SENSITIVITY ANALYSIS OF MISSING DATA: MILK PROTEIN EXAMPLE

Ariel Alonso Abad y Olga María Rodríguez, Centro Nacional Coordinador de Ensayos Clínicos, Cuba
Fabian Tibaldy, Limburgs Universitair Centrum, Bélgica
José Cortinas Abrahantes, Centro de Sanidad Vegetal, Cuba

ABSTRACT
Missing data are one of the most serious problems that should be solved in the statistical analysis of biological experiments. In this report different techniques for handling missing information are shown using an example. It is also shown how sensitive the conclusions of the study can be with respect to the way the missing data are analised. The missing data process is studied in the framework of longitudinal data.

Key words: Missing Data Process, longitudinal data, mixed models.

RESUMEN
Los datos missing son unos de los problemas más serios que es preciso enfrentar en el análisis estadístico de experimentos biológicos. En el trabajo se expone, a través de un ejemplo, diferentes técnicas de imputación y lo sensible que pueden ser las conclusiones de un estudio con respecto a la técnica empleada y al manejo que se haga de las observaciones missing. El análisis del mecanismo de generación de las observaciones missing se realiza en el contexto de un problema longitudinal.

MSC: 62P10

1. INTRODUCTION TO THE MILK PROTEIN DATA

The data considered in this paper were obtained in a longitudinal study, designed to determine the efficacy of three different diets on the protein contents of milk samples. They were taken, weekly, during 19 weeks from 79 australian cows. The cows entered the experiment after calving and were randomly alocated to one of three diets: Barley, Mixed Barley-Lupins and Lupins alone, with 25, 27 and 27 animals in the three groups, respectively. Some of the time series were shorter than others; with a proportion as high as 48 % of dropouts from week 15 onwards.

The primary objective of the Milk Protein Experiment was to describe the effects of diets on the mean response profiles of milk protein content over time.

The aim of the study was three-fold:

1. Do the three groups differ in protein content?
2. How does the evolution of the protein content in the diet groups depend on time?
3. What was the effect of the missing data on the results?

2. NUMERICAL EXPLORATION OF THE MILK PROTEIN DATA

In order to obtain a first, general understanding of the dataset at hand, descriptive statistics were used to summarize the information contained in the 1501 observations with 164 missing. The descriptive information on the relevant features of study sample is outlined in Table 1.

In the Table 1 above we can see that the number of observations starts to decrease from the 15^{th} week on, whereas in the previous weeks it was quite stable (25, 26 and 27) in each group.

The number of experimental units in each week is variable; ranging from 41 up to a maximum of 79 with a median of 78 cows per week. These findings indicate that we are dealing with a missing data problem. Additionally it can be noticed that this is a balanced design where the observations are equally spaced over time, since the measurements were taken at fixed time points (weekly).

Table 1. Descriptive Statistics.

	Barley			Mixed			Lupins		
Week	N	Mean	Std	N	Mean	Std	N	Mean	Std
1	25	3.887	0.378	27	3.861	0.378	27	3.758	0.446
2	24	3.643	0.238	27	3.540	0.278	27	3.428	0.298
3	25	3.498	0.199	27	3.346	0.250	27	3.373	0.309
4	25	3.376	0.231	27	3.278	0.233	27	3.294	0.292
5	25	3.484	0.353	27	3.338	0.249	26	3.238	0.384
6	25	3.386	0.235	27	3.393	0.303	27	3.280	0.332
7	25	3.469	0.255	27	3.333	0.213	25	3.187	0.245
8	25	3.503	0.305	26	3.398	0.234	26	3.310	0.364
9	23	3.512	0.265	27	3.435	0.251	27	3.347	0.355
10	25	3.519	0.235	27	3.437	0.283	26	3.269	0.305
11	24	3.455	0.338	27	3.355	0.245	27	3.233	0.333
12	25	3.429	0.312	27	3.374	0.272	27	3.214	0.270
13	25	3.512	0.323	26	3.411	0.258	27	3.334	0.281
14	25	3.507	0.382	27	3.372	0.310	27	3.254	0.263
15	19	3.542	0.355	20	3.446	0.319	20	3.264	0.285
16	17	3.602	0.332	17	3.571	0.374	16	3.265	0.311
17	15	3.682	0.280	16	3.511	0.330	15	3.252	0.245
18	15	3.641	0.350	16	3.451	0.255	15	3.302	0.286
19	13	3.640	0.350	14	3.396	0.259	14	3.206	0.325

3. GRAPHICAL EXPLORATION OF THE MILK PROTEIN TRIAL DATA

3.1 Response variable versus time

A basic graph was obtained to address the relationship between the response variable, which is the protein content, and the explanatory variable time of measurement (see Figure 1). The graph was constructed using three meantraces for each treatment. As a result from the latter it is now possible to distinguish some kind of tracking effect, meaning that mean-traces situated at a certain level have the tendency to remain at that level throughout the study.

