

Mixed Models

Pedro Victor Brasil Ribeiro

2021-11-22 - Last changed in 2021-12-12

Longitudinal Data

Longitudinal data, also called as panel data, is data that is collected through a series of repeated observations of the same subject overtime. Longitudinal experiments planning concern the observation of one or more variables in the same subject on different occasions or condition of evaluation, time is a common factor used in most of experiment, but it can be use on distances among other factors, but in the present work it will be implied that the subjects measures is taken over time.

Given that longitudinal data are measures of the same subject taken in a systematic way, is expected not null correlations between the measures, especially the one taken in consecutive. Furthermore is expected heterocedasticity.

About the data structure is expected 3 characteristics, but have in mind that in a real experiment ambient, not necessarily those characteristics can actually be hold:

- Regular (in respect to time): The interval of time between one measure and other are the same;
- Balanced (in respect to time): All observations are taken in the same time, on the same conditions in all subjects;
- Complete: No lost observations.

Table 1: Basic structure of balanced and complete longitudinal data

| Grupo ou Tratamento | Unidade Experimental | Condições de Avaliação | | | |
|---------------------|----------------------|------------------------|-------------|----------|-------------|
| | | 1 | 2 | ... | t |
| 1 | 1 | y_{111} | y_{112} | ... | y_{11t} |
| | 2 | y_{121} | y_{122} | ... | y_{12t} |
| | \vdots | \vdots | \vdots | \ddots | \vdots |
| | n_1 | y_{1n_11} | y_{1n_12} | ... | y_{1n_1t} |
| 2 | 1 | y_{211} | y_{212} | ... | y_{21t} |
| | 2 | y_{221} | y_{222} | ... | y_{22t} |
| | \vdots | \vdots | \vdots | \ddots | \vdots |
| | n_2 | y_{2n_21} | y_{2n_22} | ... | y_{2n_2t} |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| g | 1 | y_{g11} | y_{g12} | ... | y_{g1t} |
| | 2 | y_{g21} | y_{g22} | ... | y_{g2t} |
| | \vdots | \vdots | \vdots | \vdots | \vdots |
| | n_g | y_{gn_g1} | y_{gn_g2} | ... | y_{gn_gt} |

Mixed Model

For the analysis of longitudinal data to take into account the correlation between measurements taken in the same experimental unit, the adjustment of regression models with the inclusion of random effects can be considered, and it's oftenly used in these kind of cases. The usual models are:

- Linear Mixed Models;
- Generalized Linear Mixed Models;
- Non-Linear Mixed Models.

In this we will speak specifically about Linear Mixed Models (LMM) since is the most common model used in this kind of situation. So we'll use LMM to fit the mean profile of the subjects.

As usually done in Linear Models, the LMM is use to make the inference on the populacional mean of the populacion, but in the LMM is used the y conditioned in u, in other words $E[Y|u]$. For a LMM $X\beta$ represent the fixed effects, but is added a random effect represented by Zu , where just as X is the knew model matriz, and u the random variables used to define the condicional model, expressed as:

$$E[Y|u] = X\beta + Zu + \varepsilon \quad (1)$$

Where from the model 1 is usually assumed that $b_i \sim N_q(0, G)$ and $\varepsilon \sim N_{m_i}(0, R_i)$, where m_i is the number of instants that the subject is measured.

Table 2: Summary of the dimensions of each component

| Componente | Dimensão |
|---------------|----------------|
| y | $N \times 1$ |
| X | $N \times p$ |
| β | $p \times 1$ |
| Z | $N \times nq$ |
| b | $nq \times 1$ |
| ε | $N \times 1$ |
| Γ | $nq \times nq$ |
| R | $N \times N$ |
| Ω | $N \times N$ |
| θ | $t \times t$ |

Where in a more technicall manner the modelo 1, which $y = [y_1^T, \dots, y_n^T](N \times 1)$, with $N = \sum_{i=1}^n n_i$, have the fixed part and the random part, the parameter of the model can be write in a matter where $b \sim N_{nq}(0, \Gamma(\theta))$, where $\Gamma(\theta) = I_n \otimes G(\theta)$ is independent of $\varepsilon \sim N_N(0, R(\theta))$, $R(\theta) = \otimes_{i=1}^n R_i(\theta)$, with give the final distribution of y as $y \sim N_N(X\beta, \Omega(\theta))$, where $\Omega(\theta) = Z\Gamma(\theta)Z^T + R(\theta)$.

Which show us that the fixed effect affect only the mean of y, when the random affect affect the variance of y.

Structure of the covariance

The greater part of the effort of fit a Mixed Linear Model is to choose the structure of the covariance, which greatly affect the goodness-of-fit of the fitted model.

Theorically speaking those structure can be used on $\Omega(\theta)$, $R(\theta)$ and $G(\theta)$, but is usually used in the variance of y ($\Omega(\theta)$).

Some exemples of covariante structure, considering $n_i = 4$:

Uniform structure

$$[\theta = (\sigma^2, \tau)^T]$$

$$\begin{bmatrix} \sigma^2 + \tau & \tau & \tau & \tau \\ \tau & \sigma^2 + \tau & \tau & \tau \\ \tau & \tau & \sigma^2 + \tau & \tau \\ \tau & \tau & \tau & \sigma^2 + \tau \end{bmatrix}$$

AR(1) structure

$$[\theta = (\sigma^2, \phi)^T]$$

$$\begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix}$$

ARMA(1,1) structure

$$[\theta = (\sigma^2, \phi, \gamma)^T]$$

$$\begin{bmatrix} 1 & \gamma & \phi\gamma & \gamma\phi^2 \\ \gamma & 1 & \gamma & \gamma\phi \\ \gamma\phi & \gamma & 1 & \gamma \\ \gamma\phi^2 & \gamma\phi & \gamma & 1 \end{bmatrix}$$

Antidependence structure of 1st order

$$[\theta = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \rho_1, \rho_2, \rho_3, \rho_4)^T]$$

$$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \sigma_1\sigma_4\rho_1\rho_2\rho_3 \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 \\ \sigma_1\sigma_3\rho_1\rho_2 & \sigma_2\sigma_3\rho_2 & \sigma_3^2 & \sigma_3\sigma_4\rho_3 \\ \sigma_1\sigma_4\rho_1\rho_2\rho_3 & \sigma_2\sigma_4\rho_2\rho_3 & \sigma_3\sigma_4\rho_3 & \sigma_4^2 \end{bmatrix}$$

Toeplitz structure

$$[\theta = (\sigma^2, \sigma_1, \sigma_2, \sigma_3)^T]$$

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

Heterogeneously uniform structure

$$[\theta = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \rho)^T]$$

$$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

Not structured

$$[\theta = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_{12}, \sigma_{13}, \sigma_{14}, \sigma_{23}, \sigma_{24}, \sigma_{34})^T]$$

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}$$

Chicken diet Data

the `ChickWeight` available on R by using the code `data("ChickWeight")`, is a data that containing 578 rows and 4 variables, which measure 50 different chickens on 12 different times where the model contains some data which has some missing values that are not explicit on the data itself.

