9
Analysis of Repeated Measures Data

9.1 Introduction

The multivariate data sets considered in previous chapters have involved measurements or observations on a number of different variables for each object or individual in the study. In this chapter, however, we will consider multivariate data of a different nature, namely that resulting from the repeated measurements of the same variable on each unit in the sample. Examples of such data are common in many disciplines. Often the repeated measurements arise from the passing of time (*longitudinal data*) but this is not always so. The two data sets in Tables 9.1 and 9.2 illustrate both possibilities. The first, taken from Crowder (1998), gives the loads required to produce slippage x of a timber specimen in a clamp. There are eight specimens, each with 15 repeated measurements. The second data set in Table 9.2 reported in Zerbe (1979) and also given in Davis (2002), consists of plasma inorganic phosphate measurements obtained from 13 control and 20 obese patients 0, 0.5, 1, 1.5, 2, and 3 hours after an oral glucose challenge.

The distinguishing feature of a repeated measures study is that the response variable of interest and a set of explanatory variables are measured several times on each individual in the study. The main objective in such a study is to characterize change in the repeated values of the response variable and to determine the explanatory variables most associated with any change. Because several observations of the response variable are made on the same individual, it is likely that the measurements will be correlated rather than independent, even after conditioning on the explanatory variables. Consequently repeated measures data require special methods of analysis, and models for such data need to include parameters linking the explanatory variables to the repeated measurements, parameters analogous to those in the usual multiple regression model (see Chapter 8), and, in addition parameters that account for the correlational structure of the repeated measurements. It is the former parameters that are generally of most interest, with the latter often being regarded as *nuisance parameters*. But providing an adequate model for the correlational structure of the repeated measures is necessary to avoid misleading inferences about the parameters that are of most importance to the researcher.

Table 9.1 Data Giving Loads Needed for a Given Slippage in 8 Specimens of Timber. From Nonlinear Growth Curves, Crowder, M.J., in *Encyclopedia of Biostatistics*, Armitage, P. and Colton, T. (Eds), Vol. 4, pp 3012–3014. Copyright © John Wiley & Sons Limited. Reproduced with permission.

	Slippage														
Specimen	0.0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.20	1.40	1.60	1.80
1	0.0	2.38	4.34	6.64	8.05	9.78	10.97	12.05	12.98	13.94	14.74	16.13	17.98	19.52	19.97
2	0.0	2.69	4.75	7.04	9.20	10.94	12.23	13.19	14.08	14.66	15.37	16.89	17.78	18.41	18.97
3	0.0	2.85	4.89	6.61	8.09	9.72	11.03	12.14	13.18	14.12	15.09	16.68	17.94	18.22	19.40
4	0.0	2.46	4.28	5.88	7.43	8.32	9.92	11.10	12.23	13.24	14.19	16.07	17.43	18.36	18.93
5	0.0	2.97	4.68	6.66	8.11	9.64	11.06	12.25	13.35	14.54	15.53	17.38	18.76	19.81	20.62
6	0.0	3.96	6.46	8.14	9.35	10.72	11.84	12.85	13.83	14.85	15.79	17.39	18.44	19.46	20.05
7	0.0	3.17	5.33	7.14	8.29	9.86	11.07	12.13	13.15	14.09	15.11	16.69	17.69	18.71	19.54
8	0.0	3.36	5.45	7.08	8.32	9.91	11.06	12.21	13.16	14.05	14.96	16.24	17.34	18.23	18.87

Table 9.2 Plasma Inorganic Phosphate Levels from 33 Subjects. From *Statistical Methods for the Analysis of Repeated Measurements*, Davis, C.F., 2002. Copyright Springer-Verlag New York Inc. Reprinted with permission.

		Hours after glucose challenge							
Group control	ID	0	0.5	1	1.5	2	3	4	5
	1	4.3	3.3	3.0	2.6	2.2	2.5	3.4	4.4
	2	3.7	2.6	2.6	1.9	2.9	3.2	3.1	3.9
	3	4.0	4.1	3.1	2.3	2.9	3.1	3.9	4.0
	4	3.6	3.0	2.2	2.8	2.9	3.9	3.8	4.0
	5	4.1	3.8	2.1	3.0	3.6	3.4	3.6	3.7
	6	3.8	2.2	2.0	2.6	3.8	3.6	3.0	3.5
	7	3.8	3.0	2.4	2.5	3.1	3.4	3.5	3.7
	8	4.4	3.9	2.8	2.1	3.6	3.8	4.0	3.9
	9	5.0	4.0	3.4	3.4	3.3	3.6	4.0	4.3
	10	3.7	3.1	2.9	2.2	1.5	2.3	2.7	2.8
	11	3.7	2.6	2.6	2.3	2.9	2.2	3.1	3.9
	12	4.4	3.7	3.1	3.2	3.7	4.3	3.9	4.8
	13	4.7	3.1	3.2	3.3	3.2	4.2	3.7	4.3
	14	4.3	3.3	3.0	2.6	2.2	2.5	2.4	3.4
	15	5.0	4.9	4.1	3.7	3.7	4.1	4.7	4.9
	16	4.6	4.4	3.9	3.9	3.7	4.2	4.8	5.0
	17	4.3	3.9	3.1	3.1	3.1	3.1	3.6	4.0
	18	3.1	3.1	3.3	2.6	2.6	1.9	2.3	2.7
	19	4.8	5.0	2.9	2.8	2.2	3.1	3.5	3.6
	20	3.7	3.1	3.3	2.8	2.9	3.6	4.3	4.4
Obese	21	5.4	4.7	3.9	4.1	2.8	3.7	3.5	3.7
	22	3.0	2.5	2.3	2.2	2.1	2.6	3.2	3.5
	23	4.9	5.0	4.1	3.7	3.7	4.1	4.7	4.9
	24	4.8	4.3	4.7	4.6	4.7	3.7	3.6	3.9
	25	4.4	4.2	4.2	3.4	3.5	3.4	3.8	4.0
	26	4.9	4.3	4.0	4.0	3.3	4.1	4.2	4.3
	27	5.1	4.1	4.6	4.1	3.4	4.2	4.4	4.9
	28	4.8	4.6	4.6	4.4	4.1	4.0	3.8	3.8
	29	4.2	3.5	3.8	3.6	3.3	3.1	3.5	3.9
	30	6.6	6.1	5.2	4.1	4.3	3.8	4.2	4.8
	31	3.6	3.4	3.1	2.8	2.1	2.4	2.5	3.5
	32	4.5	4.0	3.7	3.3	2.4	2.3	3.1	3.3
	33	4.6	4.4	3.8	3.8	3.8	3.6	3.8	3.8

Over the last decade methodology for the analysis of repeated measures data has been the subject of much research and development, and there are now a variety of powerful techniques available. A comprehensive account of these methods is given in Diggle et al. (2002) and Davis (2002). Here we will concentrate on a single class of methods, *linear mixed effects models*.

9.2 Linear Mixed Effects Models for Repeated Measures Data

Linear mixed effects models for repeated measures data formalize the sensible idea that an individual's pattern of responses is likely to depend on many characteristics of that individual, including some that are unobserved. These unobserved variables are then included in the model as random variables, that is, random effects. The essential feature of the model is that correlation amongst the repeated measurements on the same unit arises from stored, unobserved variables. Conditional on the values of the random effects, the repeated measurements are assumed to be independent, the so-called *local independence* assumption.

Linear mixed effects models are introduced in Display 9.1 in the context of the timber slippage data in Table 9.1 by describing two commonly used models, the *random intercept* and *random intercept and slope* models.

