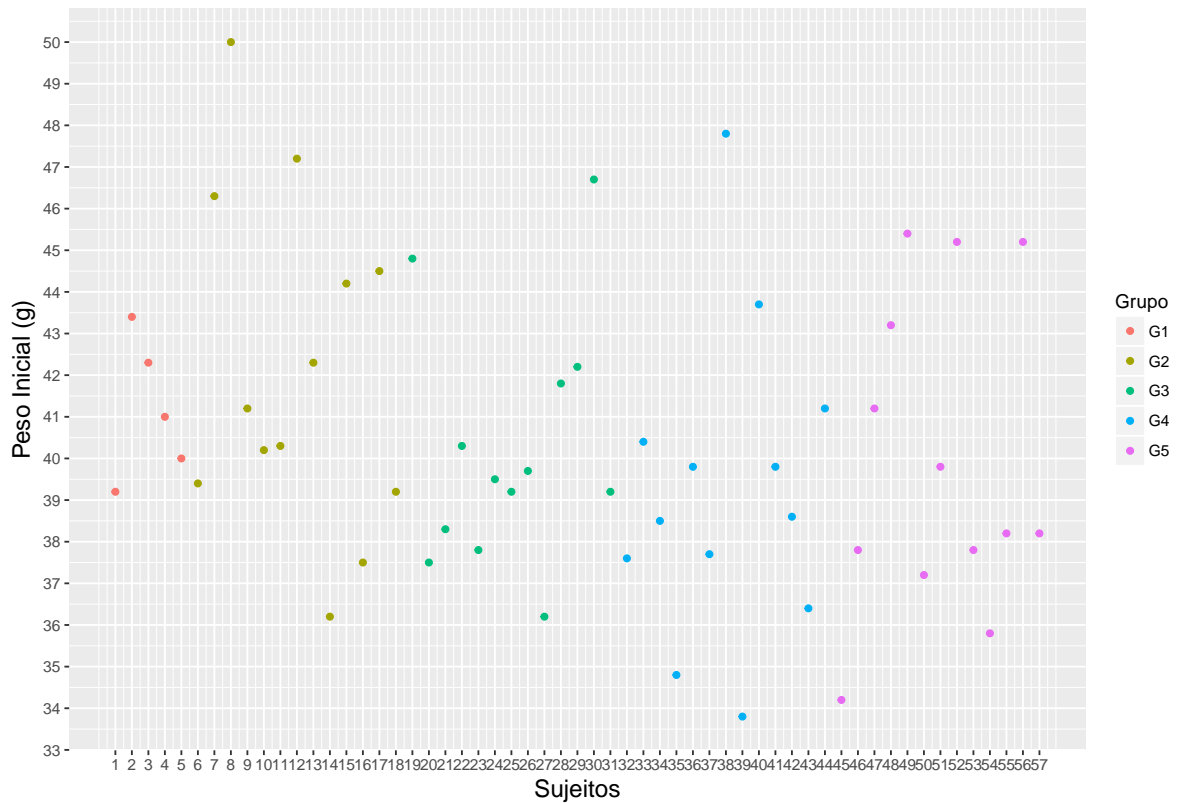


Figura 3.1.1 – Gráfico de Dispersão do Peso Inicial dos Sujeitos

A Figura 3.1.1 apresenta a o gráfico de dispersão do peso de cada sujeito no início do experimento, nota-se como os pesos dos camundongos são discrepantes. O grupo de tratamento ao qual cada sujeito pertence é indicado por cores distintas.

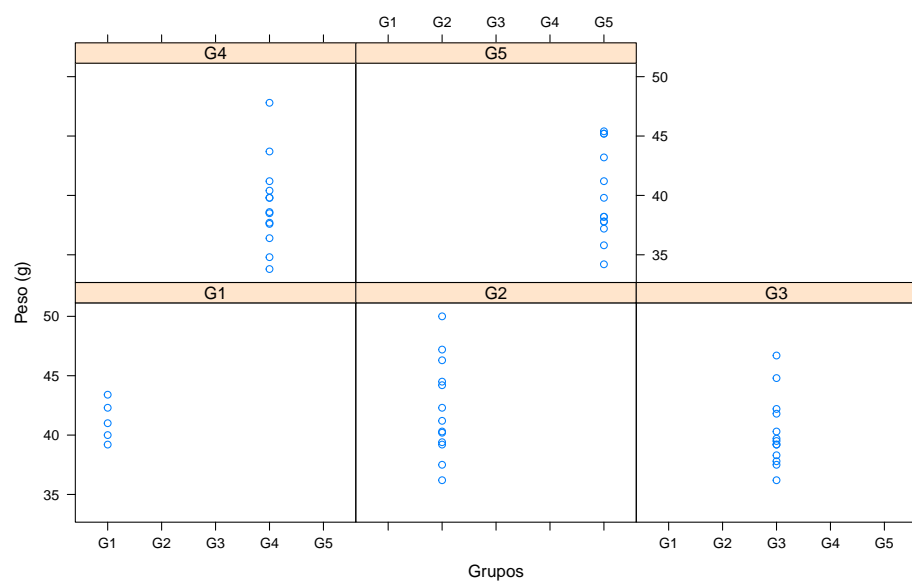


Figura 3.1.2 – Gráfico de Dispersão do Peso Inicial dos Sujeitos por Grupo

Considerando o grupo na análise descritiva, nota-se na Figura 3.1.2 que o grupo 1 apresenta o peso inicial mais próximo entre seus sujeitos, já nos demais grupos não observa-se tal proximidade inicial, os grupos 4 e 2 que apresentam as maiores diferenças entre os pesos iniciais dos sujeitos.

Já quando a gaiola é considerada na análise descritiva, observa-se na Figura 3.1.3 que as gaiolas (1, 6, 10, 13) apresentam o peso inicial mais próximo entre seus sujeitos, já as demais gaiolas não observa-se tal proximidade inicial.

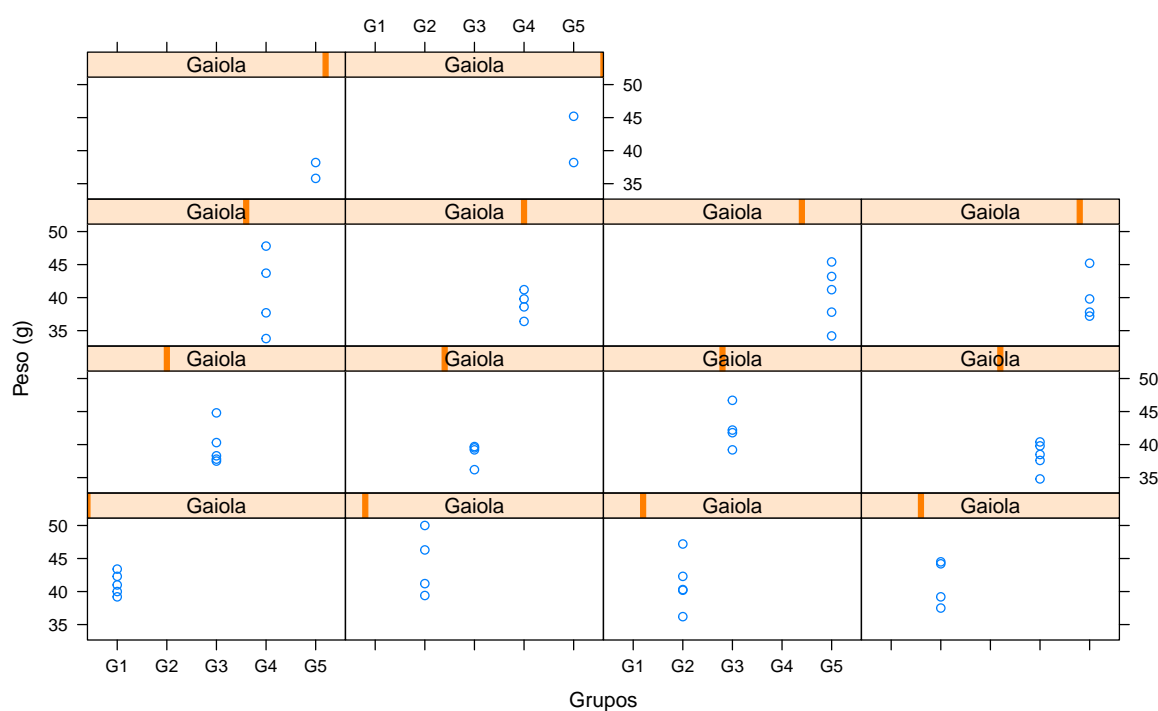


Figura 3.1.3 – Gráfico de Dispersão do Peso Inicial dos Sujeitos por Gaiola

A Figura 3.1.4 exibe os *boxplot* para o comportamento da variável resposta peso por grupo. Enquanto a Figura 3.1.5 exibe os *boxplot* para o comportamento da variável resposta peso por gaiola.

