Mixed Models Case Study

Caspar Schmeits, Ludovico Lombardo, Floor Komen & Ilse van Beelen

April 2019

Contents

1 Introduction

The aim of this case study is to observe the weight change of chickens with Macrophage Activation Syndrome (MAS) disease, which is, in fact, one of the symptoms of chickens affected by this very disease. The underlying idea is to study and compare the weight loss in two groups of chickens. The chickens are divided in two groups, which are then exposed differently to the disease. We would like to have an idea on which method is the best to get MAS diseased chickens. The chickens are housed in pens which are part of different departments. The data is collected at five different time points.

It is important to note that the data set is unbalanced due to the fact that some observations are missing and the amount of chickens per group is not the same. To tackle this problem the models are fitted with the Restricted Maximum Likelihood (REML) approach and the Maximum Likelihood (ML) method.

The data set:

- Weight Change: outcome variable indicating the weight of a chicken in grams (numeric)
- Department: variable indicating the department number in which the pens are found (factor, 4 levels)
- Group: variable indicating in which way the disease was introduced (factor, 2 levels)
- Pen: variable indicating the pen number within the department (factor, 5 levels)
- ID: variable that identifies the chicken with a unique number (factor,162 levels)
- Time: variable indicating the day at which the measurements are made (day: 3,10,20,27,34) (numeric)

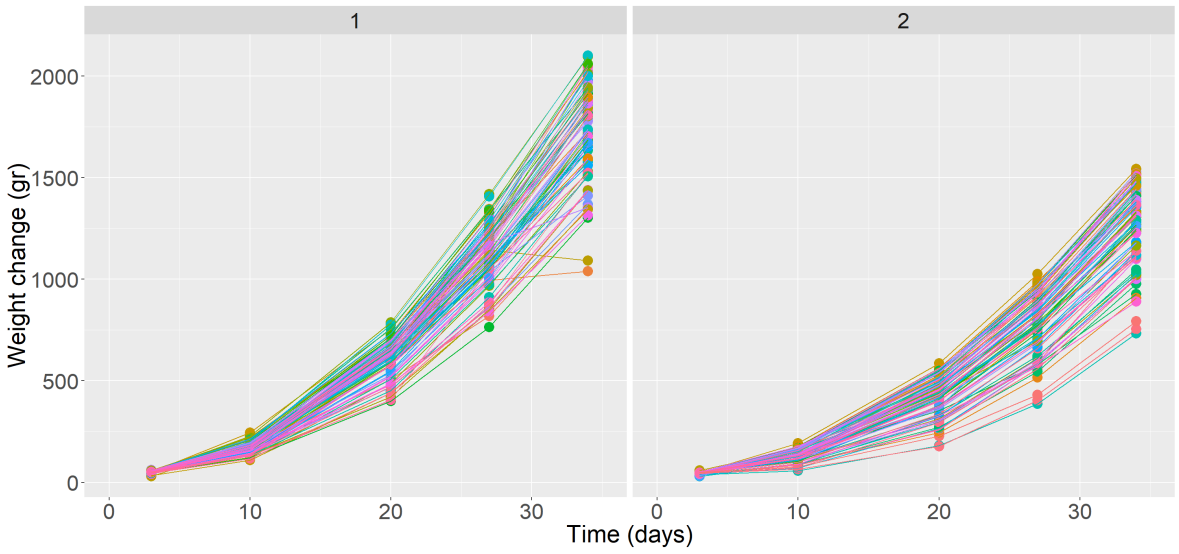

Figure 1: Weight change (gr) over time (days) per chicken in the two groups

In the first part of the study, after some descriptive statistics, we fitted a classical mixed model. We compared the fitted models using a transformed outcome variable and one without transformations. We analyzed the possible interactions and made some conclusions on which group is more likely to experience weight loss.

For the second part it is decided to use a longitudinal model, because these models capture the dependence between measurements over time better. The dependence is modeled with random intercept and possibly a random slope. Again, both a non-transformed and log transformed response are fitted.

2 Classical Mixed Modeling

The first step we took was to fit a simple classical mixed model. We introduced the following variables as fixed effects: Department, Group and Time. The variable Pen is introduced as random effect, since we assume that there is a certain dependency structure between diseased chicken in the same pen. The variable ID was also introduced as a random effect, and it represents the random chicken effect since repeated measurements on the same chicken are obviously dependent. For the variable Time we used a second order polynomial. Higher order polynomials give approximately the same result, thus we choose the second order polynomial to keep the model as simple as possible, as we can see below (figure 2).