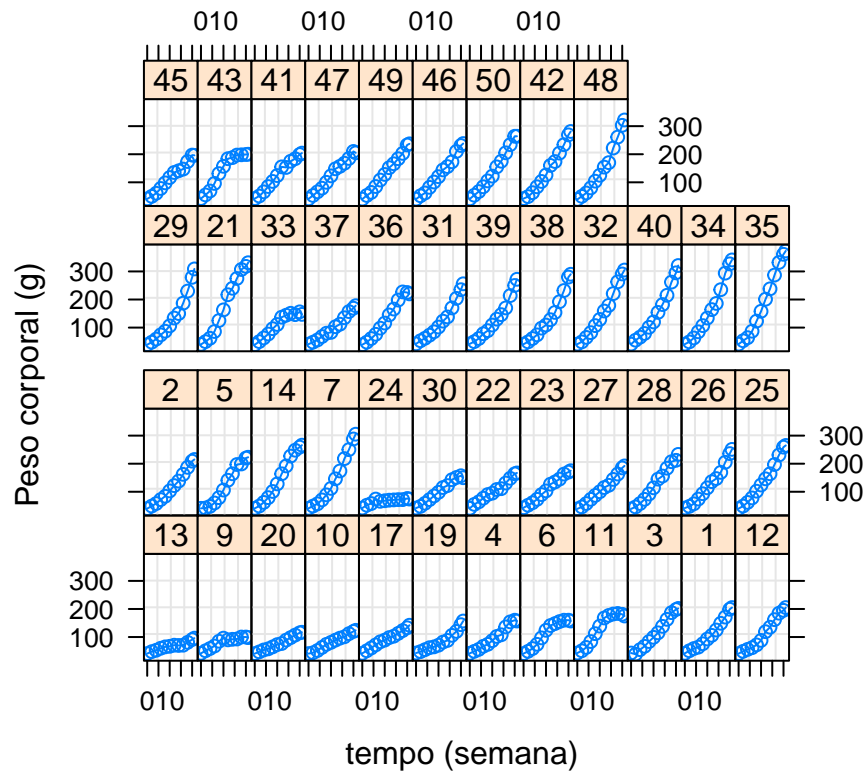
For example, the Chicken number 18 has only been measured on the time 0 and not on the subsequent time. The help (`help(data("ChickWeight"))`) of the database doesn't provide us the explanation/detail of why that occurs.

In the present work, we will be discarding those cases where the data is not complete, as defined in the beginning, since it is one of the assumptions of the model, but that problem can be solved by using a different model, such as Generalized Linear Mixed Models, considering a distribution capable of fitting data with "censor", or even by using a non-parametric analysis or Random Forest.

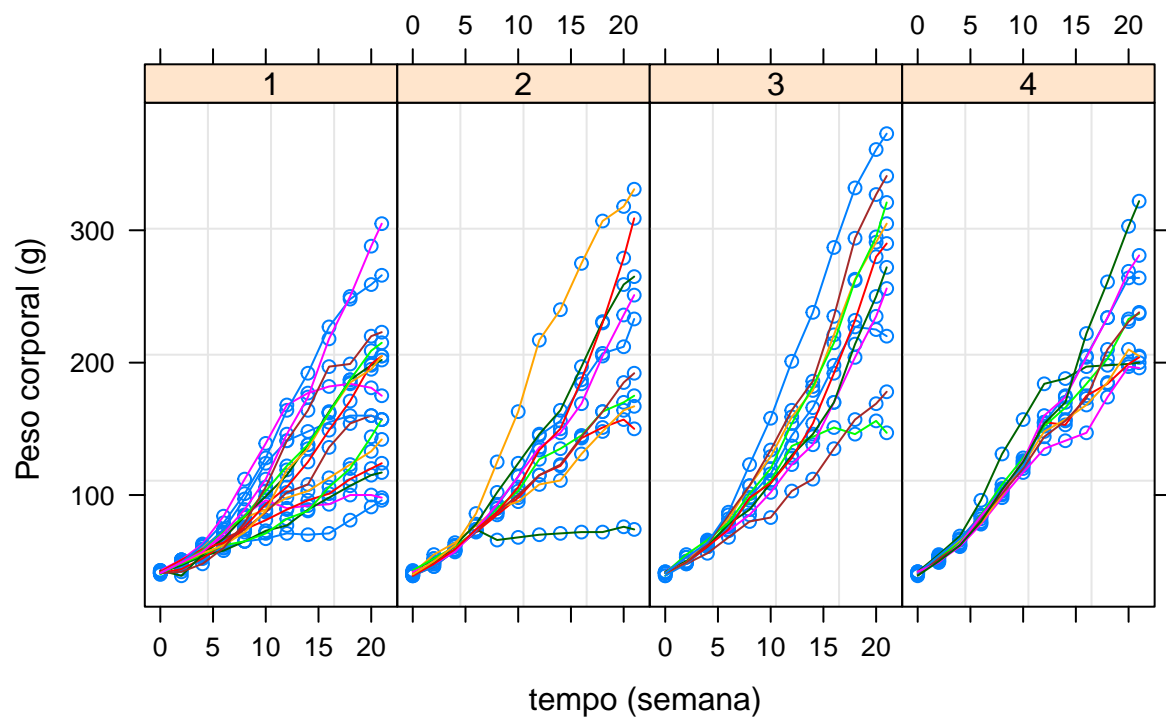
But since the presence of NA is only in 5 cases (8,15,16,18,44) it was decided to drop the data for lack of good technical features. The absence of which observation can be seen on the table below.

| Chick | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 21 |
|-------|---|---|---|---|---|----|----|----|----|----|----|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 7 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | x |
| 9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 10 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 11 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 12 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 13 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 14 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 15 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | x | x | x | x |
| 16 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | x | x | x | x | x |
| 17 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 18 | 1 | 1 | x | x | x | x | x | x | x | x | x | x |
| 19 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 20 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 21 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 22 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 23 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 24 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 25 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 26 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 27 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 28 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 29 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 30 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 31 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 32 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 33 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 34 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 35 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 36 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 37 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 38 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 39 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 40 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 41 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 42 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 43 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 44 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | x | x |
| 45 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 46 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 47 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 48 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 49 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 50 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