Display 9.1
Two Simple Linear Mixed Effects Models

- Let y_{ij} represent the load in specimen i needed to produce a slippage of x_j, with $i = 1, \ldots, 8$ and $j = 1, \ldots, 15$. A possible model for the y_{ij} might be

$$y_{ij} = \beta_0 + \beta_1 x_j + u_i + \varepsilon_{ij} \quad \text{(A)}$$

- Here the total residual that would be present in the usual linear regression model has been partitioned into a subject-specific random component u_i, which is constant over time plus a residual ε_{ij}, which varies randomly over time. The u_i are assumed to be normally distributed with zero mean and variance σ_u^2. Similarly the ε_{ij} are assumed to be normally distributed with zero mean and variance σ^2. The u_i and the ε_{ij} are assumed to be independent of each other and of the x_j.
- The model in (A) is known as a *random intercept model*, the u_i being the random intercepts. The repeated measurements for a specimen vary about that specimen's own regression line which can differ in intercept but not in slope from the regression lines of other specimens. The random effects model possible heterogeneity in the intercepts of the individuals.
- In this model slippage has a fixed effect.
- The random intercept model implies that the total variance of each repeated measurement is

$$\text{Var}(u_i + \varepsilon_{ij}) = \sigma_u^2 + \sigma^2.$$

- Due to this decomposition of the total residual variance into a between-subject component, σ_u^2, and a within-subject component, σ^2, the model is sometimes referred to as a *variance component model*.
- The covariance between the total residuals at two slippage levels j and j' in the same specimen i is

$$\text{Cov}\left(u_i + \varepsilon_{ij}, u_i + \varepsilon_{ij'}\right) = \sigma_u^2.$$

- Note that these covariances are induced by the shared random intercept; for specimens with $u_i > 0$, the total residuals will tend to be greater than the mean, for specimens with $u_i < 0$ they will tend to be less than the mean.
- It follows from the two relations above that the residual correlations are given by

$$\text{Cor}(u_i + \varepsilon_{ij}, u_i + \varepsilon_{ij'}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}.$$

- This is an *intraclass correlation* interpreted as the proportion of the total residual variance that is due to residual variability between subjects.
- A random intercept model constrains the variance of each repeated measure to be the same and the covariance between any pair of measurements to be equal. This is usually called the *compound symmetry* structure.
- These constraints are often not realistic for repeated measures data. For example, for longitudinal data it is more common for measures taken closer to each other in time to be more highly correlated than those taken further apart. In addition the variances of the later repeated measures are often greater than those taken earlier.
- Consequently for many such data sets the random intercept model will not do justice to the observed pattern of covariances between the repeated measures. A model that allows a more realistic structure for the covariances is one that allows heterogeneity in both slopes and intercepts, the *random slope and intercept model*.
- In this model there are two types of random effects, the first modelling heterogeneity in intercepts, u_{i1}, and the second modelling heterogeneity in slopes, u_{i2}.
- Explicitly the model is

$$y_{ij} = \beta_0 + \beta_1 x_j + u_{i1} + u_{i2} x_j + \varepsilon_{ij}, \tag{B}$$

  where the parameters are not, of course, the same as in (A).
- The two random effects are assumed to have a bivariate normal distribution with zero means for both variables, variances $\sigma_{u_1}^2$, $\sigma_{u_2}^2$ and covariance $\sigma_{u_1 u_2}$.
- With this model the total residual is $u_{i1} + u_{i2} x_j + \varepsilon_{ij}$ with variance

$$\text{Var}(u_{i1} + u_{i2} x_j + \varepsilon_{ij}) = \sigma_{u_1}^2 + 2\sigma_{u_1 u_2} x_j + \sigma_{u_2}^2 x_j^2 + \sigma^2,$$

  which is no longer constant for different values of x_j.

- Similarly the covariance between two total residuals of the same individual

$$\begin{aligned}&\mathrm{Cov}(u_{i1} + x_j u_{i2} + \varepsilon_{ij}, u_{i1} + u_{i2} x_{j'} + \varepsilon_{ij'})\\ &= \sigma^2_{u_1} + \sigma_{u_1 u_2}(x_j + x_{j'}) + \sigma^2_{u_2} x_j x_{j'}\end{aligned}$$

  is not constrained to be the same for all pairs j and j'.
- Linear mixed-effects models can be estimated by maximum likelihood. However, this method tends to underestimate the variance components. A modified version of maximum likelihood, known as *restricted maximum likelihood*, is therefore often recommended; this provides consistent estimates of the variance components. Details are given in Diggle et al. (2002) and Longford (1993).
- It should also be noted that re-estimating the models after adding or subtracting a constant from x_j (e.g., its mean), will lead to different variance and covariance estimates, but will not affect fixed effects.
- Competing linear mixed-effects models can be compared using a likelihood ratio test. If, however, the models have been estimated by restricted maximum likelihood this test can only be used if both models have the same set of fixed effects (see Longford, 1993).

Assuming that the data is available as shown in Table 9.1 as the matrix `timber`, we first need to rearrange it into what is known as the *long form* before we can apply the `lme` function that fits linear mixed effects models. This simply means that the repeated measurements are arranged "vertically" rather than horizontally as in Table 9.1. Suitable R and S-PLUS® code to make this rearrangement is

```
x<-c(0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0,1.2,1.4,
   1.6,1.8)
#
slippage<-rep(x,8)
loads<-as.vector(t(timber))
specimen<-rep(1:8,rep(15,8))
#
timber.dat<-data.frame(specimen,slippage,loads)
#
```

The rearranged data (`timber.dat`) are shown in Table 9.3. We can now fit the two models (A) and (B) as described in Display 9.1 and test one against the other using the `lme` function (in R the `nlme` library will first need to be loaded);

```
#in R use library(nlme)
attach(timber.dat)
#random intercept model
```

Table 9.3 Timber Data in "Long" Form

Observation	Specimen	Slippage	Load
1	1	0.0	0.00
2	1	0.1	2.38
3	1	0.2	4.34
4	1	0.3	6.64
5	1	0.4	8.05
6	1	0.5	9.78
7	1	0.6	10.97
8	1	0.7	12.05
9	1	0.8	12.98
10	1	0.9	13.94
11	1	1.0	14.74
12	1	1.2	16.13
13	1	1.4	17.98
14	1	1.6	19.52
15	1	1.8	19.97
16	2	0.0	0.00
17	2	0.1	2.69
18	2	0.2	4.75
19	2	0.3	7.04
20	2	0.4	9.20
21	2	0.5	10.94
22	2	0.6	12.23
23	2	0.7	13.19
24	2	0.8	14.08
25	2	0.9	14.66
26	2	1.0	15.37
27	2	1.2	16.89
28	2	1.4	17.78
29	2	1.6	18.41
30	2	1.8	18.97
31	3	0.0	0.00
32	3	0.1	2.85
33	3	0.2	4.89
34	3	0.3	6.61
35	3	0.4	8.09
36	3	0.5	9.72

(*Continued*)

Table 9.3 (*Continued*)

Observation	Specimen	Slippage	Load
37	3	0.6	11.03
38	3	0.7	12.14
39	3	0.8	13.18
40	3	0.9	14.12
41	3	1.0	15.09
42	3	1.2	16.68
43	3	1.4	17.94
44	3	1.6	18.22
45	3	1.8	19.40
46	4	0.0	0.00
47	4	0.1	2.46
48	4	0.2	4.28
49	4	0.3	5.88
50	4	0.4	7.43
51	4	0.5	8.32
52	4	0.6	9.92
53	4	0.7	11.10
54	4	0.8	12.23
55	4	0.9	13.24
56	4	1.0	14.19
57	4	1.2	16.07
58	4	1.4	17.43
59	4	1.6	18.36
60	4	1.8	18.93
61	5	0.0	0.00
62	5	0.1	2.97
63	5	0.2	4.68
64	5	0.3	6.66
65	5	0.4	8.11
66	5	0.5	9.64
67	5	0.6	11.06
68	5	0.7	12.25
69	5	0.8	13.35
70	5	0.9	14.54
71	5	1.0	15.53
72	5	1.2	17.38

(*Continued*)

Table 9.3 (*Continued*)

