Mixed and longitudinal modeling

Roula Tsonaka, Biomedical Data Sciences, Medical Statistics, LUMC

Case study: Modelling the progression of migration in knee prostheses

1 Introduction

The data set considered in this assignment comes from a longitudinal study on patients who underwent total knee arthroplasty. The goal is to investigate several aspects of a new design of total knee prosthesis. Specifically the migration over time for two different versions of this new design is monitored. Migration is measured by means of radiostereometric analysis and here we will study one of the recorded outcomes, the Maximum Total Point Motion (MTPM) which is measured in mm. After operation some migration of the prosthesis is expected and stabilizes afterwards. Our interest is on the progression of migration and how different it is between the two types of prostheses.

In this study 60 patients have been included between March 2009 and July 2010, and underwent total knee operation. The patients have been randomized to receive one of two different types of prosthesis and they were follow-up for 5 years post-operatively. MTPM measurements were recorded the first day after the surgery, then at 3 months, 1 year, 2 years and at 5 years postoperatively. In addition to MTPM, the age, gender and the BMI of the patients is also recorded.

A complication in this study is that not all planned measurements have been recorded for several reasons e.g. loss to follow-up, relocation, poor quality of radiograph, etc.

2 Data file

- Prostheses.txt
- Variables
  - ID: patient indicator.
  - sex: sex indicator (1: males, 2: females).
  - BMI: BMI of the patient at each visit.
  - Age: age of the patient at inclusion.
  - nMTPM: age of the patient at inclusion.
  - FU.Months: months since inclusion in the study.
  - FU.Years: years inclusion in the study.

3 Questions

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 modelling?

2. Fit a multivariate model. The main research questions are:

   (*a*) Do both operation types migrate similarly over time?

   (*b*) Report the differences in the mean migration between the two operation types at each time point and make statistical inference.

   (*c*) Report the size of the migration change from 3 months to 1 year, 3 months to 2 years, 3 months 5 years and change from 1 year to 5 years per operation type and make statistical inference. Give the corresponding 95% confidence interval and p-value.

   (*d*) Are these changes different between the two types? Give the corresponding 95% confidence interval and p-value.

   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? Previous studies have used the log10(MTPM + 1) transformation. Would you consider this transformation in this study as well? Run model diagnostics to check the validity of the model assumptions.

3. Formulate a plausible linear mixed-effects model. As in the multivariate model above, the main research questions are:

   (*a*) Do both operation types migrate similarly over time?

   (*b*) Report the differences in the mean migration between the two operation types at each time point and make statistical inference.

   (*c*) Report the size of the migration change from 3 months to 1 year, 3 months to 2 years, 3 months 5 years and change from 1 year to 5 years per operation type and make statistical inference.

   (*d*) Are these changes different between the two types?

   What is the most parsimonious random-effects and mean structure that fits best on the data? Compare the results with those from the multivariate model. Run model diagnostics to check the validity of the model assumptions.

4 General remarks

- For each question, motivate your choice of techniques, estimation methods, assumptions you make, and describe the possible advantages/disadvantages, problems.

- Carefully reflect on the parameterization of your models, as well as on the nature of the covariates: which ones will be used as time-varying, which will be used as baseline-characteristics? Is a data transformation necessary? Will you use the time variable as categorical i.e. visits or as continuous i.e. years/months since first visit?
- Do you have any recommendations with respect to future similar experiments?
- Include as an appendix the SAS syntax file and/or R-code that reproduces your analyses.
- You are encouraged to use R markdown to prepare your report. In this case you may submit the .Rmd file that reproduces your report and analyses.

Good luck!