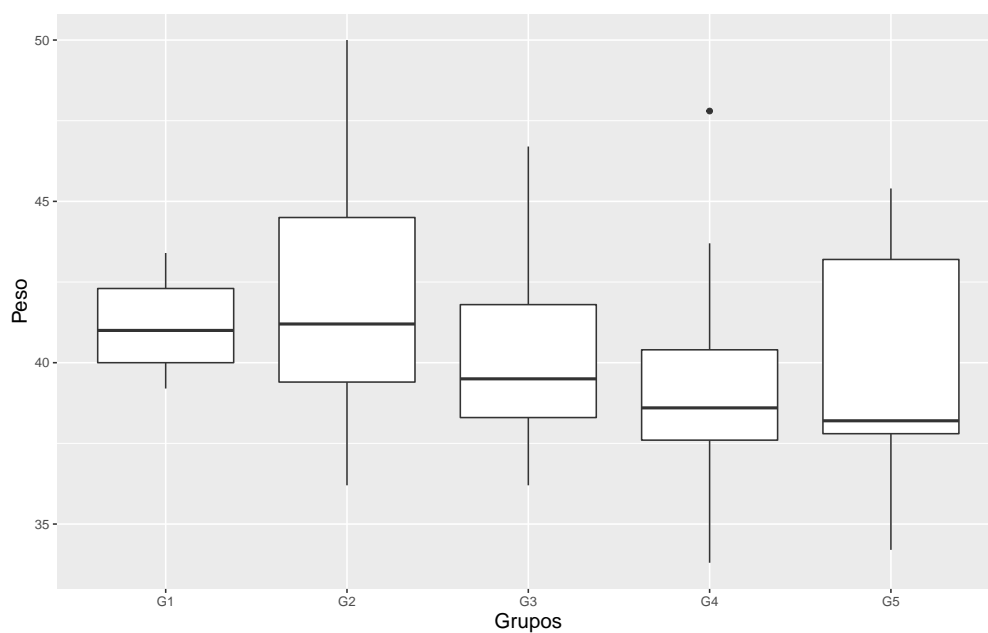


Figura 3.1.4 – Boxplot dos Pesos Iniciais por Grupo

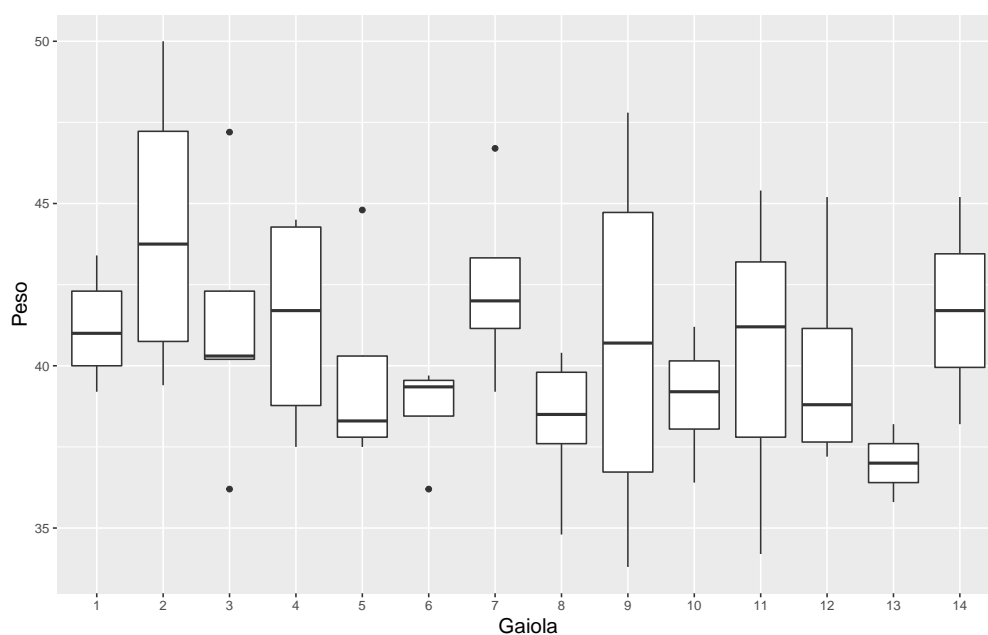


Figura 3.1.5 – Boxplot dos Pesos Iniciais por Gaiola

Ao que parece, o grupo 4 na Figura 3.1.4 é o único que apresenta uma observação *outlier* dentre os pesos iniciais, como já visto anteriormente, é o grupo com maior diferença entre os pesos. Já as gaiolas que apresentam medidas *outlier* são (3, 5, 6, 7), que pode ser visto na Figura 3.1.5, sendo a gaiola 3 com maior diferença entre os pesos iniciais dos sujeitos.

3.2 Análise Descritiva do Peso dos Camundongos Avaliados no Período da 2ª à 6ª Semana

De forma similar a análise do peso inicial, na Tabela 3 encontra-se o valor das médias para os grupos e gaiolas e os respectivos desvios padrão.

Tabela 3 – Médias e desvios padrão dos pesos dos grupo e gaiolas em cada semana

	Média	SEMANAS				
		2ª	3ª	4ª	5ª	6ª
GRUPO	1	44,46	44,62	45,18	44,92	45,72
	2	43,20	44,46	45,01	45,22	46,08
	3	41,18	42,06	42,77	42,71	43,38
	4	41,45	42,32	42,91	42,51	43,39
	5	41,35	41,86	42,09	42,02	42,80
GAIOLA	1	44,46	44,62	45,18	44,92	45,72
	2	44,75	46,10	47,27	47,30	48,07
	3	42,82	44,18	44,54	44,58	45,44
	4	42,12	43,57	43,90	44,45	45,38
	5	40,70	41,22	42,30	42,64	43,58
	6	39,70	41,50	42,00	41,83	42,17
	7	43,25	43,52	43,95	43,45	44,02
	8	40,98	42,28	42,70	42,50	43,64
	9	42,75	43,33	44,08	43,48	44,12
	10	40,72	41,03	41,70	41,23	42,00
	11	41,84	42,84	43,06	42,68	44,20
	12	39,70	41,50	42,00	41,83	42,17
	13	39,00	40,45	40,10	40,90	41,25
	14	41,65	40,10	40,30	40,50	39,20

	Desvio Padrão	SEMANAS				
		2ª	3ª	4ª	5ª	6ª
GRUPO	1	3,582	3,315	3,029	2,860	2,944
	2	4,000	4,102	4,399	4,648	4,726
	3	2,731	2,509	2,319	2,366	2,421
	4	3,814	3,985	4,221	4,198	4,095
	5	3,723	3,575	3,528	3,307	3,272
GAIOLA	1	3,581	3,314	3,028	2,859	2,943
	2	4,787	5,602	6,520	6,564	6,841
	3	3,942	4,045	3,859	4,347	4,479
	4	3,920	3,881	3,955	4,415	4,318
	5	2,570	2,608	2,789	2,987	3,224
	6	2,167	0,608	0,854	1,167	1,327
	7	2,703	3,068	2,456	2,493	2,088
	8	2,270	2,986	2,803	2,799	2,583
	9	6,460	6,388	6,874	6,896	6,812
	10	2,499	1,950	2,523	2,369	2,260
	11	4,194	4,065	4,082	3,913	3,155
	12	4,649	4,144	3,955	3,579	3,254
	13	1,697	3,323	2,687	3,676	4,313
	14	3,889	-	-	-	-

A maior média entre os grupos ocorreu na 6ª semana do grupo 2, enquanto a menor foi na 2ª semana do grupo 3. Já para as gaiolas, a maior média ocorreu na 6ª semana da gaiola 2 e a menor na 2ª semana da gaiola 13.

O maior desvio entre os grupos ocorreu também ocorreu na 6ª semana do grupo 2, enquanto a menor foi na 5ª semana do grupo 3. Já para as gaiolas, a maior média ocorreu na 6ª semana da gaiola 9 e a menor na 3ª semana da gaiola 6.

Assim como no peso inicial, foram construídos gráficos exploratórios de dispersão dos pesos dos sujeitos entre a 2ª e 6ª semana, dispersão dos pesos dos sujeitos por grupo e gaiola entre a 2ª e 6ª semana, gráficos de perfil individual e médio por grupo entre a 2ª e 6ª semana, além de gráficos para verificação da suposição de normalidade dos dados.

A Figura 3.2.1 apresenta a dispersão do peso de cada sujeito no decorrer do tratamento, nota-se como o peso dos sujeitos variam em relação a cada tratamento em função do tempo. O grupo de tratamento ao qual cada sujeito pertence é indicado por cores distintas.

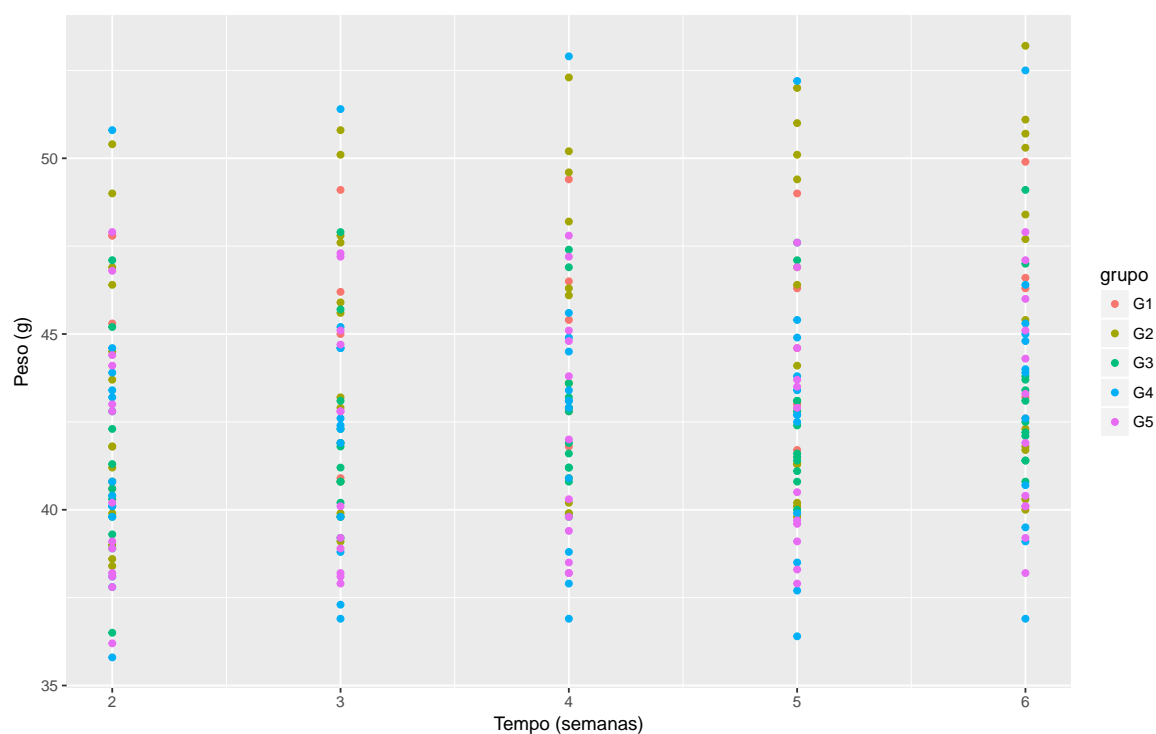


Figura 3.2.1 – Gráfico de Dispersão dos Pesos dos Sujeitos entre a 2ª e 6ª Semana

A Figura 3.2.2 traz a evolução do peso de cada sujeito em função do tempo, podendo notar um comportamento parecido no peso dos sujeitos, conforme o decorrer das semanas.