Observation	Specimen	Slippage	Load
73	5	1.4	18.76
74	5	1.6	19.81
75	5	1.8	20.62
76	6	0.0	0.00
77	6	0.1	3.96
78	6	0.2	6.46
79	6	0.3	8.14
80	6	0.4	9.35
81	6	0.5	10.72
82	6	0.6	11.84
83	6	0.7	12.85
84	6	0.8	13.83
85	6	0.9	14.85
86	6	1.0	15.79
87	6	1.2	17.39
88	6	1.4	18.44
89	6	1.6	19.46
90	6	1.8	20.05
91	7	0.0	0.00
92	7	0.1	3.17
93	7	0.2	5.33
94	7	0.3	7.14
95	7	0.4	8.29
96	7	0.5	9.86
97	7	0.6	11.07
98	7	0.7	12.13
99	7	0.8	13.15
100	7	0.9	14.09
101	7	1.0	15.11
102	7	1.2	16.69
103	7	1.4	17.69
104	7	1.6	18.71
105	7	1.8	19.54
106	8	0.0	0.00
107	8	0.1	3.36
108	8	0.2	5.45

(*Continued*)

Table 9.3 (*Continued*)

Observation	Specimen	Slippage	Load
109	8	0.3	7.08
110	8	0.4	8.32
111	8	0.5	9.91
112	8	0.6	11.06
113	8	0.7	12.21
114	8	0.8	13.16
115	8	0.9	14.05
116	8	1.0	14.96
117	8	1.2	16.24
118	8	1.4	17.34
119	8	1.6	18.23
120	8	1.8	18.87

```
timber.lme<-
   lme(loads~slippage,random=~1|specimen,data=timber.dat,
   method="ML")
#random intercept and slope model
timber.lme1<-
   lme(loads~slippage,random=~slippage|specimen,data
   =timber.dat, method="ML")
#compare two models
anova(timber.lme,timber.lme1)
```

The p-value associated with the likelihood ratio test is very small indicating that the random intercept and slope model is to be preferred over the simpler random intercept model for these data. The results from this model found from

```
summary(timber.lme1)
```

are shown in Table 9.4. The regression coefficient for slippage is highly significant. We can find the predicted values under this model and then plot them alongside a plot of the raw data using the following R and S-PLUS code:

```
predictions<-matrix(predict(timber.lme1),ncol=15,byrow=T)
par(mfrow=c(1,2))
matplot(x,t(timber),type="l",col=1,xlab="Slippage",
   ylab="Load",lty=1,
```

Table 9.4 Results of Random Intercept and Slope Model for the Timber Data

Effect	Estimated reg coeff	SE	DF	t-value	p-value
Intercept	3.52	0.26	111	13.30	<0.0001
Slippage	10.37	0.28	111	36.59	<0.0001

$\hat{\sigma}_{u_1} = 0.042, \hat{\sigma}_{u_2} = 0.014, \hat{\sigma} = 1.64.$

```
ylim=c(0,25))
title("(a)")
matplot(x,t(predictions),type="l",col=1,xlab="Slippage",
   ylab="Load",lty=1,
ylim=c(0,25))
title("(b)")
```

The resulting plot is shown in Figure 9.1 Clearly the fit is not good. In fact, under the random intercept and slope model the predicted values for each specimen are almost identical, reflecting the fact that the estimated variances of both random effects are essentially zero.

The plot of the observed values in Figure 9.1 shows that a quadratic term in slippage is essential in any model for these data. Including this as a fixed effect, the required model is

$$y_{ij} = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + u_{i1} + u_{i2} x_j + \varepsilon_{ij}. \tag{9.1}$$

The necessary R and S-PLUS code to fit this model and test it against the previous random intercept and slope model is

```
timber.lme2<-lme(loads~slippage+I(slippage*slippage),
random=~slippage|specimen,data=timber.dat,method="ML")
anova(timber.lme1,timber.lme2)
```

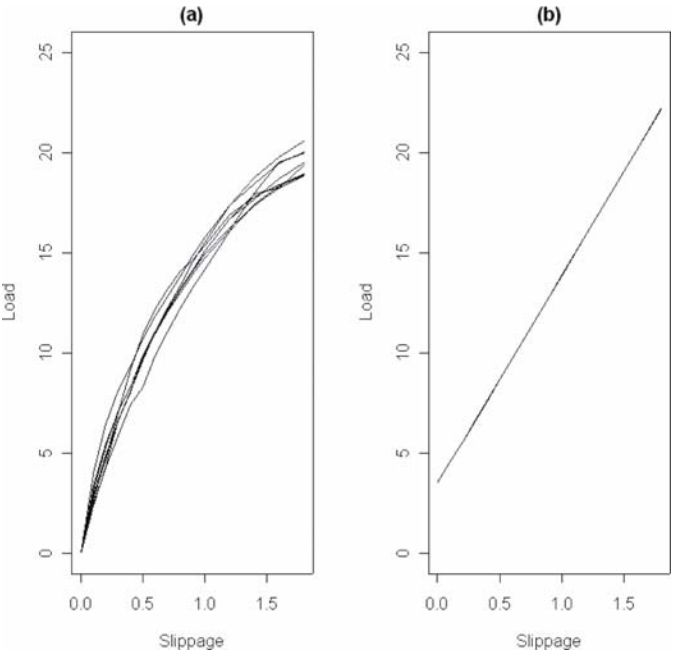

Figure 9.1 Observed timber data (a) and predicted values from random intercept and slope model (b).

Table 9.5 Results of Random Intercept and Slope Model with a Fixed Quadratic Effect for Slippage for Timber Data

Effect	Estimated reg coeff	Sd S.E	DF	t-value	p-value
Intercept	0.94	0.21	110	4.52	<0.0001
Slippage	19.89	0.33	110	61.11	<0.0001
Slippage^2	−5.43	0.17	110	−32.62	<0.0001

$\hat{\sigma}_{u_1} = 0.049, \hat{\sigma}_{u_2} = 0.032, \hat{\sigma} = 0.50.$

The p-value from the likelihood ratio test is less than 0.0001 indicating that the model that includes a quadratic term does provide a much improved fit. The results from this model are shown in Table 9.5. Both the linear and quadratic effects of slippage are highly significant.

We can now produce a similar plot to that in Figure 9.1 but showing the predicted values from the model in (9.1). The code is similar to that given above and so is not repeated again here. The resulting plot is shown in Figure 9.2. Clearly the model describes the data more satisfactorily although there remains an obvious problem which is taken up in Exercise 9.1.

Now we can move on to consider the data in Table 9.2, which we assume are available as the matrix `plasma`. Here we will begin by plotting the data so that we

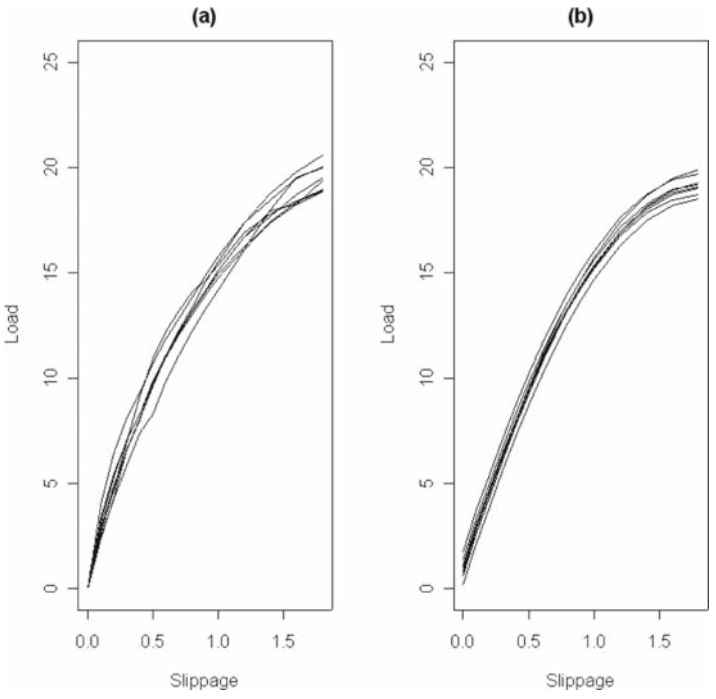

Figure 9.2 Observed timber data (a) and predicted values from random intercept and slope model that includes a quadratic effect for slippage (b).