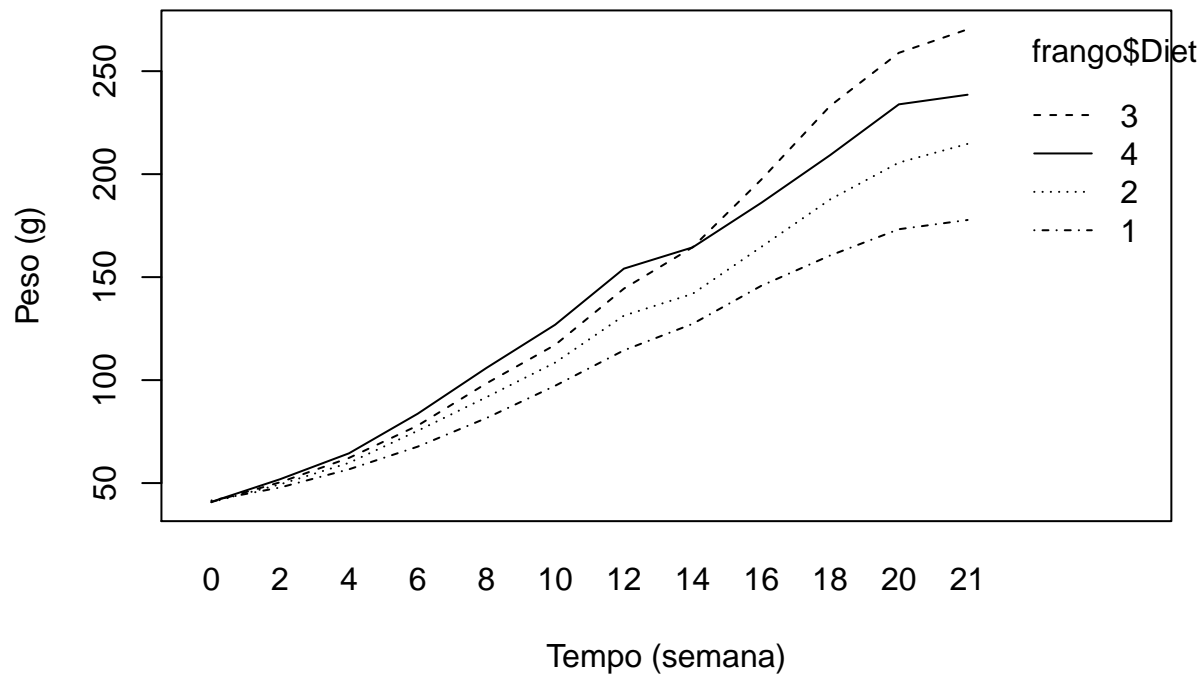
Now we'll give some descriptive analysis of the data to help us to get some help for the decision of the implementation of the random effects.



The first plot show us a different inclination for each chicken, when taken Weight in relation with Time, which shown us a possible random effects of the the inclination of the line, but is not shown a possible random effect on the intercepto. By a closer observation on the same plot can be shown a possible quadratic effect, so in the model will be added on the model, and after de decision of the model, we will make a anova to see if that quadratic coef is significative in relation with the model without it.



By the second graph, it gives us a good visual effect a high variation of the mean weight for each chicken in time for each Diet, what give us a good intuition that exist a random effect for this case.



For the last graph, is a good plot to show a possible difference on the effect for each Diet.

```
mod1 <- lme(
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,
  random = ~1,
  data = frango,
  control = lmeControl(opt="optim")
)
mod2 <- lme(
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,
  random = ~ Diet,
  data = frango,
  control = lmeControl(opt="optim")
)
mod3 <- lme(
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,
  random = ~Time+I(Time^2),
  data = frango,
  control = lmeControl(opt="optim")
)
mod4 <- lme(
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,
  random = ~Time,
  data = frango,
  control = lmeControl(opt="optim")
)
```



```

mod5 <- lme(
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,
  random = ~Time + Diet,
  data = frango,
  control = lmeControl(opt="optim")
)
mod6 <- lme(
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,
  random = ~Time + Diet + Time:Diet,
  data = frango,
  control = lmeControl(opt="optim")
)

```

So we did the modelling of the models, which can be seen in the code above. We mainly worked with the random effect part using the same fixed effect for them, based of the descriptive analysis that we made. So we made a table showing the AIC and BIC of the fitted models. Where AIC and BIC can be difined as:

$$\begin{aligned}
 AIC &= -2(\ln(\hat{L}) - 2k) \\
 BIC &= k\ln(n) - 2\ln(\hat{L})
 \end{aligned}
 \tag{2}$$

```

## # A tibble: 6 x 3
##   Modelo      AIC    BIC
##   <chr>    <dbl> <dbl>
## 1 Modelo 3 3991. 4059.
## 2 Modelo 4 4415. 4471.
## 3 Modelo 5 4433. 4540.
## 4 Modelo 6 4471. 4668.
## 5 Modelo 1 5119. 5166.
## 6 Modelo 2 5130. 5216.

```

So the model that gave the lower values for AIC and for BIC the the model 3, which the random effect is added only for the intercept and Time, but looked as a quadratic function. Which lead us to the coeficients on the table bellow:

| (Intercept) | Time | I(Time ²) | Diet2 | Diet3 | Diet4 | Time:Diet2 | Time:Diet3 | Time:Diet4 |
|-------------|------------|-----------------------|------------|---------|-----------|------------|------------|------------|
| 44.23283 | 2.0157770 | 0.0108738 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 36.39245 | 6.8166764 | -0.1847535 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 43.04724 | 2.4209261 | 0.0576192 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 40.59307 | 3.8794799 | 0.0044837 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 38.37823 | 5.1536904 | -0.0190936 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 45.75903 | 0.4185794 | 0.2200391 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 39.78235 | 3.9280085 | 0.1010582 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 28.24315 | 11.0727061 | -0.2165383 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 22.00133 | 14.7151570 | -0.3268902 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 38.02733 | 4.4858038 | 0.1792265 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 41.51242 | 2.4172066 | 0.2629587 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 39.99751 | 3.2644858 | 0.2361642 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 37.87230 | 4.5273952 | 0.1964772 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 35.38478 | 5.7866566 | 0.1812678 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 29.34284 | 9.1211018 | 0.1337274 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 38.40518 | 3.1828799 | 0.4653889 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 44.03860 | 2.6305467 | -0.1151147 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 37.74807 | 5.5050785 | -0.0379517 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 40.12662 | 4.0192851 | 0.0424911 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 36.97819 | 5.8064010 | -0.0094263 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 40.62468 | 3.3738338 | 0.1286087 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 36.29951 | 5.5533938 | 0.1309174 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 40.15285 | 3.0525819 | 0.2742729 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 36.72241 | 4.9243679 | 0.2436019 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 45.15060 | -0.6003678 | 0.5523146 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 23.65560 | 12.1501694 | 0.0967299 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 34.23868 | 7.9805664 | -0.2120190 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 46.04457 | 0.4188105 | 0.1755731 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 36.25356 | 5.7901696 | 0.0765723 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 45.89064 | -0.2440132 | 0.3715729 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 45.16007 | 0.0521081 | 0.3906814 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 45.48336 | -0.4475989 | 0.4725616 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 38.58810 | 3.6078468 | 0.3369167 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 41.45913 | 1.7534146 | 0.4396115 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 39.94182 | 2.3513985 | 0.4799223 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 31.38734 | 7.1874537 | 0.3587938 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 40.08508 | 4.0450996 | 0.0296904 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 24.59625 | 13.5773103 | -0.3739631 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 35.70593 | 6.6734419 | -0.0670478 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 37.32603 | 5.6129354 | -0.0088323 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 37.11864 | 5.4391340 | 0.0626519 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 40.14337 | 3.5452426 | 0.1493315 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 38.71929 | 4.0742510 | 0.2018616 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 39.64656 | 3.3822903 | 0.2539914 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |
| 43.28233 | 0.6874723 | 0.4641115 | -0.4449183 | -2.1266 | -1.387141 | 1.161036 | 2.369003 | 2.950291 |

It's interesting to note that the estimated coefficients for the intercept and for Time is different for each row, where that row represent each chicken looked in the model (45 chickens), but the others effect are not. Which is why we only added the random effect on these coefficients therefore the coefficients that we did not

list, there's no random effect, only fixed effects.

