A Handbook of Statistical Analyses Using R

Brian S. Everitt and Torsten Hothorn
10.1 Introduction

10.2 Analysing Longitudinal Data

10.3 Analysis Using R

We shall fit both random intercept and random intercept and slope models to the data including the baseline BDI values (pre.bdi), treatment group, drug and length as fixed effect covariates. Linear mixed effects models are fitted in R by using the lmer function contained in the lme4 package (Bates and Sarkar, 2006, Pinheiro and Bates, 2000, Bates, 2005), but an essential first step is to rearrange the data from the ‘wide form’ in which they appear in the BtheB data frame into the ‘long form’ in which each separate repeated measurement and associated covariate values appear as a separate row in a data.frame. This rearrangement can be made using the following code:

```R
R> data("BtheB", package = "HSAUR")
R> BtheB$subject <- factor(rownames(BtheB))
R> nobs <- nrow(BtheB)
R> BtheB_long <- reshape(BtheB, idvar = "subject",
+   varying = c("bdi.2m", "bdi.4m", "bdi.6m", "bdi.8m"),
+   direction = "long")
R> BtheB_long$time <- rep(c(2, 4, 6, 8), rep(nobs, 4))
```

such that the data are now in the form (here shown for the first three subjects)

```
R> subset(BtheB_long, subject %in% c("1", "2", "3"))
```

```
drug length treatment bdi.pre subject time bdi
1.2m No >6m TAU 29 1 2 2
2.2m Yes >6m BtheB 32 2 2 16
3.2m Yes <6m TAU 25 3 2 20
1.4m No >6m TAU 29 1 4 2
2.4m Yes >6m BtheB 32 2 4 24
3.4m Yes <6m TAU 25 3 4 NA
1.6m No >6m TAU 29 1 6 NA
2.6m Yes >6m BtheB 32 2 6 17
3.6m Yes <6m TAU 25 3 6 NA
1.8m No >6m TAU 29 1 8 NA
2.8m Yes >6m BtheB 32 2 8 20
3.8m Yes <6m TAU 25 3 8 NA
```

The resulting data.frame BtheB_long contains a number of missing values
and in applying the `lmer` function these will be dropped. But notice it is only the missing values that are removed, not participants that have at least one missing value. All the available data is used in the model fitting process. The `lmer` function is used in a similar way to the `lm` function met in Chapter ?? with the addition of a random term to identify the source of the repeated measurements, here `subject`. We can fit the two models (??) and (??) and test which is most appropriate using

R> library("lme4")
R> BtheB_lmer1 <- lmer(bdi ~ bdi.pre + time + treatment + drug +
+ length + (1 | subject), data = BtheB_long,
+ method = "ML", na.action = na.omit)
R> BtheB_lmer2 <- lmer(bdi ~ bdi.pre + time + treatment + drug +
+ length + (time | subject), data = BtheB_long,
+ method = "ML", na.action = na.omit)
R> anova(BtheB_lmer1, BtheB_lmer2)

Data: BtheB_long
Models:
BtheB_lmer1: bdi ~ bdi.pre + time + treatment + drug + length + (1 | subject)
BtheB_lmer2: bdi ~ bdi.pre + time + treatment + drug + length + (time | subject)

Df AIC  BIC logLik deviance Chisq Chi Df
BtheB_lmer1  8 1886.6 1915.7  -935.31  1870.6
BtheB_lmer2 10 1889.8 1926.2  -934.90  1869.8  0.8161  2
Pr(>Chisq)
BtheB_lmer1
BtheB_lmer2 0.665

R> summary(BtheB_lmer1)

Linear mixed model fit by REML ['lmerMod']
Formula:
bdi ~ bdi.pre + time + treatment + drug + length + (1 | subject)
Data: BtheB_long

REML criterion at convergence: 1866.1

Scaled residuals:
          Min     1Q Median     3Q    Max
-2.7501 -0.4755 -0.0934  0.4001  3.7377

Random effects:
Groups   Name     Variance Std.Dev.
subject  (Intercept)  51.44   7.172
Residual            25.27   5.027
Number of obs: 280, groups: subject, 97

Fixed effects:   Estimate Std. Error t value
(Intercept)  5.92148   2.30586   2.568
bdi.pre     0.63888   0.07961   8.025
time        -0.71353   0.14664  -4.866
treatmentBtheB -2.35900   1.70841  -1.381
drugYes     -2.78885   1.76594  -1.579
length>6m    0.23810   1.67537   0.142

Correlation of Fixed Effects:
(Intr) bdi.pr time trtmBdB drugYs
bdi.pre    -0.679
time       -0.258  0.023
treatmentBtheB -0.389  0.121  0.022
drugYes    -0.072 -0.236 -0.025 -0.323
length>6m   -0.239 -0.241 -0.042  0.002  0.158

Figure 10.2  R output of the linear mixed-effects model fit for the BtheB data.
Bibliography

