FINAL HOMEWORK
MIXED EFFECTS MODELS IN LIFE SCIENCE

The homework is to be done alone.

The homework is divided into two parts. A first part deals with linear mixed-effects models (class 1-3), and a second part deals with nonlinear mixed-effects models (class 3-6).

Each student is expected to provide:

- A pdf with a summary of the answers to the questions and results. It is important that you detail the reasoning that led to each analysis. Please also provide a brief description of how each algorithm used worked. This report should be no longer than 10 pages.
- A zip file containing all codes (R and/or Monolix) organized in an intelligible way.

The project is to be done in R and/or Monolix.

If you have any questions, please ask them in the Moodle forum only.
The project outputs need to be updated in the “Depot – Final Homework” section on moodle no later than the due date (to be selected in class and provided on moodle).

PART 1

This part consists in the analysis of a trial aimed to test the efficacy of a new combination antiretrovirals in HIV infected patients. It consists in comparing two nucleoside reverse transcriptase inhibitors, didanosine and stavudine (d4T+ddI), with the reference combination, zidovudine and lamivudine (AZT+3TC). This randomized trial enrolled 151 people for 24 weeks.

Patients were randomly assigned to one of three treatment groups:

- "AZT+3TC" arm: patients were treated with a combination of transcriptase inhibitors, zidovudine (AZT) and lamivudine (3TC), for 24 weeks,
- "d4T+ddI" arm: patients were treated for 24 weeks with dual therapy of transcriptase inhibitors with stavudine (d4T) and didanosine (ddI),
- Switch arm ("Switch" arm): patients were treated with a combination of stavudine and didanosine (d4T+ddI) for 12 weeks followed by zidovudine and lamivudine (AZT+3TC) for 12 weeks.

The study data are contained in the arv_hiv.txt file. The file contains the following variables:

- NUM _PAT: identifier of the patient.
- TIME: delay between baseline and visit in days
- VIS: number of the visit
- TD: delay between baseline and visit in weeks
- RAN _GRP: treatment group (1 for AZT+T3C, 2 for "switch" or 3 for d4T+ddI).
- CD4: number of CD4+ T cells in cells (cells/mm3)
- VL: HIV viral load in plasma (detection limit at 50 copies/ml).

1.1. Data Management: Import the data into R. Apply a transformation to normalize the response variables of interest (CD4 and VL). Generally, apply a log10(.) transformation to the viral loads and a (.)0.25 transformation to the cell count data. Verify that normality is improved using a qqplot. Remove non-normalized CD4 and VL variables from the table. Create a variable d4TddI that is 1 if the patient receives d4T+ddI and 0 if the patient receives AZT+3TC.

1.2. Describe the data. Plot the average trajectories.

1.3. In the following, we focus on groups "AZT+3TC" (RAN_GRP=1) and "d4T+ddI" (RAN_GRP=3). Exclude the group "switch" (RAN_GRP=2). Explain CD4 count as a function of time using linear mixed effects model (possibly using polynomial regression).

1.4. Conclude on the differential effect of the d4T+ddI treatment compared to AZT+3TC?

1.5. Bonus questions: Using the dataset arv_hiv.txt propose a mechanistic modeling of the virus and immune system dynamics. What is the mechanism of action the antiretroviral presented in these data?

PART 2

20 patients (identifier ID) received 135 mg (DOSE_mg) of an orally administered drug. The plasmatic concentration in ng/mL (CONC_ngmL) were then measured during 24 hours after the administration (TIME). Four baseline covariates are reported: SEX, AGE, WEIGHT, and RACE. The objective of this part 2 is to analyze the PK data provided in the file “final_PK_data_exam2022.txt”.

2.1. Describe and plot the data.

2.2. Build the structural model: use Monolix to find the model from the PK library that better fit the data. Explain and illustrate how you built the model. See “Monolix_PKPD_library.pdf” if you want details on the PK library.

2.3. Build the statistical model:

   i. The best covariate model
   ii. the best model for the random effects (are some random effects correlated? can some PK parameters be considered as fixed?)

2.4. Build the observation model: select the best residual error model.

2.5. Provide a general conclusion and description of the model.