1

Introduction

At the start of each chapter, we will 'set the scene' by outlining its content. In this introductory chapter, we start Section 1.1 by describing some situations where a mixed models analysis will be particularly helpful. In Section 1.2, we describe a simplified example and use it to illustrate the idea of a statistical model. We then introduce and compare fixed effects and random effects models. In the next section, we consider a more complex 'real-life' multi-centre trial and look at some of the variety of models that could be fitted (Section 1.3). This example will be used for several illustrative examples throughout the book. In Section 1.4, the use of mixed models to analyse a series of observations (repeated measures) is considered. Section 1.5 broadens the discussion on mixed models and looks at mixed models with a historical perspective of their use. In Section 1.6, we introduce some technical concepts: containment, balance and error strata.

We will assume in our presentation that the reader is already familiar with some of the basic statistical concepts as found in elementary statistical textbooks.

1.1 The use of mixed models

In the course of this book, we will encounter many situations in which a mixed models approach has advantages over the conventional type of analysis, which would be accessible via introductory texts on statistical analysis. Some of them are introduced in outline in this chapter and will be dealt in detail later on.

Example 1: Utilisation of incomplete information in a cross-over trial Cross-over trials are often utilised to assess treatment efficacy in chronic conditions, such as asthma. In such conditions, an individual patient can be tested for response to a succession of two or more treatments, giving the benefit of a 'within-patient' comparison. In the most commonly used cross-over design, just two treatments

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

are compared. If, for generality, we call these treatments A and B, then patients will be assessed either on their response to treatment A, followed by their response to treatment B, or vice versa. If all patients complete the trial, and both treatments are assessed, then the analysis is fairly straightforward. However, commonly, patients drop out during the trial and may have a valid observation from only the first treatment period. These incomplete observations cannot be utilised in a conventional analysis. In contrast, the use of a mixed model will allow all of the observations to be analysed, resulting in more accurate comparisons of the efficacy of treatment. This benefit, of more efficient use of the data, applies to all types of cross-over trial where there are missing data.

Example 2: Cross-over trials with fewer treatment periods than treatments In cross-over trials, for logistical reasons, it may be impractical to ask a patient to evaluate more than two treatments (e.g. if the treatment has to be given for several weeks). Nevertheless, there may be the need to evaluate three or more treatments. Special types of cross-over design can be used in this situation, but a simple analysis will be very inefficient. Mixed models provide a straightforward method of analysis, which fully uses the data, resulting again in more precise estimates of the effect of the treatments.

Example 3: A surgical audit A surgical audit is to be carried out to investigate how different hospitals compare in their rates of postoperative complications following a particular operation. As some hospitals carry out the operation commonly, while other hospitals perform the operation rarely, the accuracy with which the complication rates are estimated will vary considerably from hospital to hospital. Consequently, if the hospitals are ordered according to their complication rates, some may appear to be outliers compared with other hospitals, purely due to chance variation. When mixed models are used to analyse data of this type, the estimates of the complication rates are adjusted to allow for the number of operations, and rates based on small numbers become less extreme.

Example 4: Analysis of a multi-centre trial Many clinical trials are organised on a multi-centre basis, usually because there is an inadequate number of suitable patients in any single centre. The analysis of multi-centre trials often ignores the centres from which the data were obtained, making the implicit assumption that all centres are identical to one another. This assumption may sometimes be dangerously misleading. For example, a multi-centre trial comparing two surgical treatments for a condition could be expected to show major differences between centres. There could be two types of differences. First, the centres may differ in the overall success, averaged over the two surgical treatments. More importantly, there may be substantial differences in the relative benefit of the two treatments across different centres. Surgeons who have had more experience with one operation (A) may produce better outcomes with A, while surgeons with more experience with the alternative operation (B) may obtain better results

with B. Mixed models can provide an insightful analysis of such a trial by allowing for the extent to which treatment effects differ from centre to centre. Even when the difference between treatments can be assumed to be identical in all centres, a mixed model can improve the precision of the treatment estimates by taking appropriate account of the centres in the analysis.

