



MATH48132 - 2012/2013

General Information

- Title: Longitudinal Data Analysis
- Unit code: MATH48132
- Credits: 15
- Prerequisites: MATH38011 *Linear Statistical Models*.
- Co-requisite units: None
- School responsible: Mathematics
- Members of staff responsible: Prof. [J. Pan](#)

Specification

Aims

To study advanced techniques of statistical sciences, and to develop statistical skill of analyzing correlated data and cluster data. To explore a wide range of real-life examples occurring in particular in biology, medicine and social sciences.

Brief Description of the unit

In longitudinal studies, repeated measurements are made on subjects over time and responses within a subject are likely to be correlated, although responses between subjects may be independent. Data such as these are very common in practice, for example, in quality control in industry, panel data analysis in economics, growth curve analysis in biology and agriculture, randomized controlled trials in medicine and public health, etc. Longitudinal data therefore combine elements of multivariate and time series data. However, they differ from classical multivariate data in that the time series aspect of the data typical imparts a much more highly structured pattern of interdependence among measurements than for standard multivariate data sets; and they differ from classical time series data in consisting of a large number of short series, one from each subject, rather than a single long series. When modelling such data, these characteristics have to be taken into account. Otherwise, it is very likely that statistical inferences are severely biased.

The primary objective of longitudinal data analysis is to study how a response variable is related to explanatory variables of interest and how its expectation varies over time, by taking into account the within-subject correlation. The second objective is to quantify random variations in different sources and to characterize the within-subject correlation structures, which plays an important role in longitudinal and clustered data analysis arising in many areas.

Learning Outcomes

On successful completion of this course unit students will have a good understanding of

- the principles and methods for modelling of longitudinal data;
- the use of the statistical software R and S-PLUS to analyze longitudinal data;
- the implementation and interpretation of statistical models in standard applications.

Future topics requiring this course unit

This course unit is naturally related to some other 4th year courses on statistical modelling, e.g., Linear and Generalized Linear Models, Survival Analysis, etc.

Syllabus

1. Introduction: motivation examples from medical practice, fundamental problems of longitudinal data, exploring longitudinal data [2]
2. Ordinary linear regression models for longitudinal data: linear models with independent random errors, analysis of variance (ANOVA) for longitudinal data, drawbacks and limitations of the classical models [2]
3. General linear models for longitudinal data: general linear models with correlated random errors, various covariance models including compound symmetry, AR(1), exponential correlation, ante-dependence, etc., maximum likelihood estimation, restricted maximum likelihood estimation [4]

4. Linear mixed models: Fixed effects, random effects, random variation in different sources, model representation, variance components, maximum likelihood estimation, EM-algorithm, restricted maximum likelihood estimation, prediction of random effects, goodness of fit [10]
5. Non-normal longitudinal data models: a) population-averaged models: generalized estimating equations, working covariance specification, estimation and properties, b) subject-specific models: random effects models, exponential family of distributions, generalized linear mixed models, penalized quasi-likelihood estimation, variance component estimators, goodness of fit [10]
6. Statistical methods dealing with missing data: a) missing data mechanism: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR), b) simple methods of correction for missing data: single imputation and last-value-carried-forward methods, drawbacks and limitations, c) inference based methods: likelihood-based methods, multiple imputation, weighted estimating equations, sensitivity analysis [8]

Textbooks

- Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*. Springer, New York
- Diggle, P. J., Heagerty, P., Liang, K Y. and Zeger, S. L. (1994). *Analysis of longitudinal data*. 2nd Edition. Oxford University Press
- Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2004). *Applied longitudinal analysis*. New York, Wiley.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical analysis with missing data*, 2nd Edition. New York: Wiley.

Teaching and learning methods

Three lectures and one examples class each week. In addition students should expect to spend at least seven hours each week on private study for this course unit.

Assessment

Coursework 20%.

End of semester examination: three hours weighting 80%

Arrangements