Statistical Modelling: Practical 2

A. C. Davison and J. J. Forster

April 9, 2008

The data in the file `hip.txt` are taken from Crowder and Hand (*Analysis of Repeated Measures*, 1990, Chapman and Hall) and can be read into R using

```
hip <- read.table("hip.txt", col.names = c("y", "age", "sex", "subj", "time"))
```

Variable `y` represents measurements of response variable *haematocrit* on 30 patients (`subj`) on up to three occasions (`time`), one before a hip-replacement operation, and two afterwards. The `age` and `sex` (0=male, 1=female) of the patients is also recorded.

Plot the time profiles of the response variable for each subject on a single plot (equivalent plot to slide 98). It seems likely that models allowing for intra-subject dependence will be required.

Investigate these data using linear mixed models of the form: $y_{ij} \overset{\text{ind}}{\sim} N(\mu_{ij}, \sigma^2)$ where y_{ij} is the response for subject i, time j and

$$\mu_{ij} = x_{ij}^{\mathrm{T}}\beta + z_{ij}^{\mathrm{T}} b_i + \epsilon_{ij}, \quad b_i \overset{\text{ind}}{\sim} N(0, \Sigma_b).$$

You should consider including `age`, `sex` and `time` (and possibly interactions) within x_{ij} and `time` within z_{ij}.

LMMs for clustered data can be fitted in R using the `lme` function from the `nlme` library, so you first need to load this library using `library(nlme)`. [A more recent alternative is the function `lmer` from the `lme4` library, but this is not available in the lab we are using in Oxford].

For example

```
hip.lmm1 <- lme(y ~ age + sex + factor(time), random = ~ 1 | subj, data = hip)
```

fits the model with `1`, `age`, `sex` and `I(time=2)` and `I(time=3)` in x_{ij}, and just the intercept `1` in z_{ij}.

The default estimation method is REML. If you want to obtain maximum likelihood estimates (for example, for use in model comparison), they can be obtained using the additional argument `method = "ML"`.

You might find the following functions useful – they all take an `lme` fit as their first argument: `summary`, `fitted`, `residuals` (obvious), `fixef` (fixed effects estimates), `ranef` (random effects estimates), `VarCorr` (variance estimates) `coef` (coefficient estimates at cluster level, incorporating fixed and random effects), `AIC`, `BIC` (obvious), `intervals` (confidence intervals for fixed effects and variance parameters), `qqnorm` (normal probability plot of residuals or random effects) and `plot` (type `?plot.lme` to see the details of this flexible function).

Investigate these functions. For example, produce a plot equivalent to slide 113 (illustrating shrinkage in a random effects model) for your chosen model.

If you have time, reproduce the results given in lectures for the rat growth data (obtained as `rat.growth` by loading `library(SMPracticals)`).