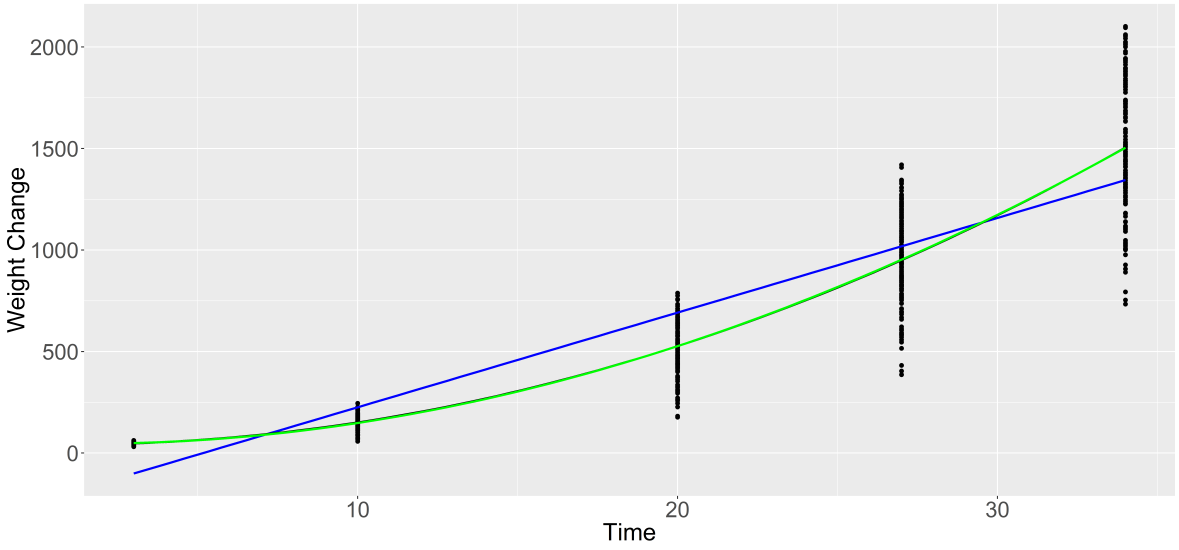

Figure 2: Relation between outcome variable and time (green line second order polynomial)

2.1 Model without transformed outcome variable

Initially, we did not apply transformations to the outcome variable and we did not consider interactions, so the fitted model is the following:

$$y_{ijklm} = \beta_0 + \beta_1 \cdot Department_m + \beta_2 \cdot Time_j + \beta_3 \cdot Group_k + \beta_4 \cdot Pen_{mi} + \varepsilon_{ijklm} \tag{1}$$

Let i = ID, with $i = 1, 2, 3, \ldots, 162$, $Time_j$, with $j = 3, 10, \ldots, 34$, $Department_m$, with $m = 21, 23, 34, 36$, $Group_k$, with $k = 1, 2$, Pen_i, with $i = 1, 2, \ldots, 5$.It is assumed that $\varepsilon_{ij} \sim N(0, \sigma^2)$ and that $Pen \sim N(0, \sigma^2_{pen})$

The results of the aforementioned model are shown below table (1). The model was fitted using REML due to the fact that we have unbalanced data. The tests carried out on the parameters of the model are all significant (considering an $\alpha = 0.05$) expect for the factor Department 34. Having a look at the coefficients we can see that we expect the biggest reduction in weight change for chickens that are in department 23 and part of group 2. Already from this very simple classical mixed model we could think that chickens in group 2 and department 23 are more likely to experience weight loss,

thus suggesting to take in great consideration this very department and group when deciding to spread MAS disease.

