Acta Scientiarum
http://periodicos.uem.br/ojs
ISSN on-line: 1807-8648
Doi: 10.4025/actascihealthsci.v42i1.49916

Linear mixed model for weight analysis in mice infected by *Trypanosoma cruzi*

Roney Peterson Pereira[1*], Terezinha Aparecida Guedes[2], Érika Cristina Ferreira[2], Silvana Marques de Araújo[2], Larissa Aparecida Ricardini[2] and Larissa Ciupa[3]

[1]Programa de Pós-Graduação em Engenharia Agrícola, Universidade Estadual do Oeste do Paraná, Rua Universítaria, 2069, 85519-110, Cascavel, Paraná, Brazil. [2]Universidade Estadual de Maringá, Maringá, Paraná, Brazil. [3]Centro Universitário Ingá, Maringá, Paraná, Brazil. *Author for correspondence. E-mail: roney_peterson@hotmail.com

ABSTRACT. The use of linear mixed models for nested structure longitudinal data is called hierarchical linear modeling. This modeling takes into account the dependence of existing data within each level and between hierarchical levels. The process of modeling, estimating and analyzing diagnoses was 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 their body weight as a result of using different types of biotherapeutics produced from *Gallus gallus domesticus* (chicken) serum to treat *Trypanosoma cruzi*. Through the model selection criteria AIC and BIC and the likelihood ratio test, a model was chosen to describe the data correctly. Model diagnoses were then performed by means of residual analysis for both levels and an analysis of influential observations to verify if any observations were signaled as influencing the fixed effects, the components of variance and the adjusted values. After the analysis, it was possible to notice that the observations that were signaled as influential had little impact on the Model chosen initially, so it was maintained, with no differences being evidenced between the treatments with the biotherapeutics tested; only the Time variable and the Random intercept were necessary to describe the weight of the mice.

Keywords: epidemiology and biostatistics; statistical models; chagas disease.

Received on September 10, 2019.
Accepted on May 29, 2020.

Introduction

There has been a significant increase worldwide in the capacity to produce, store and transmit information; the latter is characterized as data and demands more and more advances from statistics, both as to the development of methodologies and as to new, ever-complex indicators that require modern equipment, statistical software and trained technicians (Ignácio, 2010). Mixed Models were widely studied by Fisher in 1918, with major impacts on quantitative genetics studies, and referred to by the author as components of variance models (Scheffe, 1999). The development of linear mixed models combined in one single equation is a result of primordial investigations (Harville, 1976; 1977) that facilitated this achievement; later, their use would be discussed in (Laird & Ware, 1982) for longitudinal data, which are data characterized by a time sequence of two or more observations on each individual.

The applications of Hierarchical Linear Models have been growing due to the great extension of problems that have hierarchically structured data, just as in this study, which analyzed body weight data of individuals infected by *Trypanosoma cruzi*, the agent of Chagas disease. The latter is one of the most widely distributed pathologies in the American continent. Vectors of the disease can be found from the south of the United States to Argentina. There are more than one hundred species responsible for the natural transmission of the infection by *Trypanosoma cruzi*, directly helping it spread in the home environment or participating in the maintenance of chagasic enzooty. It is estimated that 16 to 18 million individuals are infected, and that approximately 80 million people are at risk of contamination in Latin America (Schmunis, 1997; WHO, 1991 apud Vinhaes & Dias, 2000). Hence the importance of the several studies conducted in this field.

Material and methods

The experiment described in this research was carried out by (Ferreira et al., 2018) through blind, controlled and randomized assays. The objective was to verify the efficiency of using variations of a

biotherapeutics produced from a serum of *Gallus gallus domesticus* (chicken), under parasitological, clinical and immunological parameters, in mice experimentally infected with *Trypanosoma cruzi*. The expected efficiency of each biotherapeutics would be observed in the subjects' body weight.

Also, according to the authors, the experiment used 57 Swiss mice, aged 56 days old, all male and sourced from the Central Vivarium of the State University of Maringá. The animals were distributed into treatment groups and housed in cages with a maximum of 5 animals. With polysulfone (ALESCO®), 20 × 32 × 21 cm^3 in dimension, controlled temperature (22 ± 2ºC), a 12-hour light/dark cycle, and water and feed being supplied *ad libitum*, the cages were transformed into a microenvironment. All groups were subjected to the same experimental conditions.

The animals were divided into 5 groups and subdivided into 14 cages: group 1 was composed of 5 animals allocated in cage 1; groups 2, 3 and 4 were composed of 13 animals and divided into 3 cages each, respectively in cages 2, 3, 4, 5, 6, 7, 8, 9 and 10; and group 5 was composed of 13 animals as well, but they were divided into 4 cages – 11, 12, 13 and 14, respectively. The infected animals were inoculated intraperitoneally with 1,400 blood trypomastigotes (infectious forms) of *Trypanosoma cruzi* (Y strain) (Nussenzweig et al., 1953).

Here is a description of the experimental groups (treatments):
1: NIC - (Non-infected control) Non-infected and non-treated animals (n = 5);
2: G13cH - Animals treated with 13cH chicken serum biotherapeutics (n = 13);
3: G6cH - Animals treated with 6cH chicken serum biotherapeutics (n = 13);
4: ICG - (Infection control) The animals were infected and received no treatment (n = 13);
5: G3cH - Animals treated with 3cH chicken serum biotherapeutics (n = 13).

The medicine was diluted in water (1mL/10mL) and supplied *ad libitum* in a sterile amber bottle, in accordance with Aleixo *et al*. (2013), for 16 consecutive hours (medicine available for the animals from 16:00 to 8:00), on the 4th, 7th and 10th days after infection (totaling 3 doses). The treatment scheme is based on the action of the drug, which is linked to its immunological effects and to the specific evolution of the Y strain of *Trypanosoma cruzi* in Swiss mice (Aleixo et al., 2013; Ferraz et al., 2016). The project from which the experiments were run was approved by the Ethics Committee on Research Involving Animals of the State University of Maringá, Paraná, CEUA Opinion 2401220716/2016.

Body weight evolution was monitored over 12 weeks using a semi-analytical balance (Balance BEL®). In this study, the assessment performed at the beginning of the treatment was used to analyze the initial weight of the mice. Due to a great loss of information as of the 7th study week, the modeling was based on the longitudinal measures observed up to the 6th week.

Hierarchical Linear Models

For modeling the weight of the mice (y) as a function of the predictor variables (Z) at the individual level and/or at a higher level (W), the dataset is assumed to be multilevel; theoretically, the model can be presented as a hierarchical system of regression equations.

