CHAPTER 13

Analyzing Longitudinal Data I: Computerized Delivery of Cognitive Behavioral Therapy – Beat the Blues

13.1 Introduction

13.2 Analyzing Longitudinal Data

13.3 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, 2014; 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> data("BtheB", package = "HSAUR3")
R> BtheB$subject <- factor(rownames(BtheB))
R> nobs <- nrow(BtheB)
R> BtheB_long <- reshape(BtheB, idvar = "subject",
+ varying = c("bdi.2m", "bdi.3m", "bdi.5m", "bdi.8m"),
+ direction = "long")
R> BtheB_long$time <- rep(c(2, 3, 5, 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"))

<table>
<thead>
<tr>
<th>drug</th>
<th>length</th>
<th>treatment</th>
<th>bdi.pre</th>
<th>subject</th>
<th>time</th>
<th>bdi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2m</td>
<td>No</td>
<td>&gt;6m</td>
<td>TAU</td>
<td>29</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2.2m</td>
<td>Yes</td>
<td>&gt;6m</td>
<td>BtheB</td>
<td>32</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>3.2m</td>
<td>Yes</td>
<td>&lt;6m</td>
<td>TAU</td>
<td>25</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>1.3m</td>
<td>No</td>
<td>&gt;6m</td>
<td>TAU</td>
<td>29</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2.3m</td>
<td>Yes</td>
<td>&gt;6m</td>
<td>BtheB</td>
<td>32</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>3.3m</td>
<td>Yes</td>
<td>&lt;6m</td>
<td>TAU</td>
<td>25</td>
<td>3</td>
<td>NA</td>
</tr>
<tr>
<td>1.5m</td>
<td>No</td>
<td>&gt;6m</td>
<td>TAU</td>
<td>29</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2.5m</td>
<td>Yes</td>
<td>&gt;6m</td>
<td>BtheB</td>
<td>32</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>3.5m</td>
<td>Yes</td>
<td>&lt;6m</td>
<td>TAU</td>
<td>25</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>1.8m</td>
<td>No</td>
<td>&gt;6m</td>
<td>TAU</td>
<td>29</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2.8m</td>
<td>Yes</td>
<td>&gt;6m</td>
<td>BtheB</td>
<td>32</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>3.8m</td>
<td>Yes</td>
<td>&lt;6m</td>
<td>TAU</td>
<td>25</td>
<td>3</td>
<td>NA</td>
</tr>
</tbody>
</table>

The resulting data.frame BtheB_long contains a number of missing values
and in applying the \texttt{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 \texttt{lmer} function is used in a similar way to the \texttt{lm} function met in Chapter 6 with the addition of a random term to identify the source of the repeated measurements, here \texttt{subject}. We can fit the two models (??) and (??) and test which is most appropriate using

\begin{verbatim}
R> library("lme4")
\end{verbatim}
ANALYSIS USING R

R> BtheB_lmer1 <- lmer(bdi ~ bdi.pre + time + treatment + drug + + length + (1 | subject), data = BtheB_long, + REML = FALSE, na.action = na.omit)
R> BtheB_lmer2 <- lmer(bdi ~ bdi.pre + time + treatment + drug + + length + (time | subject), data = BtheB_long, + REML = FALSE, 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)

npar AIC BIC logLik deviance Chisq Df
BtheB_lmer1 8 1887.5 1916.6 -935.75 1871.5
BtheB_lmer2 10 1891.0 1927.4 -935.52 1871.0 0.4542 2

Pr(>Chisq)
BtheB_lmer1
BtheB_lmer2 0.7969
R> summary(BtheB_lmer1)

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

AIC BIC logLik deviance df.resid
1887.5 1916.6 -935.7 1871.5 272

Scaled residuals:
Min 1Q Median 3Q Max
-2.6975 -0.5026 -0.0638 0.4124 3.8203

Random effects:
Groups Name Variance Std.Dev.
subject (Intercept) 48.78 6.984
Residual 25.14 5.014
Number of obs: 280, groups: subject, 97

