Exercises for course day 1

Statistical analysis of repeated measurements and clustered data

In this exercise we will work with the data from the gastric bypass study exemplified in the lecture and further described in tutorial # *1. Introduction to linear mixed models and analysis of single group studies* (available from the course webpage).

To do the exercise you will need the program file `gastricbypass.R`.

Note: If you use other software than R, you can use the data in `gastricbypass.txt`, which is located in the txt_files folder in the Files menu at the course webpage.

Exercise 1

Open the program file `gastricbypass.R` and load the data by following the instructions.

- Check that the dataset `wide` has been succesfully created.
- What variables does the data contain? Is it in the wide or in the long format?

Next use the code in `gastricbypass.R` to transform the data from the wide to the long format and proceed with the analysis of the bodyweights that was exemplified in the lectures. Make sure to notice what each step in the program does.

Exercise 2

In this exercise, write your own code to make an analysis with `glucagonAUC` as the outcome. Please note that you can use the data `long` which was created when you did exercise 1, you don't need to transform the data once more.

1. Before making a formal statistical analysis, you should make descriptive statistics to investigate the distribution of the AUCs.
    - Make a spaghettiplot to visualize the individual time evolutions in the AUCs of all the study participants. What trends do you see?

- Make scatterplots to visualize the joint distribution of the AUCs. Does it seem that the repeated measurements follow a joint normal distribution?
- Compute summary statistics over time. Is the data complete? Do you find a trend in the sample means? Or in the standard deviations?
- Compute the correlations between the repeated measurements. Any trends?

2. Fit a oneway ANOVA-like linear mixed model with visit (categorical) as fixed effect and an unstructured covariance pattern. Make visit 1 the reference visit in the analysis.

   - Verify that the numerical optimisation has converged.
   - Estimate the mean change in AUC from baseline to each follow-up time with 95% confidence intervals.
   - Take a look at the estimated residual correlation matrix. How does this compare with the correlations you computed in question 1?
   - What are the estimated standard deviations from the four time points? Compare these to the summary statistics from question 1.

3. For model validation:

   - Make a plot of the predicted mean AUCs over time. How do these compare with the summary statistics you computed in question 1?
   - Make diagnostic plots of the residuals. Anything worth noticing?

4. Estimate the mean change in AUC from one week before (2nd visit) to one week after the gastric bypass (3rd visit), either by changing the reference time point and re-running the linear mixed model.

5. Use the wide data to make paired t-tests that estimates:

   - The mean change in AUC from baseline to end of follow-up.
   - The mean change in AUC from one week before to one week after the gastric bypass.

   Compare these estimates and their 95% confidence intervals with the corresponding results you obtained from the linear mixed model analysis in questions 2 and 4.

   - **Challenge:** Note that the estimated mean change differ between the linear mixed model and the second of the paired t-tests. Can you guess why?