The Hierarchical Linear Model, in its general form, can be formulated through two equations:

$$y_i = Z_i\beta_i + e_i \quad (1)$$

$$\beta_i = W_i\gamma + b_i \quad (2)$$

in which (1) represents the model equation within the group – level 1 –, and (2) represents the model between groups – level 2. The explanatory variables at the subject level are represented by index (q),while the explanatory variables at the group level are represented by index (p), so (q) and (p) represent the number of variables. Thus:
y_i is a result vector.
Zi is a matrix of explanatory variables;
β_i is one of the components of the vector of unknown fixed parameters;
e_i is the vector of level-1;
W_i is a matrix of level-2 explanatory variables;
γ is a vector of fixed effects;
b_i Is a vector of random effects.

This model can be combined to result in a single Linear Mixed Model, with $X_i = Z_iW_i$ and $\beta = \gamma$.

$$y_i = X_i\beta + Z_i b_i + e_i \quad (3)$$

Generally referred to as Hierarchical Model in (Verbeke & Molenberghs, 2000). $X_i\beta$ corresponds to the fixed part, with β representing the vector of components β_i, while $Z_i b_i + e_i$ corresponds to the random part of the Model.

From the model given in (3), errors will be assumed as independent between groups, and different between levels, that is, it is assumed that random effects and errors follow a normal distribution, with mean equal to 0, and have correlated residuals, where ϕ is the covariance between b_i and e_i, and (D) and (R) are the covariance matrices, with both matrices being positive and defined by hypothesis, therefore not singular:

$$V_i = Var(Y) = \begin{bmatrix} b_i \\ e_i \end{bmatrix} = \begin{bmatrix} D & \phi \\ \phi & R_i \end{bmatrix} \quad (4)$$

where

$$Var(b_i) = E[b_i b_i^T] = D \quad (5)$$

and

$$Var(e_i) = E[e_i e_i^T] = R_i \quad (6)$$

with the following assumptions, $e_i \sim N(0, R_i)$, $b_i \sim N(0, D)$, $Cov(e_i, b_j) = 0 \; \forall \; i, j$; $Cov(b_i, b_j)$ can be zero for $i \neq j$.

Frequently, it is assumed that $R_i = \sigma^2 I_{ni}$. Thus, the total variance of the model (3) for response vector y_i is given as:

$$V(y_i) = Z_i D Z_i^T + R_i \quad (7)$$

where such considerations mean that:

$$y_i \sim N(X_i\beta, V_i) \quad (8)$$

Parameter Estimation

When there is a Hierarchical Linear Model, in the form given in (3), with matrix of variances and covariances such as those presented in (7), there is usually interest in predicting fixed and random effects and in estimating the components of variance. Commonly, as a procedure for estimating Hierarchical Linear Models, Maximum Likelihood Estimation (MLE) is used. The Maximum Likelihood Estimation method consists of maximizing the likelihood function of observations in relation to the fixed effects and to the components of variance, requiring data normality assumption. Assuming that the $\boldsymbol{y_i}$ vector of the observations has $X_i\beta_i$ as mean, and V_i as matrix of variances and covariances, the likelihood function (L) of y_i is:

$$L_{MV}(y_n; \beta_q) = \prod_{i=1}^{N} \left[(2\pi)^{\left(-n_i/2\right)} |V_i|^{(1/2)} \times exp\left[-\frac{1}{2}(y_i - X_i\beta)^T V_i^{-1} (y_i - X_i\beta) \right] \right] \quad (9)$$

where $|V_i|$ represents the determinant of the V_i matrix.

Model diagnoses

The Hierarchical Linear Model, just as in Ordinary Linear Regression, has distributive assumptions that may or may not be valid when used in practice. However, the diagnoses to assess these assumptions and consequent alternatives, for assumption violation suspicions, have not been fully developed for this model, mainly because the analysis tool is relatively recent.

The diagnoses were divided into two stages, residual analysis and influential point analysis; for the residual analysis, the model given in (3) encompasses the uncertainty both at the individual level e_i and and at the group level b_i. A residual term should allow assessing the distributional assumptions of the model. Thus, level-1 (individual) residuals $\widehat{e_i}$ and level-2 (group) residuals $\widehat{b_i}$ are deemed essential; therefore, regardless of the form of the covariance matrix, these residuals are of interest. All upper-level residuals are the best linear unbiased predictors (BLUPs) of random effects. Following the convention established in (Goldstein, 2011) and (Raudenbush & Bryk, 2002) for level specification, the residuals are titled by the level at which they were introduced in the Model.

Checking if their definitions are interrelated, level-1 and level-2 residuals are fundamental for modeling, since a deficiency at one level of the model can be perceived in the residual analysis at another level. An

ascending residual analysis is recommended for verifying the validation of level-1 residuals; this way, having concluded the appropriate model for this level, move on to level 2. By doing so, the impacts on the residual analysis, resulting from a possible confusion when modeling, will be minimized (Loy & Hofmann, 2013).

One way of indicating whether the residual variation within the group is constant between groups is to use the programming protocol presented by (Buja et al., 2009), which employs several sets of simulated data from which null graphs are built.

In model fitting and parameter estimation, not all observations, or groups, have the same effect. Some observations or groups stand out from the others, and the fit of the model detects these differences. These observations or groups are referred to as influential or leverage points. We are especially interested in the points of influence on fitted values, fixed effects (estimates) and components of variance.

Considering that, in addition to fixed effects, there are random effects influencing the result of a studied phenomenon, Hierarchical Linear Models can be used to study the best covariance structure. For this case, we assume that the covariance structure, V_i, is fixed, which is a generalization of the linear regression, with $H_i = \partial \widehat{y_i} / \partial y_i$ denoting leveraging at level i.

The multiple statistics that can define 'leverage points' for fitted values in a Hierarchical Linear Model are described by general leverage points (H), leverage points in fixed effects (H_1), leverage points in random effects (H_2), and leverage points in non-confounding random effects (H_2^*). Following the definition provided by (Demidenko & Stukel, 2005), the leveraging of group i is the sum of the leverages for fixed effects H_{1i} and random effects H_{2i}, where

$$H_{1i} = X_i(X_i^T V_i^{-1} X_i)^{-1} X_i^T V_i^{-1} \quad (10)$$

$$H_{2i} = Z_i D Z_i^T V_i^{-1}(I - H_{1i}) \quad (11)$$

Here is some confusion as to the diagnosis of influential points, which occurs between levels. Because the leveraging of the random effects (11) results from the leveraging of the fixed effects (10). Optionally, according to (Nobre & Singer, 2011), the leveraging for the random effects can be defined as

$$H_{2i}^* = Z_i D Z_i^T \quad (12)$$

which solves such a confusion.

Influential observations in fixed-effect estimations can be made using Cook's distance, or also MDFFITS statistics, which is a multivariate version of DFFITS statistics (Belsley, Kuh, & Welsch, 2004). Both statistics determine the distance between fixed-effect estimations deriving from complete data, and those from reduced data, and are generalized by (Christensen, Pearson, & Johnson, 1992) and (Schabenberger, 2005) for the Hierarchical Linear Model as follows

$$C_i(\hat{\beta}) = (\hat{\beta} - \hat{\beta}_{\iota})^T \widehat{Var(\hat{\beta})}^{-1} (\hat{\beta} - \hat{\beta}_{\iota})/p \quad (13)$$

$$MDFFITS_i(\hat{\beta}) = (\hat{\beta} - \hat{\beta}_{\iota})^T \widehat{Var(\hat{\beta})}^{-1} (\hat{\beta} - \hat{\beta}_{\iota})/p \quad (14)$$

These statistics present great values for influential observation and, because $\hat{V}$ *is* used, there is no exact reference distribution for it.