Figure 1 suggests different effects of the 3 diets, where Barley has the highest and Mixed lies in between the values of the Barley and the Lupins diets. It is also remarkable that there is a distinctive behavior of the traces. In the three first weeks the traces have a linear decreasing trend, whereas from the fourth week up to the end of the study, a slight rise is observed.

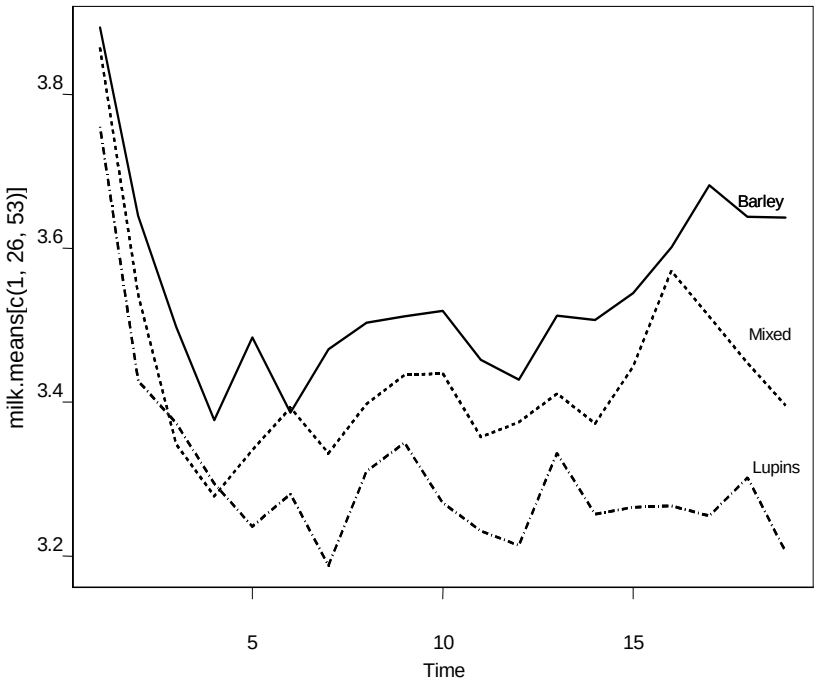


Figure 1. Mean traces for the Milk Data.

In Figure 2, the variance function is displayed together with a fitted Loess smooth curve. The graph highlighted in Figure 3 suggested that a constant pattern for the variance function is present; and this fact supports the idea of no random slope.

3.2 Correlation matrix

The Analysis of the correlations shows that the correlations appear to be smaller when the distance between the time units becomes larger. However, caution should be taken when interpreting the results since the number of observations on which they are based is not very high.

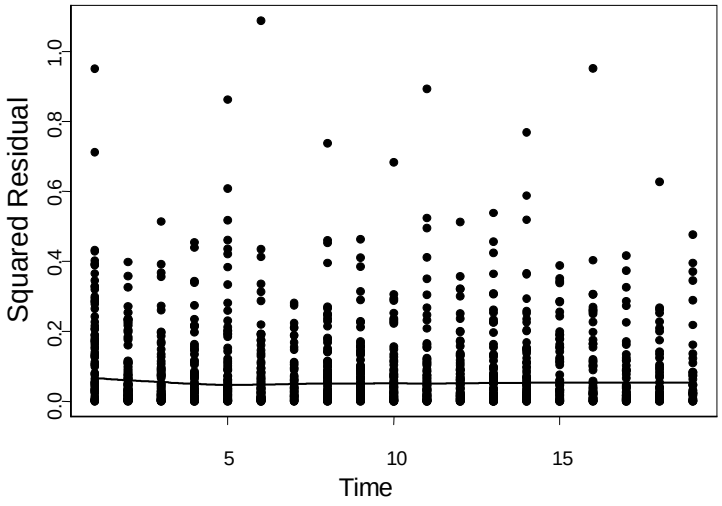


Figure 2. Variance function with Loess smooth.

