

Mixed Models

Pedro Victor Brasil Ribeiro

2021-11-22 - Last changed in 2021-12-08

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.

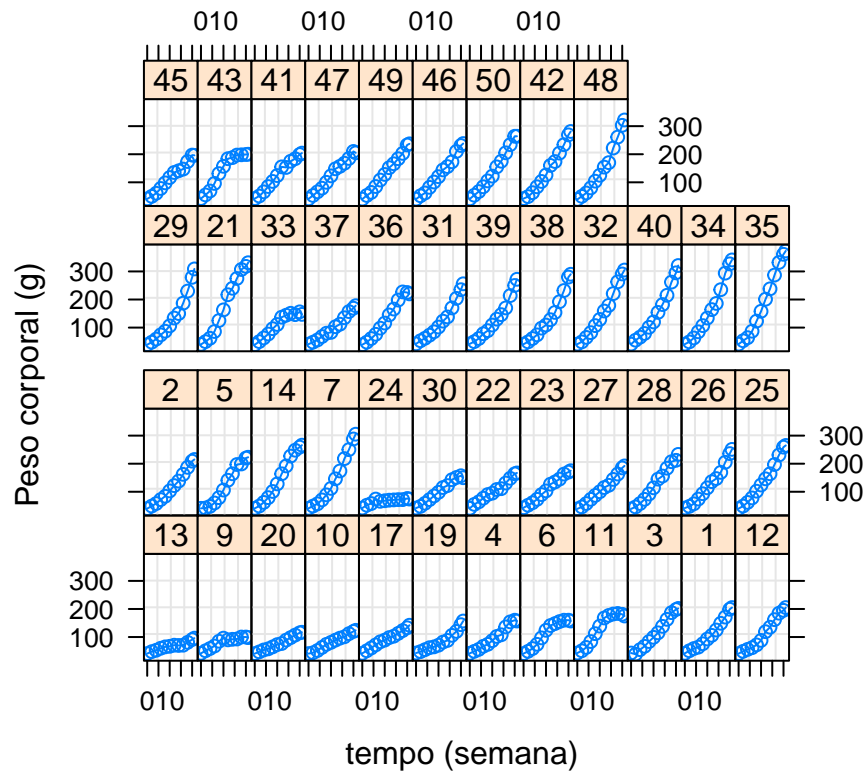
For example, the Chicken number 18 has only been measured on the time 0 and not on the subsequent time. The help (`help(data("ChickWeight"))`) of the database doesn't provide us the explanation/detail of why that occurs.

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.

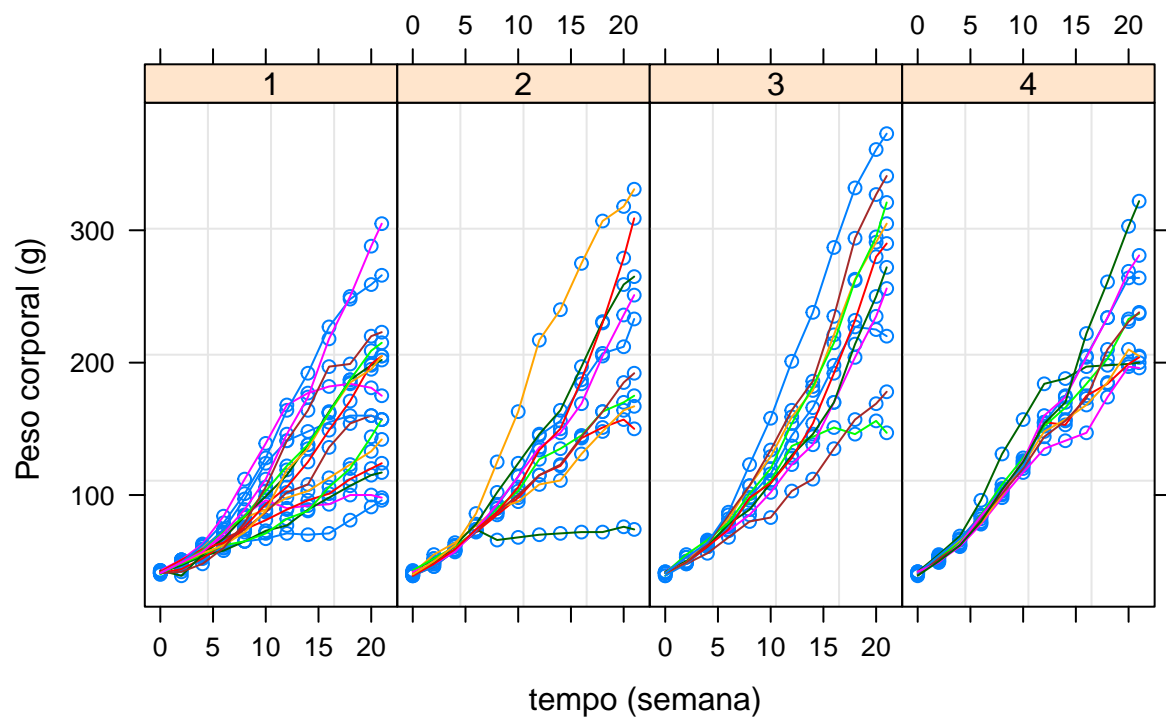
But since the presence of NA is only in 5 cases (8,15,16,18,44) it was decided to drop the data for lack of good technical features. The absence of which observation can be seen on the table below.