Changing the correlation matrix structure

Here we'll give a quick look on how we can add a different correlation structure on the selected models.

```
mod3_corAR1 <- lme(  
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,  
  random = ~Time+I(Time^2),  
  data = frango,  
  control = lmeControl(opt="optim"),  
  correlation = corAR1()  
)  
mod3_corARMA <- lme(  
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,  
  random = ~Time+I(Time^2),  
  data = frango,  
  control = lmeControl(opt="optim"),  
  correlation = corARMA(p = 0, q = 2)  
)  
mod3_corCAR1 <- lme(  
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,  
  random = ~Time+I(Time^2),  
  data = frango,  
  control = lmeControl(opt="optim"),  
  correlation = corCAR1(form = ~Time)  
)  
mod3_corCompSymm <- lme(  
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,  
  random = ~Time+I(Time^2),  
  data = frango,  
  control = lmeControl(opt="optim"),  
  correlation = corCompSymm()  
)  
mod3_corExp <- lme(  
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,  
  random = ~Time+I(Time^2),  
  data = frango,  
  control = lmeControl(opt="optim"),  
  correlation = corExp(form = ~Time)  
)  
mod3_corGaus <- lme(  
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,  
  random = ~Time+I(Time^2),  
  data = frango,  
  control = lmeControl(opt="optim"),  
  correlation = corGaus(form = ~Time)  
)  
mod3_corLin <- lme(  
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,  
  random = ~Time+I(Time^2),  
  data = frango,  
  control = lmeControl(opt="optim"),
```

```

    correlation = corLin(form = ~Time)
  )
mod3_corRatio <- lme(
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,
  random = ~Time+I(Time^2),
  data = frango,
  control = lmeControl(opt="optim"),
  correlation = corRatio(form = ~Time)
)
mod3_corSpher <- lme(
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,
  random = ~Time+I(Time^2),
  data = frango,
  control = lmeControl(opt="optim"),
  correlation = corSpher(form = ~ Time)
)
mod3_corSymm <- lme(
  fixed = weight~Time+I(Time^2)+Diet+Time:Diet,
  random = ~Time+I(Time^2),
  data = frango,
  control = lmeControl(opt="optim"),
  correlation = corSymm(form = ~ 1|Chick, fixed = T)
)

```

As said before the selection of the models used can be (and usually is) a big problem and might require a huge theoretical statistical background, so in these document we'll talk about someways, to select those models, and we'll not enter on the specifics on why the selection of the specific parameters were made.

Because the explanation will depend for the specific structure, but in a nutshell, the selection can be think on how the correlation matrix behave and if it "fits", on the structures shown bellow.

Note that not aways the models converge, problem that is more commom for the correlation structure that can correlate to the "Time Series" problems, such as the AR and ARMA models. That behavior is shown when the roots of the models cannot be but inside a circle on ratio 1. For more details i highly sugest a good read on stationarity and unit roots for ARIMA Process ¹