Table 1: Coefficients of model without transformation and interactions

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	802.13	15.77	41.28	50.87	<0.001
Department23	-135.52	16.70	154.74	-8.12	<0.001
Department34	-19.30	19.01	154.77	-1.02	0.31
Department36	-46.90	16.95	154.49	-2.77	0.01
Group2	-212.57	11.94	154.71	-17.80	0.00
poly(Time, 2)1	14833.43	136.44	645.59	108.72	<0.001
poly(Time, 2)2	3724.83	136.44	645.43	27.30	<0.001

2.2 Model with transformed outcome variable and test for interaction

From the plot of the residuals against fitted values we can see that there is a pattern in the residuals(figure3). When comparing the sample quantiles to the theoretical quantiles of a standardized normal distributions we can see how the sample quantiles start to diverge from the theoretical ones. These plots suggest that we have to transform the outcome variable in a certain way, since some of the hypothesis on which linear models are build, residuals normally distributed and constant variance, are not totally respected. We can also see the dependency from time in the residuals, each batch refers to the 5 different time points (day:3,10,20,27,34)

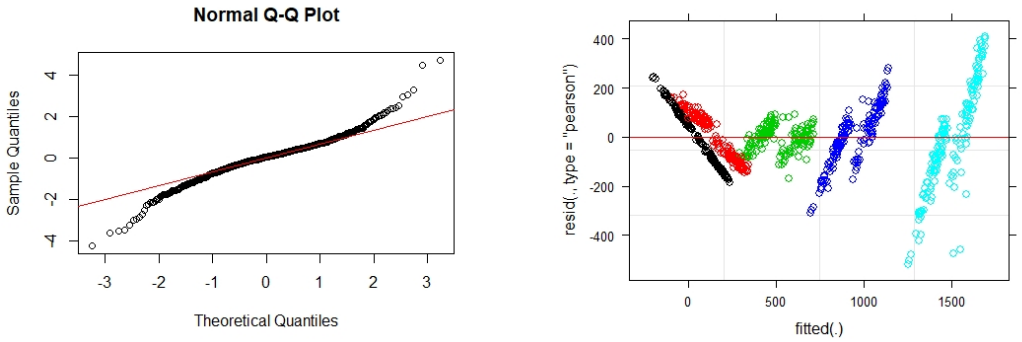

Figure 3: Quantile plot (left) residuals vs fitted values (right)

To fix the problem regarding heteroscedasticity is to transform the outcome variable. We decided to fit a model with the logarithm transformation of the outcome variable. We choose the logarithm since there is a positive skewness in the distribution of the weight change and also because it keeps the interpretation of the regression coefficients relatively simple.

Figure 4 visualized the histogram of the log transformed response variable. The transformation causes the distribution of the weight change to approximate a normal distribution.

With this in mind, we decided to fit a model with the logarithm transformation of the outcome variable. We then added the interaction between time and group because it was the only interaction that resulted significant and that might make sense from a logical point of view. The results are shown below.(table 2)