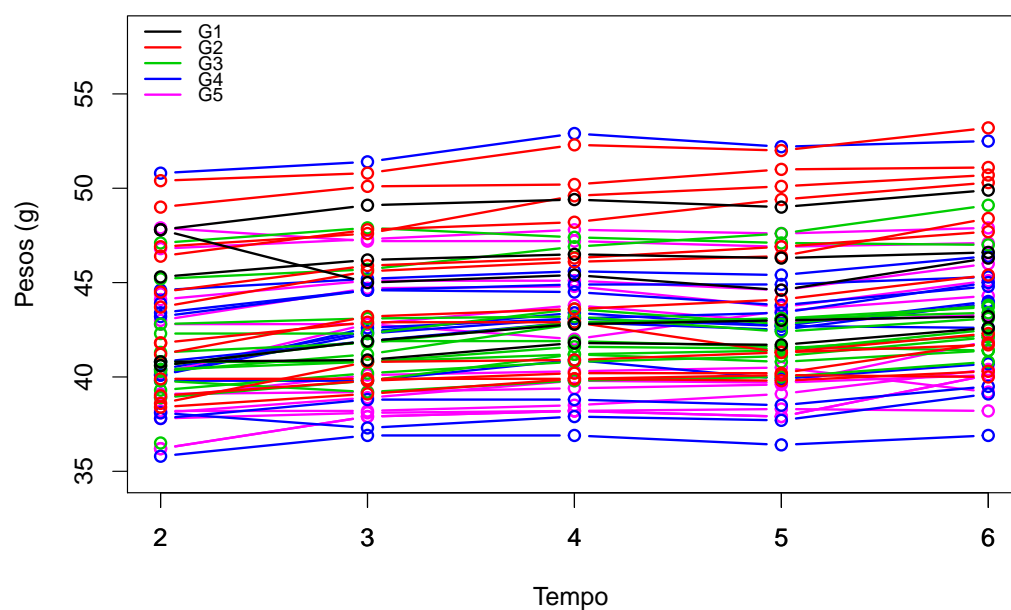


Figura 3.2.2 – Gráfico de Perfil de Cada Sujeito no Tempo

Considerando o grupo na análise descritiva, nota-se na Figura 3.2.3 que aparentemente os grupos 1 e 2 apresentam o maior peso médio entre os grupos.

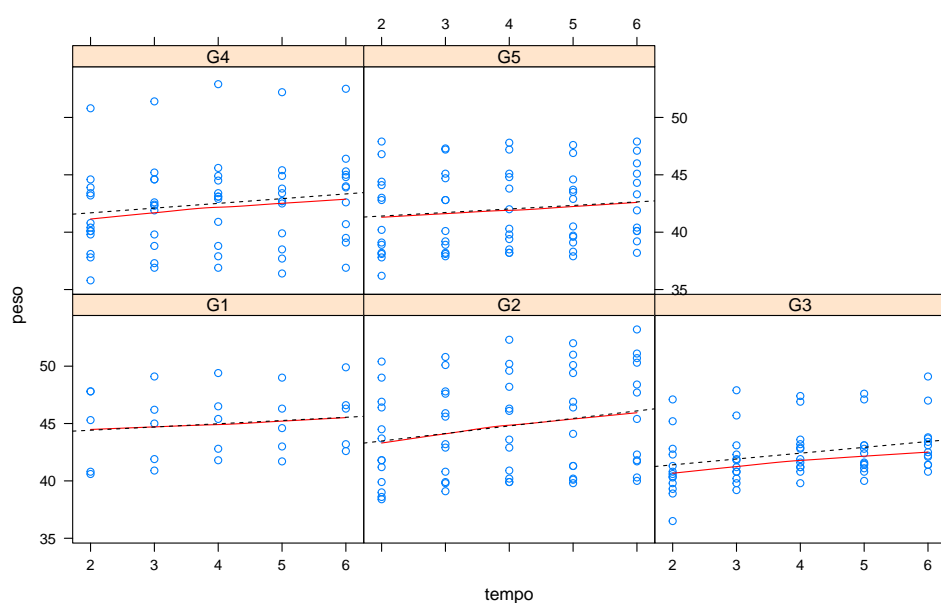


Figura 3.2.3 – Gráfico de Dispersão do Peso dos Sujeitos por Grupo Durante o Tratamento

Além do mais se observa na Figura 3.2.4 que todos os grupos obtiveram um crescimento médio do peso em função do tempo.

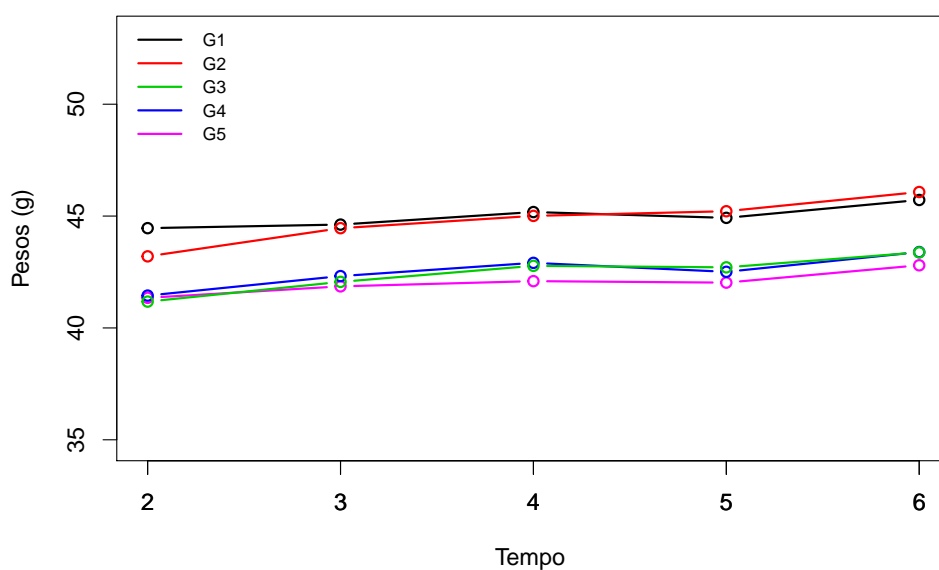


Figura 3.2.4 – Peso Médio dos Grupos em Função do Tempo

Assim como na Figura 3.2.3 observa-se na Figura 3.2.4 que, aparentemente há um crescimento médio do peso em função do tempo em todos os grupos.

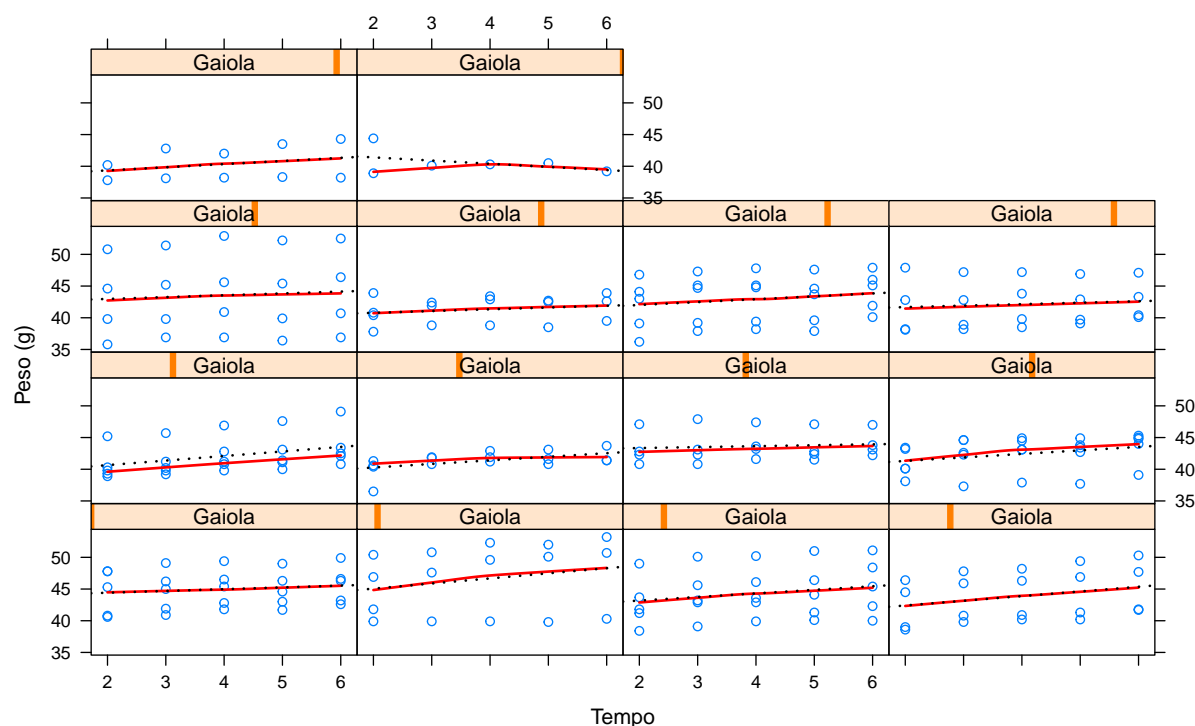


Figura 3.2.5 – Gráfico de Dispersão do Peso dos Sujeitos por Gaiola entre a 2ª e 6ª Semana

A Figura 3.2.5 apresenta a dispersão dos pesos dos sujeitos por gaiola entre a 2ª e 6ª semana, nota-se que praticamente todas as gaiolas apresentaram evolução no peso dos seus sujeitos, com exceção da gaiola 15, que aparentemente teve um crescimento do peso dos seus sujeitos até a 4ª semana e, em seguida, iniciou-se uma redução de peso dos sujeitos que a ela pertenciam.

Na Figura 3.2.6 temos o histograma, o gráfico da Função de Distribuição Acumulada empírica e teórica, QQ-Plot e PP-Plot todos em função das observações do peso de todos os tratamentos no período da 2ª semana à 6ª semana, considerando a distribuição Normal como referência.