Components of variance estimation is another attribute that allows analyzing the influence of possible model observations. Even though components of variance estimations are not of primary interest for researchers, it is fundamental to investigate this part of the model because components of variance impact its fixed part (Loy & Hofmann, 2013). For the Variance Components, one can directly compare the relative change for each variance component, θ_l

$$RVC_i(\theta_l) = \frac{\widehat{\theta_{l\iota}}}{\widehat{\theta_l}} - 1 \quad (15)$$

Note that the relative variance change (RVC) will be close to zero when the i-th unit does not influence the component of variance in question.

All decision measures for choosing the best proposed fit, estimating parameters and diagnosing the model were obtained from software R version 3.4.4 (R Core Team, 2019) with the following packages (ggplot2, HLMdiag, lme4, nlme, mlmRev, nullabor, fitdistrplus, dbplyr , plyr).

The model selection was based on the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the likelihood-ratio test.

Results

Initially, a descriptive analysis was performed to ascertain the behavior of the subjects' weight during the experiment; this analysis was divided in two parts – subjects' initial weight (1st week), and subjects' weight in the following experiment weeks (2nd to 6th week). The dataset used for both the descriptive analysis and the fittings is made up of 342 observations, of which 16 were lost (not recorded).

Mice weight descriptive analysis

To study the initial weight of the mice, a statistical summary was produced, observing that the means of the groups range from 39.24 g for group 4 to 42.19 g for group 2, with standard deviations equal to 3.68 and 4.04 g, respectively.

Figure 1 shows the scatter of the weight of each subject at the beginning of the experiment and how discrepant the weights of the mice are. The treatment group to which each subject belongs is identified with different colors.

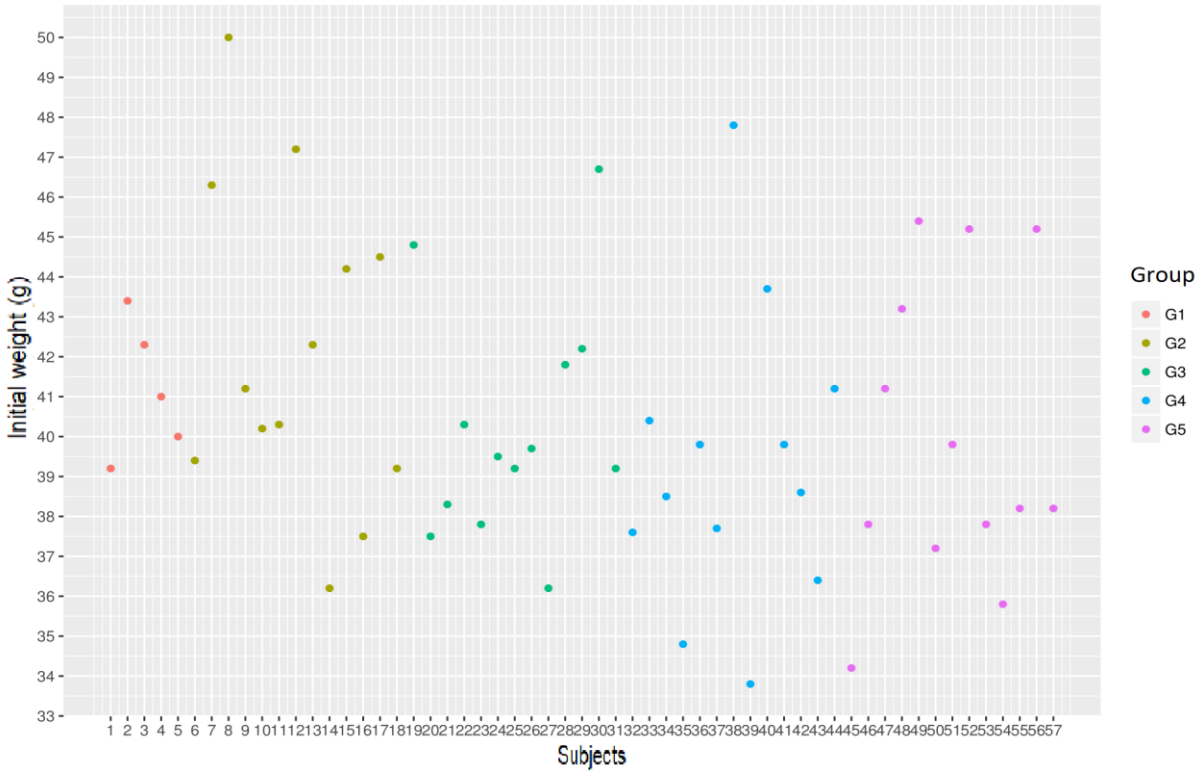

Figure 1. Scatter plot of the subjects' initial weight.

Figure 2 shows the weight evolution of each subject as a function of time, evidencing a similar behavior as the weeks go by. In addition, Figure 3 shows that all groups had an average weight increase as a function of time.

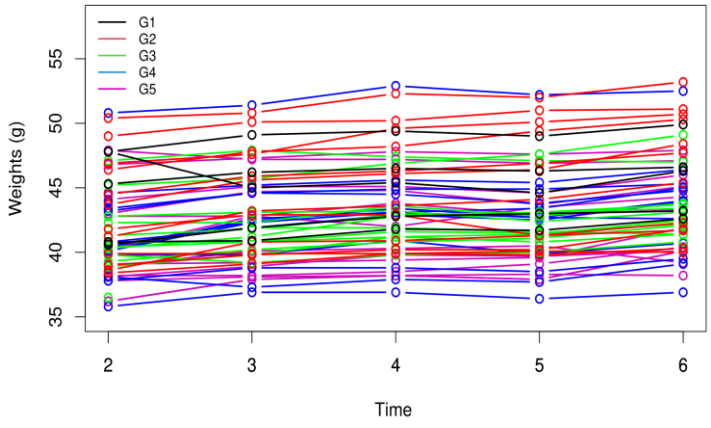

Figure 2. Graph for each subject's profile over time.