Example 5: Repeated measurements over time In a clinical trial, the response to treatment is often assessed as a series of observations over time. For example, in a trial to assess the effect of a drug in reducing blood pressure, measurements might be taken at two, four, six and eight weeks after starting treatment. The analysis will usually be complicated by a number of patients failing to appear for some assessments or withdrawing from the study before it is complete. This complication can cause considerable difficulty in a conventional analysis. A mixed models analysis of such a study does not require complete data from all subjects. This results in more appropriate estimates of the effect of treatment and their standard errors (SEs). The mixed model also gives great flexibility in analysis, in that it can allow for a wide variety of ways in which the successive observations are correlated with one another.

1.2 Introductory example

We consider a very simple cross-over trial using artificial data. In this trial, each patient receives each of treatments A and B for a fixed period. At the end of each treatment period, a measurement is taken to assess the response to that treatment. In the analysis of such a trial, we commonly refer to treatments being *crossed* with patients, meaning that the categories of 'treatments' occur in combination with those of 'patients'. For the purpose of this illustration, we will suppose that the response to each treatment is unaffected by whether it is received in the first or second period. The table shows the results from the six patients in this trial.

	Treatment		Difference	
Patient	**A**	**B**	**A − B**	**Patient mean**
1	20	12	8	16.0
2	26	24	2	25.0
3	16	17	−1	16.5
4	29	21	8	25.0
5	22	21	1	21.5
6	24	17	7	20.5
Mean	22.83	18.67	4.17	20.75

1.2.1 Simple model to assess the effects of treatment (Model A)

We introduce in this section a very simple example of a statistical model using this data. A model can be thought of as an attempt to describe quantitatively the effect of a number of factors on each observation. Any model we describe is likely to be a gross oversimplification of reality. In developing models, we are seeking ones which are as simple as possible but which contain enough truth to ask questions of interest. In this first simple model, we will deliberately be oversimplistic in order to introduce our notation. We just describe the effect of the two treatments. The model may be expressed as

$$y_{ij} = \mu + t_j + e_{ij},$$

where

$j =$ A or B,
$y_{ij} =$ observation for treatment j on the ith patient,
$\mu =$ overall mean,
$t_j =$ effect of treatment j,
$e_{ij} =$ error for treatment j on the ith patient.

The constant μ represents the overall mean of the observations. $\mu + t_A$ corresponds to the mean in the treatment group A, while $\mu + t_B$ corresponds to the mean in the treatment group B. The constants μ, t_A and t_B can thus be estimated from the data. In our example, we can estimate the value of μ to be 20.75, the overall mean value. From the mean value in the first treatment group, we can estimate $\mu + t_A$ as 22.83, and hence our estimate of t_A is $22.83 - 20.75 = 2.08$. Similarly, from the mean of the second treatment group, we estimate t_B as -2.08. The term t_j can therefore be thought of as a measure of the relative effect that treatment j has had on our outcome variable.

The error term, e_{ij}, or *residual* is what remains for each patient in each period when $\mu + t_j$ is deducted from their observed measurement. This represents random variation about the mean value for each treatment. As such, the residuals can be regarded as the result of drawing random samples from a distribution. We will assume that the distribution is Gaussian or normal, with standard deviation σ, and that the samples drawn from the distribution are independent of each other. The mean of the distribution can be taken as zero, since any other value would simply cause a corresponding change in the value of μ. Thus, we will write this as

$$e_{ij} \sim \mathrm{N}(0, \sigma^2),$$

where σ^2 is the variance of the residuals. In practice, checks should be made to determine whether this assumption of normally distributed residuals is reasonable. Suitable checking methods will be considered in Section 2.4.6. As individual

observations are modelled as the sum of $\mu + t_j$, which are both constants, and the residual term, it follows that the variance of individual observations equals the residual variance:

$$\text{var}(y_{ij}) = \sigma^2.$$

The covariance of any two separate observations y_{ij} and $y_{i'j'}$ can be written as

$$\begin{aligned}\text{cov}(y_{ij}, y_{i'j'}) &= \text{cov}(\mu + t_i + e_{ij}, \quad \mu + t_{i'} + e_{i'j'}) \\ &= \text{cov}(e_{ij}, e_{i'j'}) \text{ (since other terms are constants).}\end{aligned}$$

Since all the residuals are assumed independent (i.e. uncorrelated), it follows that

$$\text{cov}(y_{ij}, y_{i'j'}) = 0.$$

The residual variance, σ^2, can be estimated using a standard technique known as analysis of variance (ANOVA). The essence of the method is that the total variation in the data is decomposed into components that are associated with possible causes of this variation, for example, that one treatment may be associated with higher observations, with the other being associated with lower observations. For this first model, using this technique, we obtain the following ANOVA table:

Source of variation	Degrees of freedom	Sums of squares	Mean square	*F*	*p*
Treatments	1	52.08	52.08	2.68	0.13
Residual	10	194.17	19.42		

Note: F = value for the F test (ratio of mean square for treatments to mean square for residual).
p = significance level corresponding to the F test.

