Analysis of Longitudinal Data SoSe 2015 | **Sheet 2**

Prof. Dr. Sonja Greven and Dipl. Stat. Jona Cederbaum | 04.05.2015

This exercise sheet will introduce you to the estimation of **L**inear **M**ixed **M**odels. The exercises refer to the content of the third and fourth lecture slides.

Exercise 1:

In this exercise, we are working with the data set `rats` once more (see sheet 1, exercise 1). Use the code available on the homepage to import `rats` in `R` as well as to prepare the data set for the following analysis.

a) Fit a linear model based on all data with a linear slope in `logT` and a fixed effect for each rat. Which disadvantages do you see in this approach?

b) Now estimate a linear mixed model with a linear slope in `logT` and subject-specific random intercepts using the function `lme()`.

- **(i)** Formulate the underlying model for the response vector $\boldsymbol{Y}_i$ of the i-th rat and specify the dimensions of all components.
- **(ii)** What is the estimated marginal correlation between two measurements on the same rat?
  Note: Consider slides 37 – 41 from the third lecture slides. The function `getVarCov()` might be helpful.
- **(iii)** What is the estimated conditional correlation between two measurements on the same rat?
- **(iv)** What is the estimated correlation between two measurements at the same time on different rats?

c) In order to check graphically whether subject-specific slopes would be useful as well, estimate a separate linear model for each rat with at least 3 measurements using the function `lmList()` and plot the estimators and the confidence intervals for the intercept and for `logT` using the function `plot(intervals())`.
Note: The data set has to be a `groupedData` object with grouping variable `SUBJECT`.

- **i)** Why do we only consider rats with at least 3 measurements here?
- **ii)** Why are separate linear models for the single rats only suitable for such an illustration?
- **iii)** As a fairly large variation can also be seen in the estimates of `logT`, fit a linear mixed model with a linear slope in `logT` and with subject-specific random intercepts and slopes. For better comparability only use rats with at least 3 measurements.

(iv) Determine the estimated covariance matrix $\hat{\mathbf{D}}$ of the random effects. What is the estimated correlation between the random intercepts and slopes?

(v) How could the model be simplified with respect to this covariance structure? Estimate the corresponding model.

(vi) Now compare the coefficient estimates and the fitted values of the subject-specific linear models and of the linear mixed model suggested in v) using `plot(compareFits())` and `plot(comparePred())`. Describe what you notice. What is the name of the observed effect and how can it be explained?

d) Consider now the following model:

$$\begin{aligned}\texttt{RESPONSE}_{ij} &= \beta_0 + \beta_1\texttt{logT}_{ij} + \beta_2\texttt{GROUP1}_i + \beta_3\texttt{GROUP2}_i + \beta_4\texttt{GROUP1}_i \cdot \texttt{logT}_{ij} \\ &\quad + \beta_5\texttt{GROUP2}_i \cdot \texttt{logT}_{ij} + b_{0i} + \varepsilon_{ij}.\end{aligned}$$

Is it reasonable to assume that the random effects assumption is fulfilled here? Give reasons for your answer.