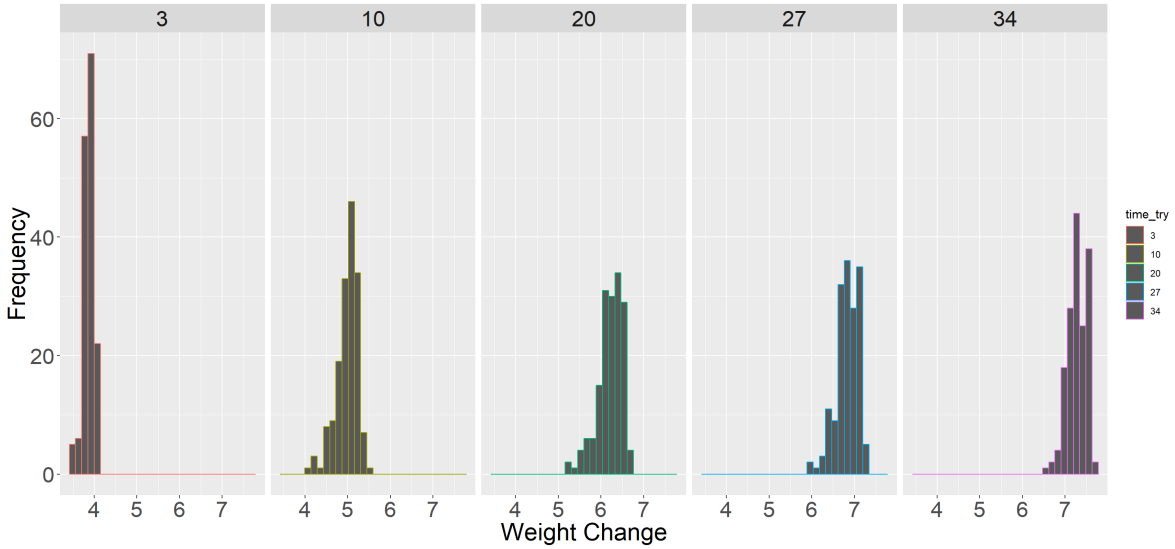

Figure 4: Histograms per time point of the log transformed Weight change

Table 2: Model with transformation and interaction output

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.09	0.03	39.63	227.30	<0.001
Department23	-0.22	0.03	154.40	-7.68	<0.001
Department34	-0.02	0.03	154.49	-0.77	0.44
Department36	-0.10	0.03	154.17	-3.35	<0.001
Group2	-0.29	0.02	154.66	-14.60	0.00
poly(Time, 2)1	36.23	0.14	643.29	250.37	<0.001
poly(Time, 2)2	-6.41	0.14	643.21	-44.30	<0.001
Group2:poly(Time, 2)1	-2.24	0.21	643.17	-10.92	<0.001
Group2:poly(Time, 2)2	1.95	0.21	643.13	9.48	<0.001

Considering the logarithm transformation we expect more accurate results. The results obtained are similar to the ones obtained by fitting the model without logarithm. Thus, higher weight loss for chickens in department 23 and in group 2. The plot of the residuals looks much better now, after applying the transformation. The same can be said when comparing with the theoretical quantiles. However, we can still notice a pattern related to the time(even though less accentuated)(Figure 5)

We used the Kenward Roger F-test (table3) to find out if the interaction between group and time is significant. We fitted a model without interaction in order to test it against the one with interaction.

Given a significant interaction,we decided to compare the mean differences between the two groups at different time levels. To do so we used a t-test with Satterthwaite's approximation to calculate the (broken) degrees of freedom of the test statistic distribution. We considered the values of Time = 10 and Time = 34.(Table 5 and 4)

The more the days the bigger the difference between the two groups. What was found in the previous models is confirmed again by the test, i.e. we expect a bigger weight decrease for chickens in group 2. (the bigger the amount of time from when the disease is injected the bigger the mean loss of weight between the two groups)

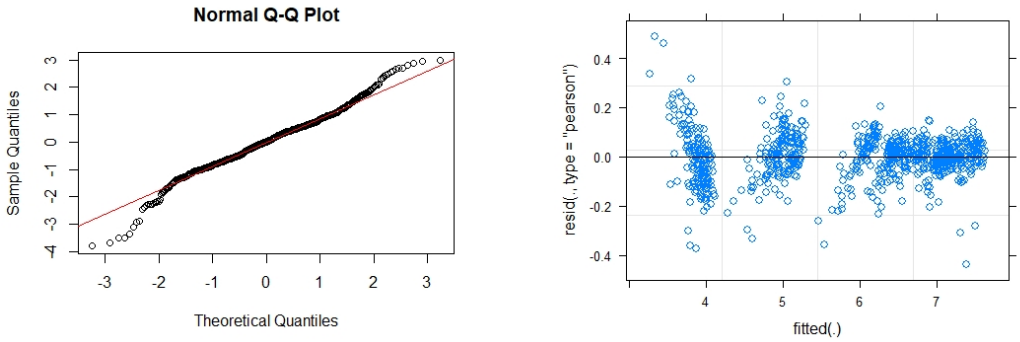

Figure 5: Quantile plot (left) residuals vs fitted values (right)

Table 3: Kenward Roger F-test

	stat	ndf	ddf	F.scaling	p-value
F-test	104.48	2.00	643.10	1	<0.001

Table 4: T-test a time point day 34

	stat	df	p-value
t-test	14.049	156.71	<2.2e-16
95% confidence interval	[408.7261,542.4538]		
Mean group 1	1737.402		
Mean group 2	1261.812		

Table 5: T-test a time point day 10

	stat	df	p-value
t-test	8.6595	154.85	5.82e-15
95% confidence interval	[30.29136,4819583]		
Mean Group 1	170.7561		
Mean Group 2	131.5125		

2.3 Classical Mixed Model limits

The following plot(figure 6) shows the weight change plotted against time for every animal. The transformation affects the variance at each time point, i.e. it gets constant(or close to being it). Given the impact of time on the model and the time dependency structure of the data we thought of a better model to analyze this data set. Longitudinal models can be very handy in this case. These very models are helpful to study changes over time and to separate cross-sectional effects from longitudinal effects. In longitudinal models the correlation structure between repeated measurements is modelled via the correlation matrix In the second part of the case study we analyzed different longitudinal models in order to get a better idea of time effect on weight change, the second plot, for example, might suggest

the use of a random intercept model.