get some ideas as to what form of linear mixed effect model might be appropriate. First we plot the raw data separately for the control and the obese groups using the following code:

```
par(mfrow=c(1,2))
matplot(matrix(c(0.0,0.5,1.0,1.5,2.0,2.5,3.0,4.0),ncol=1),
t(plasma[1:13,]),type="l",col=1,lty=1,
xlab="Time (hours after oral glucose challenge)",
   ylab="Plasma inorganic phosphate",ylim=c(1,7))
title("Control")
matplot(matrix(c(0.0,0.5,1.0,1.5,2.0,2.5,3.0,4.0),ncol=1),
t(plasma[14:33,]),type="l",col=1,lty=1,
xlab="Time (hours after glucose challenge)",ylab="Plasma
   inorganic phosphate",ylim=c(1,7))
title("Obese")
```

This gives Figure 9.3. The profiles in both groups show some curvature, suggesting that a quadratic effect of time may be needed in any model. There also appears to be some difference in the shape of the curves in the two groups, suggesting perhaps the need to consider a group $\times$ time interaction.

Next we plot the scatterplot matrices of the repeated measurements for the two groups using;

```
pairs(plasma[1:13,])
pairs(plasma[14:33,])
```

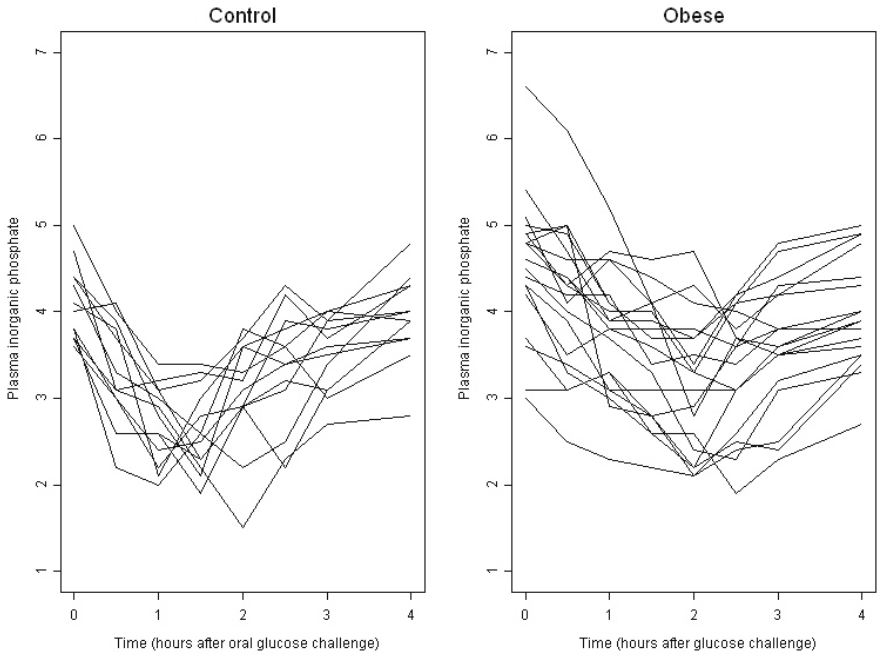

Figure 9.3 Glucose challenge data for control and obese groups.

The results are shown in Figures 9.4 and 9.5. Both plots indicate that the correlations of pairs of measurements made at different times differ so that the compound symmetry structure for these correlations is unlikely to be appropriate.

On the basis of the plots in Figure 9.3–9.5 we will begin by fitting the model in (9.1) with the addition, in this case, of an extra covariate, namely a dummy variable coding the group, control or obese, to which a subject belongs. We first need to put the data into the long form and combine with the appropriate group coding, subject number, and time. The necessary R and S-PLUS code for this is:

```
#
group<-rep(c(0,1),c(104,160))
#
time<-c(0.0,0.5,1.0,1.5,2.0,3.0,4.0,5.0)
time<-rep(time,33)
#
subject<-rep(1:33,rep(8,33))
plasma.dat<-cbind(subject,time,group,as.vector(t(plasma)))
dimnames(plasma.dat)<-list(NULL,c("Subject","Time","Group",
   "Plasma"))
plasma.df<-as.data.frame(plasma.dat)
plasma.df$Group<-factor(plasma.df$Group,levels=c(0,1),
   labels=c("Control","Obese"))
attach(plasma.df)
```

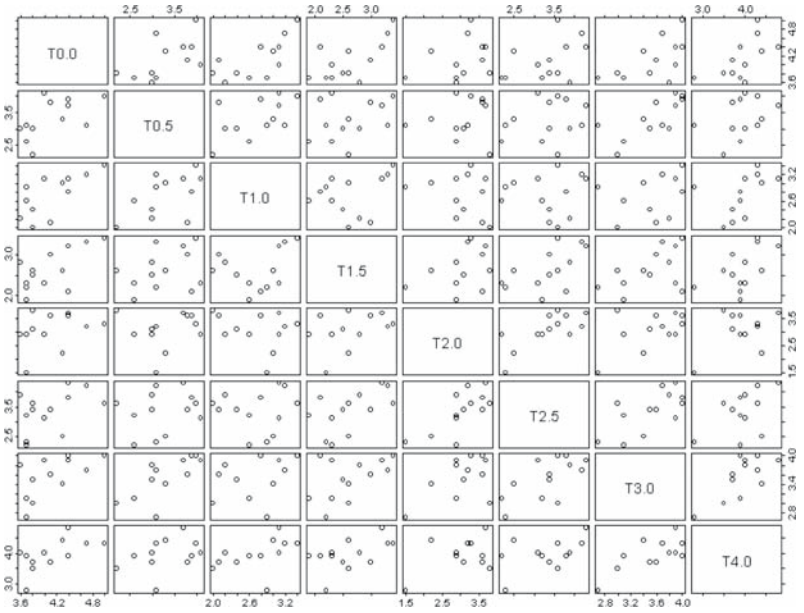

Figure 9.4 Scatterplot matrix for control group in Table 9.2.

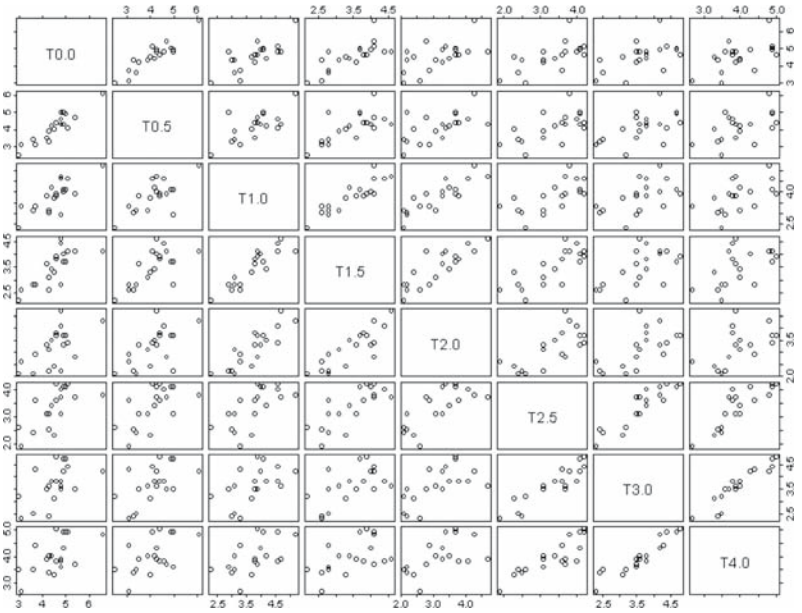

Figure 9.5 Scatterplot matrix for obese group in Table 9.2.

