Lab 1 - Linear Mixed Models of Longitudinal Data

1 Introduction

In this first lab you should use SAS to explore a dataset of growth data of 11 girls and 16 boys. At each time of measurement the distance from the centre of the pituitary to the maxillary fissure is measured, and this was recorded at the ages 8, 10, 12 and 14. The data is unbalanced in the sense of unequal number of boys and girls. There are three data files for this data, *growthcc.sd2* contains the complete data, while in *growthax.sd2* is the dataset resulted from the values removed by Little and Rubin (1987). And in the file *growth2.sd2* individuals with "missing values" are completely removed. For a more detailed description of the data see section 2.6 and 17.4 in Verbeke and Molenberghs.

2 Modeling

First you should make separate linear profiles for boys and girls, with random intercepts and slopes, with a general covariance matrix D.

$$Y_{ij} = \beta_0 + \beta_{01}x_i + \beta_{10}(1 - x_i)t_{ij} + \beta_{11}x_i t_{ij} + b_0 + b_1 t_{ij} + \epsilon_{ij},$$

where x_i indicates the sex of the subject. You should try and fit the model using the options of

- untransformed age versus the transformation

$$age2 = \frac{age - 11}{3},$$

- with and without the 'nobound' option.

To see the effect of the missing values, fit the models to all three data sets.

The questions which you need to answer is

- In what statement should nobound be placed? What implication does it have on the resulting model(s)? What is the pros and cons of using this option?
- What assumption is made for Σ if nothing is specified in the repeated statement?
- What is the implication of the time transformation? When and why should it be used?
- How are the different models affected by the missing values?

- Which are the X_i and Z_i matrices for the complete data? (I.e. the design matrices which can be obtained from the model above, not the estimates.)
- Do the 4 different models yield any obvious difference in the mean response profiles?
- Consider the covariance matrix of the 4 repeated measures for the different models and discuss.

2.1 Modeling serial correlation between repeated measures

Here we start with model 7 (page 253 in Verbeke and Molenberghs), but make two new models with another variance-covariance structure,

- random intercept and AR(1) serial structure,
- random intercept, AR(1) serial structure and measurement error.

In these models it is enough to consider the complete data.

Questions:

- Is there any differences between the two models? Why? What does it mean to have both serial correlation and measurement error?
- Are these models more suitable to the data than the previous models?
- In which situations might the different models be appropriate (i.e random effects versus serial correlation)?
- Have you observed any effect of the unbalanced data on the estimated models?

3 Summary

Please write a short report where you describe your work in this lab and discuss the questions above.