Table 2. Correlation matrix for the Milk Data.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19
R1	1.00																		
R2	0.40	1.00																	
R3	0.42	0.63	1.00																
R4	0.20	0.36	0.58	1.00															
R5	0.19	0.45	0.53	0.58	1.00														
R6	0.35	0.21	0.42	0.45	0.54	1.00													
R7	0.33	0.39	0.28	0.41	0.43	0.53	1.00												
R8	0.01	0.15	0.22	0.33	0.33	0.47	0.49	1.00											
R9	0.23	0.22	0.33	0.32	0.37	0.34	0.45	0.45	1.00										
R10	0.20	0.21	-0.01	0.05	0.09	0.23	0.48	0.58	0.46	1.00									
R11	0.05	0.29	0.12	0.26	0.27	-0.03	0.37	0.51	0.45	0.57	1.00								
R12	0.16	0.21	0.23	0.25	0.32	0.15	0.42	0.50	0.53	0.69	0.58	1.00							
R13	0.23	0.32	0.29	0.40	0.30	0.50	0.59	0.69	0.57	0.69	0.54	0.65	1.00						
R14	0.24	0.46	0.33	0.44	0.41	0.35	0.52	0.67	0.55	0.74	0.60	0.72	0.84	1.00					
R15	-0.02	0.38	0.31	0.47	0.34	-0.04	0.29	0.40	0.32	0.31	0.56	0.47	0.42	0.52	1.00				
R16	0.07	0.48	0.30	0.39	0.34	0.03	0.08	-0.01	0.15	0.07	0.17	0.25	0.16	0.28	0.42	1.00			
R17	0.03	0.50	0.33	0.44	0.24	0.17	0.13	0.22	0.03	0.17	0.26	0.14	0.28	0.39	0.34	0.70	1.00		
R18	0.00	0.22	0.32	0.61	0.39	0.14	0.22	0.21	0.09	0.06	0.23	0.16	0.28	0.34	0.47	0.61	.72	1.00	
R19	0.02	0.31	0.37	0.42	0.44	0.17	0.11	0.10	0.08	0.03	0.03	0.26	0.20	0.24	0.38	0.72	0.67	0.81	1.00

The empirical variogram was constructed and shown in a plot. It can be observed in Figure 3 that the pattern of the correlation has a decreasing trend. This finding is in concordance with the previous results that were obtained from the analysis of the correlation matrix.

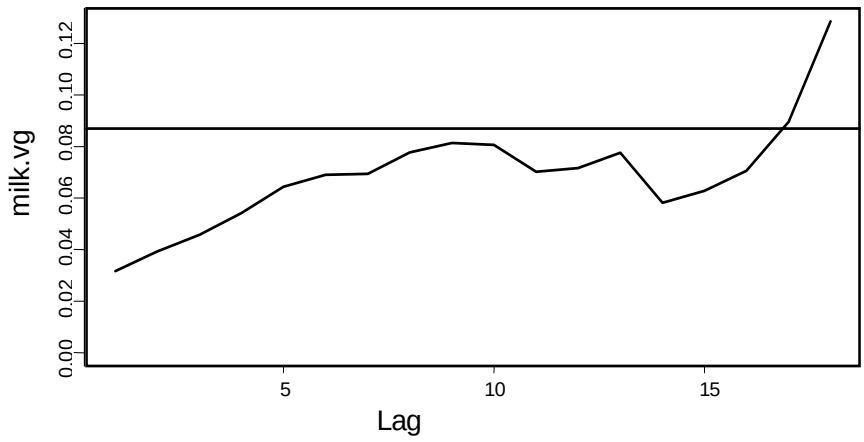


Figure 3. Variogram of the Milk Data.

3.3 Conclusions obtained from the exploratory analysis

In summary, from the previous exploratory data analysis the following conclusions can be drawn:

1. The mean response profiles are approximately parallel. In the three first weeks the traces have a linear decreasing trend, followed by an approximately constant mean response over most of the time and a possible increasing trend to the end of the study.

2. The 3 diets appear to have different effects. The Barley diet seems to give the highest values and the values obtained with the Mixed diet lie between the values of the Barley and the Lupins diets.

3. The variance seems to follow a constant pattern. The intercept seems to be random, but from the estimated variance function it can be suspected that the slope is not random.

4. The degree of correlation seems to decreases as the observations are moved further from one another in time.

All the previous points that have been highlighted lead us to consider a linear mixed model with random intercepts.

4. FITTING AND CHECKING OF A PROPOSED STATISTICAL MODEL FOR THE MILK PROTEIN DATA

In view of the results obtained from the previous exploratory data analysis, the following model for the mean response was fitted:

$$\mu_g(t) = \begin{cases} \beta_{0g} + \beta_1 t & t \le 3 \\ \beta_{0g} + 3\beta_1 + \beta_2(t-3) + \beta_3(t-3)^2 & t > 3 \end{cases}$$

where $g = 1,2,3$ denotes the treatment group, and *t* the time measured in weeks.