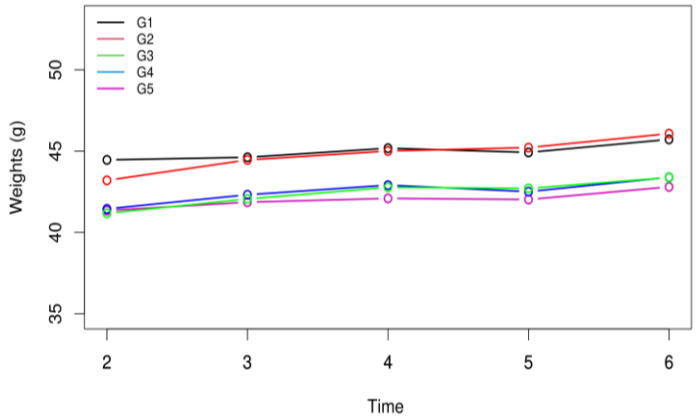

Figure 3. Mean weight of the groups as a function of time.

The highest mean between groups was found in the 6th week, for group 2, while the lowest one was found in the 2nd week, for group 3. The largest deviation between groups also occurred in the 6th week of group 2, while the lowest one occurred in the 5th week of group 3.

Figure 4 displays a histogram, a Cumulative Distribution Function (CDF) graph, a QQ-Plot and a PP-Plot, all as a function of weight observations for all treatments from the 2nd to the 6th week, considering normal distribution as reference.

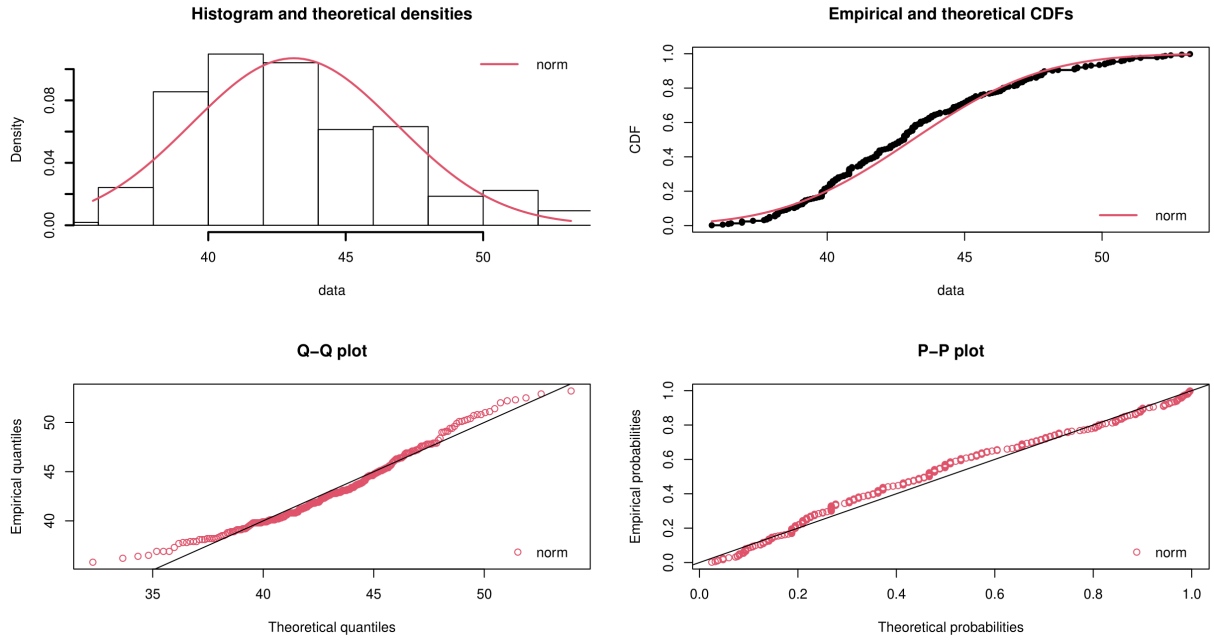

Figure 4. Histogram, fitted density function, cumulative empirical distribution, fitted distribution function, theoretical quantiles, empirical quantities, theoretical probabilities and empirical probabilities.

The Kolmogorov-Smirnov test presented D = 0.077755 and p-value = 0.077 as statistics, indicating that the subjects' weights follow normal distribution.

Model fit

The analysis of the models sought to ascertain the behavior of the subjects infected by *Trypanosoma cruzi* by relating the response variable 'weight' to possible influencing factors, cage and time, which belong to the fixed part of the model. The treatment group and the subjects themselves belong to the random part.

The models were fitted for response variable , in which $i = 1, 2, \ldots, 5$ (group number), $j = 1, 2, \ldots, n_i$ (number of subjects, where if i = 1, $n_i = 5$, and if i = 1, $n_i = 13$), k = 1, . . . , m_i (cage number, where if 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 $m_i = 11, 12, 13, 14$) and t = 2, . . . , 6 (treatment week).

The proposed models are described in Table 1; the first one is given only by the intercept of the fixed part, while the effect of the subjects is given by means of a random intercept for each one of them. In the

second, the effect of time is added to the fixed part; in the third model, the effect of the groups is added to the fixed part; in the fourth model, in addition to time and group, the effect of the cages is added to the fixed part.

Table 1. Proposed models.

	Model
1	$y_j = \beta_0 + b_{0j} + e_j$
2	$y_{jt} = \beta_0 + Time_t\lambda_t + b_{0j} + e_{jt}$
3	$y_{ijt} = \beta_0 + Time_t\lambda_t + Group_i\alpha_i + b_{0j} + e_{ijt}$
4	$y_{ijkt} = \beta_0 + Time_t\lambda_t + Group_i\alpha_i + Cage_k\theta_k + b_{0j} + e_{ijkt}$

α_i represents the effects of the i-th group, θ_k represents the effects of the k-th cage, λ_t represents the effects of the t-th time, b_j and e_j represent the effect and random error, respectively, of the j-th subject.

All proposed models presented violation in the residual normality assumption. The graphs in Figure 5 show how observation 16 stands out in 3 of the 4 fittings, so it is an observation worth investigating.

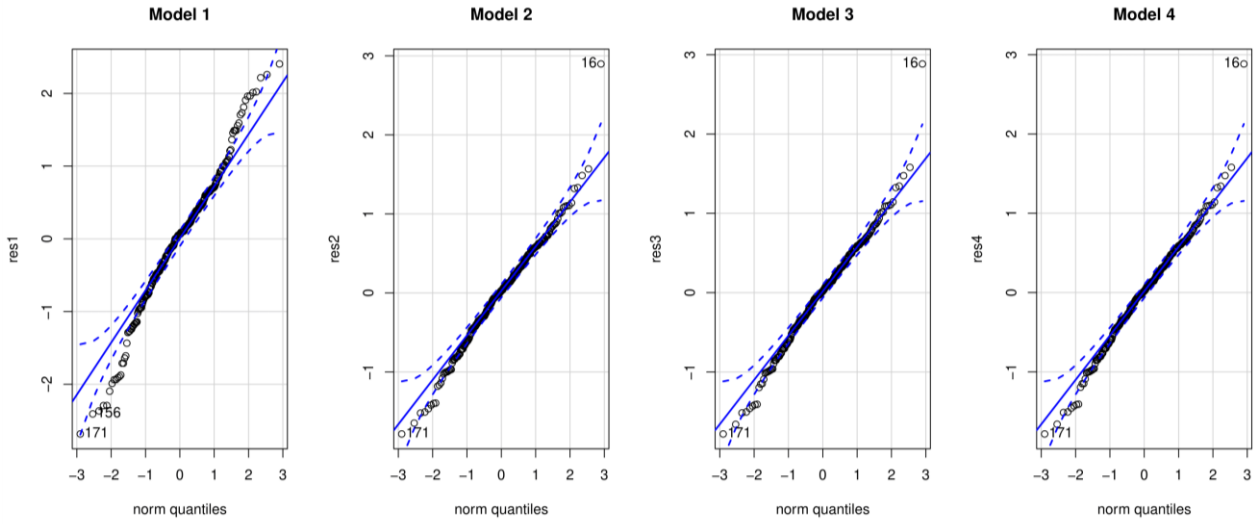

Figure 5. QQ-Plot for the fittings in the proposed models.

