This assignment does not count toward the final grade.

Seminar 1

Start Assignment

- Due Friday by 23:59
- Points 4
- Submitting a file upload
- File types doc, docx, pdf, and odf

General instructions

Written solutions to all the tasks must be submitted before the deadline (Sep 6, at 23:59; as pdf files on Canvas, preferably including the most important parts of your scripts; make sure to add the names of all group members in the file).
You are also expected to prepare an oral presentation of your solution for each of the tasks. The presentation should aim at taking 15 minutes to leave room for questions and discussion. At the seminar, a member of your group (randomly chosen by the teacher) will be asked to present the solution.
We expect each of you to spend around 15h on these tasks; please plan your time and meetings keeping this in mind.
Please note that the course book (and the lectures) are not your only sources of information. There are lots of information available on the Internet about statistics as well as R and its possibilities, and you are very much encouraged to try and find methods not mentioned in the book or the lectures.

Background: drug development for the treatment of rheumatoid arthritis

Rheumatoid arthritis (RA) is a chronic autoimmune disorder that causes inflammation, swelling, tenderness and stiffness of the joints. Clinically, disease activity and treatment response are measured through composite scores based on swelling of the joints, such as the DAS28 (Disease Activity Score 28), and inflammatory biomarkers, such as CRP (C-reactive protein) and ESR (erythrocyte sedimentation rate).

Conventional treatment is centred around anti-inflammatory treatment and immunosuppression, *e.g.*, methotrexate, hydroxychloroquine and sulfasalazine. These medicines have many side effects and show high variability in treatment response. In recent years, biological medicines have started to become available for the treatment of RA with improved efficacy and tolerability.

Statistical and mathematical modelling are widely used to support the development and regulatory assessment of new medicinal products. As members of the clinical team at a large pharmaceutical company (named KTH), you have been tasked with supporting data analysis on disease and treatment response in RA. The company have several candidate drugs in development.

Task 1: the DAS28 dataset

A parallel clinical study was conducted to test the efficacy of a new experimental treatment (internal name, KTH001) as compared to control (standard treatment with placebo) treatment in 30 RA patients per study arm. The data set, '**data_task1.csv (https://canvas.kth.se/courses/49059/files/8220344?wrap=1)** ↓ **(https://canvas.kth.se/courses/49059/files/8220344/download?download_frd=1)** ', contains DAS28 measures in the two study groups.

- Develop and present a data analysis pipeline (a set of relevant assessments of the data along with a method selection approach).
- Analyse the clinical data set. Present your approach in a stepwise manner and the results of your chosen method(s).

Task 2: clinical study design

Due to the favourable safety profile of KTH001 (data not provided), a larger clinical study is being planned for the candidate drug. Assuming that the data in task 1 is normally distributed, carry out an analysis to assess the impact of sample size (number of patients included per study arm) and the impact of effect size on study power.

- Present your analysis of sample size and effect size.
- What suggestions can you make regarding the design of the upcoming larger clinical study?

Task 3: the CRP dataset

Additional biomarker data on CRP (C-reactive protein) has been made available for the two study arms in the clinical study detailed in Task 1 (KTH001 vs. control and placebo). In addition, CRP data has been made available for a separate study investigating an alternative candidate drug in clinical development (internal name, KTH002). The data set is available in '**data_task3_crp.csv' (https://canvas.kth.se/courses/49059/files/8220345?wrap=1)** ↓ **(https://canvas.kth.se/courses/49059/files/8220345/download?download_frd=1)** .

- Compare the CRP data across the three treatment arms. Carry out an appropriate statistical analysis to test the differences between the three groups.
- What conclusions can be drawn?

Task 4: drug concentration-time data

Individual drug concentration-time data for KTH002 (for a total of 11 study participants) has been sent over from the bioanalysis unit ('**conctimedata_reduced.csv**

(https://canvas.kth.se/courses/49059/files/8220343?wrap=1) ⬇ (https://canvas.kth.se/courses/49059/files/8220343/download?download_frd=1) ').

The dataset includes the following variables, ID: study participant ID number; TIME: time in hours after drug dosing; DV: dependent variable - the concentration data for KTH002 at the given observed time points (mg/L); WGT: bodyweight (kg); BSA: body surface area (m^2); AGE: age (years); HGT: height (cm); DOSE: dose (mg); GFR: glomerular filtration rate (ml/min; a measure of kidney function, the main drug eliminating organ).

- Explore and present different ways of visualising the data in the csv-file data_drugconcstime.csv.
- Can you come up with ways of summarising the individual concentration-time profiles for easy comparison to other variables?
- Do you find any trends in the data? What are the potential implications?