The first part of the rearranged data is shown in Table 9.6. We can fit the required model using

```
plasma.lme1<-lme(Plasma~Time+I(Time*Time)+Group,random
   =~Time|Subject,
data=plasma.df,method="ML")
summary(plasma.lme1)
```

The results are shown in Table 9.7. The regression coefficients for linear and quadratic time are both highly significant. The group effect just fails to reach significance at the 5% level. A confidence interval for the group effect is obtained from $0.38 \pm 2.04 \times 0.19$ giving $[-0.001, 0.767]$. (In S-PLUS the group effect and its standard error will be half those given in R corresponding to the group levels being coded by default as -1 and 1. This can be changed by use of the `contr.treatment` function.)

Here to demonstrate what happens if we make a very misleading assumption about the correlational structure of the repeated measurements, we will compare the results in Table 9.7 with those obtained if we assume that the repeated measurements are independent. The independence model can be fitted in the usual way with the `lm` function (see Chapter 8):

```
summary(lm(Plasma~Time+I(Time*Time)+Group,data=plasma.df))
```

The results are shown in Table 9.8. We see that under the independence assumption the standard error for the group effect is about one-half of that given in Table 9.7 and

Table 9.6 Part of Glucose Challenge Data in "Long" Form

	Subject	Time	Group	Plasma
1	1	0.0	Control	4.3
2	1	0.5	Control	3.3
3	1	1.0	Control	3.0
4	1	1.5	Control	2.6
5	1	2.0	Control	2.2
6	1	3.0	Control	2.5
7	1	4.0	Control	3.4
8	1	5.0	Control	4.4
9	2	0.0	Control	3.7
10	2	0.5	Control	2.6
11	2	1.0	Control	2.6
12	2	1.5	Control	1.9
13	2	2.0	Control	2.9
14	2	3.0	Control	3.2
15	2	4.0	Control	3.1
16	2	5.0	Control	3.9
17	3	0.0	Control	4.0
18	3	0.5	Control	4.1
19	3	1.0	Control	3.1
20	3	1.5	Control	2.3

if used would lead to the claim of strong evidence of a difference between control and obese patients.

We will now plot the predicted values from the fitted linear mixed effects model for each group using

```
predictions<-matrix(predict(plasma.lme1),ncol=8,byrow=T)
par(mfrow=c(1,2))
matplot(matrix(c(0.0,0.5,1,1.5,2,3,4,5),ncol=1),
t(predictions[1:13,]),type="l",lty=1,col=1,
xlab="Time (hours after glucose challenge)",ylab="Plasma
   inorganic phosphate",ylim=c(1,7))
title("Control")
matplot(matrix(c(0.0,0.5,1,1.5,2,3,4,5),ncol=1),
```

Table 9.7 Results from Random Slope and Intercept Model with Fixed Quadratic Time Effect Fitted to Glucose Challenge Data

Effect	Estimated reg coeff	SE	DF	t-value	p-value
Intercept	3.95	0.17	229	23.74	<0.0001
Time	−0.83	0.06	229	−13.34	<0.0001
Time^2	0.16	0.01	229	14.47	<0.0001
Group	0.38	0.19	31	2.03	0.051

$\hat{\sigma}_{u_1} = 0.61, \hat{\sigma}_{u_2} = 0.12, \hat{\sigma} = 0.42.$

Table 9.8 Results from Independence Model Fitted to Glucose Challenge Data

Effect	Estimated reg coeff	Sd S.E	t-value	p-value
Intercept	3.91	0.11	36.25	<0.0001
Time	−0.83	0.10	−8.65	<0.0001
Time^2	0.16	0.02	8.80	<0.0001
Group	0.46	0.09	5.24	<0.0001

```
t(predictions[14:33,]),type="l",lty=1,col=1,
xlab="Time (hours after glucose challenge)",ylab="Plasma
   inorganic phosphate",ylim=c(1,7))
title("Obese")
```

This gives Figure 9.6. We can see that the model has captured the profiles of the control group relatively well but not those of the obese group. We need to consider a further model that contains a group × time interaction.

The required model can be fitted and tested against the previous model using

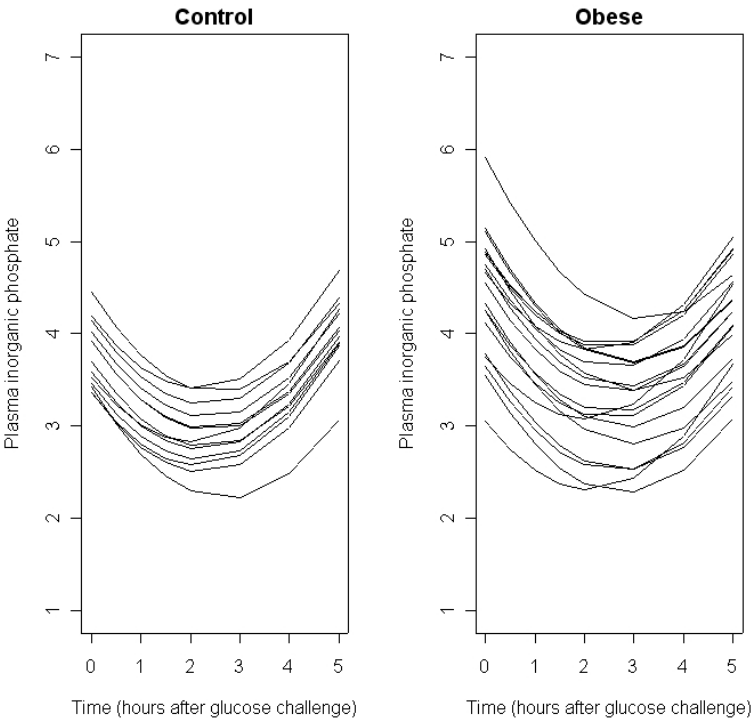

Figure 9.6 Fitted values from random intercept and slope model with fixed quadratic effect for glucose challenge data.

```
plasma.lme2<-lme(Plasma~Time*Group+I(Time*Time),random
   =~Time|Subject,
data=plasma.df,method="ML")
#
anova(plasma.lme1,plasma.lme2)
```

The p-value associated with the likelihood ratio test is 0.0011, indicating that the model containing the interaction term is to be preferred. The results for this model are given in Table 9.9. The interaction effect is highly significant. The fitted values from this model are shown in Figure 9.7 (the code is very similar to that given for producing Figure 9.6). The plot shows that the new model has produced predicted values that more accurately reflect the raw data plotted in Figure 9.3. The predicted profiles for the obese group are "flatter" as required.

We can check the assumptions of the final model fitted to the glucose challenge data, that is, the normality of the random effect terms and the residuals by first using the `random.effects` function to *predict* the former and the `resid` function 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 Display 9.2. The necessary R and S-PLUS code to obtain the effects, residuals and plots is as follows:

```
res.int<-random.effects(plasma.lme2)[,1]
res.int
res.slope<-random.effects(plasma.lme2)[,2]
par(mfrow=c(1,3))
qqnorm(res.int,ylab="Estimated random intercepts",
   main="Random intercepts")
qqnorm(res.slope,ylab="Estimated random slopes",
   main="Random slopes")
resids<-resid(plasma.lme2)
qqnorm(resids,ylab="Estimated residuals",main="Residuals")
```

The resulting plot is shown in Figure 9.8. The plot of the residuals is linear as required, but there is some slight deviation from linearity for each of the predicted random effects.