Observation 16 corresponds to the body weight of subject 4 in the second week; said subject belongs to group G1, which is the control with non-infected animals. Table 2 specifically shows this subject's data; it is possible to see that, in the second week, its weight had a sharp rise in relation to its initial weight, which was 41.00 g, then a drop in the 3rd week, presenting small variations until the 6th week.

Table 2. Observations for Subject 4.

Obs	Group	Subject	Cage	Cage	Time
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

This way, there is evidence that this observation has gone through a collection error, so the fittings were redone without observation 16.

Through the Shapiro-Wilk residual normality test, Model 2, Model 3 and Model 4 obtained, as p-value, 0.267, 0.266 and 0.265, respectively, thus presenting residual normality.

The values of all ANOVA decision measures for each selected model are displayed in Table 3.

Table 3. ANOVA decision measures.

Model	Df	AIC	BIC	logLik	GD	Chisq	Chi.Df.	Pr(>Chisq)
2	4	826.29	840.65	-409.14	818.29	167.43	1	<0.000
3	8	828.42	857.15	-409.21	812.42	5.87	4	0.21
4	9	830.24	862.56	-406.12	812.24	0.18	1	0.67

Models 2, 3 and 4 are statistically equivalent but, considering the model selection criteria, AIC and BIC, and the likelihood ratio test, Model 2 was chosen as the one the best explains the data. Another factor that led it to be chosen is that Model 2 is more parsimonious than the other equivalent ones.

Table 4 presents estimations of fixed-effect parameters for Model 2 with their respective confidence intervals.

Table 4. Estimations of fixed-effect parameters.

Variable	Estimation	Standard Error	t Value	Pr(>|t|)	95% CI
Intercept	41.232	0.49	84.43	<0.000	40.26 ; 42.20
Time	0.462	0.03	15.96	<0.000	0.405 ; 0.519

The validity of Model 2 can be confirmed by verifying its assumptions, which were assessed through diagnosis analysis.

Residual analysis

The level-1 and level-2 residuals of Model 2 were analyzed in an ascending manner. This analysis mode is necessary because the residuals are inter-related, therefore confounding, and can make the model diagnosis difficult if not analyzed correctly.

Residuals resulting from level 1 of least squares (LS) are not confounded with the residuals at level 2. By fitting the LS regression models separately, random effects are treated as fixed. Figure 6 presents a graph for the level-1 residuals of the LS, Time. It is suggested that time may not be linearly related to weight. In order to address the assumption of level-1 homoscedastic residuals, we used semi-standardized residuals, just as shown in (Snijders & Berkhof, 2008). Figure 7 display this graph, which indicates no linearity violation by said assumption.

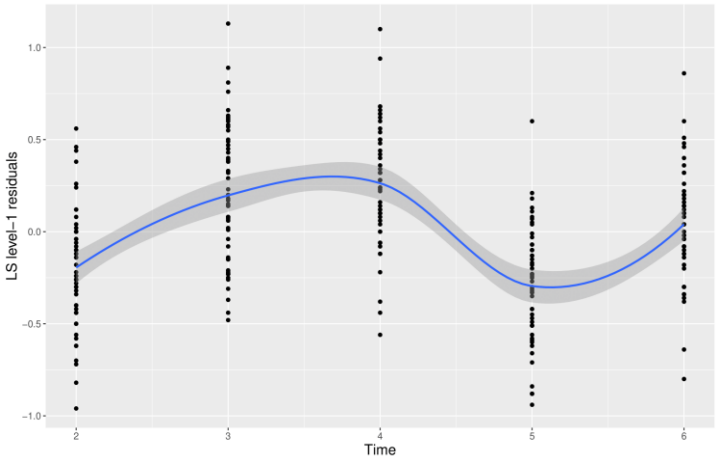

Figure 6. Graph for level-1 LS residuals by time.

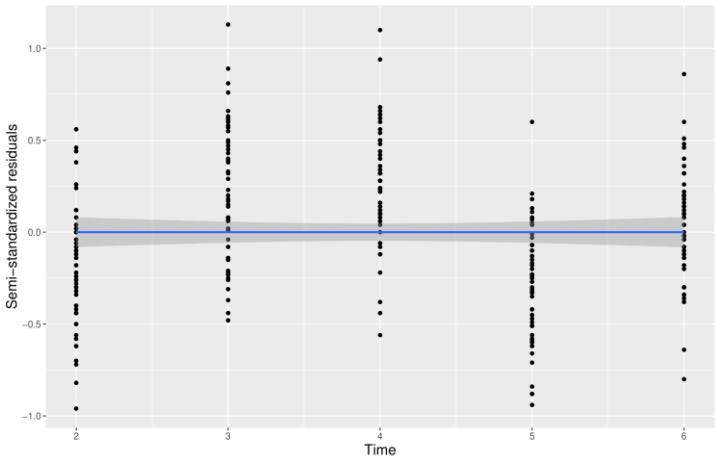

Figure 7. Graph for semi-standardized residuals by time.

Figure 8 shows the normal quantile graph of the level-1 semi-standardized residuals; visually, the semi-standardized residuals seem normal, thus showing no evidence against their normality assumption. The normality assumption of the residuals was assessed by the normal quantile graph presented in Figure 9.