Interpretation of the models and model selection

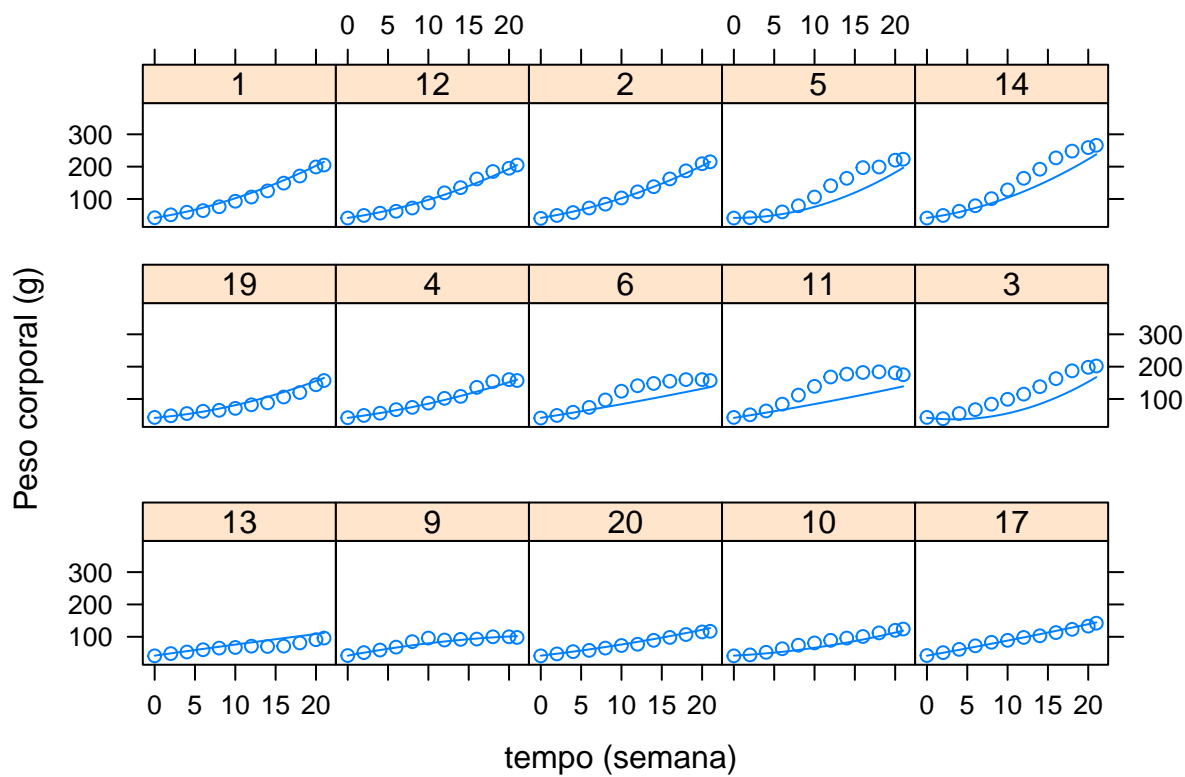
```

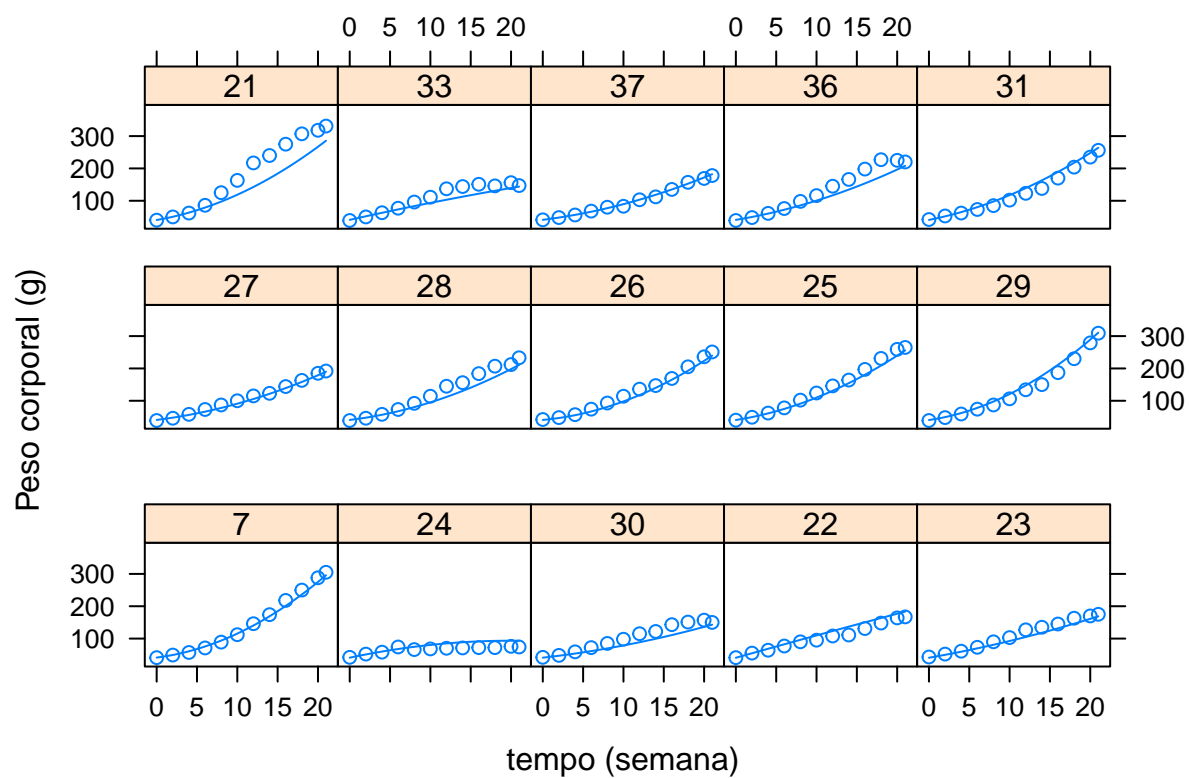
## # A tibble: 10 x 3
##   modelo      AIC    BIC
##   <chr>      <dbl> <dbl>
## 1 mod3_corRatio 3770. 3843.
## 2 mod3_corSpher 3778. 3851.
## 3 mod3_corARMA02 3781. 3858.
## 4 mod3_corLin 3792. 3864.
## 5 mod3_corAR1 3801. 3874.
## 6 mod3_corExp 3817. 3889.
## 7 mod3_corCAR1 3817. 3889.
## 8 mod3_corGaus 3857. 3929.
## 9 mod3_corCompSymm 3985. 4057.
## 10 mod3_corSymm 3991. 4059.

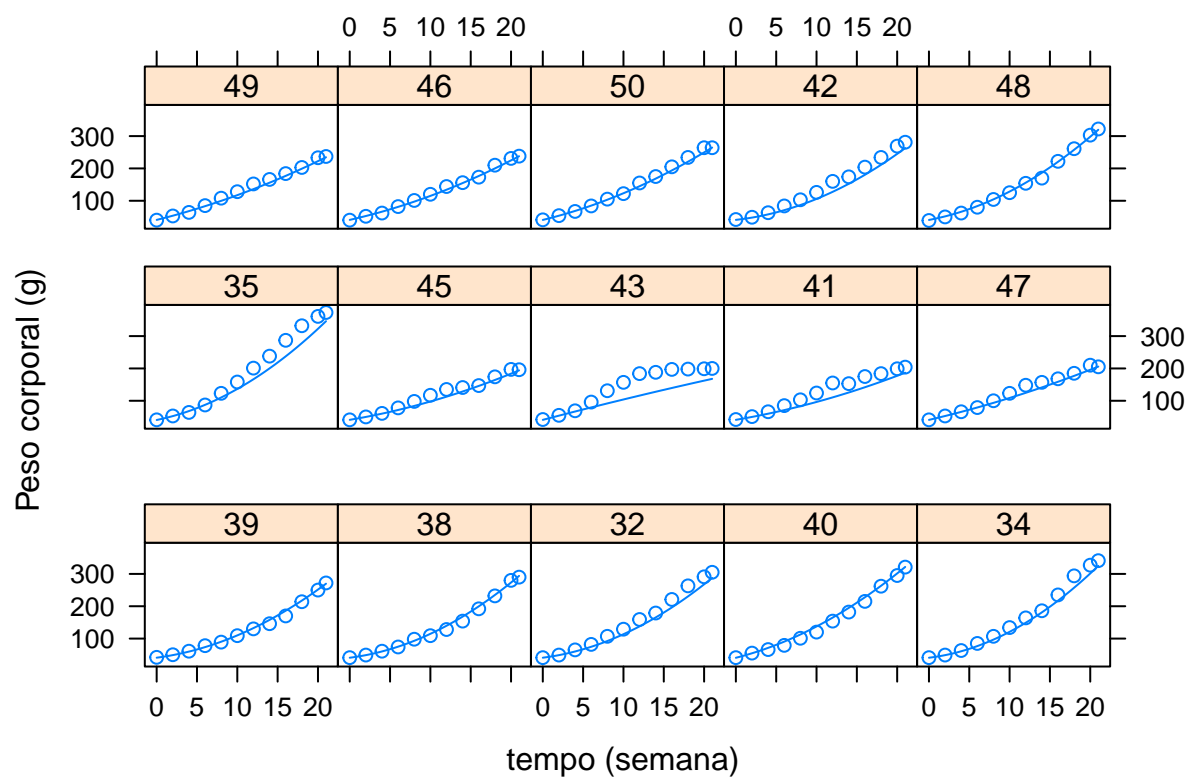
```

¹Some read sugestion: [Link 1](#) and [Link 2](#)