Table 9.9 Results from Random Intercept Slope and Model with Quadratic Time Effect and Group × Time Interaction Fitted to Glucose Challenge Data

Effect	Estimated reg coeff	SE	DF	t-value	p-value
Intercept	3.70	0.18	228	20.71	<0.0001
Time	−0.73	0.07	228	−10.90	<0.0001
Time^2	0.16	0.01	228	14.44	<0.0001
Group	0.81	0.22	31	3.60	0.0011
Group × time	−0.16	0.05	228	−3.51	0.0005

$\hat{\sigma}_{u_1} = 0.57, \hat{\sigma}_{u_2} = 0.09, \hat{\sigma} = 0.42.$

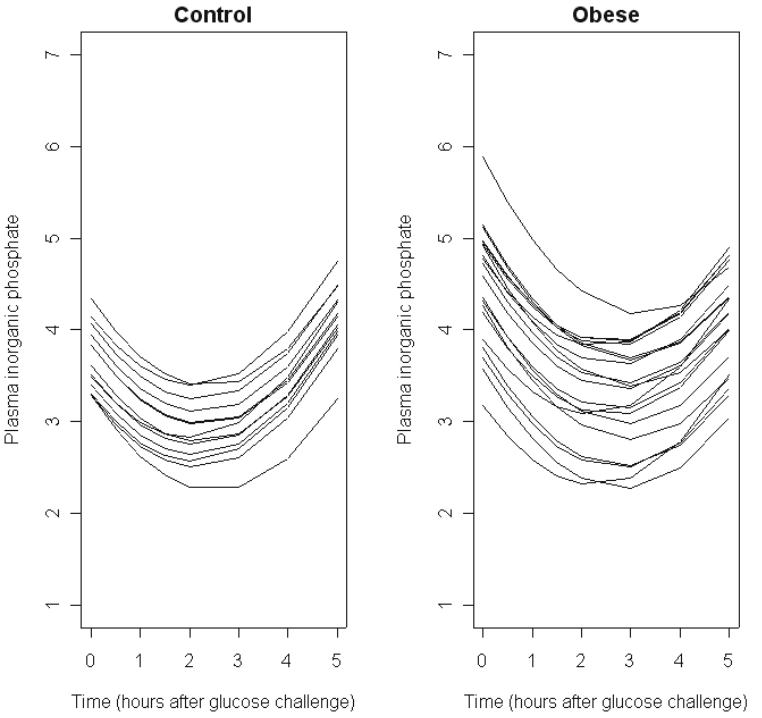

Figure 9.7 Fitted values from random intercept and slope model with fixed quadratic effect and group × time interaction for glucose challenge data.

Display 9.2
Prediction of Random Effects

- The random effects are not estimated as part of the model. However, having estimated the model, we can *predict* the values of the random effects.
- According to Bayes theorem, the *posterior probability* of the random effects is given by

$$\Pr(\mathbf{u}|\mathbf{y}, \mathbf{x}) = f(\mathbf{y}|\mathbf{u}, \mathbf{x})g(\mathbf{u}),$$

where $f(\mathbf{y}|\mathbf{u}, \mathbf{x})$ is the conditional density of the responses given the random effects and covariates (a product of normal densities) and $g(\mathbf{u})$ is the *prior* density of the random effects (multivariate normal). The means of this posterior distribution can be used as estimates of the random effects and are known as *empirical Bayes* estimates.

- The empirical Bayes estimator is also known as a shrinkage estimator because the predicted random effects are smaller in absolute value than their fixed-effect counterparts.
- *Best linear unbiased predictions* (BLUPs) are linear combinations of the responses that are unbiased estimators of the random effects and minimize the mean square error.

9.3 Dropouts in Longitudinal Data

A problem that frequently occurs when collecting longitudinal data is that some of the intended measurements are, for one reason or another, not made. In clinical trials, for example, some patients may miss one or more protocol scheduled visits after treatment has begun and so fail to have the required outcome measure taken. There will be other patients who do not complete the intended follow-up for some reason and drop out of the study before the end date specified in the protocol. Both situations result in missing values of the outcome measure. In the first case these are intermittent, but dropping out of the study implies that once an observation at a particular time point is missing so are all the remaining planned observations.

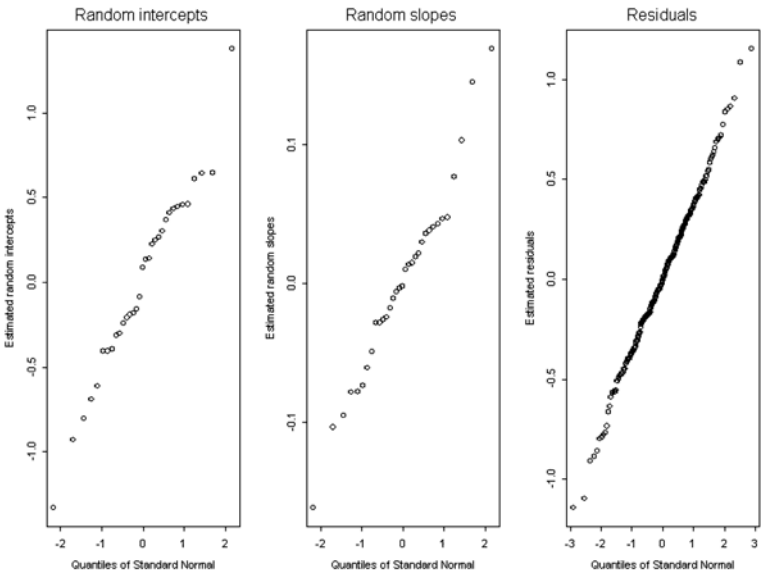

Figure 9.8 Probability plots of predicted random intercepts, random slopes, and residuals for final model fitted to glucose challenge data.

Many studies will contain missing values of both types, although in practice it is dropouts that cause most problems when turning to analyzing the resulting data set.

An example of a set of longitudinal data in which a number of patients have dropped out is given in Table 9.10. These data are essentially a subset of those collected in a clinical trial that is described in detail in Proudfoot et al. (2003). The trial was designed to assess the effectiveness of an interactive program using multimedia techniques for the delivery of cognitive behavioral therapy for depressed patients and known as Beating the Blues (BtB). In a randomized controlled trial of the program, patients with depression recruited in primary care were randomized to either the BtB program, or to Treatment as Usual (TAU). The outcome measure used in the trial was the Beck Depression Inventory II (Beck et al., 1996) with higher values indicating more depression. Measurements of this variable were made on five occasions, one prior to the start of treatment and at two monthly intervals after treatment began. In addition whether or not a participant in the trial was already taking antidepressant medication was noted along with the length of time they had been depressed.

To begin we shall graph the data here by plotting the boxplots of each of the five repeated measures separately for each treatment group. Assuming the data are available as the data frame `btb.data` the necessary code is

```
par(mfrow=c(2,1))
boxplot(btb.data[Treatment=="TAU",4],btb.data
   [Treatment=="TAU",5],btb.data[Treatment=="TAU",6],
btb.data[Treatment=="TAU",7],btb.data[Treatment=="TAU",8],
   names=c("BDIpre","BDI2m","BDI4m","BDI6m",
"BDI8m"),ylab="BDI",xlab="Visit",col=1)
title("TAU")
boxplot(btb.data[Treatment=="BtheB",4],btb.data
   [Treatment=="BtheB",5],btb.data[Treatment=="BtheB",6],
btb.data[Treatment=="BtheB",7],btb.data
   [Treatment=="BtheB",8],names=c("BDIpre","BDI2m","BDI4m",
"BDI6m","BDI8m"),ylab="BDI",xlab="Visit",col=1)
title("BtheB")
```

The resulting diagram is shown in Figure 9.9.