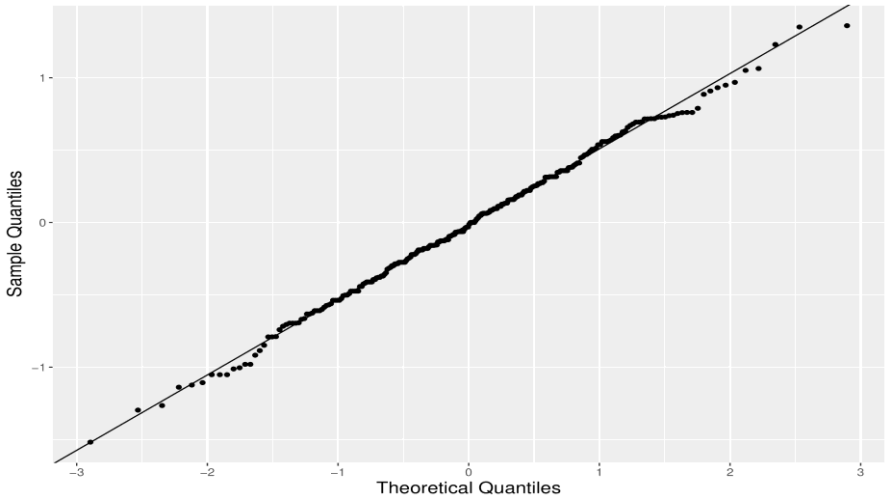

Figure 8. Normal quantile graph for level-1 semi-standardized residuals.

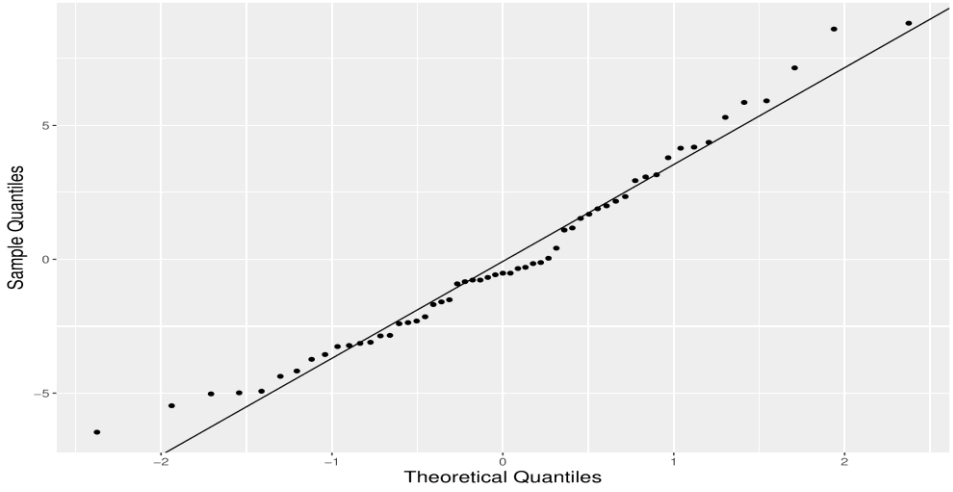

Figure 9. Normal quantile graph for level-2 Bayes EB empirical residuals for the intercept.

The programming protocol proposed by (Buja et al., 2009) was also used to corroborate with the verification of the assumption of level-1 homoscedastic residuals. It is displayed in Figure 10. It is not possible to distinguish the real-data graph from the simulated graph. Therefore, the analysis can move on without need for corrective measures.

The random effects, commonly referred to as level-2 effects, are defined by Zb, or b only. As previously said, an ascending residual analysis is performed, so residuals EB at level 2 are used. This choice was made because least square residuals are more variable than EB residuals, and EB estimates of b were obtained directly.

Analysis of influential observations

This sub-section will present the use of diagnoses to assess changes in the components of variance estimation using RVC, the fixed-effect estimation using Cook's distance, and fitted values using leveraging. These quantities are used to assess influences in level-1 and level-2 units.

For fixed effects, we have two statistics commonly used to measure whether there has been any changes, namely Cook's distance and MDFFITS, already mentioned. Figure 11 shows that no subject was identified by both statistics (which present the same values) as influential in the fixed effects, considering an internal scale. Figure 11 displays graphs for the abovementioned statistics in the form of dot plot and modified dot plot.

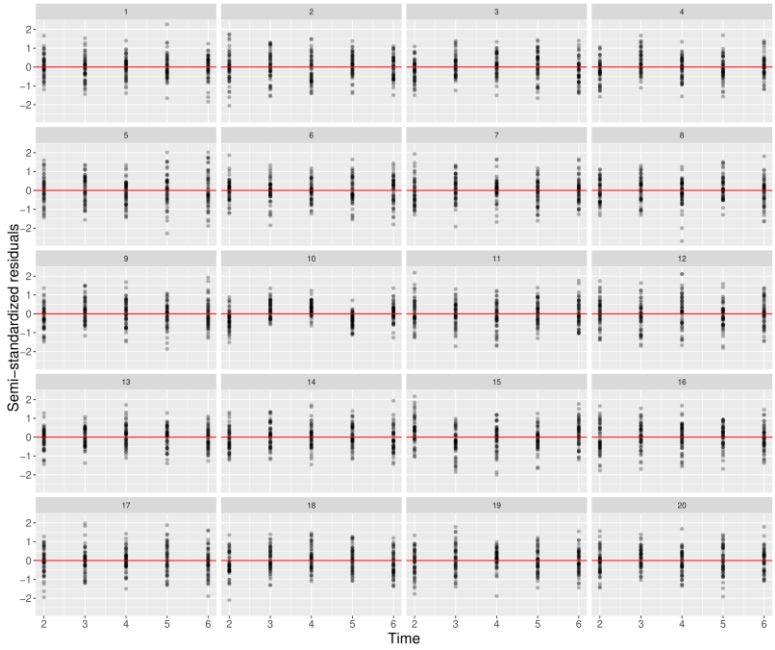

Figure 10. Panel with the twenty graphs that present semi-standardized residuals from a hierarchical model in relation to the 'time' predictor variable. The plotting of the real data was randomly incorporated into nineteen simulated graphs.