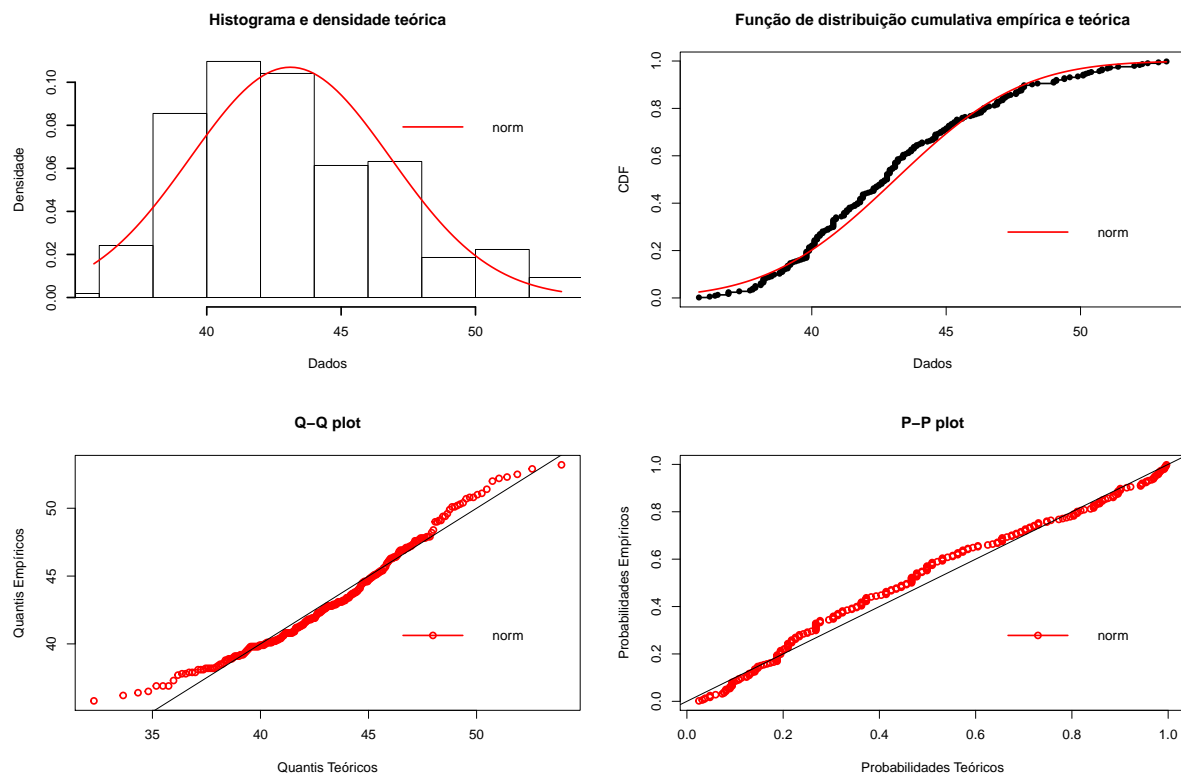


Figura 3.2.6 – Histograma, Função de Densidade Ajustadas, Distribuição Acumulada Empírica, Função de Distribuição Ajustadas, Quantis Teóricos, Quantis Empíricos e Probabilidades Teóricas, Probabilidades Empíricas

A verificação de normalidade dos dados se deu pelo teste de Kolmogorov-Smirnov, onde a Tabela 4 apresenta os resultados do teste.

Tabela 4 – Teste de Kolmogorov-Smirnov

D	p-valor
0,077755	0,07734

Como p-valor é maior que 0,05 no teste, temos que as observações dos pesos dos sujeitos apresentam normalidade, como desejado.

3.3 Ajuste

No ajuste dos modelos se buscou analisar o comportamento dos sujeitos infectados por *Trypanosoma cruzi* através da relação da variável resposta peso aos possíveis fatores de influência, gaiola e tempo sendo estes pertencentes à parte fixa do Modelo. Já o grupo de tratamento e o próprio sujeito pertencem à parte aleatória.

Os modelos foram ajustados na Variável Resposta y_{ijkt} onde, $i = 1, 2, \dots, 5$ (número do grupo), $j = 1, 2, \dots, n_i$ (número do sujeitos, onde se $i = 1$, $n_i = 5$ e se $i \neq 1$, $n_i = 13$), $k = 1, \dots, m_i$ (número da gaiola, onde se $i = 1$, $m_i = 1$ $i = 2$, $m_i = 2, 3, 4$ $i = 3$, $m_i = 5, 6, 7$ $i = 4$, $m_i = 8, 9, 10$ $i = 5$, $m_i = 11, 12, 13, 14$) e $t = 2, \dots, 6$ (semana de tratamento).

A seleção do Modelo foi baseada no Critério de Informação de Akaike (AIC), no Critério bayesiano de Schwarz (BIC) e no Teste de Razão da Verossimilhança.

Os Modelos ajustados estão descritos como o primeiro sendo dado apenas pelo intercepto da parte fixa e o efeito dos sujeitos por meio de um intercepto aleatório para cada um deles. No segundo é acrescentado o efeito do tempo à parte fixa, no terceiro Modelo é acrescento o efeito dos grupos à parte fixa, no quarto Modelo além da parte do tempo e grupo é acrescento o efeito das gaiolas à parte fixa.

Tabela 5 – Modelos Propostos

	Modelo
1	$y_j = \beta_0 + b_{0j} + e_j$
2	$y_{jt} = \beta_0 + \text{Tempo}_t \lambda_t + b_{0j} + e_{jt}$
3	$y_{ijt} = \beta_0 + \text{Tempo}_t \lambda_t + \text{Grupo}_i \alpha_i + b_{0j} + e_{ijt}$
4	$y_{ijkt} = \beta_0 + \text{Tempo}_t \lambda_t + \text{Grupo}_i \alpha_i + \text{Gaiola}_k \theta_k + b_{0j} + e_{ijkt}$

Onde α_i representa os efeitos devido ao i -ésimo grupo, θ_k os efeitos devido a k -ésima gaiola, λ_t ao t -éssimo tempo, b_j e e_j representam o efeito e erro aleatório respectivamente para o j -éssimo sujeito.

Todos os Modelos propostos tiveram violação no pressuposto de normalidade de resíduos, como pode-se ver na Tabela 6, que apresenta o teste de Shapiro-Wilk.

Tabela 6 – Teste de Shapiro-Wilk para Normalidade de Resíduos

Modelo	W	p-valor
1	0,983	0,0032
2	0,984	0,0042
3	0,984	0,0047
4	0,984	0,0048

Observando os gráficos da Figura 3.3.1, podemos notar como a observação 16 se destaca em 3 dos 4 ajustes, sendo assim é uma observação que merece investigação.

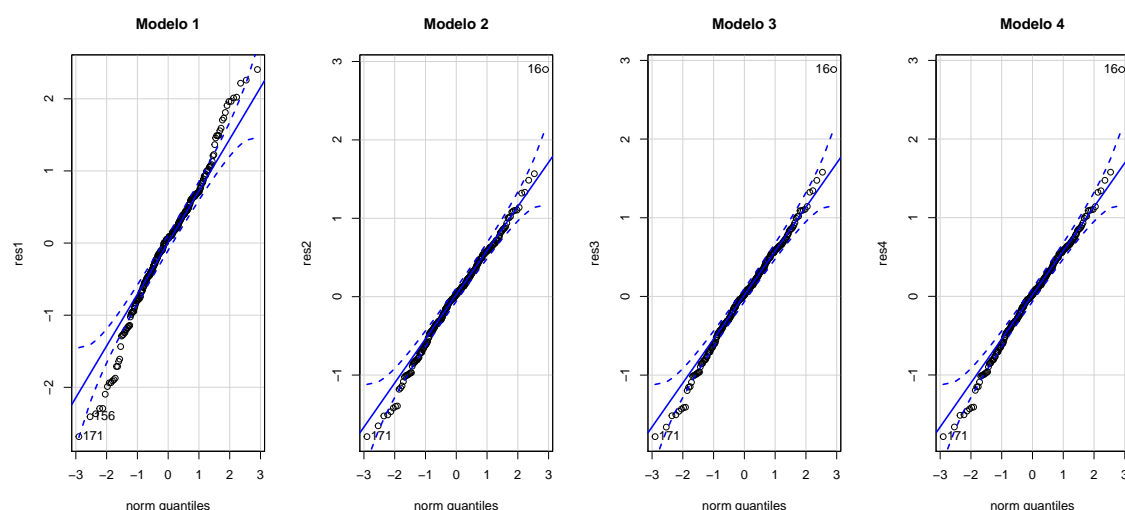


Figura 3.3.1 – QQ-Plot dos ajustes nos modelos propostos

Temos que a observação 16 corresponde ao peso corporal do sujeito 4 na segunda semana, onde o mesmo pertence ao grupo G1 que é o controle dos não infectados. Seu peso era de 41,00g na semana inicial. Observando especificamente os dados desse sujeito na Tabela 7 nota-se que na segunda semana seu peso teve um salto em relação ao inicial e em seguida uma queda que se manteve com pequenas variações até a 6ª semana.

Tabela 7 – Observações do Sujeito 4

Obs	Grupo	Sujeito	Gaiola	Peso	Tempo
16	G1	4	1	47,80	2
17	G1	4	1	45,00	3
18	G1	4	1	45,40	4
19	G1	4	1	44,60	5
20	G1	4	1	46,30	6

Portanto, tem-se indício que tal observação tenha sofrido um erro de coleta. Dessa forma, os ajustes foram refeitos excluindo a observação 16, onde podemos ver na Tabela 8 que o problema de normalidade de resíduos foi resolvido para todos os Modelos propostos, com exceção do modelo 1.

Tabela 8 – Teste de Shapiro-Wilk para Normalidade de Resíduos, Sem a Observação 16

Modelo	W	p-valor
1	0,982	0,002
2	0,993	0,267
3	0,993	0,266
4	0,993	0,265

Assim sendo o Modelo 1, foi desconsiderado.

Os valores de todas as medidas de decisão da ANOVA de cada Modelo proposto são apresentados na tabela 7.

Tabela 9 – Medidas de Decisão da ANOVA.

	Df	AIC	BIC	logLik	GD	Chisq	Chi.Df.	Pr(>Chisq)
2	4	826,29	840,65	-409,14	818,29	167,43	1	<2e-16
3	8	828,42	857,15	-406,21	812,42	5,87	4	0,21
4	9	830,24	862,56	-406,12	812,24	0,18	1	0,67

Os Modelos 2,3,4 são estatisticamente equivalentes, mas considerando os Critérios de Seleção de Modelo (AIC), (BIC) e no Teste de Razão da Verossimilhança, o Modelo 2 foi escolhido como o que melhor explica os dados. Outro fator que corroborou na sua escolha foi que o Modelo 2 é mais parcimonioso em relação aos demais equivalentes.