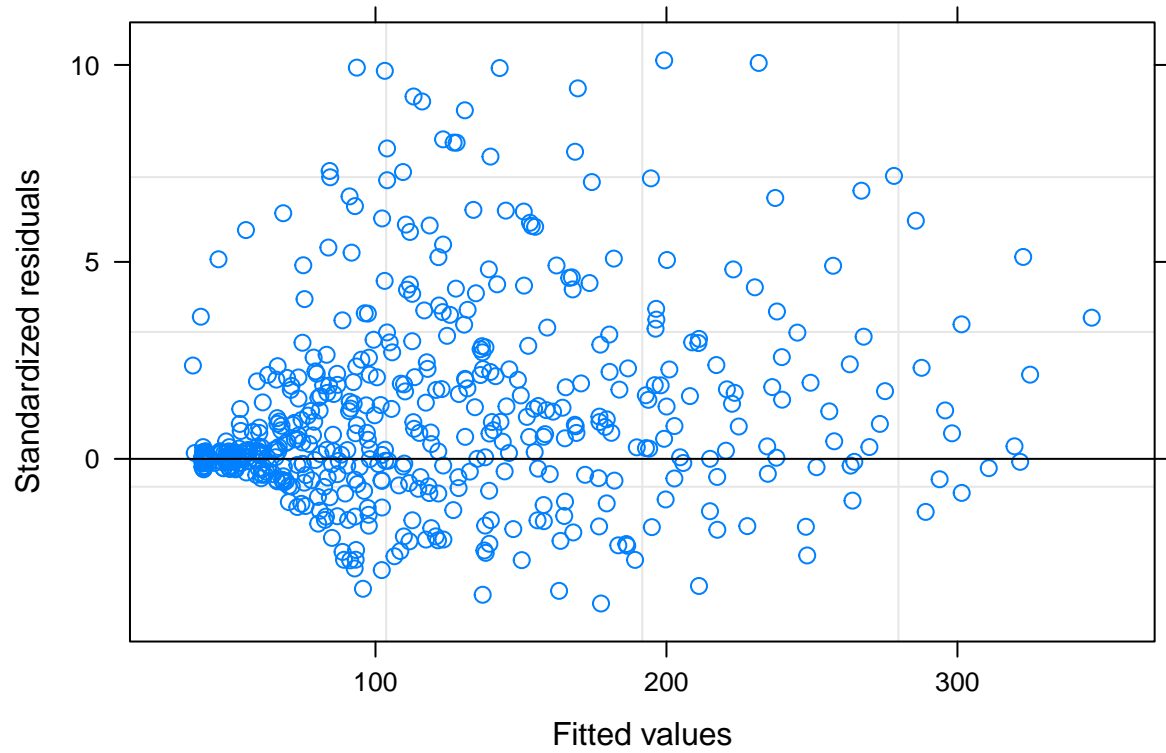
Using the AIC and BIC values to select the model we have that the lower value is related to “mod3_corRatio”, but we can see some problems with the fitted values and the residual of the model.



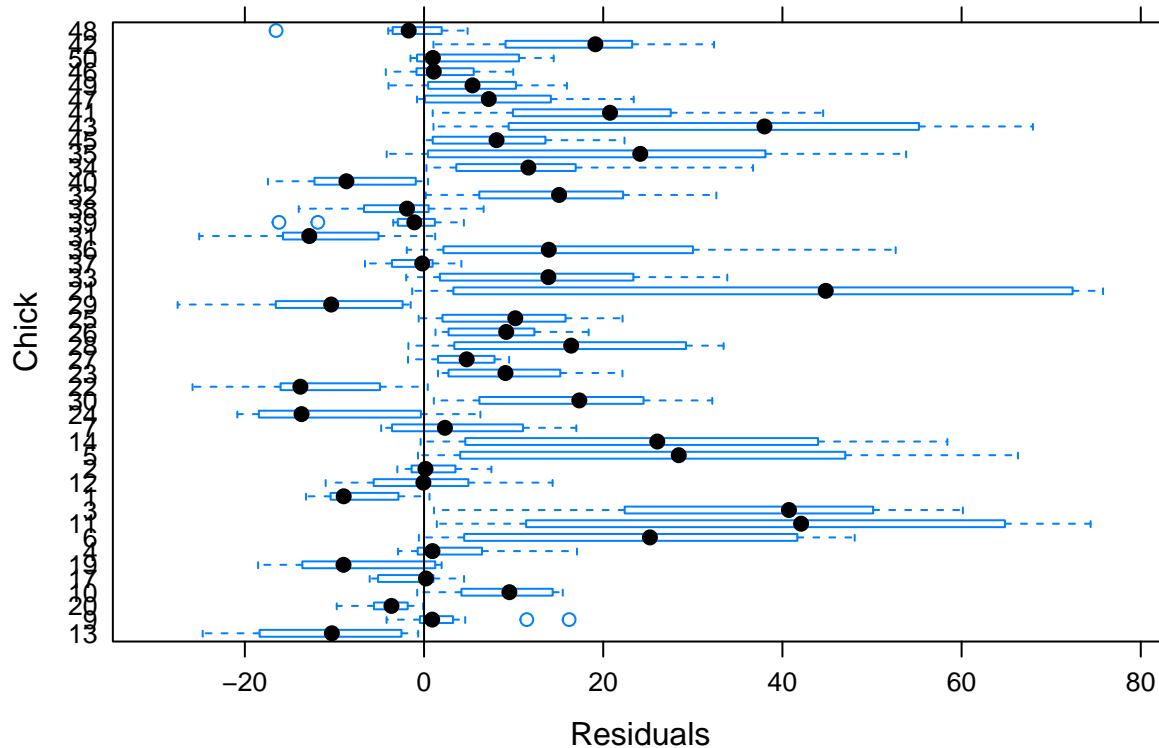




In the plot above we can see some cases which the fitted line does not have a good fit, mainly on the cases where the quadratic coefficient could be relevant.



In the graph above is made a dispersion point for the fitted values and the standardized residuals, in here we can clearly see a pattern in a shape of a funnel on the left side of the plot.



Lastly we see some cases which the value 0 is not inside the variation of the boxplot, which is a indication that the coefficient for the specific chicken is not significative.

So for the selected models we do not have good residuals, but why that occurs? Since we show that the specific data might have a good response for the Mixed Models, a short answer is that the AIC and BIC values they are not good for the selection of this type of model, since the models have a great ammount of parameters and both values (AIC and BIC) penalized the model for the number of parameters they have, which can be shown in the equation 2, they're not good for the selection for Mixed Models ².

