Longitudinal Data Analysis

Assignment 1: Continuous Longitudinal Data
2024–2025

1 Introduction on Hemodialysis data

The data considered in this assignment are longitudinal measurements of hemoglobin (Hb, g/dl) for patients with renal deficiency receiving hemodialysis. Patients are closely monitored and are treated with Erythropoietin (EPO) in order to have their hemoglobin level within the desired interval (not too low, not too high). Hemoglobin is measured on a monthly basis, and the measured hemoglobin level is used to decide about the level of EPO to be administered during the next month. Every month, it is also checked whether there is indication of iron deficiency defined as serum ferritin below 100μg and/or transferrin saturation below 20%.

The objective of the study is to investigate how hemoglobin changes over time, and how it relates to the amount of EPO administered, the presence/absence of iron deficiency, and the age and sex of the patient.

2 Data file

- SAS file Hemodialysis.sas7bdat
- Variables:

  1. ID : identification number of the subject
  2. MONTH: month at which the measurement was taken (1–6)
  3. AGE : age of the patient at the time of entering the study (years)
  4. SEX: sex of the patient (1: male; 2: female)
  5. Hb: hemoglobin level (g/dl)
  6. DOSE: dose of EPO to be administered during the following month (IU/kg/week, IU: International Unit)
  7. IRON: indicator for iron deficiency (0: inadequate iron stores, 1: adequate iron stores)

3 Assignments

1. Describe the data, and use graphical techniques to explore the mean structure, the variance structure and the correlation structure. Summarize your conclusions. What are the implications with respect to statistical modeling ?

2. What summary statistics are appropriate for the analysis of these data ? Why ? Do they yield the same results ? Summarize your conclusions.

3. Fit a multivariate model, and find the most parsimonious mean structure which can be used to describe the average evolutions in the data. What covariance structures are applicable in this case ? What is the most parsimonious structure you can find ?

4. Use an explicit two-stage analysis to get an initial impression about trends and the effect of age on those trends.

5. Formulate a plausible random-effects model. Fit your model, and compare the results with those from the multivariate model. Check the appropriateness of your random-effects model. Calculate the subject-specific intercepts/slopes and compare them to the ones you obtained from a two-stage analysis. What do you conclude ?

4 General remarks

- For each question, motivate your choice of techniques, estimation methods, assumptions you make, and describe possible advantages/disadvantages, problems.
- For each of the above questions, summarize your conclusions and report them to a clinician.
- Carefully reflect on the parameterization of your models.
- Do you have any recommendations with respect to future similar experiments ?
- The deadline for submitting your report is **November 12, 2024, 9am**. Submit your report by mailing it to **both instructors** (Geert Verbeke: geert.verbeke@kuleuven.be, Geert Molenberghs: geert.molenberghs@uhasselt.be).
- As soon as your team composition is final, submit the composition to **both instructors**, and indicate for each member of your team whether he/she takes the course on campus or online, and whether he/she will present on November 12 or on December 5.