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

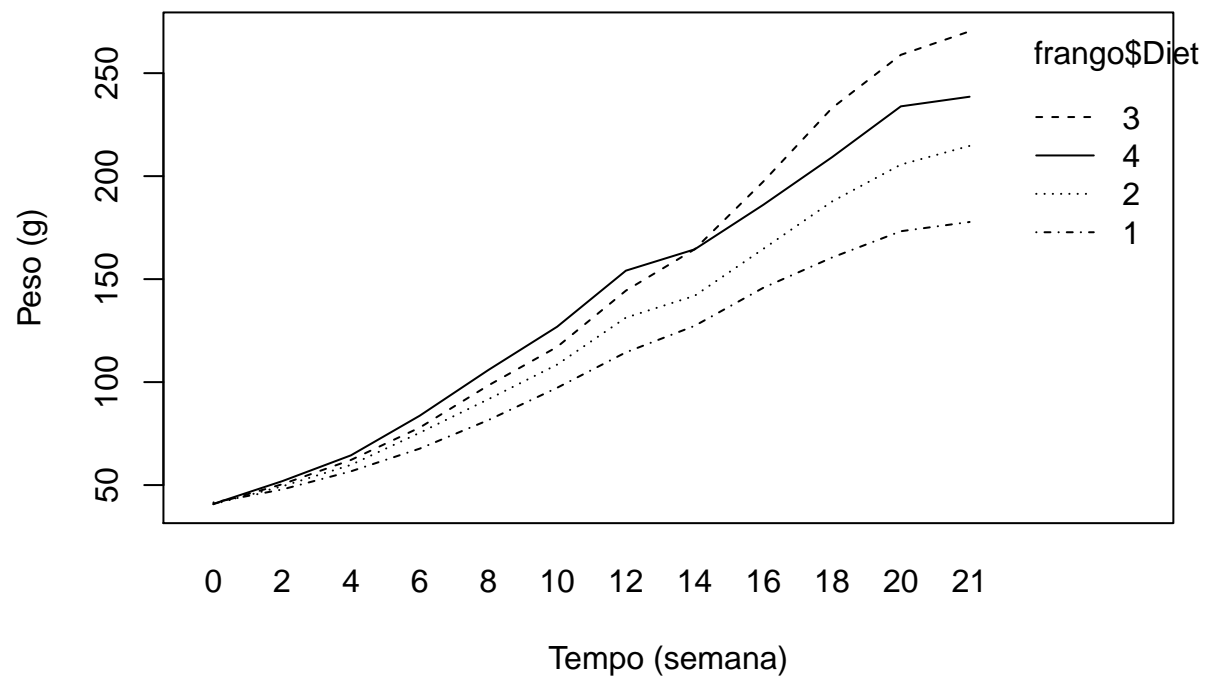
Now we'll give some descriptive analysis of the data to help us to get some help for the decision of the implementation of the random effects.



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.

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


(Intercept)	Time	I(Time ²)	Diet2	Diet3	Diet4	Time:Diet2	Time:Diet3	Time:Diet4
55.78377	-0.8512330	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
56.19710	0.0540489	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
50.85541	0.5835137	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
50.15888	1.0149735	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
49.69516	1.7572823	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
46.16140	1.8240599	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
43.81863	3.0507655	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
45.03148	3.9252436	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
42.71834	5.4063396	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
35.81301	5.3314896	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
37.35181	4.8213331	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
35.71856	5.2181491	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
35.47261	5.6723583	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
30.58829	6.8246446	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
26.15733	9.2073084	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
20.36782	9.8999564	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
65.11435	-3.3167626	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
50.33275	1.3836599	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
50.55702	1.4071933	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
48.14220	2.2530658	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
44.81652	2.6340633	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
37.67500	5.0134157	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
36.04990	5.3129617	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
32.69992	6.6408367	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
27.27072	7.2904048	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
21.02260	11.2188073	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
60.84347	-1.0640901	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
55.83855	-0.9050481	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
46.46767	2.7095964	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
45.19511	2.4262223	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
42.95852	3.1396315	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
38.36295	4.3487368	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
35.94614	5.7778479	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
34.75857	5.9396518	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
29.65294	7.4466299	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
23.65708	10.0184677	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
48.02740	1.7242040	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
49.25946	3.2451420	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
47.42498	2.4433399	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
46.36005	2.5523748	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
41.98263	3.8579661	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
41.24632	3.6938483	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
36.35424	5.3279901	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
34.34260	5.7162615	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101
27.31523	7.2927547	0.1380097	-3.151904	-13.53517	-1.261474	1.660575	4.47431	2.927101

Changing the correlation matrix structure

```
## [1] 3990.842
```

```
## # 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.
```