The residual mean square of 19.42 is our estimate of the residual variance, σ^2, for this model. The key question often arising from this type of study is as follows: 'do the treatment effects differ significantly from each other?' This can be assessed by the F test, which assesses the null hypothesis of no mean difference between the treatments (the larger the treatment difference, the larger the treatment mean square and the higher the value of F). The p value of 0.13 is greater than the conventionally used cutoff point for statistical significance of 0.05. Therefore, we cannot conclude that the treatment effects are significantly different. The difference between the treatment effects and the SE of this difference provides a measure of the size of the treatment difference and the accuracy with which it is estimated:

$$\text{difference} = t_A - t_B = 2.08 + 2.08 = 4.16.$$

The SE of the difference is given by the formula

$$\begin{aligned}\mathrm{SE}(t_\mathrm{A} - t_\mathrm{B}) &= \sqrt{\sigma^2(1/n_\mathrm{A} + 1/n_\mathrm{B})} \\ &= \sqrt{(2 \times \sigma^2/6)} = \sqrt{6.47} = 2.54.\end{aligned}$$

Note that a t test can also be constructed from this difference and SE, giving $t = 4.16/2.54 = 1.63$. This is the square root of our F statistic of 2.68 and gives an identical t test p value of 0.13.

1.2.2 A model taking patient effects into account (Model B)

Model A as discussed previously did not utilise the fact that pairs of observations were taken on the same patients. It is possible, and indeed likely, that some patients will tend to have systematically higher measurements than others, and we may be able to improve the model by making allowance for this. This can be done by additionally including patient effects into the model:

$$y_{ij} = \mu + p_i + t_j + e_{ij},$$

where p_i are constants representing the ith patient effect.

The ANOVA table arising from this model is as follows:

Source of variation	Degrees of freedom	Sums of squares	Mean square	*F*	*p*
Patients	5	154.75	30.95	3.93	0.08
Treatments	1	52.08	52.08	6.61	0.05
Residual	5	39.42	7.88		

The estimate of the residual variance, σ^2, is now 7.88. It is lower than in Model A because it represents the 'within-patient' variation, as we have taken account of patient effects. The F test p value of 0.05 indicates that the treatment effects are now significantly different. The difference between the treatment effects is the same as in Model A, 4.16, but its SE is now as follows:

$$\mathrm{SE}(t_\mathrm{A} - t_\mathrm{B}) = \sqrt{(2 \times \sigma^2/6)} = \sqrt{2.63} = 1.62.$$

(Note that the SE of the treatment difference could alternatively have been obtained directly from the differences in patient observations.)

Model B is perhaps the 'obvious' one to think of for this dataset. However, even in this simple case, by comparison with Model A we can see that the statistical

modeller has some flexibility in his/her choice of model. In most situations, there is no single 'correct' model, and, in fact, models are rarely completely adequate. The job of the statistical modeller is to choose that model which most closely achieves the objectives of the study.

1.2.3 Random effects model (Model C)

In the Models A and B, the only assumption we made about variation was that the residuals were normally distributed. We did not assume that patient or treatment effects arose from a distribution. They were assumed to take constant values. These models can be described as *fixed effects models*, and all effects fitted within them are *fixed effects*.

An alternative approach available to us is to assume that some of the terms in the model, instead of taking constant values, are realisations of values from a probability distribution. If we assumed that patient effects also arose from independent samples from a normal distribution, then the model could be expressed as

$$\begin{aligned} y_{ij} &= \mu + p_i + t_j + e_{ij}, \\ e_{ij} &\sim \mathrm{N}(0, \sigma^2) \\ p_i &\sim \mathrm{N}(0, \sigma_{\mathrm{p}}^2). \end{aligned}$$

The p_i are now referred to as *random effects*. Such models, which contain a mixture of fixed and random effects, provide an example of a *mixed model*. In this book, we will meet several different types of mixed model, and we describe in Section 1.5 the common feature that distinguishes them from fixed effects models. To distinguish the class of models we have just met from those we will meet later, we will refer to this type of model as a *random effects model*.