Na Tabela 10 são apresentados as estimativas dos parâmetros de efeitos fixos do modelo 2 com seus respectivos intervalos de confiança.

Tabela 10 – Estimativas dos Parâmetros de Efeitos Fixos

Variável	Estimativa	Erro Padrão	Valor t	Pr(> t)	IC 95%
Intercepto	41,232	0,49	84,43	<2e-16	40,26 ; 42,20
Tempo	0,462	0.03	15,96	<2e-16	0,405 ; 0,519

A validade do Modelo 2 pode ser confirmada verificando os pressupostos do mesmo, os quais foram avaliados através da análise de diagnósticos.

3.3.1 Análise dos Resíduos

Os resíduos de nível 1 e nível 2 do Modelo 2 foram analisados de uma forma ascendente como sugerido por (HILDEN-MINTON, 1995). Essa forma de análise se faz necessária pois os resíduos estão interrelacionados e, assim ditos confusos, podem levar à complicações no diagnóstico do Modelo se não forem analisados de uma forma correta.

3.3.1.1 Nível 1

Segundo Hilden-Minton (1995) os resíduos resultantes do nível 1 do MQ, não são confundidos pelos resíduos no nível 2. Ao ajustar modelos de regressão MQ separados, tratamos os efeitos aleatórios como fixos.

A Figura 3.3.2 apresenta o gráfico dos resíduos do nível 1 do MQ pelo Tempo, sugere que o Tempo pode não estar linearmente relacionado ao peso.

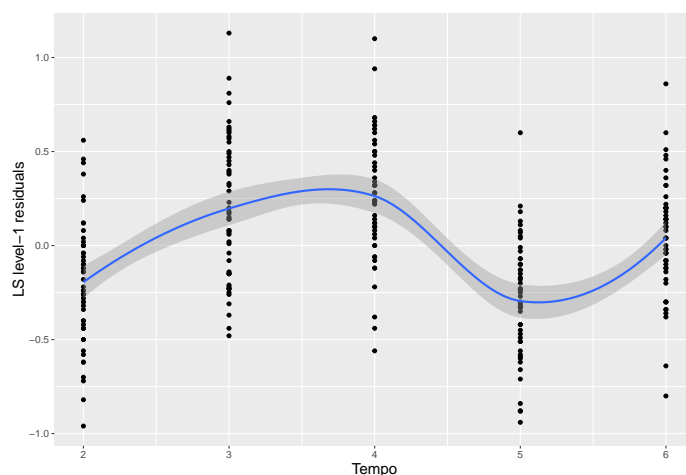


Figura 3.3.2 – Gráfico dos Resíduos MQ de nível 1 pelo Tempo

A fim de direcionar a suposição de resíduos homocedásticos de nível 1, fazemos uso dos resíduos semi-padronizados de (SNIJDERS; BERKHOF, 2008). A Figura 3.3.3 apresenta tal gráfico, onde não há indicação de violação da linearidade pela mesma.

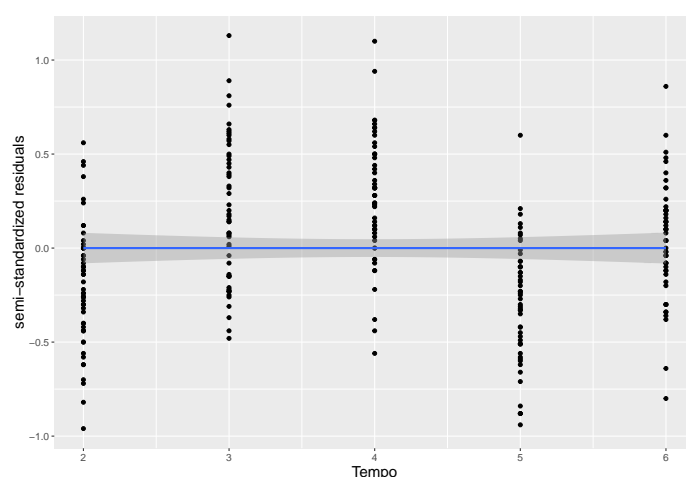


Figura 3.3.3 – Gráfico dos Resíduos Semi-padronizados pelo Tempo

Também foi utilizado o protocolo de programação proposto por (BUJA et al., 2009), para corroborar na verificação da suposição de resíduos homocedásticos de nível 1. Onde o mesmo é exibido na Figura 3.3.4. Nota-se que não é possível distinguir o gráfico dos dados reais em relação aos gráficos simulados. Assim sendo, podemos prosseguir com a análise sem a necessidade de medidas corretivas.

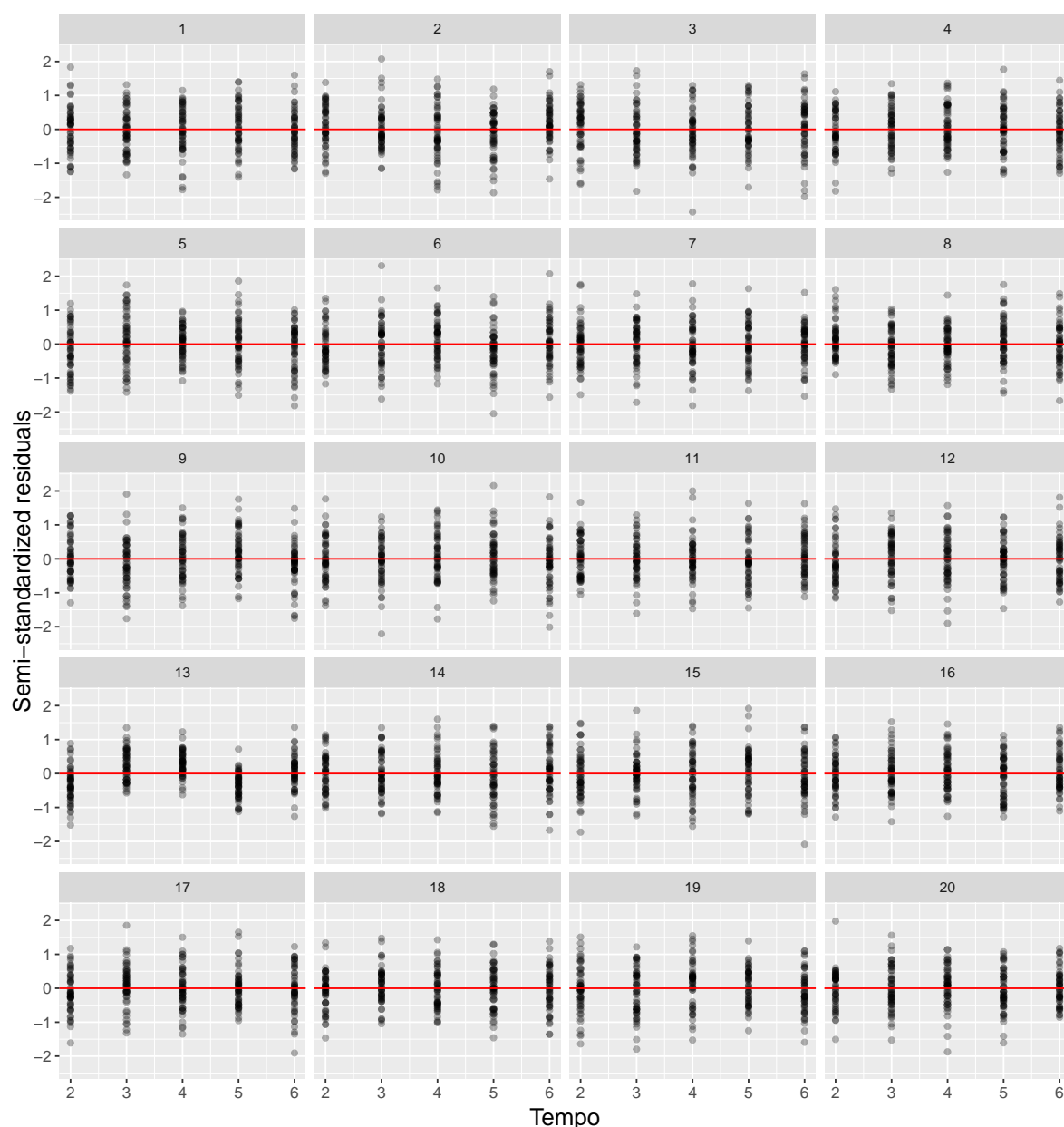


Figura 3.3.4 – Painel dos Vinte Gráficos que Apresentam Resíduos Semi-padronizados de um Modelo Hierárquico em Relação à Variável Preditora Tempo. A Plotagem dos Dados Reais foi Incorporada Aleatoriamente em Dezenove Gráficos Simulados

A Figura 3.3.5 apresenta o gráfico quantil normal dos resíduos semi-padronizados de nível 1, visualmente os resíduos semi-padronizados parecem normais, não apresentando, assim, evidências contra a suposição de normalidade.

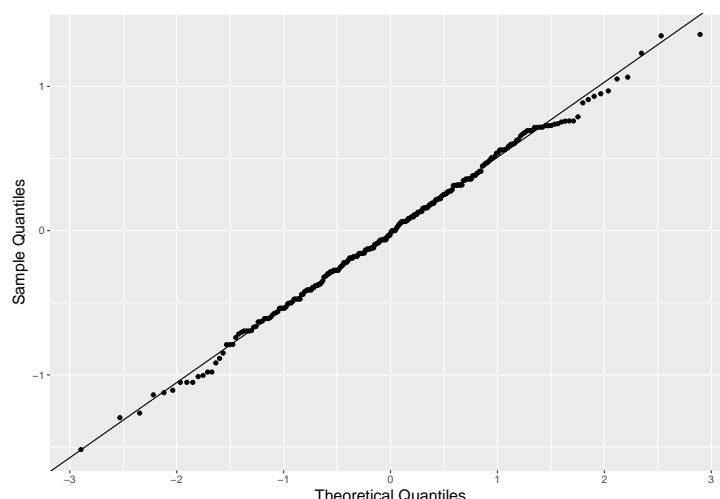


Figura 3.3.5 – Gráfico de Quantil Normal dos Resíduos Semi-padronizados de Nível 1

Os efeitos aleatórios ou comumente referidos como nível 2 são definidos por Zb ou, apenas b . Como já dito, uma análise residual crescente é seguida, então optamos pelo uso de resíduos EB no nível 2. Essa opção se deu pois os resíduos de mínimos quadrados são mais variáveis que os resíduos EB e estimativas EB de b foram obtidas de forma direta.

