Introduction to Modelling Multilevel Data
EPFL – Lausanne – Friday 16th December 2005 am

James Carpenter

London School of Hygiene & Tropical Medicine *and* Institute for Medical Biometry, Freiburg

Email: jrc@imbi.uni-freiburg.de

Who should come, and why

This course consists of two 1.5h lecture sessions and is designed for quantitative researchers and statisticians, particularly those working with medical and social data.

Motivated by medical and educational examples, the aim is to introduce the key concepts of multilevel and cross-classified models, and demonstrate accessible software for fitting them.

By the end of the morning, the objectives are that participants will (i) understand the key concepts of multilevel modelling; (ii) have an overview of the research area; (iii) be able to relate multilevel modelling to their particular applications, and (iv) have the confidence to download software and further materials for use in their own work.

There will be very limited time available to analyse participants' data. Anyone interested should email James Carpenter before 14th December.

Outline

Multilevel (also known as random effect, hierarchical, and mixed) models are an extensive and flexible class of models for correlated data, which occur naturally in many social science, medical and economic applications. For example, a child's achievement at school may be correlated with the achievement of other children in their class who share the same teacher; family members may share common traits. Further, many studies involve repeated measurements of the same subject or unit over time. Such longitudinal data are usually quite highly correlated.

We begin by illustrating the key ideas using data from an international multi-centre clinical trial with longitudinal follow-up and a continuous response. Initially we view this as a '2 level data set', with repeated measures at level 1 and patients at level 2. We discuss various models for the correlation structure, and their relevance for various settings. We then show how to extend the model to allow for possible correlation induced by patients being followed up by the same clinical centers (level 3) and living in the same country (level 4).

We describe how this approach can be generalised for discrete data, and compare likelihood based models with moment based models (i.e. generalised estimating equations). We outline various algorithms available for fitting multilevel models to both continuous and discrete data, discussing their advantages and disadvantages, and we review the capabilities of some common software packages.

We end with a discussion of the correlation structure that arises when two multilevel structures cross each other. For example, a child's educational achievement is likely to be correlated with that of other children in the same class and school. This is one multilevel structure. However, children who share the same neighbourhood (but who may not go to the same school) are also likely to have correlated achievement, if only because they are likely to share a common social background. This is another multilevel structure. Frequently, a data set will have both these structures co-existing. We call such data *cross-classified*. We discuss the implications of ignoring cross-classified structure when it is present, and outline how multilevel models can be extended to cope with cross classified data, illustrating the ideas with an educational example.

Background reading

Carpenter, J. R. (2005) Multilevel models. In *Encyclopaedic Companion to Medical Statistics*, eds. Everitt, B. and Palmer, C. London: Hodder & Stoughton.