This model can be expressed as follows

$$Y_{ij} = \mu_{ij} + Z_{ij} + U_i + W_i(t_{ij})$$

where Y_{ij} represents the j^{th} responses from the ith subject

μ_{ij} is the mean response

Z_{ij} is the measurement error, which is assumed to follow a normal distribution with mean 0 and variance τ^2

U_i is the random effect, which is assumed to be normally distributed with mean 0 and variance ν^2

$W_i(t_{ij})$ is the realization of a stationary gaussian process, which depends on the time at which Y_{ij} is measured and is normally distributed with mean 0 and variance σ^2

The variance of Y_{ij} has the following form

$$Var(Y_{ij}) = \sigma^2 + \tau^2 + \nu^2$$

and

$$Cov(Y_{ij,} Y_{ik}) = \nu^2 + \sigma^2 \rho(| t_{ij} - t_{ik} |)$$

In matrix notation, the model takes the following form for each subject

$$Y_i = \mu_i + Z_i + 1 \cdot U_i + W_i$$

It will be assumed that a multivariate normal is the distribution of the response variable, $Y_i \sim MVN(\mu_i, V_i)$, with $V_i = \sigma^2 H_i + \tau^2 I + \nu^2 J$ where $H_i = \left[h_{jk} = \rho(|t_j - t_j|)\right]$ and J is a matrix of ones.

Under the model described above it can be shown that

$$\gamma(u) = \tau^2 + \sigma^2(1 - \rho(u))$$

5. MISSING DATA PROCESS AND MODEL FITTING

The present study deals with some missing observations. Therefore, the distribution of the full data has to be considered rather than confining attention to the observed data distribution:

$$f(Y_i, R_i \mid X_i, Z_i, W_i, \theta, \beta)$$

where X_i, Z_i, W_i are matrices grouping all the available covariate information

θ parameterizes the measurement distribution

β parameterizes the missingness process.

Most of the models are based in the following factorization (Rubins, 1976):

$$f(Y_i, R_i \mid X_i, Z_i, W_i, \theta, \beta) = f(Y_i \mid X_i, Z_i, \theta) \cdot f(R_i \mid Y_i, W_i, \beta) \quad (1)$$

where the first factor is the marginal density of the measurement process and the second factor is the density of the missingness process, given the outcomes.

The missingness process can be classified in three different ways as follows:

1. *Missing Completely at Random* (MCAR) where missingness is independent of the measurements. This means that the second factor equals $f(R_i \mid W_i, \beta)$.

2. *Missing at Random* (MAR) where missingness is independent of the unobserved measurements, possibly depending on the observing measurements i.e, $f(R_i \mid Y_i^O, W_i, \beta)$.

Informative where missingness_depends on the missing values. In order to illustrate the missingness imputation process the following three different imputation methods will be considered; the *Complete Cases Analysis*, the *Last Observation Carried Forward Analysis* and finally the *Unconditional Mean Analysis*. These methods are based on filling in the missing values by "matching" subjects where an appropriate matching criterion can be used.

A complete case analysis includes only those cases for analysis for which all the measurements were recorded. It is important to remark that with this kind of analysis two severes drawbacks are involved. First there can be a substantial loss of information and in addition severe biases can resolve when the missingness mechanism is MAR but not MCAR. Indeed, should an estimator be consisted in the complete data problem, then the derived complete case analysis is consistent only if the missingness process is MCAR.

The model which was proposed previously was fitted assuming MCAR. The obtained results can be found in the following table (Table 3).

It can be noticed from Table 3 that the parameters associated with $\hat{a}_2$ and $\hat{a}_3$ are very small. Appropriate contrasts were used to test formally the hypothesis whether H_0: $\hat{a}_2 = \hat{a}_3 = 0$, and whether the three diets affect the mean response profiles; that is H_0: $\hat{a}_{01} = \hat{a}_{02} = \hat{a}_{03}$.

Table 3. Results of the model fitting assuming MCAR.

Parameters	Est.	SE	Prob.
β_{01}	4.2085	0.0718	0.0001
β_{02}	4.0275	0.0693	0.0001
β_{03}	3.8469	0.0693	0.0001
β_1	-0.2600	0.0219	0.0001
β_2	0.0265	0.0102	0.0101
β_3	-0.0007	0.0006	0.2083

The obtained results are outlined in Table 4.

Table 4. Contrast statement results for the Complete Case.

Source	NDF	DDF	F	Prob.
$\beta_2 = \beta_3 = 0$	2	663	9.65	0.0001
$\beta_3 = 0$	2	663	1.59	0.2083
Treatment Effect	2	663	15.68	0.0001

It can be seen from Table 4 that both the first and the second hypothesis can be rejected. This finding indicates the existence of a treatment effect and a linear trend over time after the 3rd week. Our model for the data is summarized by the parameter estimates in Table 3 but with the simplification that $\hat{a}_3 = 0$. Figure 4 below, compares the empirical and the fitted mean response profiles.