A suposição de normalidade dos resíduos foi avaliada pelo gráfico quantílico normal apresentado na Figura 3.3.6. Cujo mesmo não apresenta comprovação de um desvio da normalidade.

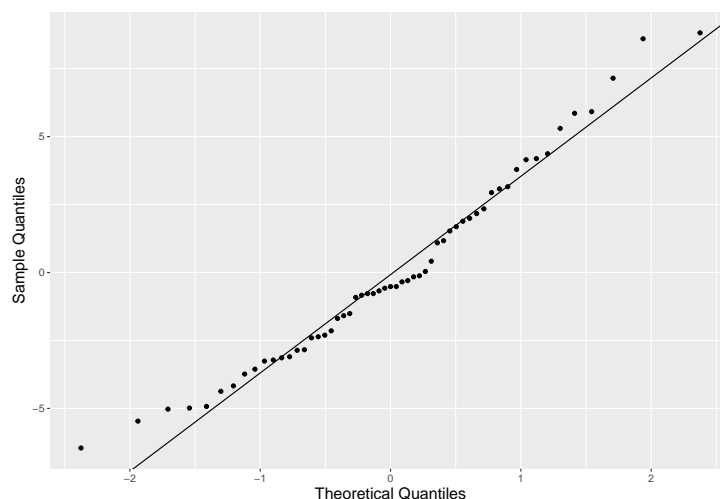


Figura 3.3.6 – Gráfico de Quantil Normal dos Resíduos do Nível 2 do EB para o Intercepto

3.3.2 Análise de Observações Influentes

Nesta seção, descrevemos a implementação de diagnósticos para avaliar mudanças na estimativa dos componentes de variância usando mudança de variância relativa, a esti-

mativa dos efeitos fixos usando a distância de Cook e os valores ajustados usando alavancagem. Essas quantidades são usadas para avaliar a influência nas unidades de nível 1 e nível 2.

3.3.2.1 Efeitos Fixos

Utilizamos duas estatísticas que são comumente aplicadas para medir se houve alguma mudança nos efeitos fixos, sendo elas distância de Cook e MDFFITS, já referidas anteriormente. Pode-se observar na Figura 3.3.7 que nenhum sujeito foi apontado pelas duas estatísticas (que apresentam os mesmos valores) como influente nos efeitos fixos considerando uma escala interna. A Figura 3.3.7 trás os gráficos das estatísticas supracitadas na forma de diagrama de pontos e diagrama de pontos modificados.

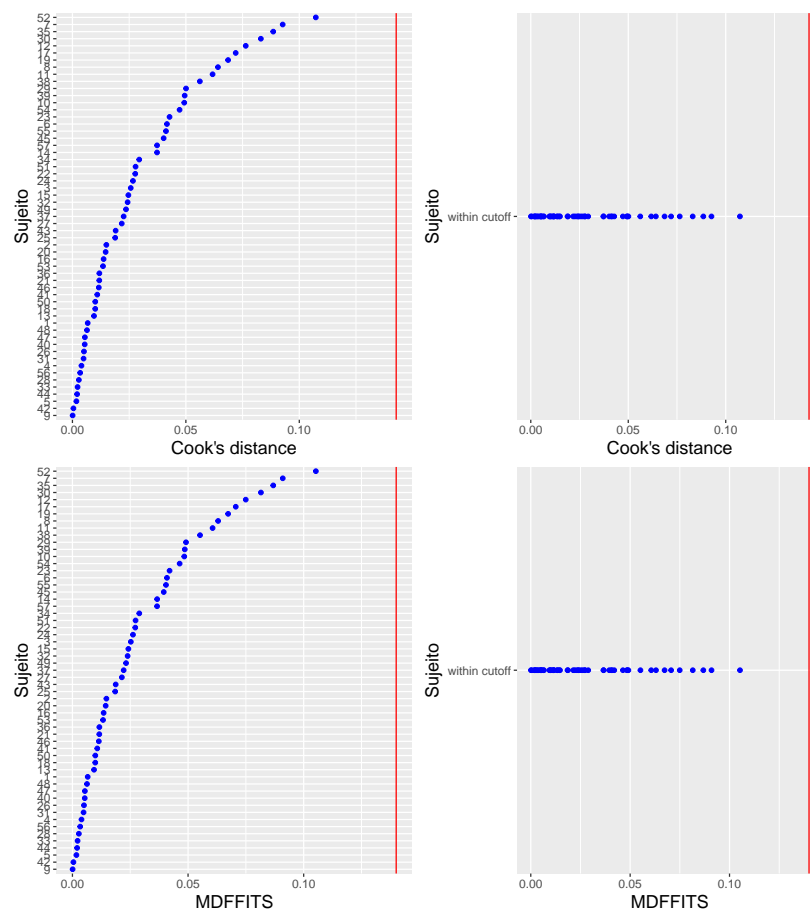


Figura 3.3.7 – À Esquerda um Diagrama de Pontos da Distância de Cook e MDFFITS, à Direita um Diagrama de Pontos Modificado da Distância de Cook e MDFFITS, Ambos para o Nível 2 (sujeito)

3.3.2.2 Componentes de Variância

Para os componentes de variância concentramo-nos em diagnósticos que não requerem estimativa da matriz de covariância, possibilitando assim o uso do **lmer**. O diagnóstico que atendeu a este quesito foi apresentado na equação 2.4.9, onde o mesmo dito RVC mede a mudança nas estimativas da ℓ -ésima componente de variância, θ_ℓ , com e sem unidade i .

Tabela 11 – Matriz de Mudança de Variância Relativa

n	$\widehat{\sigma^2}$	n	$\widehat{\sigma^2}$	n	$\widehat{D11}$	n	$\widehat{D11}$
1	0,0141317079	30	-0,0160375532	58	0,0357492791	87	0,0108306308
2	0,0163940719	31	0,0208422053	59	0,0225738157	88	0,0322969531
3	0,0159141670	32	-0,0118458439	60	-0,0150452115	89	0,0363928740
4	0,0059051539	33	0,0192541151	61	0,0307452692	90	0,0330994896
5	0,0216088128	34	0,0051724188	62	0,0331598212	91	-0,0006586536
6	0,0058874054	35	-0,0304985253	63	0,0223740677	92	0,0369814078
7	-0,0163933737	36	0,0055999098	64	-0,0136163844	93	0,0348479887
8	0,0126815677	37	0,0072950812	65	-0,0723556501	94	0,0246597221
9	0,0048782209	38	0,0116638987	66	0,0359999322	95	-0,0780134695
10	-0,0081946241	39	0,0074558217	67	0,0366383234	96	-0,0249156581
11	-0,0274348674	40	0,0205563144	68	0,0357444701	97	0,0284375554
12	-0,0214010398	41	0,0112454681	69	0,0242903801	98	0,0359983611
13	0,0199285839	42	-0,0073800383	70	0,0178806123	99	0,0358731808
14	0,0210740470	43	0,0156550439	71	-0,0387873929	100	0,0083623379
15	0,0129694618	44	0,0048518840	72	0,0219869512	101	0,0319886346
16	0,0048188451	45	-0,0048160403	73	0,0281969840	102	0,0001717857
17	-0,0033117866	46	0,0017156558	74	-0,0045857476	103	0,0214498574
18	0,0159505913	47	-0,0007846893	75	0,0245428183	104	0,0346538485
19	-0,0075102610	48	0,0124035179	76	0,0156939145	105	0,0313344677
20	0,0145593521	49	0,0174293047	77	0,0220005946	106	0,0083949227
21	0,0190195656	50	0,0204410881	78	0,0283591648	107	0,0159722642
22	0,0034713960	51	0,0034493797	79	0,0355362854	108	0,0365794387
23	0,0016682711	52	-0,0257833253	80	0,0289721522	109	0,0113745040
24	0,0011085579	53	0,0169276215	81	0,0328707769	110	0,0109199864
25	0,0142724860	54	0,0090400096	82	0,0297218191	111	0,0007265276
26	0,0201309807	55	-0,0176092099	83	0,0361894653	112	0,0365619143
27	0,0045730567	56	0,0048338085	84	-0,0078482255	113	0,0292426143
28	0,0141850009	57	-0,0113729758	85	0,0365396664	114	0,0210282370
29	-0,0054484352	-	-	86	0,0365806731	-	-

A Tabela 11 ilustra o RVC que apresenta como saída uma matriz onde cada coluna representa uma componente de variância, para o nosso caso temos σ^2 que é a variância residual e D11 que é a variância associada ao intercepto aleatória dos sujeitos. Note que

temos o valor mais longe de zero em D11 para $n=95$, caracterizando esta unidade como influente na componente de variância.

Pode-se observar na Figura 3.3.8 o gráfico diagrama de pontos modificado do RVC de nível 2 para o intercepto aleatório dos sujeitos. Por meio do uso de escala interna, $n=95$ foi identificado como unidade influente assim como na Tabela 11 e merece atenção.