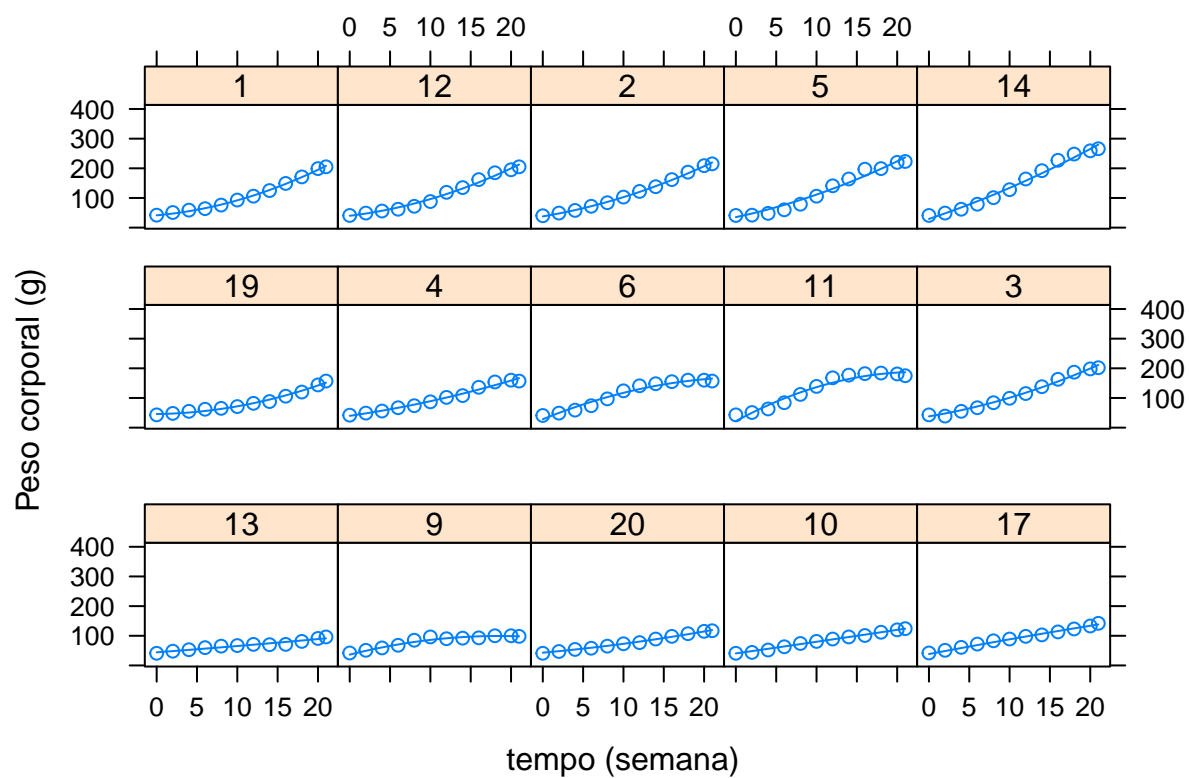
Model Comparison

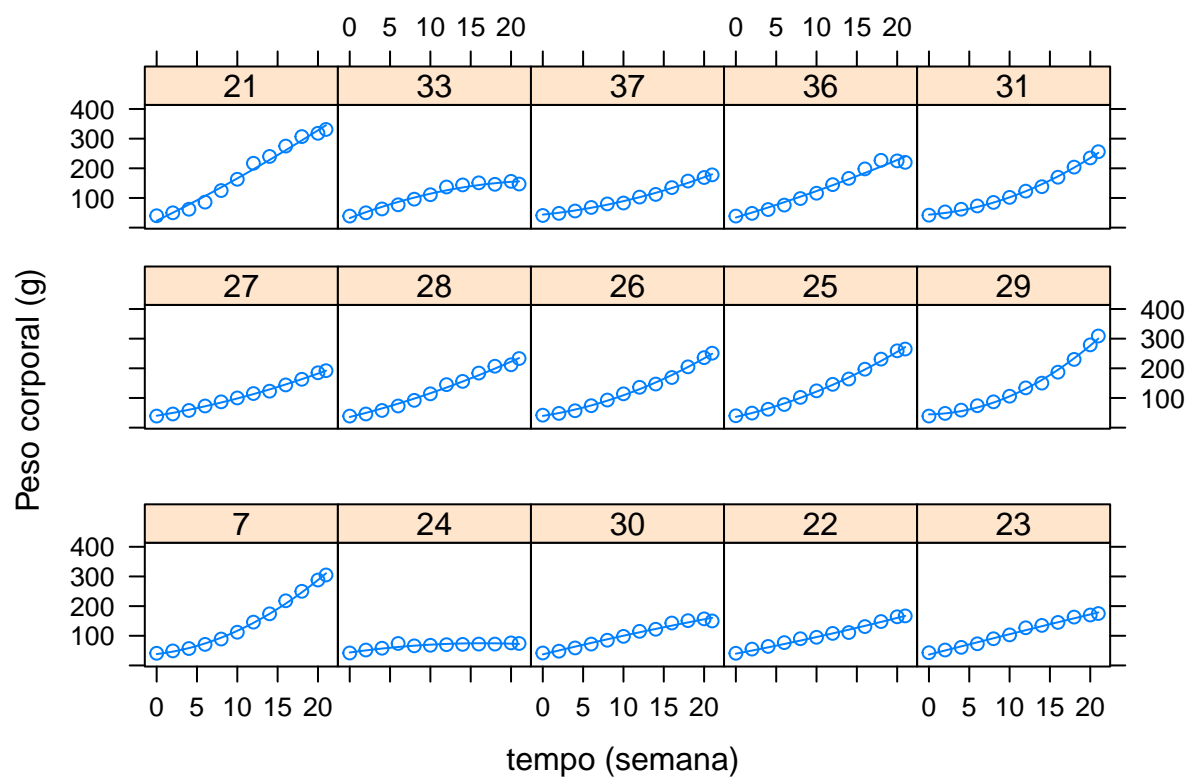
A correlation structure that usually have good results is the Not structured one, som we will used it to make some comparisons with the model selected for the AIC and BIC values. First of all we will make a Nested Model Test.

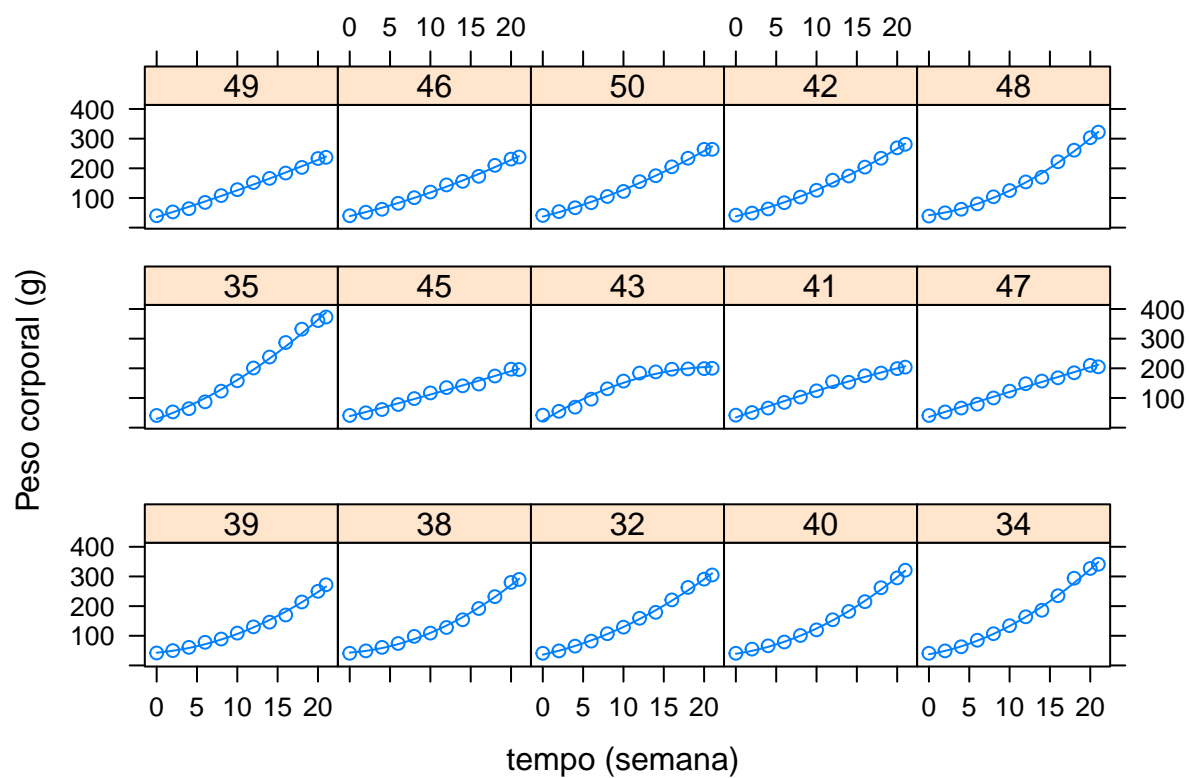
```
anova(mod3_corSymm,mod3_corRatio)
```

| ## | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|----|---------------|------|----------|----------|-----------|--------|----------|---------|
| ## | mod3_corSymm | 1 16 | 3990.842 | 4059.238 | -1979.421 | | | |
| ## | mod3_corRatio | 2 17 | 3770.268 | 3842.939 | -1868.134 | 1 vs 2 | 222.5742 | <.0001 |

