This exercise sheet will familiarize you with testing for random effects, as well as with flexible extensions of the linear mixed model (LMM). The exercises refer to the content of the fifth and sixth lecture slides.

Exercise 1: Testing for random effects

In the following, we will test the requirement of certain random effects using the data set antibiotics. The data set (in `groupedData` format) and a short description are available on the homepage.

(a) What is tested when we test the requirement of a random effect?

(b) Estimate a model `m_RI` with random intercepts for each child and a model `m_RIRS` with random intercepts and random slopes for each child, respectively, both including `time` as fixed effect.
Note: Use the function `update()`.

(c) Run a likelihood ratio (LR) test using the function `anova()` (`?anova.lme()`) to test the requirement for subject-specific slopes.

- **(i)** Formulate the corresponding null hypothesis.
- **(ii)** What is the test decision?
- **(iii)** Why is the use of this test, as it is implemented in `R`, problematic?

(d) A common solution is to use a mixture of two χ^2-distributions as an approximate distribution of the LR statistic under H_0. Calculate the resulting p value in this way and compare it with the results of **c)**.

(e) Is it problematic in this test that the models were estimated using REML? Give reasons for your answer.

(f) Now estimate the model `m_RIRS_unc`, which only differs from `m_RIRS` in the assumption, that the random intercepts and the random slopes are uncorrelated.

- **(i)** Formulate the null hypothesis.
- **(ii)** Perform a likelihood ratio test, which checks whether the random intercept and the random slope are correlated.
- **(iii)** Is the common distribution assumption of the LR statistic valid in this case?

(Bonus) Consider an alternative method to the χ^2-mixture used in **d)** to determine an approximate p value (googling allowed!).

Exercise 2: Flexible extensions of the mean

In this exercise, we focus on flexible assumptions of the relationship between covariates and response, which does not exclusively have to be linear.

a) Read the `R` help of the function `gamm` included in the package `mgcv`.

b) Now consider the data set antibiotics once more and estimate an LMM including uncorrelated random intercepts and random slopes for each child as well as a smooth function of `time`.

c) Represent the effect of `time` graphically and interpret it.
Note: Use the function `plot.gam`.

d) Write down the estimated covariance matrix of the random effects.
Note: Use the function `summary`.

Exercise 3: Flexible extensions of the error variance

In this exercise, we are looking at the assumptions of the error variance $\text{Cov}(\epsilon_i) = \mathbf{\Sigma}_i$ in the LMM. Hereby, we use the already known data set rats, which is linked to the homepage.

a) What does the frequently made assumption $\mathbf{\Sigma}_i = \sigma^2 \mathbf{I}_{n_i}$ imply? What are the consequences in a random intercept model?

b) How can serial correlation be formally integrated in the LMM? Write down the respective model components and the corresponding assumptions.

c) In how far does the consideration of serial correlation in `R` (using `corStruct`, cf. last exercises) differ from the form in **b)**?

d) What is the marginal correlation between two measurements on one subject for a model including random intercepts and serial correlation?

e) Although including serial correlation in a model can be reasonable, this also means that more parameters need to be estimated. It therefore makes sense to check whether serial correlation is really necessary.

(i) How can serial correlation be assessed? Outline the procedure briefly.

(ii) Use the code available on the homepage to estimate the following linear model for the data set rats

$$\texttt{REPONSE}_{ij} = \beta_0 + \beta_1 \texttt{logT}_{ij} + \beta_2 \texttt{GROUP1}_i \cdot \texttt{logT}_{ij} + \beta_3 \texttt{GROUP2}_i \cdot \texttt{logT}_{ij} + \varepsilon_{ij}.$$

Determine the empirical semi-variogram and plot it.
Note: Code for plotting an empirical semi-variogram can be found in the sixth lecture slides.

(iii) What can you see from the plot?