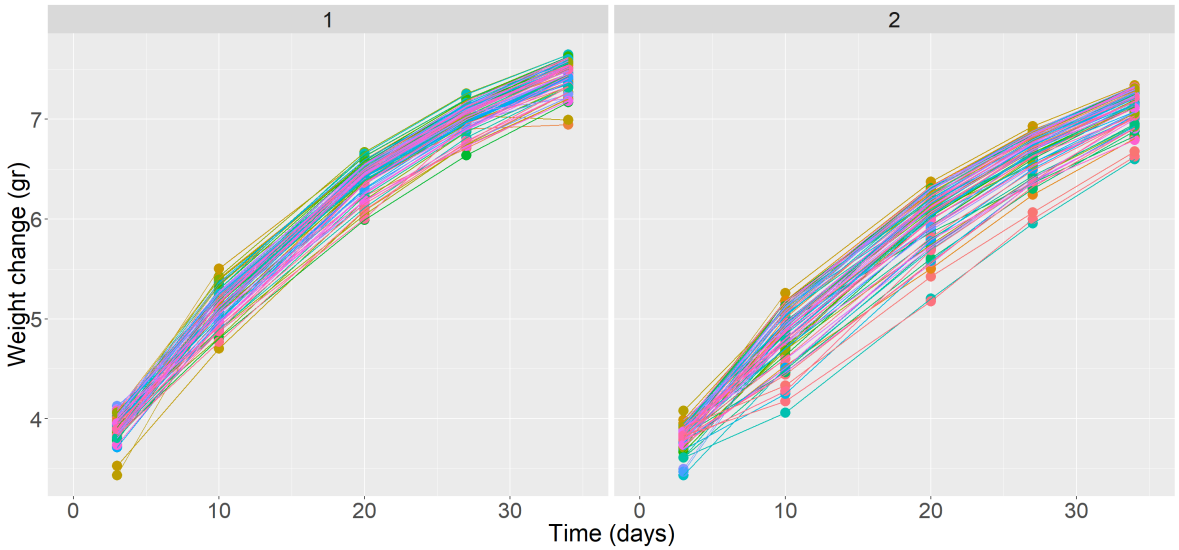

Figure 6: Log transformed weight change (gr) over time (days) per chicken in the two groups

3 Longitudinal Mixed Modeling

A longitudinal model would be a better fitting model, because the measurements are performed over time. This results in dependence between measurements in the same chicken. The dependence can be modeled with a random intercept and slope. This chapter will describe which steps are taken to find the best fitting longitudinal model:

- Adding fixed effects
- Adding random effects
- Adding correlation structures
- Visualize final result

3.1 Fixed effects

As already discussed in the classical mixed model, the response is again fitted on a log scale. As figure 2 already visualized, a random intercept describes more variance on the log scale than on the normal scale.

The ML method is used in this part of the analysis. The reason for this is two fold. First, this data set is unbalanced and secondly, because of the comparison of nested models with different $X_i\beta$ parts. At first, the minimal model is fitted. This model has Group and Time as fixed effects and a random intercept. The fixed effects are added, one at a time. This eventually least to the maximum model with the interaction between time and group and the different departments. The different nested models are compared to each other with the Likelihood Ratio Test (LRT).

3.2 Random effects

As figure 2 already visualized, the weight of each chicken follows a similar pattern over time. Furthermore, there is a wide variaty of weights at baseline. Therefore, it hypothesized that a random intercept model is sufficient for this data set. Nonetheless, multiple other random effects are fitted as well. The REML method is used for this part of the analysis, because different random effect structures are compared. The LRT is used the compare models to each other.

The model with a random intercept only has a $log(L) = -47.97$ and the model with a random intercept and slope has a $log(L) = -47.60$. However, adding a random intercept and slope resulted in a model with a singular fit. This can indicate that the model is overfitted. In other words, the random effect structure is too complex to be supported by the data. Furthermore, the between group and within group variance is practically zero and the correlation between the random intercept and slope is 1.00.