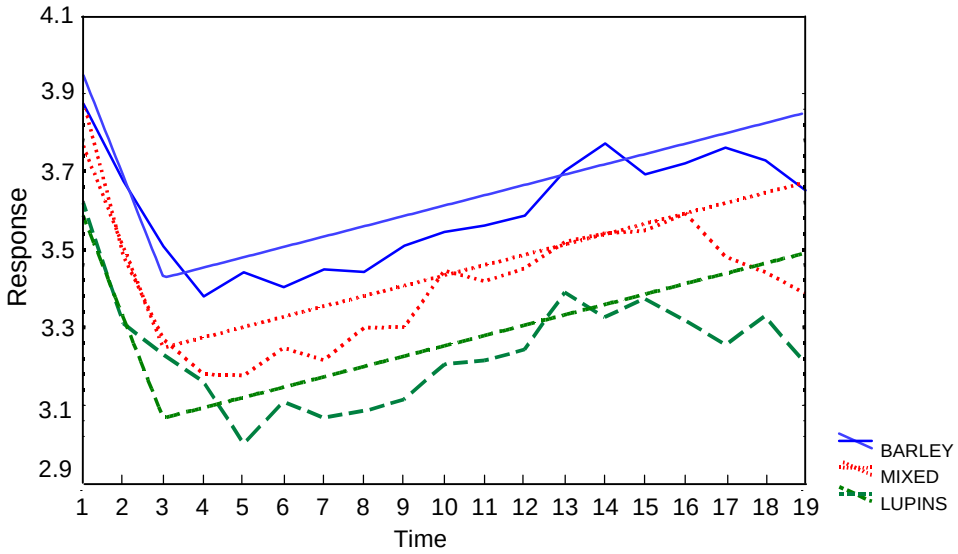


Figure 4. Mean traces for the Milk data - Complete Case.

It is important to point out that Figure 4 shows a stronger increasing trend compared with Figure 1. This gives evidence of the fact that the drop out cows had the lowest values of milk protein content. Furthermore, Figure 4 suggests that the model does not fit well. Some lack of fit can be observed, indicating that another model should be considerd in order to get a better fit for the mean structure. When the lack of fit of the model was checked using the saturated one, results showed no evidence of lack of fit.

In the Last Observation Carried Forward Analysis the missing value is substituted for the last observed value. This method assumes that a subjects' measurement states at the same level from the moment of drop out onwards (or during the period they are unobserved in then case of intermittent missingness).

This analysis was performed on the data set at hand and the results of the fitting procedure are shown in Table 5.

Table 5.Results of the model fitting using LOCF analysis.

Parameters	Est.	SE	Prob.
β_{01}	4.1453	0.0547	0.0001
β_{02}	4.0465	0.0536	0.0001
β_{03}	3.9456	0.0536	0.0001
β_1	-0.2294	0.0149	0.0001
β_2	0.0132	0.0076	0.0857
β_3	-0.0012	0.0004	0.0053

From Table 5 it can be noticed that the parameters associated with the treatment effect are quite similar to the ones obtained in the complete case imputation. However, it should be pointed out that the standard errors for the estimated parameters in the LOCF analysis are smaller than the previous ones. It is also remarkable that $\hat{a}_2$ and $\hat{a}_3$ are very small.

To test formally the hypothesis whether H_0: $\hat{a}_2 = \hat{a}_3 = 0$ and whether the three diets affect the mean response profiles that is H_0: $\hat{a}_{01} = \hat{a}_{02} = \hat{a}_{03}$, appropriate contrasts were used. The obtained results are presented in Table 6.