Figure 9.9 shows that there is decline in BDI values in both groups with perhaps the values in the BtheB group being lower at each postrandomization visit. We shall fit both random intercept and random intercept and slope models to the data including the pre-BDI values, treatment group, drugs, and length as fixed-effect covariates. First we need to rearrange the data into the long form using the following code:

```
n<-length(btb.data[,1])
#
BDI<-as.vector(t(btb.data[,c(5,6,7,8)]))
#
treat<-rep(btb.data[,3],rep(4,n))
subject<-rep(1:n,rep(4,n))
preBDI<-rep(btb.data[,4],rep(4,n))
drug<-rep(btb.data[,1],rep(4,n))
```

Table 9.10 Subset of Data from the Original BtB Trial

Sub	DRUG	Duration	Treatment	BDIpre	BDI2m	BDI3m	BDI5m	BDI8m
1	n	>6 m	TAU	29	2	2	NA	NA
2	y	>6 m	BtheB	32	16	24	17	20
3	y	<6 m	TAU	25	20	NA	NA	NA
4	n	>6 m	BtheB	21	17	16	10	9
5	y	>6 m	BtheB	26	23	NA	NA	NA
6	y	<6 m	BtheB	7	0	0	0	0
7	y	<6 m	TAU	17	7	7	3	7
8	n	>6 m	TAU	20	20	21	19	13
9	y	<6 m	BtheB	18	13	14	20	11
10	y	>6 m	BtheB	20	5	5	8	12
11	n	>6 m	TAU	30	32	24	12	2
12	y	<6 m	BtheB	49	35	NA	NA	NA
13	n	>6 m	TAU	26	27	23	NA	NA
14	y	>6 m	TAU	30	26	36	27	22
15	y	>6 m	BtheB	23	13	13	12	23
16	n	<6 m	TAU	16	13	3	2	0
17	n	>6 m	BtheB	30	30	29	NA	NA
18	n	<6 m	BtheB	13	8	8	7	6
19	n	>6 m	TAU	37	30	33	31	22
20	y	<6 m	BtheB	35	12	10	8	10
21	n	>6 m	BtheB	21	6	NA	NA	NA
22	n	<6 m	TAU	26	17	17	20	12
23	n	>6 m	TAU	29	22	10	NA	NA
24	n	>6 m	TAU	20	21	NA	NA	NA
25	n	>6 m	TAU	33	23	NA	NA	NA
26	n	>6 m	BtheB	19	12	13	NA	NA
27	y	<6 m	TAU	12	15	NA	NA	NA
28	y	>6 m	TAU	47	36	49	34	NA
29	y	>6 m	BtheB	36	6	0	0	2
30	n	<6 m	BtheB	10	8	6	3	3
31	n	<6 m	TAU	27	7	15	16	0
32	n	<6 m	BtheB	18	10	10	6	8
33	y	<6 m	BtheB	11	8	3	2	15
34	y	<6 m	BtheB	6	7	NA	NA	NA
35	y	>6 m	BtheB	44	24	20	29	14
36	n	<6 m	TAU	38	38	NA	NA	NA

(*Continued*)

Table 9.10 (*Continued*)

Sub	DRUG	Duration	Treatment	BDIpre	BDI2m	BDI3m	BDI5m	BDI8m
37	n	<6 m	TAU	21	14	20	1	8
38	y	>6 m	TAU	34	17	8	9	13
39	y	<6 m	BtheB	9	7	1	NA	NA
40	y	>6 m	TAU	38	27	19	20	30
41	y	<6 m	BtheB	46	40	NA	NA	NA
42	n	<6 m	TAU	20	19	18	19	18
43	y	>6 m	TAU	17	29	2	0	0
44	n	>6 m	BtheB	18	20	NA	NA	NA
45	y	>6 m	BtheB	42	1	8	10	6
46	n	<6 m	BtheB	30	30	NA	NA	NA
47	y	<6 m	BtheB	33	27	16	30	15
48	n	<6 m	BtheB	12	1	0	0	NA
49	y	<6 m	BtheB	2	5	NA	NA	NA
50	n	>6 m	TAU	36	42	49	47	40
51	n	<6 m	TAU	35	30	NA	NA	NA
52	n	<6 m	BtheB	23	20	NA	NA	NA
53	n	>6 m	TAU	31	48	38	38	37
54	y	<6 m	BtheB	8	5	7	NA	NA
55	y	<6 m	TAU	23	21	26	NA	NA
56	y	<6 m	BtheB	7	7	5	4	0
57	n	<6 m	TAU	14	13	14	NA	NA
58	n	<6 m	TAU	40	36	33	NA	NA
59	y	<6 m	BtheB	23	30	NA	NA	NA
60	n	>6 m	BtheB	14	3	NA	NA	NA
61	n	>6 m	TAU	22	20	16	24	16
62	n	>6 m	TAU	23	23	15	25	17
63	n	<6 m	TAU	15	7	13	13	NA
64	n	>6 m	TAU	8	12	11	26	NA
65	n	>6 m	BtheB	12	18	NA	NA	NA
66	n	>6 m	TAU	7	6	2	1	NA
67	y	<6 m	TAU	17	9	3	1	0
68	y	<6 m	BtheB	33	18	16	NA	NA
69	n	<6 m	TAU	27	20	NA	NA	NA
70	n	<6 m	BtheB	27	30	NA	NA	NA
71	n	<6 m	BtheB	9	6	10	1	0
72	n	>6 m	BtheB	40	30	12	NA	NA

(*Continued*)

Table 9.10 (*Continued*)

Sub	DRUG	Duration	Treatment	BDIpre	BDI2m	BDI3m	BDI5m	BDI8m
73	n	>6 m	TAU	11	8	7	NA	NA
74	n	<6 m	TAU	9	8	NA	NA	NA
75	n	>6 m	TAU	14	22	21	24	19
76	y	>6 m	BtheB	28	9	20	18	13
77	n	>6 m	BtheB	15	9	13	14	10
78	y	>6 m	BtheB	22	10	5	5	12
79	n	<6 m	TAU	23	9	NA	NA	NA
80	n	>6 m	TAU	21	22	24	23	22
81	n	>6 m	TAU	27	31	28	22	14
82	y	>6 m	BtheB	14	15	NA	NA	NA
83	n	>6 m	TAU	10	13	12	8	20
84	y	<6 m	TAU	21	9	6	7	1
85	y	>6 m	BtheB	46	36	53	NA	NA
86	n	>6 m	BtheB	36	14	7	15	15
87	y	>6 m	BtheB	23	17	NA	NA	NA
88	y	>6 m	TAU	35	0	6	0	1
89	y	<6 m	BtheB	33	13	13	10	8
90	n	<6 m	BtheB	19	4	27	1	2
91	n	<6 m	TAU	16	NA	NA	NA	NA
92	y	<6 m	BtheB	30	26	28	NA	NA
93	y	<6 m	BtheB	17	8	7	12	NA
94	n	>6 m	BtheB	19	4	3	3	3
95	n	>6 m	BtheB	16	11	4	2	3
96	y	>6 m	BtheB	16	16	10	10	8
97	y	<6 m	TAU	28	NA	NA	NA	NA
98	n	>6 m	BtheB	11	22	9	11	11
99	n	<6 m	TAU	13	5	5	0	6
100	y	<6 m	TAU	43	NA	NA	NA	NA

```
length<-rep(btb.data[,2],rep(4,n))
time<-rep(c(2,4,6,8),n)
#
#
btb.bdi<-data.frame(subject,treat,drug,length,preBDI,
   time,BDI)
#
attach(btb.bdi)
```

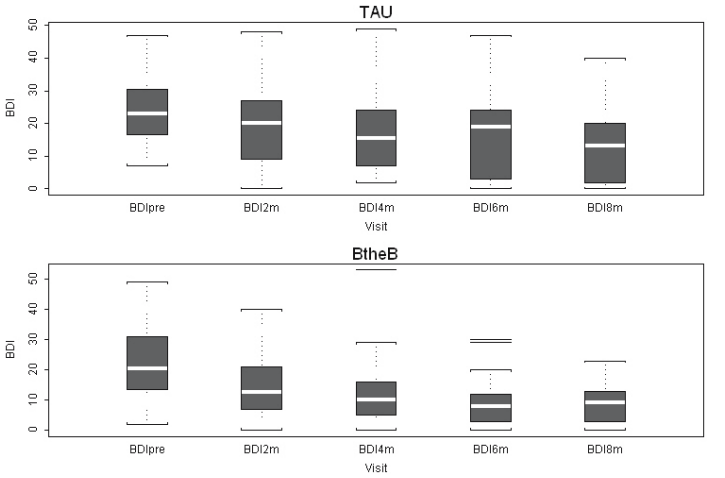

Figure 9.9 Boxplots for the repeated measures by treatment group for the BtheB data.