Fixed effects:
Estimate Std. Error t value
(Intercept) 5.59239 2.24244 2.494
bdi.pre 0.63968 0.07789 8.212
time -0.70476 0.14639 -4.814
treatmentBtheB -2.32908 1.67036 -1.394
drugYes -2.82495 1.72684 -1.636
length>6m 0.19708 1.63832 0.120

Correlation of Fixed Effects:
(Intr) bdi.pr time trtmTB drugTs
bdi.pre -0.682

time -0.238 0.020
treatmentBtheB -0.390 0.121 0.018
drugYes -0.073 -0.237 -0.022 -0.323
length>6m -0.243 -0.242 -0.036 0.002 0.157

Figure 13.2 R output of the linear mixed-effects model fit for the BtheB data.

The summary method for lmer objects doesn’t print p-values for Gaussian mixed models because the degrees of freedom of the t reference distribution are
not obvious. However, one can rely on the asymptotic normal distribution for computing univariate $p$-values for the fixed effects using the \texttt{cftest} function from package \texttt{multcomp}. The asymptotic $p$-values are given in Figure [13.3]

\begin{verbatim}
R> cftest(BtheB_lmer1)

Simultaneous Tests for General Linear Hypotheses
Fit: lmer(formula = bdi ~ bdi.pre + time + treatment + drug + length +
(1 | subject), data = BtheB_long, REML = FALSE, na.action = na.omit)

Linear Hypotheses:

(Intercept) == 0  5.59239  2.24244  2.494  0.0126
bdi.pre == 0     0.63968  0.07789  8.212  2.22e-16
time == 0        -0.70476  0.14639 -4.814  1.48e-06
treatmentBtheB == 0 -2.32908  1.67036 -1.394  0.1632
drugYes == 0     -2.82495  1.72684 -1.636  0.1019
length>6m == 0   0.19708  1.63832  0.120  0.9043
(Univariate p values reported)
\end{verbatim}

Figure 13.3  R output of the asymptotic $p$-values for linear mixed-effects model fit for the BtheB data.

We can check the assumptions of the final model fitted to the BtheB data, i.e., the normality of the random effect terms and the residuals, by first using the \texttt{ranef} method to \textit{predict} the former and the \texttt{residuals} method to calculate the differences between the observed data values and the fitted values, and then using normal probability plots on each. How the random effects are predicted is explained briefly in Section ???. The necessary R code to obtain the effects, residuals, and plots is shown with Figure [13.4]. There appear to be no large departures from linearity in either plot.
ANALYSIS USING R

R> layout(matrix(1:2, ncol = 2))
R> qint <- ranef(BtheB_lmer1)$subject["(Intercept)"]
R> qres <- residuals(BtheB_lmer1)
R> qqnorm(qint, ylab = "Estimated random intercepts",
+ xlab = c(-3, 3), ylab = c(-20, 20),
+ main = "Random intercepts")
R> qqline(qint)
R> qqnorm(qres, xlim = c(-3, 3), ylim = c(-20, 20),
+ ylab = "Estimated residuals",
+ main = "Residuals")
R> qqline(qres)

Figure 13.4  Quantile-quantile plots of predicted random intercepts and residuals for the random intercept model BtheB_lmer1 fitted to the BtheB data.


```r
R> bdi <- BtheB[, grep("bdi", names(BtheB))]
R> plot(1:4, rep(-0.5, 4), type = "n", axes = FALSE,
+     ylim = c(0, 50), xlab = "Months", ylab = "BDI")
R> axis(1, at = 1:4, labels = c(0, 2, 3, 5))
R> axis(2)
R> for (i in 1:4) {
+     dropout <- is.na(bdi[,i + 1])
+     points(rep(i, nrow(bdi)) + ifelse(dropout, 0.05, -0.05),
+            jitter(bdi[,i]), pch = ifelse(dropout, 20, 1))
+ }
```

Figure 13.5 Distribution of BDI values for patients that do (circles) and do not (bullets) attend the next scheduled visit.
Bibliography