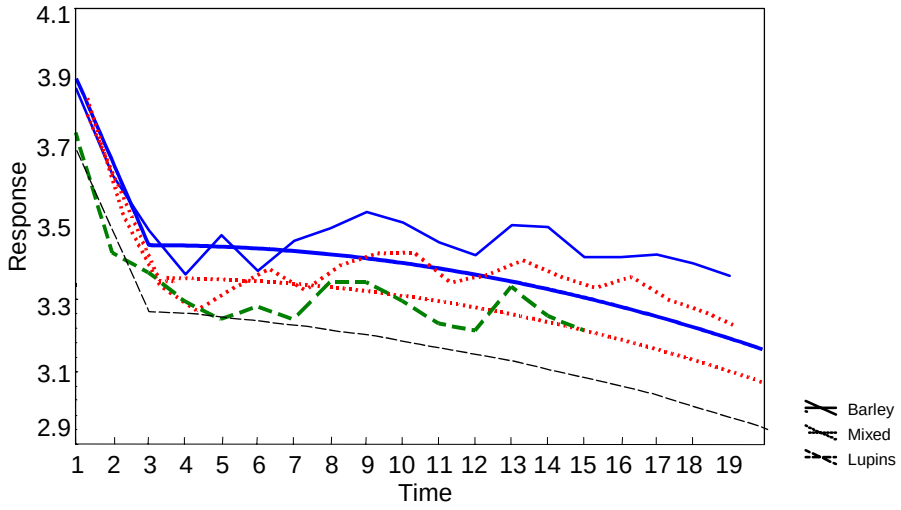
Table 6. Contrast statement results for the LOCF Case.

Source	NDF	DDF	F	Prob.
$\beta_2 = \beta_3 = 0$	2	1419	13.54	0.0012
$\beta_2 = 0$	1	1419	2.96	0.0857
$\beta_3 = 0$	1	1419	7.78	0.0053
Treatment Effect	2	1419	13.36	0.0013

It can be seen from Table 6 that three hypothesis can be rejected whereas the second hypothesis ($\hat{a}_2 = 0$) can not be rejected. This suggests that there seems to be a treatment effect and a quadratic trend over time after the 3rd week.In contrast to the complete case imputation; now the presence of a quadratic decreasing tendency can be noticed. This finding confirms the evidence that was obtained with, namely that the dropout cows have the lowest responses. That is why the application of this method reduces the mean profiles towards the end of the study.

Our model for the data is summarized by the parameter estimates in Table 5, but with the simplification that $\hat{a}_2 = 0$. Figure 6 below, compares the empirical and the fitted mean response profiles.

Figure 5. Mean traces for the Milk data - LOCF Case.



It is clear that the fitted mean profiles underestimate the empirical ones. The decreasing tendency here, confirms the previous findings with respect to the quadratic pattern of the response. It suggests that another model should be consider in this case.

It is remarkable that the results that were obtained here, are quite different from the results obtained in the previous imputation process.

The idea behind unconditional mean imputation is to replace a missing value with the average of the observed values on the same variable over the other subjects. The following table (Table 7) shows the results of the fitting process after the unconditional mean imputation method.

Table 7. Results of the model fitting using unconditional mean imputation.

Parameters	Est.	SE	Prob.
β_{01}	4.1633	0.0491	0.0001
β_{02}	4.0475	0.0483	0.0001
β_{03}	3.9160	0.0483	0.0001
β_1	-0.2281	0.0155	0.0001
β_2	0.0030	0.0073	0.6764
β_3	0.0001	0.0004	0.7522

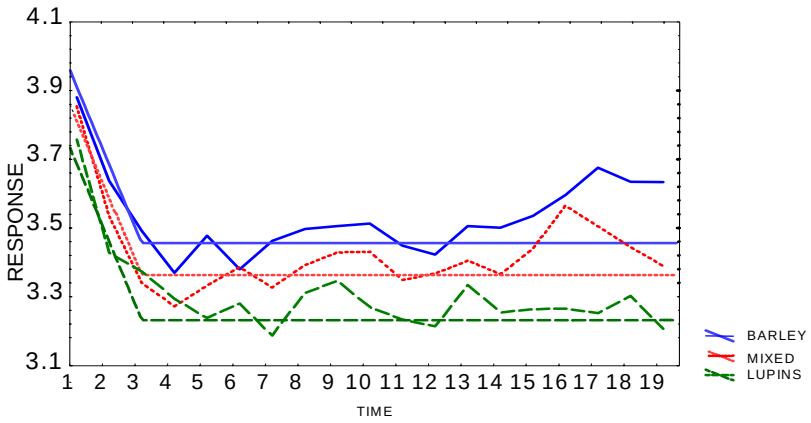
The parameters associated with the treatment effect are quite similar to the ones obtained in the LOCF imputation method. It must be pointed out, that the standard errors for the estimated parameters in this case were the smallest of the three imputation analysis. Furthermore it is remarkable that $\hat{a}_2$ and $\hat{a}_3$ are also very small in this case. To test formally the hypothesis whether H_0: $\hat{a}_2 = \hat{a}_3 = 0$ and whether the three diets affect the mean response profiles, that is H_0: $\hat{a}_{01} = \hat{a}_{02} = \hat{a}_{03}$, appropriate contrasts were used. The results are summarized in Table 8.