Each random effect in the model gives rise to a *variance component*. This is a model parameter that quantifies random variation due to that effect only. In this model, the patient variance component is σ_{p}^2. We can describe variation at this level (between patients) as occurring within the patient *error stratum* (see Section 1.6 for a full description of the error stratum). This random variation occurs in addition to the residual variation (the residual variance can also be defined as a variance component.)

Defining the model in this way causes some differences in its statistical properties compared with the fixed effects model met earlier.

The variance of individual observations in a random effects model is the sum of all the variance components. Thus,

$$\mathrm{var}(y_{ij}) = \sigma_{\mathrm{p}}^2 + \sigma^2.$$

This contrasts with the fixed effects models where we had

$$\text{var}(y_{ij}) = \sigma^2.$$

The effect on the covariance of pairs of observations in the random effects model is interesting and perhaps surprising. Since $y_{ij} = \mu + p_i + t_j + e_{ij}$, we can write

$$\begin{aligned}\text{cov}(y_{ij}, y_{i'j'}) &= \text{cov}(\mu + p_i + t_j + e_{ij}, \mu + p_{i'} + t_{j'} + e_{i'j'}) \\ &= \text{cov}(p_i + e_{ij}, p_{i'} + e_{i'j'}).\end{aligned}$$

When observations from different patients are being considered (i.e. $i \neq i'$), because of the independence of the observations, $\text{cov}(y_{ij}, y_{i'j'}) = 0$. However, when two samples from the same patient are considered (i.e. $i = i'$), then

$$\begin{aligned}\text{cov}(y_{ij}, y_{i'j'}) &= \text{cov}(p_i + e_{ij}, p_i + e_{ij'}) \\ &= \text{cov}(p_i, p_i) = \sigma_{\text{p}}^2.\end{aligned}$$

Thus, observations on the same patient are correlated and have covariance equal to the patient variance component, while observations on different patients are uncorrelated. This contrasts with the fixed effects models where the covariance of any pair of observations is zero.

The ANOVA table for the random effects model is identical to that for the fixed effects model. However, we can now use it to calculate the patient variance component using results from the statistical theory that underpins the ANOVA method. The theory shows the expected values for each of the mean square terms in the ANOVA table, in terms of σ^2, σ_{p}^2 and the treatment effects. These are tabulated in the following table. We can now equate the expected value for the mean squares expressed in terms of the variance components to the observed values of the mean squares to obtain estimates of σ^2 and σ_{p}^2.

Source of variation	**Degrees of freedom**	**Sums of squares**	**Mean square**	**E(MS)**
Patients	5	154.75	30.95	$2\sigma_{\text{p}}^2 + \sigma^2$
Treatments	1	52.08	52.08	$\sigma^2 + 6\Sigma t_i^2$
Residual	5	39.42	7.88	σ^2

Note: E(MS) = expected mean square.

Thus, from the residual line in the ANOVA table, $\widehat{\sigma}^2 = 7.88$. In addition, by subtracting the third line of the table from the first we have:

$$2\widehat{\sigma}_{\text{p}}^2 = (30.95 - 7.88), \quad \text{and} \quad \widehat{\sigma}_{\text{p}}^2 = 11.54.$$

(We are introducing the notation $\widehat{\sigma}_{\text{p}}^2$ to denote that this is an estimate of the unknown σ_{p}^2, and $\widehat{\sigma}^2$ is an estimate of σ^2.)

In this example, we obtain identical treatment effect results to those for the fixed effects model (Model B). This occurs because we are, in effect, only using within-patient information to estimate the treatment effect (since all information on treatment occurs in the within-patient residual error stratum). Again, we obtain the treatment difference as −4.16 with a SE of 1.62. Thus, in this case, it makes no difference at all to our conclusions about treatments whether we fit patient effects as fixed or random. However, had any of the values in the dataset been missing, this would not have been the case. We now consider this situation.

Dataset with missing values

We will now consider analysing the dataset with two of the observations set to missing.

