APTS Statistical Modelling: Practical 2

Helen Ogden

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

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

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

You should consider including `age`, `sex` and `time` (and possibly interactions) within x_{ij} and `time` within z_{ij}. We will treat `time` as a categorical variable.

LMMs for clustered data can be fitted in R using the `lmer` function from the `lme4` library:

```
library(lme4)
```

```
## Loading required package: Matrix
```

For example

```
hip_lmm1 <- lmer(y ~ age + sex + factor(time) + (1 | subj), data = hip)
```

fits the model with `1`, `age`, `sex`, `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 `REML = FALSE`.

You might find some of the following functions useful – they all take an `lmer` fit as their first argument: `summary`, `fitted`, `residuals`, `fixef` (fixed effects estimates), `ranef` (random effects estimates), `VarCorr` (variance estimates) `coef` (coefficient estimates at cluster level, incorporating fixed and random effects), `AIC`, `BIC` and `predict`.

Tasks

1. Plot the time profiles of the response variable for each subject on a single plot (as we did for the rat growth data in Example 2.4 in the lecture notes). Do you think you think it will be necessary to include a random intercept for the subject? What about a random slope for time?
2. Find your preferred LMM for this data.
3. For your preferred LMM, plot the predicted haematocrit levels for each subject against time.