It can be seen from Table 8 that there is evidence of treatment effect, however there does not seem to be evidence of a linear or quadratic trend in the response. This result indicates the presence of a constant pattern in the response after the third week. Just as in the previous cases the goodness of fit of this model was checked, and there was no evidence of lack of fit comparing this model with the saturated one. In order to investigate the validity of the model, the empirical mean curves and the fitted mean profiles were plotted together in Figure 6.

Table 8. Contrast statement results for the LOCF Case.

Source	NDF	DDF	F	Prob.
$\beta_2 = \beta_3 = 0$	2	1419	2.32	0.0988
Treatment Effect	2	1419	16.80	0.0001

Figure 6. Mean traces for the unconditional mean imputation method.



Some of the previous methods have been subject to heavy criticism in the scientific literature. A likelihood-based ignorable analysis is said to be preferable, since it uses all available information whithout the need to delete or impute measurements or entire subjects. It is theoretical justified whenever the missing data mechanism is MAR, which is a more relaxed assumption than MCAR.

The likelihood-based ignorable analysis was applied to this data set and the findings are discussed in the following section.

Under a MAR process, the likelihood factorizes in two components of the same functional form as the general factorization of the complete data (1). If further è and ø satisfy the separability condition, then the missing data process should be ignorable in the likelihood inference sense. This implies, that a module with likelihood estimation facilities such as PROC MIXED manipulates the correct likelihood and leads to valid likelihood ratios. The proposed model was fitted assuming the previous statements. The obtained results are outlined in Table 9.

Table 9. Results of the model fitting using likelihood-based ignorable analysis.

Parameters	Est.	SE	Prob.
β_{01}	4.1482	0.0538	0.0001
β_{02}	4.0467	0.0529	0.0001
β_{03}	3.9361	0.0529	0.0001
β_1	-0.2286	0.0156	0.0001
β_2	0.0079	0.0080	0.3251
β_3	-0.0006	0.0005	0.2402

It can be seen from Table 9 that the estimated parameters obtained via the likelihood-based ignorable analysis are quite similar to the parameters obtained using the unconditional mean imputation. Furthermore, it can be noticed that the parameters associated with $\hat{a}_2$ and $\hat{a}_3$, are very small and non-significant. The latter suggests a constant trend over time after the third week.

To test formally the hypothesis whether H_0: $\hat{a}_2 = \hat{a}_3 = 0$ and whether the diets affect the mean response profiles, that is H_0: $\hat{a}_{01} = \hat{a}_{02} = \hat{a}_{03}$, appropriate contrasts were used. The results are presented in Table 10.

Table 10. Contrast statement results.

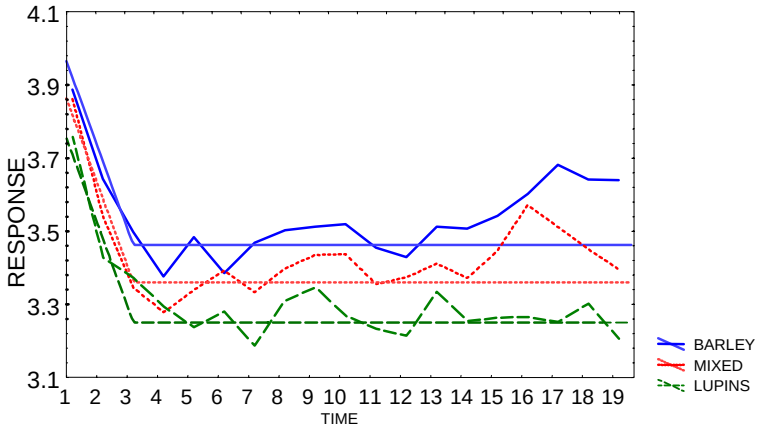
Source	NDF	DDF	F	Prob.
$\beta_2 = \beta_3 = 0$	2	1255	0.73	0.4826
Treatment Effect	2	1255	8.39	0.0002

It can be seen from the table above that the first hypothesis was accepted whereas the second one was rejected. Thus, also here there seems to be evidence of a treatment effect as was determined using all the previous methods. Constancy trend over time after the 3rd week was obtained as a result of this test.

Our model for the data is summarized by the parameter estimates in Table 9 but with the simplification that $\hat{a}_2 = \hat{a}_3 = 0$. The Figure 7 compares the empirical and the fitted mean response profiles.

Figure 7 indicates that the model is fitting quite well up to the final weeks, where some lack of fit can be observed. However, it must be taken into account that at the end of the study almost half of the animals are missing. As a consequence of the previous results the variability in the observed mean responses is increased.

Figure 7. Empirical and the fitted mean response profiles.