The resulting data frame `btb.bdi` contains a number of missing values and in applying the `lme` function these will need to 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. We can fit the two models and test which is most appropriate using

```
btbbdi.fit1 <- lme(BDI ~ preBDI + time + treat + drug
   + length, method = "ML", random
= ~ 1 | subject, data= btb.bdi, na.action = na.omit)
btbbdi.fit2 <- lme(BDI ~ preBDI + time + treat + drug
   + length, method = "ML", random
= ~ time | subject,data = btb.bdi, na.action = na.omit)
anova(btbbdi.fit1, btbbdi.fit2)
```

This results in

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
btbbdi.fit1	1	8	1886.624	1915.702	−935.3121			
btbbdi.fit2	2	10	1889.808	1926.156	−934.9040	1 vs 2	0.8160734	0.665

Clearly the simpler random intercept model is adequate for these data. The results from this model can be found using

```
summary(btbbdi.fit1)
```

and they are given in Table 9.11. Only time and the pre-BDI regression coefficients are significantly different from zero. In particular there is no occurring evidence of a treatment effect.

Table 9.11 Results from Random Intercept Model Fitted to BtheB Data

Effect	Estimated reg coeff	SE	DF	t-value	p-value
Intercept	5.94	2.10	182	2.27	0.0986
Pre BDI	0.64	0.08	92	8.14	<.0001
Time	−0.72	0.15	182	−4.86	<.0001
Treatment	−2.37	1.68	92	−1.41	0.1616
Drug	−2.80	1.74	92	−1.61	0.1110
Duration	0.26	1.65	92	0.16	0.8769

$\hat{\sigma}_{u_1} = 6.95, \hat{\sigma} = 5.01.$

We now need to consider briefly how the dropouts may affect the analyses reported above. To understand the problems that patients dropping out can cause for the analysis of data from a longitudinal trial we need to consider a classification of dropout mechanisms first introduced by Rubin (1976). The type of mechanism involved has implications for which approaches to analysis are suitable and which are not. Rubin's suggested classification involves three types of dropout mechanism:

- *Dropout completely at random* (DCAR): Here the probability that a patient drops out does not depend on either the observed or missing values of the response. Consequently the observed (nonmissing) values effectively constitute a simple random sample of the values for all subjects. Possible examples include missing laboratory measurements because of a dropped testtube (if it was not dropped because of the knowledge of any measurement), the accidental death of a participant in a study, or a participant moving to another area. Intermittent missing values in a longitudinal data set, whereby a patient misses a clinic visit for transitory reasons ("went shopping instead" or the like) can reasonably be assumed to be DCAR. Completely random dropout causes the least problem for data analysis, but it is a strong assumption.
- *Dropout at random* (DAR): The dropout-at-random mechanism occurs when the probability of dropping out depends on the outcome measures that have been observed in the past, but given this information is conditionally independent of all the future (unrecorded) values of the outcome variable following dropout. Here "missingness" depends only on the observed data with the distribution of future values for a subject who drops out at a particular time being the same as the distribution of the future values of a subject who remains in at that time, if they have the same covariates and the same past history of outcome up to and including the specific time point. Murray and Findlay (1988) provide an example of this type of missing value from a study of hypertensive drugs in which the outcome measure was diastolic blood pressure. The protocol of the study specified that the participant was to be removed from the study when his/her blood pressure got too large. Here blood pressure at the

time of dropout was observed before the participant dropped out, so although the dropout mechanism is not DCAR since it depends on the values of blood pressure, it *is* DAR, because dropout depends only on the observed part of the data. A further example of a DAR mechanism is provided by Heitjan (1997), and involves a study in which the response measure is body mass index (BMI). Suppose that the measure is missing because subjects who had high body mass index values at earlier visits avoided being measured at later visits out of embarrassment, regardless of whether they had gained or lost weight in the intervening period. The missing values here are DAR but *not* DCAR; consequently methods applied to the data that assumed the latter might give misleading results (see later discussion).

- *Nonignorable* (sometimes referred to as *informative*): The final type of dropout mechanism is one where the probability of dropping out depends on the unrecorded missing values—observations are likely to be missing when the outcome values that would have been observed had the patient not dropped out, are systematically higher or lower than usual (corresponding perhaps to their condition becoming worse or improving). A nonmedical example is when individuals with lower income levels or very high incomes are less likely to provide their personal income in an interview. In a medical setting possible examples are a participant dropping out of a longitudinal study when his/her blood pressure became too high and this value was not observed, or when their pain become intolerable and we did not record the associated pain value. For the BDI example introduced above, if subjects were more likely to avoid being measured if they had put on extra weight since the last visit, then the data are nonignorably missing. Dealing with data containing missing values that result from this type of dropout mechanism is difficult. The correct analyses for such data must estimate the dependence of the missingness probability on the missing values. Models and software that attempt this are available (see, e.g., Diggle and Kenward, 1994) but their use is not routine and, in addition, it must be remembered that the associated parameter estimates can be unreliable.

Under what type of dropout mechanism are the mixed effects models considered in this chapter valid? The good news is that such models can be shown to give valid results under the relatively weak assumption that the dropout mechanism is DAR (see Carpenter et al., 2002). When the missing values are thought to be informative, any analysis is potentially problematical. But Diggle and Kenward (1994) have developed a modeling framework for longitudinal data with informative dropouts, in which random or completely random dropout mechanisms are also included as explicit models.

The essential feature of the procedure is a logistic regression model for the probability of dropping out, in which the explanatory variables can include previous values of the response variable, and, in addition, the *unobserved* value at dropout as a *latent* variable (i.e., an unobserved variable). In other words, the dropout probability is allowed to depend on both the *observed* measurement history and the unobserved

value at dropout. This allows both a formal assessment of the type of dropout mechanism in the data, and the estimation of effects of interest, for example, treatment effects under different assumption about the dropout mechanism. A full account technical account of the model is given in Diggle and Kenward (1994) and a detailed example that uses the approach is described in Carpenter et al. (2002).

One of the problems for an investigator struggling to identify the dropout mechanism in a data set is that there are no routine methods to help, although a number of largely ad hoc graphical procedures can be used as described in Diggle (1998), Everitt (2002), and Carpenter (2002). Exercise 9.4 considers one of these.

9.4 Summary

Linear mixed effects models are extremely useful for modelling longitudinal data in particular and repeated measures data more generally. The models allow the correlations between the repeated measurements to be accounted for so that correct inferences can be drawn about the effects of covariates of interest on the repeated response values. In this chapter we have concentrated on responses that are continuous and conditional on the explanatory variables and random effects have a normal distribution. But random effects models can also be applied to nonnormal responses, for example, binary variables; see, for example, Everitt (2002).

The lack of independence of repeated measures data is what makes the modelling of such data a challenge. But even when only a single measurement of a response is involved, correlation can, in some circumstances, occur between the response values of different individuals and cause similar problems. As an example consider a randomized clinical trial in which subjects are recruited at multiple study centers. The multicenter design can help to provide adequate sample sizes and enhance the generalizability of the results. However factors that vary by center, including patient characteristics and medical practice patterns, may exert a sufficiently powerful effect to make inferences that ignore the "clustering" seriously misleading. Consequently it may be necessary to incorporate random effects for centers into the analysis.

Exercises

9.1 The final model fitted to the timber data did not constrain the fitted curves to go through the origin although this is clearly necessary. Fit an amended model where this constraint is satisfied and plot the new predicted values.

9.2 Investigate a further model for the glucose challenge data that allow a random quadratic effect.

9.3 Fit an independence model to the BtheB data and compare the estimated treatment effect confidence interval with that from the random intercept model described in the text.

9.4 Investigate whether there is any evidence of an interaction between treatment and time for the Beat the Blues data.

9.5 One very simple procedure for assessing the dropout mechanism suggested in Carpenter et al. (2002) involves plotting the observations for each treatment group, at each time point, differentiating between two categories of patients; those who do and those who do not attend their next scheduled visit. Any clear difference between the distributions of values for these two categories indicates that dropout is not completely at random. Produce such a plot for the Beat the Blues data.