

Assignment 4

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 29 November 2023 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 and 9.
- 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

In a study of the corn yield response to nitrogen fertilizer, one variety of corn were grown at two locations in Tennessee over five years. The yield was recorded for each location for six different amounts of nitrogen fertilizer for each year.

The variable `nitro` indicates the amount of nitrogen fertilizer used, `loc` the location (Jackson, Knoxville), `year` the year the corn was grown (1962-1966), and `yield` denotes the measured yield. The data is available in the file `assignment4.txt`.

Statistical analysis and report writing

Using R, perform a statistical analysis in which you investigate the effect of nitrogen fertilizer and location on the corn yield. The inference is supposed to be valid not only for the five years observed, but also for the successive years. You should consider to include higher order terms of the nitrogen amount as well as (up to) three way interactions in your model.

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 conclusion section should be used to briefly reflect on the results of the analysis in the context of the experiment and the data.