When testing if the variance of a random effect is zero, the null hypothesis is not χ^2 distributed. In this case, the tested null hypothesis is: $H_0 : \sigma^2_{int,time} = \sigma^2_{int} = 0$. And the $H_A : \sigma^2_{int,time} > 0$. This results in the following distribution of the LRT under H_0:

$$0.5 \cdot \chi^2_1 + 0.5 \cdot \chi^2_2 \tag{2}$$

Applying equation 2 results in a p-value of 1. It is concluded that a random slope does not improve the model. Next the coefficients and conditional variance-covariance matrix of the random intercept model are visualized. The formula for this final model is:

$$log(y_{ij}) = (\beta_0 + b_0) + \beta_1 \cdot Group_i + \beta_2 \cdot Time_{ij} + \beta_3 \cdot Department_i + \beta_4 \cdot Time_{ij} \cdot Group_i + \varepsilon_{ij} \quad (3)$$

Let i = ID, with $i = 1, 2, 3, \ldots, 162$ and j = Time, with $j = 3, 10, \ldots, 34$. It is assumed that $\varepsilon_{ij} \sim N_{n_i}(0, \sum_i)$. Let b_0 the random intercept for each ID and $b_0 \sim N(0, D)$. Where D is the appropriate co-variance matrix. b_0 and ε_{ij} are assumed independent.

Table 6: Coefficients of random intercept model, fitted with REML

	Value	Std.Error	DF	t-value	p-value
(Intercept)	3.95	0.03	645.00	123.12	<0.001
Time	0.11	0.00	645.00	106.94	<0.001
Group2	-0.16	0.04	157.00	-4.65	<0.001
Department_factor23	-0.22	0.03	157.00	-7.65	<0.001
Department_factor34	-0.03	0.03	157.00	-0.79	0.43
Department_factor36	-0.10	0.03	157.00	-3.39	<0.001
Time:Group2	-0.01	0.00	645.00	-4.61	<0.001

Table 6 shows the coefficients of the random intercept model. All coefficients are significant, except for department level 34, with a p-value of 0.43. All coefficients are transformed to the log scale and need to be back transformed for interpretation. Time has a coefficient of 0.11 on the log scale. This translates to $exp(0.11) = 1.11$, meaning that an increase of one unit of time (one day), increases the weight with 11% in Group 1. Furthermore Group 2 has an effect of exp(-0.16) = 0.85. This can be interpreted as a decrease of the mean weight with 15% at baseline if the chickens are in Group 2. Also the departments have an effect on the weight change of the chickens, $exp(-0.22) = 0.80$, $exp(-0.03) = 0.97$ and $exp(-0.10) = 0.90$, respectively. At last, the change of the weight in Group 2 over time is $exp(-0.16 + -0.01) = 0.84$

The amount of variance explained by the random intercept, also called the between-group variance is 0.005 on the log scale. The residual variance, also called the within-group variance is 0.058 on the log scale. Table **??** visualizes the conditional variance-covariance matrix of chicken with ID = 20. The covariance over time is equal to zero.

Figure 7 visualizes the estimates of the random intercept. The estimates are in the range from -0.2 till 0.2 and their expected value is close to zero. Each estimate has a 95 % confidence interval around it.

3.3 Correlation structures

The random intercepts model implies constant variance and constant correlation between any two measurements within subjects. However, this is not always realistic. Therefore, different correlation structures are compared to account for the remaining correlation. In other words, finding the appropriate variance-covariance matrix D. The following correlation structures are fitted:

- Unstructured correlation, but the algorithm did not converge
- Compound symmetry correlation (CS), the same correlation over all time points
- Autoregressive (AR1) correlation, pairwise correlation decrease with time

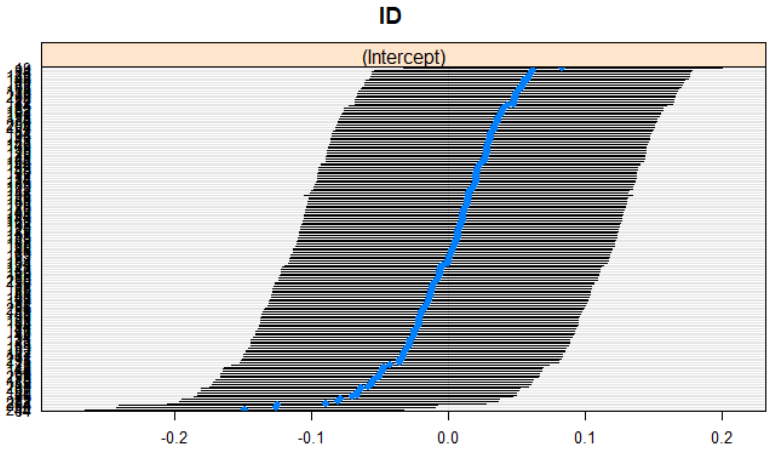

Figure 7: Estimates (blue dots) and 95 % confidence interval of the random intercepts for each chicken