7. EXPLORING THE MISSING DATA PROCESS

Even though an ignorable analysis using Proc MIXED is to be prefered over complete case analysis and some forms of single imputation, this however does not preclude that there are situations where the missingness model needs to be considered explicitly. On the other hand, unless there are good prior grounds to believe otherwise, the presences of an informative response process can not be excluded.

In the following section, the issue of assessing the non-response mechanism for our data is discussed. To this end, three different approaches are given to model the drop-out process. In our problem the measurements were taken up till 19 weeks but it can be observed that 38 cows dropped out from the 15th week onwards. The aim of the following analysis is to study whether the dropout process is completly random, random or informative. The triple logistic model was used to model the probability of dropout for the weeks 15, 16, 17 and 19 (no drop-outs were observed at week 18).

Thus let us consider

$$\text{logit}(P_{ik}(y_{i1},\ldots,y_{ik})) = \theta_0 \cdot X_i + \theta_1 y_{ik} + \theta_2 y_{ik-1}$$

where

$$k = 15, 16, 17, 19 \text{ and } i = 1,\ldots,79$$

To test whether the process is informative or not, is equivalent to test if the parameter $\grave{e}_1$ is zero or not. Hence, in the case that we test if $\grave{e}_1$ and $\grave{e}_2$ are zero then we are testing that the missing is completely at random. The obtained results after applying the previous model can be found in Table 11.

We may now want to compare the dropout model. The likelihood ratio test statistic to compare MAR with MCAR is 214.014 ($p < 0.0001$); which indicates that dropouts are no completely random. More interestingly, there is overwhelming evidence in favour of informative dropout since the likelihood ratio statistics to test the MAR assumption is 10.3 ($p = 0.0013$). Even though the MAR assumption was rejected; the maximun likelihood estimates of the parameters are quite similar. This is not surprising, as most of the information about this parameters is contained in the 14 weeks of dropout-free data.

With regard to the possibility of an increase in the mean response towards the end of the experiment; the maximum likelihood estimates of $\hat{a}_2$ and $\hat{a}_3$ are both close to zero in all the cases. In this study the reassessment of the dropout process has not led to any substantive changes in our inference concerning the mean response profiles for the underlying drop-out free process.

Table 11.

Parameter	Dropout Modeled			
	Ignorable	**MCAR**	**MAR**	**Informative**
β_{01}	4.148248	4.148720	4.146976	4.156046
β_{02}	4.046674	4.046605	4.047612	4.040500
β_{03}	3.936123	3.936404	3.936042	3.934007
β_1	-0.228617	-0.228684	-0.228493	-0.230638
β_2	0.007950	0.007956	0.007836	0.009386
β_3	-0.000595	-0.000595	-0.000588	-0.000523
ν^2	0.000000	0.000000	0.000000	0.000000
σ^2	0.072104	0.072082	0.072118	0.069056
τ^2	0.023778	0.023782	0.023769	0.024720
ϕ	0.151799	0.151891	0.151955	0.153484
$\theta_{0,15}$			19.180760	18.477570
$\theta_{0,16}$			18.469300	17.341170
$\theta_{0,17}$			17.732170	16.612910
$\theta_{0,19}$			18.454760	17.933730
θ_1				4.941772
θ_2			-6.2420080	-11.132280
Deviance	5733.34	6081.0	5867.0	5856.7

CONCLUSIONS FOR THE MILK PROTEIN TRIAL

- There seems to be a difference in the three diets with respect ot the milk protein contents. The diet which performed best was Barley, followed by Mixed and eventually by Lupins.
- The evolution of the protein content was parallel for the three diets. A decreasing linear pattern in the first three weeks was observed, which may be due to the adaptation process of the animals involved in the experiments. This process did not change over time.
- In our case any approach gave the same conclusions in terms of the efficacy of the three diets.
- The results show that when we are modelling tendency over time, different imputations methods can lead to completely differents conclusions.

REFERENCES

DIGGLE, P.J. (ED) (1990): **Time series: A Biostatistical Introduction**, Oxford University Press, Oxford.

DIGGLE, P.J. and M.G. KENWARD (1994): "Informative dropout in longitudinal data analysis (with discussion)", **Applied Statistics**, 43, 49-93.

DIGGLE, P.J.; K.Y. LIANG and S.L. ZEGER (1994): **Analysis of Longitudinal Data**: Oxford Science Publications, Oxford.

HOGAN, J.W. and N.M. LAIRD (1997): "Mixture models for the joint distribution of repeated measures and event times", **Statistics in Medicine**, 16, 293-258.

VERBYLA, A.P. and B.R. CULLIS (1990): "Modelling in repeated measures experiments", **Applied Statistics**, 39, 341-356.