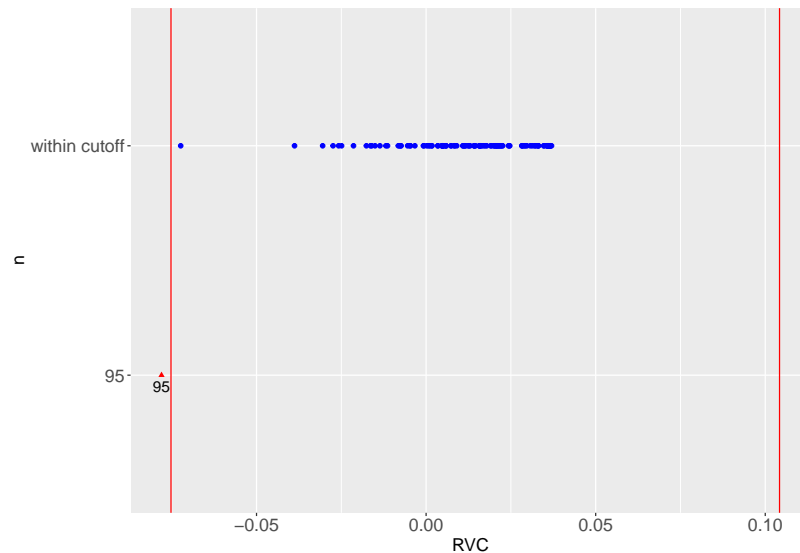


Figura 3.3.8 – Diagrama de Pontos Modificado do RVC de Nível 2 para o Intercepto Sujeito

Tabela 12 – Peso dos Sujeitos na Segunda Semana

Grupo	Sujeito	Peso	Grupo	Sujeito	Peso	Grupo	Sujeito	Peso
G1	1	40,60	G3	20	39,80	G4	39	35,80
G1	2	45,30	G3	21	39,30	G4	40	44,60
G1	3	47,80	G3	22	40,30	G4	41	40,80
G1	4	47,80	G3	23	38,90	G4	42	40,40
G1	5	40,80	G3	24	40,60	G4	43	37,80
G2	6	39,90	G3	25	40,40	G4	44	43,90
G2	7	46,90	G3	26	41,30	G5	45	36,20
G2	8	50,40	G3	27	36,50	G5	46	39,10
G2	9	41,80	G3	28	42,30	G5	47	43,00
G2	10	41,20	G3	29	42,80	G5	48	44,10
G2	11	41,80	G3	30	47,10	G5	49	46,80
G2	12	43,70	G3	31	40,80	G5	50	38,10
G2	13	38,40	G4	32	40,10	G5	51	42,80
G2	14	49,00	G4	33	43,40	G5	52	47,90
G2	15	44,50	G4	34	38,10	G5	53	38,20
G2	16	38,60	G4	35	40,10	G5	54	37,80
G2	17	46,40	G4	36	43,20	G5	55	40,20
G2	18	39,00	G4	37	39,80	G5	56	44,40
G3	19	45,20	G4	38	50,80	G5	57	38,90

$n=95$ corresponde ao peso sujeito 38 no início do tratamento, como se pode observar na Tabela 12, tal sujeito apresenta o maior peso entre os demais. Assim sendo, justifica-se ter sido marcado como influente no gráfico da Figura 3.3.8.

3.3.2.3 Valores Ajustados

Além de verificar como as observações de um mesmo sujeito impactam diretamente os parâmetros do modelo adotado, foi de interesse explorar se essas observações são atípicas em relação aos valores ajustados e às variáveis explicativas de tal Modelo. Essa exploração se deu através de múltiplas estatísticas.

As múltiplas estatísticas que podem definir "pontos de alavanca" para valores ajustados em um Modelo Linear Hierárquico são descritas por pontos de alavanca geral (H), pontos de alavanca nos efeitos fixos (H_1), ponto de alavanca nos efeitos aleatória (H_2) e ponto de alavanca nos efeitos aleatórios não confundíveis (H_2^*). Ambos descritos nas equações (2.4.10), (2.4.11) e (2.4.12).

Tais estatísticas são apresentados na Tabela 13 que ao observar verifica-se que os sujeitos 9, 27, 44 e 56 têm alta alavancagem nos efeitos fixos dos valores ajustados. Da mesma forma, tais sujeitos são sinalizados como tendo alta alavancagem nos efeitos aleatórios, o que é natural já que H_2 depende de H_1 . Considerando H_2^* todos os sujeitos apresentam a mesma métrica, assim sendo, todos têm a mesma influência sob os efeitos aleatórios dos valores ajustados do modelo 2.

Tabela 13 – Pontos de Alavanca para os Efeitos Fixos, Efeitos Aleatórios e Efeitos Aleatórios Não Confundidos dos Sujeitos

	H	H_1	H_2	H_2^*		H	H_1	H_2	H_2^*
1	0,2025	0,0073	0,1951	28,9860	30	0,2025	0,0073	0,1951	28,9860
2	0,2025	0,0073	0,1951	28,9860	31	0,2025	0,0073	0,1951	28,9860
3	0,2025	0,0073	0,1951	28,9860	32	0,2025	0,0073	0,1951	28,9860
4	0,2503	0,0068	0,2435	28,9860	33	0,2025	0,0073	0,1951	28,9860
5	0,2025	0,0073	0,1951	28,9860	34	0,2025	0,0073	0,1951	28,9860
6	0,2025	0,0073	0,1951	28,9860	35	0,2025	0,0073	0,1951	28,9860
7	0,2025	0,0073	0,1951	28,9860	36	0,2025	0,0073	0,1951	28,9860
8	0,2025	0,0073	0,1951	28,9860	37	0,2025	0,0073	0,1951	28,9860
9	0,9672	0,0173	0,9499	28,9860	38	0,2025	0,0073	0,1951	28,9860
10	0,2025	0,0073	0,1951	28,9860	39	0,2025	0,0073	0,1951	28,9860
11	0,2025	0,0073	0,1951	28,9860	40	0,2025	0,0073	0,1951	28,9860
12	0,2025	0,0073	0,1951	28,9860	41	0,2025	0,0073	0,1951	28,9860
13	0,2025	0,0073	0,1951	28,9860	42	0,2025	0,0073	0,1951	28,9860
14	0,2025	0,0073	0,1951	28,9860	43	0,2025	0,0073	0,1951	28,9860
15	0,2025	0,0073	0,1951	28,9860	44	0,9672	0,0173	0,9499	28,9860
16	0,2025	0,0073	0,1951	28,9860	45	0,2025	0,0073	0,1951	28,9860

17	0,2025	0,0073	0,1951	28,9860	46	0,2025	0,0073	0,1951	28,9860
18	0,2025	0,0073	0,1951	28,9860	47	0,2025	0,0073	0,1951	28,9860
19	0,2025	0,0073	0,1951	28,9860	48	0,2025	0,0073	0,1951	28,9860
20	0,2025	0,0073	0,1951	28,9860	49	0,2025	0,0073	0,1951	28,9860
21	0,2025	0,0073	0,1951	28,9860	50	0,2025	0,0073	0,1951	28,9860
22	0,2025	0,0073	0,1951	28,9860	51	0,2025	0,0073	0,1951	28,9860
23	0,2025	0,0073	0,1951	28,9860	52	0,2025	0,0073	0,1951	28,9860
24	0,2025	0,0073	0,1951	28,9860	53	0,2025	0,0073	0,1951	28,9860
25	0,2025	0,0073	0,1951	28,9860	54	0,2025	0,0073	0,1951	28,9860
26	0,2025	0,0073	0,1951	28,9860	55	0,2025	0,0073	0,1951	28,9860
27	0,9672	0,0173	0,9499	28,9860	56	0,9672	0,0173	0,9499	28,9860
28	0,2025	0,0073	0,1951	28,9860	57	0,2025	0,0073	0,1951	28,9860
29	0,2025	0,0073	0,1951	28,9860	-	-	-	-	-

Como o intercepto do sujeito de número 38, as observações dos sujeitos de número 9, 27, 44, 56 que são únicas, foram sinalizadas como influentes nos componentes de variância e valores ajustados respectivamente, todas foram removidas e o ajuste do modelo escolhido foi repetido, afim de averiguar se alguma mudança significativa se deu nas estimativas dos parâmetros.

As estimativas dos parâmetros sem as observações influentes, são apresentadas na Tabela 14, note que em comparação com a Tabela 15 os parâmetros sofreram pequenas mudanças.

Tabela 14 – Estimativas dos Parâmetros sem as Observações Influentes

Variável	Estimativa	Erro Padrão	Valor t	Pr(> t)	IC 95%
(Intercept)	41,275	0,510	80,82	<2e-16	40,25 ; 42,29
Tempo	0,461	0,029	15,81	<2e-16	0,403 ; 0,518

Tabela 15 – Estimativas dos Parâmetros com Todas Observações

Variável	Estimativa	Erro Padrão	Valor t	Pr(> t)	IC 95%
(Intercept)	41,232	0,49	84,43	<2e-16	40,26 ; 42,20
Tempo	0,462	0,03	15,96	<2e-16	0,405 ; 0,519

Portanto, como as observações influentes geram pouco impacto no modelo escolhido, optamos por mantê-las, afim de preservar os dados. Assim consideramos para o Modelo 2 as estimativas iniciais.

Capítulo 4

Conclusão

O Modelo Linear Misto proposto se mostrou satisfatório para o conjunto de dados com estrutura hierárquica, conseguiu descrever o comportamento do peso dos sujeitos nas semanas de experimento. As variáveis Grupo e Gaiola não apresentaram significância para o Modelo, a variável Tempo apresentou uma significância pequena mas considerável, deixando o intercepto (peso inicial) como maior responsável por descrever o peso dos indivíduos no Modelo. Tal importância do intercepto no Modelo, se deve pelo fato do mesmo ter sido considerado como aleatório, já que cada sujeito apresentava um peso inicial diferente. Assim, o Modelo somente com Intercepto e Tempo, foi capaz de descrever os pesos dos sujeitos, com estimativas aceitáveis.

Nos diagnósticos, algumas observações foram sinalizadas como influentes nos componentes de variância e valores ajustados, após uma nova estimação dos Parâmetros do Modelo sem essas observações, constatou-se que as mesmas pouco impactavam o Modelo proposto. Assim, foram mantidas a fim de preservar os dados iniciais.

Considerando os resultados da análise, sugerimos como recomendações aos pesquisadores, procurarem iniciar seus experimentos com o peso inicial dos sujeitos o mais próximo possível, dividir os sujeitos igualmente entre os grupos e entre as gaiolas. Testar novos bioterápicos para o tratamento, visto que os analisados no presente trabalho (G13cH, G6cH e G3cH) não se mostraram eficientes.

Capítulo 5

Trabalhos Futuros

Em trabalhos futuros almejamos a melhora e complementação do presente trabalho, através de algumas propostas como:

1. Análise de diferentes estruturas de covariâncias para o Modelo escolhido, a fim de aprimorar as estimativas por ele geradas;
2. Ajustar outras distribuições para o mesmo conjunto de dados através da Metodologia de Modelos Lineares Generalizados;
3. Aplicar a Metodologia de Modelos Mistos Lineares a outros conjuntos de dados que possuam estrutura hierárquica.

Referências

AITKIN, M.; ANDERSON, D.; HINDE, J. Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society. Series A (General)*, JSTOR, p. 419–461, 1981. Citado na página 16.