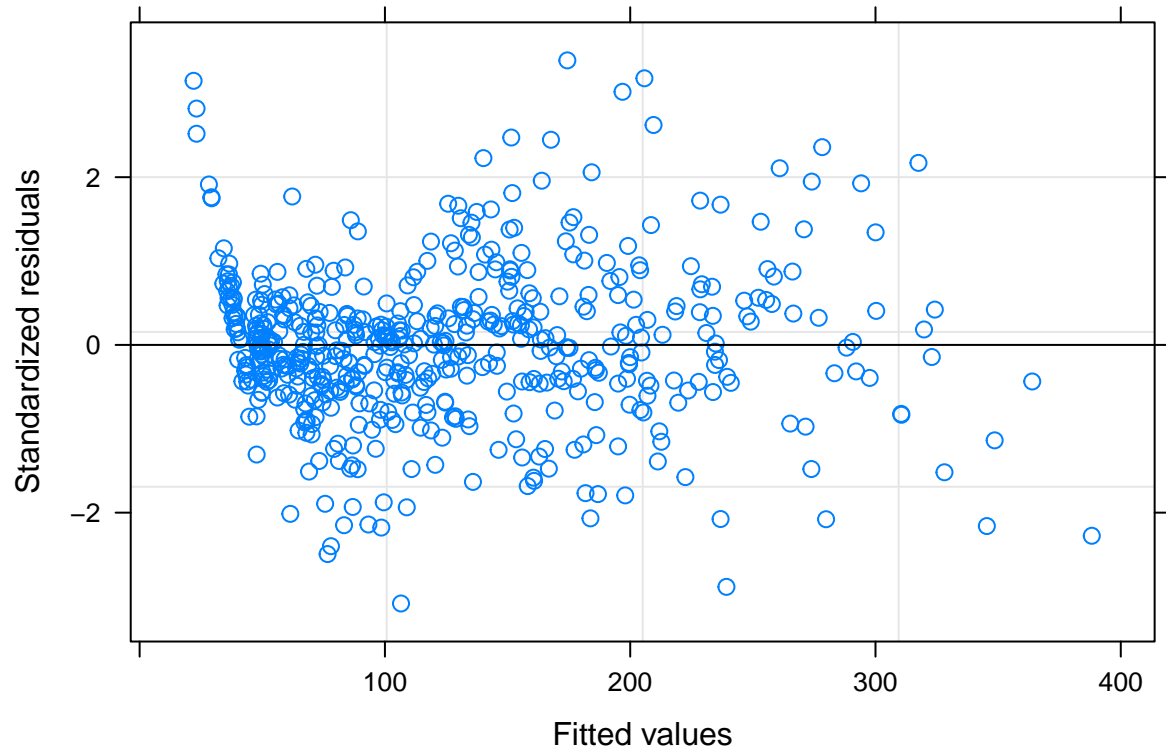
²More details can be seen in Singer & Rocha (2018), Selection of terms in random coefficient regression models; and Singer, Nobre & Rocha Análise de dados longitudinais (Portuguese)



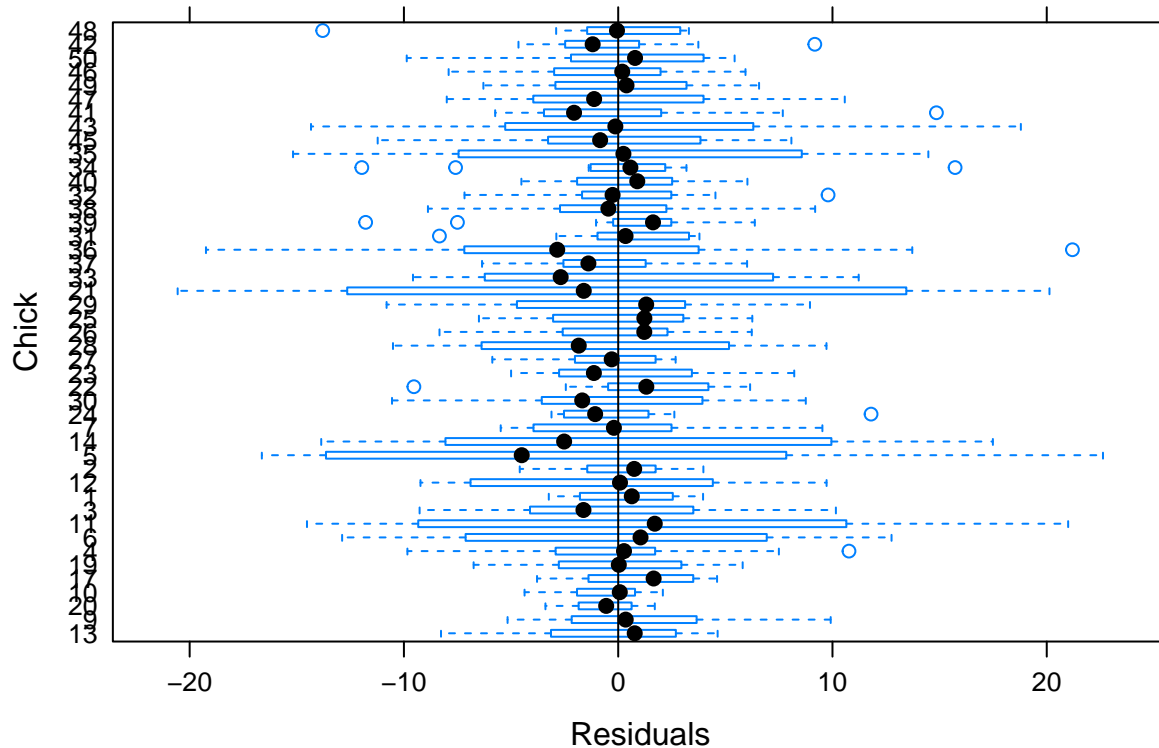




Comparing the fit of the data with the other model we can clearly see a better fit for the data in all chickens, which also means that these model fit the quadratic effect better.



For the Fitted Values x Standardized residuals plot, we can see an more random distribution of the point, that are mainly centered in the -2, 2 interval.



At last we see that all interval have the value 0, which is a good for the model.

Conclusion

Given the analysis and all explanation above we can say that the model used all the parameter, plus a quadratic function for time, in the fixed model and the quadratic function for time in the random value, with a not structured correlation structure is a good model for fit the weight of a chicken depending on a Diet that they have.

The structure and theory used in the present document can be used in a great variation of cases that can be seen as a Longitudinal data or the analyst have a perception of a presence of a random effect that alter the variance over time (second plot of the descriptive section). As said before i higly sugest a more deep dive on the theory to better understanding of the model, the link and articles suggested can be a good start.