- Autoregressive-moving average (ARMA) correlation with autoregressive terms p and q moving-average terms

Table 7: Comparison between models with different correlation structures. Each model has a random intercept.

	Model	df	AIC	BIC	Log(L)
Random int	1	9	113.94	152.13	-47.97
Random int + CS	2	10	115.94	162.81	-47.97
Random int + AR	3	10	33.26	80.13	-6.63
Random int + ARMA	4	14	-449.12	-383.50	238.56

Table 7 shows the comparison of the different correlation structures performed with the AIC. The AIC uses the Log likelihood and adds a penalty for complexity. A lower AIC generally implies a better fitting model. It is concluded that the random intercept model and ARMA correlation structure is the best fitting model, because its AIC is much lower than for the other models.

The ARMA correlation structure has p=*4* auto regressive terms for the fixed effects and q=*1* moving-average term for the random effects. In equation 4 is c a constant and ε_t is a i.i.d. random variable. $\phi_1, \ldots, \phi_p$ are the fixed stationary variables. The auto regressive model specifies that the output variable depends linearly on its own previous values and the error-term, ε_t. $\theta_1, \ldots, \theta_q$ is the random variables. The moving-average model specifies that the output variable depends linearly on the current θ and past values of ε_{t-1}.

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi X_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{4}$$

By applying the ARMA correlation structure, the variance covariance matrix changes as well. Table 8 shows the conditional variance covariance matrix for chicken with ID = 20. The variance per time point is very small and the covariance is alternately negative and positive.

Table 8: Variance covariance matrix of ID = 20, with ARMA correlation structure

	1	**2**	**3**	**4**	**5**
1	0.094	0.058	-0.013	-0.066	-0.063
2	0.058	0.094	0.058	-0.013	-0.066
3	-0.013	0.058	0.094	0.058	-0.013
4	-0.066	-0.013	0.058	0.094	0.058
5	-0.063	-0.066	-0.013	0.058	0.094

3.4 Final model

The final model has a random intercept and the ARMA correlation structure. During the analysis, each model is assessed for homogeneous variance, normality and influential data points. However, only the model diagnostics of the final model will be discussed in this report.

Applying the log transformation causes the residuals to be roughly homogeneously distributed. Before the transformation several clusters of residuals were observed. Figure 8 visualizes the normalized residuals per time point. The residuals are colored by department. The reason to choose the normalized residuals is two fold. First, the normalized residuals deal with heteroscedasticity and secondly, measurement are no longer correlated. It is clear that the residuals per time point are not all spread around zero. Especially, the residuals at time point 10 in group 1 are spread around +2. A few possible influential points are also visible, with values close to -3. The residuals are also inspected with a QQ-plot and cooks distance plot. However, no strong deviations were observed and therefore, these plot are not shown in the report.