ALEIXO, D. L.; VEIGA, F. K. da; CIUPA, L.; PORTOCARRERO, A. R.; SANDRI, P. F.; FER-RAZ, F. N.; BENVENUTTI, M. J.; ARAÚJO, S. M. de. Highly diluted medication modifies parasitological, immunological and histopathological parameters of trypanosoma cruzi infection. *International Journal of High Dilution Research*, v. 12, n. 44, p. 107–108, 2013. Citado na página 19.

BELSLEY, D. A. *Regression diagnostics; identifying influential data and sources of collinearity*. [S.l.], 1980. Citado na página 30.

BENNETT, N. Teaching styles and pupil progress. 1977. Citado na página 16.

BERGAMO, G. C. *Aplicação de modelos multiníveis na análise de dados de medidas repetidas no tempo*. Dissertação (Mestrado) — Universidade de São Paulo, 2002. Citado na página 16.

BUJA, A.; COOK, D.; HOFMANN, H.; LAWRENCE, M.; LEE, E.-K.; SWAYNE, D. F.; WICKHAM, H. Statistical inference for exploratory data analysis and model diagnostics. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, The Royal Society, v. 367, n. 1906, p. 4361–4383, 2009. Citado 2 vezes nas páginas 27 e 46.

CHRISTENSEN, R.; PEARSON, L. M.; JOHNSON, W. Case-deletion diagnostics for mixed models. *Technometrics*, Taylor & Francis, v. 34, n. 1, p. 38–45, 1992. Citado na página 30.

COOK, R. D. Detection of influential observation in linear regression. *Technometrics*, Taylor & Francis Group, v. 19, n. 1, p. 15–18, 1977. Citado na página 30.

COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 248–275, 1968. Citado na página 27.

DEMIDENKO, E.; STUKEL, T. A. Influence analysis for linear mixed-effects models. *Statistics in medicine*, Wiley Online Library, v. 24, n. 6, p. 893–909, 2005. Citado na página 31.

DIEZ-ROUX, A. V. Multilevel analysis in public health research. *Annual review of public health*, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 21, n. 1, p. 171–192, 2000. Citado na página 16.

DILLANE, D. M. *Deletion diagnostics for the linear mixed model*. Tese (Doutorado) — Trinity College Dublin, 2006. Citado na página 31.

FAUSTO, M. A.; CARNEIRO, M.; ANTUNES, C. M. d. F.; PINTO, J. A.; COLOSIMO, E. A. O modelo de regressão linear misto para dados longitudinais: uma aplicação na análise de dados antropométricos desbalanceados. *Cadernos de Saúde Pública*, SciELO Public Health, v. 24, p. 513–524, 2008. Citado 2 vezes nas páginas 13 e 15.

FERRAZ, F. N.; ALEIXO, D. L.; GRUENDLING, A. P.; GOMES, M. L.; TOLEDO, M. J. de O.; ARAUJO, S. M. D. Trypanosoma cruzi: Evaluation of pcr as a laboratory tool to follow up the evolution of parasite load. *Iranian journal of parasitology*, Tehran University of Medical Sciences, v. 11, n. 3, p. 389, 2016. Citado na página 19.

FERREIRA, É. C.; CIUPA, L.; PORTOCARRERO, A. R.; BRUSTOLIN, C. F.; MASSINI, P. F.; ALEIXO, D. L.; ARAÚJO, S. M. de. Phosphorus protects cardiac tissue by modifying the immune response in rats infected by trypanosoma cruzi. *Cytokine*, Elsevier, v. 102, p. 102–106, 2018. Citado 3 vezes nas páginas 13, 14 e 18.

FILHO, J. A. C. *Modelos lineares mistos: estruturas de matrizes de variâncias e covariâncias e seleção de modelos*. Tese (Doutorado) — Universidade de São Paulo, 2002. Citado 3 vezes nas páginas 22, 24 e 25.

GOLDSTEIN, H. *Multilevel statistical models*. [S.l.]: John Wiley & Sons, 1995. Citado 2 vezes nas páginas 17 e 29.

GURKA, M. J.; EDWARDS, L. J. 8 mixed models. *Handbook of statistics*, Elsevier, v. 27, p. 253–280, 2007. Citado 3 vezes nas páginas 22, 26 e 29.

HARVILLE, D. Extension of the gauss-markov theorem to include the estimation of random effects. *The Annals of Statistics*, JSTOR, p. 384–395, 1976. Citado na página 15.

HARVILLE, D. A. Bayesian inference for variance components using only error contrasts. *Biometrika*, Oxford University Press, v. 61, n. 2, p. 383–385, 1974. Citado 2 vezes nas páginas 25 e 26.

HARVILLE, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 72, n. 358, p. 320–338, 1977. Citado na página 15.

HENDERSON, C. R.; KEMPTHORNE, O.; SEARLE, S. R.; KROSIGK, C. V. The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, JSTOR, v. 15, n. 2, p. 192–218, 1959. Citado na página 25.

HILDEN-MINTON, J. A. *Multilevel diagnostics for mixed and hierarchical linear models*. Tese (Doutorado) — University of California, 1995. Citado 2 vezes nas páginas 28 e 45.

HOX, J. J. *Applied multilevel analysis*. [S.l.]: TT-publikaties, 1995. Citado na página 23.

- IGNÁCIO, S. A. Importância da estatística para o processo de conhecimento e tomada de decisão. *Revista Paranaense de Desenvolvimento-RPD*, n. 118, p. 175–192, 2012. Citado na página 12.
- LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. *Biometrics*, JSTOR, p. 963–974, 1982. Citado 2 vezes nas páginas 15 e 29.
- LINDLEY, D. V.; SMITH, A. F. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 1–41, 1972. Citado na página 15.
- LOY, A. M. M. *Diagnostics for mixed/hierarchical linear models*. Tese (Doutorado) — Iowa State University, 2013. Citado 6 vezes nas páginas 13, 17, 27, 28, 31 e 32.
- MARCELINO, S. D. do R.; IEMMA, A. F. Métodos de estimação de componentes de variância em modelos mistos desbalanceados. *Scientia agricola*, v. 57, n. 4, p. 643–652, 2000. Citado na página 26.
- NATIS, L. Modelos lineares hierárquicos. *Estudos em avaliação educacional, São Paulo*, v. 1, n. 23, p. 3–29, 2001. Citado na página 17.
- NOBRE, J. S.; SINGER, J. *Métodos de diagnóstico para modelos lineares mistos*. Tese (Doutorado), 2004. Citado 2 vezes nas páginas 27 e 31.
- NOBRE, J. S.; SINGER, J. M. Leverage analysis for linear mixed models. *Journal of Applied Statistics*, Taylor & Francis, v. 38, n. 5, p. 1063–1072, 2011. Citado na página 32.
- NUSSENZWEIG, V.; SONNTAG, R.; BIANCALANA, A.; FREITAS, J. d.; NETO, V. A.; KLOETZEL, J. Ação de corantes trifenil-metânicos sobre o trypanosoma cruzi "in vitro". emprego da violeta de genciana na profilaxia da transmissão da moléstia de chagas por transfusão de sangue. *O hospital*, v. 44, n. 6, p. 731–744, 1953. Citado na página 18.
- ORGANIZATION, W. H. et al. Control of chagas disease (who-technical report series, 811). Geneva: WHO, 1991. Citado 2 vezes nas páginas 13 e 14.
- PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, Oxford University Press, v. 58, n. 3, p. 545–554, 1971. Citado na página 25.
- PINHEIRO, S. *Modelo Linear Hierárquico: Um método alternativo para análise de desempenho escolar*. Dissertação (Mestrado) — Dissertação de mestrado defendida no Departamento de Estatística da Universidade Federal de Pernambuco, 2005. Citado 3 vezes nas páginas 15, 20 e 23.
- PINHO, L. G. B.; NOBRE, J. S.; FREITAS, S. M. de. On linear mixed models and their influence diagnostics applied to an actuarial problem. *Chilean Journal of Statistics*, v. 3, n. 1, p. 57–73, 2012. Citado na página 17.
- RAUDENBUSH, S. W.; BRYK, A. S. *Hierarchical linear models: Applications and data analysis methods*. [S.l.]: Sage, 2002. v. 1. Citado 2 vezes nas páginas 17 e 20.
- SALSBURG, D. S. *Uma senhora toma chá... como a estatística revolucionou a ciência no século XX*. [S.l.]: Zahar, 2009. Citado na página 12.

- SANTOS, C.; FERREIRA, L. D.; OLIVEIRA, N. F.; DOURADO, M. I. C.; BARRETO, M. L. Modelagem multinível. *Sitientibus, Feira de Santana*, n. 22, p. 89–98, 2000. Citado na página 13.
- SCHABENBERGER, O. Mixed model influence diagnostics. In: CITESEER. *SUGI*. [S.l.], 2005. v. 29, p. 189–29. Citado na página 30.
- SCHEFFE, H. The analysis of variance. 1959. *New York*, p. 331–67, 1959. Citado na página 15.
- SCHMUNIS, G. Tripanossomíase americana: seu impacto nas américas e perspectivas de eliminação. *Clínica e Terapêutica da Doença de Chagas. Um Manual Prático para o Clínico Geral*, Fiocruz Rio de Janeiro, p. 11–24, 1997. Citado 2 vezes nas páginas 13 e 14.
- SEARLE, S. *Linear models for unbalanced data*. [S.l.], 1987. Citado na página 24.
- SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. *Variance components*. [S.l.]: John Wiley & Sons, 2009. v. 391. Citado na página 26.
- SNIJDERS, T. A.; BERKHOF, J. Diagnostic checks for multilevel models. In: *Handbook of multilevel analysis*. [S.l.]: Springer, 2008. p. 141–175. Citado na página 46.
- VERBEKE, G.; MOLENBERGHS, G. Linear mixed models for longitudinal data. 2000. Citado 3 vezes nas páginas 21, 22 e 26.
- VINHAES, M. C.; DIAS, J. C. P. Doença de chagas no brasil chagas disease in brazil. *Cad. Saúde Pública*, SciELO Brasil, v. 16, n. Sup 2, p. 7–12, 2000. Citado na página 14.
- WEST, B. T.; GALECKI, A. T. An overview of current software procedures for fitting linear mixed models. *The American Statistician*, JSTOR, p. 274–282, 2011. Citado na página 17.
- WOLFINGER, R. Covariance structure selection in general mixed models. *Communications in statistics-Simulation and computation*, Taylor & Francis, v. 22, n. 4, p. 1079–1106, 1993. Citado na página 25.