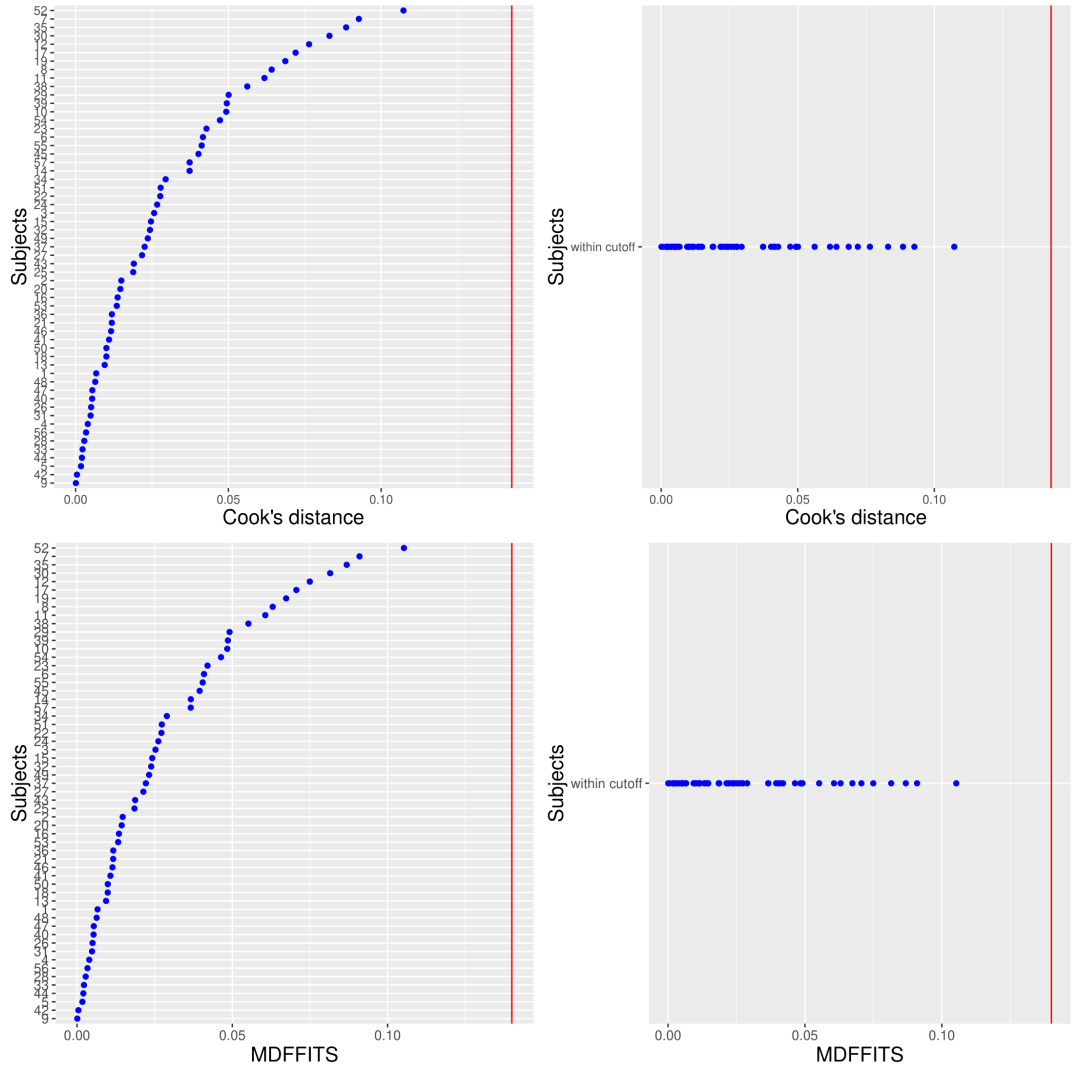

Figure 8. On the left, a dot plot for Cook's distance and MDFFITS; on the right, a modified dot plot for Cook's distance and MDFFITS, both for level 2 (subject).

For the components of variance, we used the RVC, which was presented in equation 15; it measures changes in the estimations of the A-th component of variance, θ_A, with and without unit i.

Table 5 illustrates the RVC, which presents as output a matrix where each column represents a component of variance, σ2, which is the residual variance, and D11, which is the variance associated with the random intercept of the subjects. Note that the value farthest from zero in D11 for n=95, characterizing this unit as influential in the component of variance.

Table 5. Fragment of the relative variance change matrix.

$\widehat{\sigma^2}$	$\widehat{\sigma^2}$	$\widehat{D11}$	$\widehat{D11}$
10.014131707930	-0.0160375532	580.0357492791870	.0108306308
.	.. .	.	.. .
.	.. .	.	.. .
90.004878220938	0.0116638987	660.035999932295	**-0.0780134695**
.	. . .	.	.. .
.	.. .	.	. . .
280.014185000957	-0.0113729758	850.03653966641140	.0210282370
29	-0.0054484352	860.0365806731	

Figure 9 shows the modified dot plot of the level-2 RVC for the random intercept of the subjects. By using the internal scale, n=95 was identified as influential unit, just as shown in Table 8, and is worth attention.

Subject n=95 corresponds to the weight of subject 38 at the beginning of the treatment; as shown in Table 6, said subject is the heaviest among the animals. This justifies the fact that it was identified as influential in the graph of Figure 9.

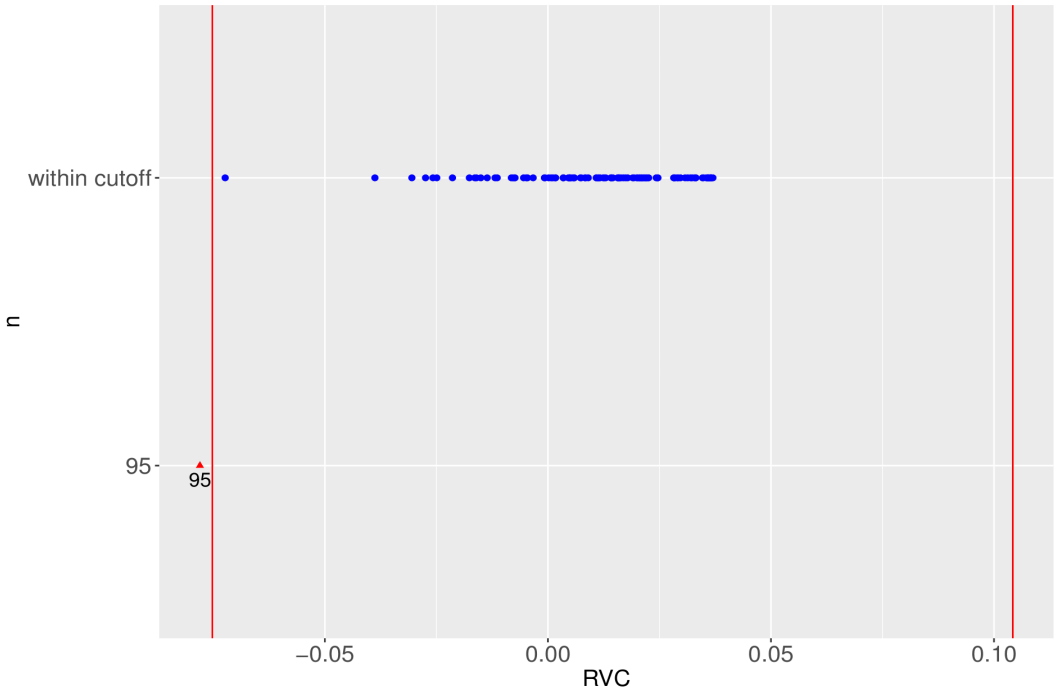

Figure 9. Modified dot plot of the level-2 RVC for the subject intercept.

In the fitted values, in addition to checking how the observations of one same subject directly impact the parameters of the adopted model, it was interesting to explore whether these observations are atypical in relation to the fitted values and to the explanatory variables of such model. This exploration was done through multiple statistics.

These statistics are displayed in Table 7, which reveals that subjects 9, 27, 44 and 56 have high leveraging in the fixed effects of the fitted values. Likewise, these subjects are identified as having high leveraging in the random effects, which is natural, since H2 depends on H1. Considering H2, all subjects present the same metrics, so all of them have the same influence under the random effects of the fitted values of Model 2.

Table 6. Subjects' weight in the second week.

