E X A M

Concepts of multilevel, longitudinal, and mixed models

G. Verbeke, 2024-2025

1 General Remarks

1. The exam is a team project in which teams analyse a particular dataset based on a number of research questions (see below). The project report is uploaded in Toledo and will be the basis for the individual oral examination during which only questions about the project will be asked. Don't expect theoretical questions. Of course, sufficient knowledge of the theory is required in order to successfully complete the data analyses and reporting and to successfully answer questions about the project during the oral exam.

2. Teams need to have **4-6** students. No exceptions are allowed unless approved in advance, in writing, by G.Verbeke. Each team needs to be registered in OneButton (`https://icts.kuleuven.be/apps/onebutton/registrations/847114`) where you can see which students already joined a particular team. **It is important that everyone who registered for the course also registers in OneButton:**

   (a) In case you intend to participate in the exams organized for this course during the current academic year (first or second chance):

      i. If you already have a team, then all members of that team should register themselves for the same team:

         - If you are the first of your team to register, register yourself for one of the teams with names from A to Z and choose a team which does not contain any members yet
         - If someone else of your team already registered, register yourself for the same team

      ii. If you do not have a team yet and want the instructor of the course to assign you to a team with other students without a team, then register yourself for the team named **'I wish to be assigned to a team'**. In any case, don't join an existing team without first discussing this with the members of that team. If you want to get in contact with other students who are also looking for team members, a designated discussion board on Blackboard/Toledo can be used to post messages and get in contact with each other.

   (b) In case you participate in the course but you do not intend to participate in any of the exams organised for this course during the current academic year (first or second chance), then register yourself for the team named **'I will not participate in the exam during this academic year'**

   The deadline for registering is **April 4 2025 (midnight)**. After the deadline, the instructor will assign students without a team to existing teams with less than the required minimum number of 4 members, or to separate teams, and the final team compositions will be posted on Blackboard/Toledo. **Afterwards, no changes can be made anymore.**

3. Report of **maximum** 20 pages, not including the appendix, if any. Additional pages will be ignored !

4. Describe in detail the techniques used in the analyses, as well as the main results (= statistical part of the report).

5. Summarize the results in a non-statistical way, to be understood by a subject-matter scientist.

6. Give and discuss **relevant** graphs and/or tables.

7. **If necessary,** (SAS-)programs can be added in an appendix.

8. The final report is due **June 1 2025 (midnight)**, and should be uploaded **in pdf format** in Blackboard/Toledo. It is sufficient for one team member to upload the report on behalf of the entire team. Once the report has been uploaded, all students in the team will receive a confirmation email. Uploading after the deadline will not be possible. Several attempt are possible but only the last one will be graded.

9. The assignments will not be discussed over email. Questions about the course can be asked during lectures and during the Question&Answer online sessions. A discussion board has also been opened on Toledo to allow for interactions between students or teams. You can subscribe to the discussion board to receive a message each time a student has posted a question. The first objective is for other students to respond. If no response is found, or a wrong response is given, the teacher will post a reaction.

10. Marks: Project: 8pt
    Oral defense: 12pt

    The project is counted for the full 8 points on the condition that at least 6/12 is obtained for the oral part. In case the score for the oral part is less than 6/12, the score for the report is reduced to at most 4/8. In case the score for the oral part is less than 3/12, the score for the report is reduced to at most 2/8.

11. Arrangements for the September exams: No new assignment will be given, but a new, revised project report can be uploaded in Toledo (due date **August 3 2025, midnight**). Each student will have the opportunity to upload his/her individual document **in pdf format** in Blackboard/Toledo. If you do so, clearly indicate (e.g., using a different color) what has changed in comparison to the original report. Once the report has been uploaded, a confirmation email will be received. Uploading after the deadline will not be possible. Several attempt are possible but only the last one will be graded. If no new report is submitted, the original score on the project remains.

2 Introduction of the data

The data set considered in this homework results from a longitudinal observational study, the aim of which is to study the post-operative evolution of the physical ability of elderly hipfracture patients and their pre-operative cognitive status, and to study the effects of housing situation and age on these evolutions.

The physical ability is measured using the ADL (Activities of Daily Living) score, with values between 6 and 24, where low values correspond to little physical dependency of the patient, while high scores correspond to high physical dependency. The cognitive status is measured through the so-called 'neuro-status' which is a binary indicator for being neuro-psychiatric.

3 Data file

- SAS file adl.sas7bdat
- Variables
    1. id: patient identification number
    2. age: age of the patient at entry
    3. neuro: neuro-psychiatric status of the patient (1: neuro-psychiatric, 0: not neuro-psychiatric)
    4. adl: ADL score
    5. time: day after operation at which the ADL score has been measured (1, 5 or 12)
    6. housing situation: the housing situation prior to the hip fracture (1: alone, 2: with family or partner, 3: nursing home)

4 Questions

1. Describe the data, and explore the mean structure, the variance structure and the correlation structure. Summarize your conclusions. What are the implications with respect to statistical modelling ?
2. Fit a linear mixed model assuming a linear evolution for the response of every subject, on a log-scale for time (i.e., use $\log(t)$ as 'time scale'). Allow for subject-specific intercepts as well as slopes. Compare the average evolution of the two neuro groups. Interpret all estimated parameters.
3. Estimate the random effects, and produce a scatterplot of the slopes versus the intercepts.
4. In order to correct for important baseline differences between subjects, include the factor 'housing situation', and the covariate 'age'. Re-fit the model and check what results are different, when compared to the original model. Do the estimates for the random effects change ? Why (not) ?

5. Do older patients show different ADL evolutions than younger patients ? Is the evolution of ADL different for the different housing situations ?

6. What would be a meaningful dichotomization of the ADL score ?

7. Fit a meaningful logistic mixed model to compare the evolution of your dichotomized version of ADL over time, between the two neuro groups. Interpret all parameters in the model, including the random effects. What is the average evolution for neuro-psychiatric patients and for non-neuro-psychiatric patients, respectively. Compare your results with those obtained for the original ADL score.