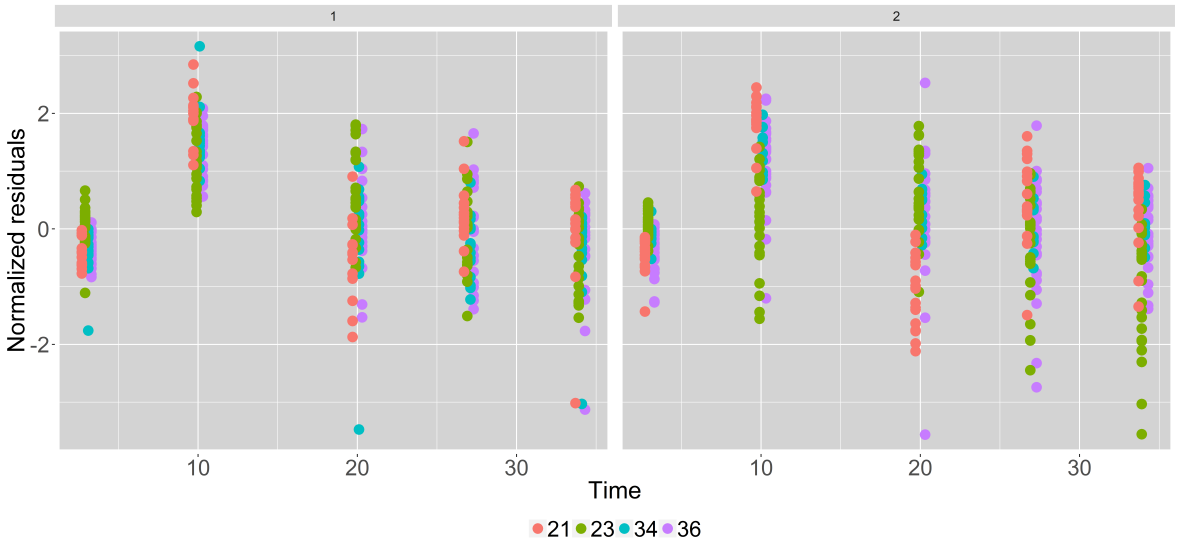

Figure 8: Residuals of final model plotted per time point and group. The colors represent the departments

At last, the estimated marginal mean (EMM) per group is calculated over all time points. This final result is visualized in table 9. The two estimates are on the log scale. It is possible to transform the contrast back between the two estimated. On average the weight difference is $exp(5.779) - exp(5.549) = 66.45$ gram. The estimate of group 1 has a 95% CI from 5.754 till 5.804 and group 2 from 5.523 till 5.574. Furthermore, the contrast between the two groups is significant, with a t-statistic of 12.653.

Table 9: Estimated Marginal Mean per group, contrast between the groups and corresponding standard error and p-value

Estimated Marginal Mean			SE	df	t-statistic	P-value
Group 1	Group 2	Contrast				
5.779	5.549	0.115	0.009	159	12.653	<0.001

4 Conclusion

After inspecting the results of the classical mixed model we can definitely conclude that there is a difference in the chickens' weight change depending on how the disease is introduced, group two, in average, experienced a bigger weight loss than group 1. Furthermore, given the significant interaction between group and time, we can conclude that the differences between groups is much more stronger with time passing. It might be ideal, when studying MAS disease, to look at chickens that have been injected with the disease after a certain amount of days since the symptoms of it, i.e. weight loss, get more accentuated after a certain amount of time.

Both the classical as the longitudinal approach result in a model with the same significant fixed effects. Chickens in department 23 have a bigger weight loss than chickens in other departments. This significant effect of the department is probably not something the researchers intended. A possible explanation could be that the chickens are not infected with the disease in a consistent way. This might be useful to consider when carrying out a subsequently analysis on MAS.

After visual inspection of figure 6, it is hypothesized that random intercept model would suffice for this data set. Multiple random effect models are fitted and the hypothesis turned out to be correct. The $Log(L)$ of the random intercept and slope model did not improve according to the LRT. Furthermore, fitting the random slope resulted in a model with a singular fit. This is due to the correlation being exactly 1.0 between the random intercept and slope. Therefore, the results of this model can be misleading. It is concluded the use the random intercept only model.

In practise, a random intercept can not describe all the variation found in longitudinal data. Therefore, different correlation structures are also added to the model. The ARMA correlation structures fits much better than the other structured, with an AIC of -449.12. The ARMA structures allows the covariance to be negative as well, as displayed in table 8. There is not a clear decrease of correlation as time points are further away from each other, due to the alternating pattern of negative and positive co-variance. However, decreasing correlation over time would have been expected for longitudional data.

The longitudinal mixed model also concludes that there is an effect of the groups on weight change. The EMM of group 1 is 5.779 and group 2 is 5.549. The results in a constrast of 0.115 on the log scale. After back transformation, is it concluded that the chickens in group 2 are on average 60 grams lighter than the ones in group 1. The t-statistic for this contrast is 12.653, which results in a p-value <0.001.

However, the results from the log transformed longitudinal model are in contradiction with the raw data. Figure 1 displays the raw data and here it would be concluded that the average weight in group 1 and 2 is roughly 500 and 350 gram, respectively. This would mean a contrast of roughly 150 gram between the two groups. It is not clear why the model result is much smaller.