Group	Subject	Cage	Group	Subject	Cage	Group	Subject	Cage
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

Table 7. Leverage points for the subjects' fixed effects, random effects and non-confounding random effects.

H	H_1	H_2	H_2^*		H	H_1	H_2	H_2^*
10.2025	0.0073	0.1951	28.9860	30	0.2025	0.0073	0.1951	28.9860
20.2025	0.0073	0.1951	28.9860	31	0.2025	0.0073	0.1951	28.9860
90.9672	**0.0173**	**0.9499**	28.9860	38	0.2025	0.0073	0.1951	28.9860
100.2025	0.0073	0.1951	28.9860	39	0.2025	0.0073	0.1951	28.9860
150.2025	0.0073	0.1951	28.9860	**44**	0.9672	**0.0173**	**0.9499**	28.9860
160.2025	0.0073	0.1951	28.9860	45	0.2025	0.0073	0.1951	28.9860
270.9672	**0.0173**	**0.9499**	28.9860	**56**	0.9672	**0.0173**	**0.9499**	28.9860
280.2025	0.0073	0.1951	28.9860	57	0.2025	0.0073	0.1951	28.9860
290.2025	0.0073	0.1951	28.9860					

Because the intercept of subject # 38 was identified as being influential in the components of variance, and the observations for subjects # 9, 27, 44 and 56 proved to be influential in the fitted values, both were removed, and the fitting of the chosen model was repeated in order to indicate whether there was any significant change in the estimations of the parameters.

The estimations of the parameters without influential observations are displayed in Table 8.

Table 8. Estimations of the parameters without influential observations.

Variable	Estimation	Standard Error	t Value	Pr(>|t|)	95% CI
Intercept	41.275	0.510	80.82	<2e-16	40.25 ; 42.29
Time	0.461	0.029	15.81	<2e-16	0.403 ; 0.518

Note that, in comparison with Table 4, the parameters went through small changes. Therefore, since the influential observations have little impact on the selected model, we chose to keep them in order to preserve the data. Thus, we consider the initial estimations for Model 2.

Conclusion

The proposed Linear Mixed Model proved to be satisfactory for the dataset with hierarchical structure, besides managing to describe the behavior of the subjects' weight during the experiment weeks. The Group and Cage variables showed no significance for the Model; only the Time variable and the intercept (initial weight) were capable of describing the weight of the individuals in the Model. Such an importance of the

intercept in the Model is due to it being considered as random, since each subject presented a different initial weight. Thus, the Model with only intercept and time was capable of describing the subjects' weight, with acceptable estimations.

Taking the analysis results into account, we suggest, as recommendation, that researchers should start their experiments with the subjects' initial weight as similar as possible and divide the subjects equally into groups. It is also advisable to test new biotherapeutics for the treatment, as the ones analyzed in the present study (G13cH, G6cH and G3cH) did not prove to be efficient.

References

Aleixo, D. L., Veiga, F. K., Ciupa, L., Portocarrero, A. R., Sandri, P. F., Ferraz, F. N., ... Araújo, S. M. (2013). Highly diluted medication modifies parasitological, immunological and histopathological parameters of *Trypanosoma cruzi* infection. *International Journal of High Dilution Research*, *12*(44), 107-108.

Belsley, D. A., Kuh, E., & Welsch, R. E. (2004). *Regression diagnostics: Identifying influential data and sources of collinearity* (Wiley series in probability and statistics). Hoboken, NJ: John Wiley & Sons, Inc.

Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E. K., Swayne, D. F., & Wickham, H. (2009). Statistical inference for exploratory data analysis and model diagnostics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1906), 4361-4383. doi: 10.1098/rsta.2009.0120

Christensen, R., Pearson, L. M., & Johnson, W. (1992). Case-deletion diagnostics for mixed models. *Technometrics*, *34*(1), 38-45. doi: 10.2307/1269550

Demidenko, E., & Stukel, T. A. (2005). Influence analysis for linear mixed-effects models. *Statistics in Medicine*, *24*(6), 893-909. doi:10.1002/sim.1974

Ferraz, F. N., Aleixo, D. L., Gruendling, A. P., Gomes, M. L., Toledo, M. J. O., & Araújo, S. M. (2016). *Trypanosoma cruzi*: Evaluation of PCR as a laboratory tool to follow up the evolution of parasite load. *Iranian Journal of Parasitology*, *11*(3), 389-395. PMID: 28127346

Ferreira, É. C., Ciupa, L., Portocarrero, A. R., Brustolin, C. F., Massini, P. F., Aleixo, D. L., & Araújo, S. M. (2018). Phosphorus protects cardiac tissue by modifying the immune response in rats infected by *Trypanosoma cruzi*. *Cytokine*, *102*, 102-106. doi: 10.1016/j.cyto.2017.07.017

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Bristol, UK: John Wiley & Sons

Harville, D. A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *The Annals of Statistics*, *4*(2), 384-395. Retrieved from www.jstor.org/stable/2958209

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*(358), 320-338. doi: 10.2307/2286796

Ignácio, S. A. (2010). Importância da estatística para o processo de conhecimento e tomada de decisão. *Revista Paranaense de Desenvolvimento*, (118), 175-192.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963-974. PMID: 7168798

Loy, A., & Hofmann, H. (2013). Diagnostic tools for hierarchical linear models. *WIREs Computational Statistics*, *5*(1), 48-61. Doi 10.1002/wics.1238

Nobre, J. S., & Singer, J. M. (2011). Leverage analysis for linear mixed models. *Journal of Applied Statistics*, *38*(5), 1063-1072. doi: 10.1080/02664761003759016

Nussenzweig, V., Sonntag, R., Biancalana, A., Freitas, J. D., Amato Neto, V., & Kloetzel, J. (1953). 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*, *44*(6), 731-744.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing - Version 3.4.4*. Vienna, AU: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Advanced quantitative Techniques in the Social Sciences, Vol 1, 2nd ed.). Thousand Oaks, CA: Sage Publications

Schabenberger, O. (2005). Mixed model influence diagnostics. In *SAS Users Group International Conference* (Vol. 29, paper 189). Retrieved from https://support.sas.com/resources/papers/proceedings/proceedings/sugi29/189-29.pdf

Scheffe, H. (1999). *The analysis of variance*. New York: Wiley-Interscience Publication.

Snijders, T. A., & Berkhof, J. (2008). Diagnostic checks for multilevel models. In J. Leeuw, & E. Meijer (Eds.), *Handbook of multilevel analysis* (p. 141-175). New York, NY: Springer. Doi: 10.1007/978-0-387-73186-5_3

Verbeke, G., & Molenberghs, G. (2009). *Linear mixed models for longitudinal data*. New York, NY: Springer.

Vinhaes, M. C., & Dias, J. C. P. (2000). Doença de chagas no Brasil. *Cadernos de Saúde Pública, 16*(Suppl. 2), S7-S12. Doi: 10.1590/S0102-311X2000000800002