Assignment 3

02429 – Analysis of Correlated Data: Mixed Linear Models

DTU Compute, Autumn 2023

Formalities

In this assignment, you must perform a statistical analysis using the statistical software R (see details below), and present your analysis and conclusions in a written report. The report is to be submitted via DTU Inside no later than Wednesday 8 November 2019 at 23.59.

The report and all R code must be written individually, and the R code must be included as an appendix to the report. The code should be neat and readable, with a sufficient amount of comments. The report may be no longer than 10 pages excluding the front page, table of contents, figures, references, and appendices. The report must be written in English. All figures and tables included in the report must be referred to and discussed in the text. There should be no R code (or raw R output) in the report other than in the R code appendix. Some additional comments:

- The report should not contain excessive material. The overall presentation of your work including its coherence and the relevance of included material is assessed as well - not just the correctness of the statistical analysis.

- Pay special attention to the teaching material from week 6, 7 and 8.

- Remember that much of statistics is about quantifying and describing variability and uncertainty: When presenting, e.g., parameter estimates and Least-Squares means (LS-means), remember to include confidence intervals.

- Think of yourself as a consulting statistician when writing your report. You should include enough detail and explanation in the body of your report:

  a. to allow a statistician with no knowledge of R to replicate your analysis and validate your results using other statistical software *and*
  b. for someone with little or no knowledge of statistics to understand the overall purpose and conclusions of your report in the context of the data.

Data

A field trial was conducted to test six levels of nitrogen fertilizer and four rice varieties in terms of the rice yield. The field was first divided into three blocks. Each of the blocks was then devided into 6 rectangles, where the 6 fertilizer treatments were assigned at random, such that each treatment was used in 1 rectangle in each block. Afterwards, each rectangle was devided into 4 squares, and in each square a rice variety was allocated at random, such that all 4 varieties were represented in each ''fertilizer rectangle''.

The variable `N` indicates the nitrogen fertilizer level (1-6), `Var` the rice variety (1-4), `Block` the block (1-3), and `Yield` denotes the measured yield. The data is available in the file `assignment3.txt`.

Statistical analysis and report writting

Using R, perform statistical analyses in which you investigate the effect of nitrogen fertilizer and rice variety on the yield. You are expected to do two full analyses:

- one where you treat the nitrogen fertilizer as a categorical variable (factor).
- and one where you treat the nitrogen fertilizer as a numerical variable (covariate) by using the actual amount of fertilizer : (`1` = 0, `2` = 60, `3` = 90, `4` = 120, `5` = 150, `6` = 180).

Your report should contain the following elements:

- Front page
- Table of contents
- Introduction
    - Description of the data
- Descriptive/exploratory analysis of the data
- *Some sections of your choice*
- Discussion and conclusions
- Appendices

The description of the data should include a short description of the experiment and present all relevant variables together with their type (factor or numeric) and values (factor levels or range observed in the data). It should be indicated whether each factor is balanced or unbalanced, considered fixed or random in the analysis, and which factors are nested or crossed.

The descriptive/exploratory analysis of the data should include relevant plots, tables, and summary measures.

The "sections of your choice" should (when applicable) include:

- a relevant factor structure diagram annotated with numbers of levels and degrees of freedom.
- descriptions of relevant statistical models (including their assumptions) using appropriate mathematical notation and accompanying text.
- documentation of hypothesis tests performed in connection with model reduction.
- parameter estimates, documentation of post-hoc analyses/tests of particular interest, and illustrations related to the final model.
- model diagnostics.

The discussion and conclusions section should be used to briefly reflect on the results of the analyses in the context of the experiment and the data. *This includes a discussion of the different advangetages and disadvangetages of the considered models.*

Hints

The following line of R code might be useful:

```
data$Nc <- rep(rep(c(0,60,90,120,150,180),each = 4),3)
```

In the model, where you treat the nitrogen fertilizer as a numerical variable you could consider to include higher order terms, i.e. to use e.g. the squared nitrogen amount as a covariate. This is done in R by writing `I(Nc^2)` instead of `Nc` in the model formula.