**Generalized mixed effects models for the analysis of experimental data**

Christoph Scheepers

University of Glasgow

This course will introduce and explain state of the art techniques for the analysis of experimental data in psycholinguistics and psychology of language. The focus is on confirmatory analysis and how to perform experimental hypothesis tests in the most accurate and generalizable manner. The course will be *problem-oriented* in the sense of trying to provide best practises for the most common challenges in the analysis of experimental data. To this end, I will draw upon a range of representative examples (based on actual or simulated data), while keeping explanation of formal mathematical background to a tolerable minimum. The course will be based on packages for generalized linear mixed effects modelling in R (specifically, *lme4* and *ordinal*) and there will be four two-hour sessions (see below). Appropriate readings will be provided as the course progresses.

**Session 1 – Regression**

Since linear mixed models can be seen as an extension of basic regression, the first session will give a refresher of the principles behind *regression analysis*, and its relationship to other commonly used methods such as *t-tests* and *ANOVA*. Various predictor-coding schemes for the specification of hypothesis-relevant contrasts will be illustrated by appropriate examples, and their implications for parameter interpretation will be discussed.

**Session 2 – Generalized Linear Models**

The second session will introduce the concept of *generalised linear models* (function glm() in R), which are capable of adjusting modelling assumptions about error distributions and IV-DV relationships so as to handle various types of data within the same (generalised) regression framework. Using appropriate examples, three model families will be discussed in more detail: *binary logistic regression* for the analysis of dichotomous DVs, *ordinal logistic regression* for the analysis of rank-scale DVs (e.g. ratings), and *gamma regression* for the analysis of positively skewed continuous DVs (e.g. response times).

**Session 3 – Generalized Linear Mixed Models**

The third session will focus on repeated-measures designs (probably the most commonly used type of design in psycholinguistics and cognitive psychology), leading to the introduction of *generalised linear* mixed *models* (GLMMs). I will discuss the basic concept behind random versus fixed effects and how repeated-measures dependencies are handled in GLMMs to allow for appropriate generalization of experimental findings beyond the current sample of participants and items. Specifically, I will make an argument for specifying the *maximal random effects structure justified by the design* (for confirmatory analysis) and provide examples to illustrate its implementation and interpretation. Potential convergence issues associated with maximal GLMMs will also be discussed.

**Session 4 – Control Predictors in a (Maximal) GLMM**

Since psycholinguistic experiments often require control of potential confound variables (e.g. *lexical frequency* in a word-recognition experiment), the last session will specifically focus on how to handle control predictors (sometimes referred to as *covariates*) in a maximal GLMM. I will present results from data simulations showing that simple ‘matching’ of confound variables between, say, different groups of items in the stimulus set is often not enough to avoid anticonservative inferences. Using appropriate examples, I will illustrate how to tackle this problem while at the same time keeping model complexity at a tolerable level to avoid convergence problems.