	Treatment		Difference	
Patient	**A**	**B**	**A − B**	**Patient mean**
1	20	12	8	16.0
2	26	24	2	25.0
3	16	17	−1	16.5
4	29	21	8	25.0
5	22	–	–	22.0
6	–	17	–	17.0
Mean			4.25	

As shown previously, there are two ways we can analyse the data. We can base our analysis on a model where the patient effects are regarded as fixed (Model B) or can regard patient effects as random (Model C).

The fixed effects model For this analysis, we apply ANOVA in the standard way, and the result of that analysis is summarised as follows:

Source of variation	**Degrees of freedom**	**Sums of squares**	**Mean square**	***F***	***p***
Patients	5	167.90	33.58	3.32	0.18
Treatments	1	36.13	36.13	3.57	0.16
Residual	3	30.38	10.12		

In the fitting of Model B, it is interesting to look at the contribution that the data from patient 5 are making to the analysis. The value of 22 gives us information that will allow us to estimate the level in that patient, but it tells us nothing at all

about the difference between the two treatments, nor does it even tell us anything about the effect of treatment A, which was received, because all the information in the observed value of 22 is used up in estimating the patient effect. The same comment applies to the data from patient 6.

Thus, in this fixed effects model, the estimate of the mean treatment difference, $\widehat{t}_{\mathrm{FE}}$, will be calculated only from the treatment differences for patients 1–4 who have complete data:

$$\widehat{t}_{\mathrm{FE}} = 4.25.$$

The variance of $\widehat{t}_{\mathrm{FE}}$ can be calculated from the residual variance, $\widehat{\sigma}^2 = 10.12$, as

$$\mathrm{var}(\widehat{t}_{\mathrm{FE}}) = \widehat{\sigma}^2(1/n_{\mathrm{p}} + 1/n_{\mathrm{p}}) = 10.12 \times (1/4 + 1/4) = 5.06,$$

where n_{p} is the number of observations with data on treatments A and B. The SE of the treatment difference is $\sqrt{5.06} = 2.25$.

The random effects model When patient effects are fitted as random, the variance components cannot be derived in a straightforward way from an ANOVA table since the data are unbalanced. They are found computationally (using `PROC MIXED`, a `SAS` procedure, which is described in more detail in Chapter 9) as

$$\widehat{\sigma}^2_{\mathrm{p}} = 12.63,$$
$$\widehat{\sigma}^2 = 8.90.$$

The treatment difference is estimated from the model to be 4.32, with a SE of 2.01. Thus, the SE is smaller than that of 2.25 obtained in the fixed effects model. This is not only due to a fortuitously lower estimate of σ^2, but also due to the fact that the random effects model utilises information on treatment from both the patient error stratum (between patients) and the residual stratum (within patients). As noted previously, the SE of the estimates is less than that in the fixed effects model, which only uses information from within patients. The use of this extra information compared with the fixed effects model can be referred to as the *recovery* of between-patient information.

In practice, we would recommend that random effects models are always fitted computationally using a procedure such as `PROC MIXED`. However, in our simple example given in this chapter, it may be of help to the understanding of the concept of recovery of information if we illustrate how the treatment estimates can be obtained manually.

Manual calculation In this example, the estimate of the treatment difference for the random effects model may be obtained by combining estimates from the between-patient and within-patient (residual) error strata. It is calculated by a weighted average of the two estimates, with the inverses of the variances of

the estimates used as weights. The within-patient estimate, $\widehat{t}_W$, is obtained as in the fixed effects model from patients 1–4 as 4.25. However, its variance is now calculated from the new estimate of σ^2 as

$$\text{var}(\widehat{t}_W) = \sigma^2(1/n_p + 1/n_p) = 8.90 \times (1/4 + 1/4) = 4.45.$$

The between-patient estimate, $\widehat{t}_B$, is simply the difference between the single values for patients 5 and 6

$$\widehat{t}_B = 22 - 17 = 5$$

and has variance as

$$\text{var}(\widehat{t}_B) = (\sigma^2 + \sigma_p^2) \times (1/1 + 1/1) = (8.90 + 12.63) \times 2 = 43.06.$$

The combined random effects model estimate, $\widehat{t}_{RE}$, is obtained as a weighted average of $\widehat{t}_W$ and $\widehat{t}_B$:

$$\widehat{t}_{RE} = K \times (\widehat{t}_W/\text{var}(\widehat{t}_W) + \widehat{t}_B/\text{var}(\widehat{t}_B)),$$

where

$$K = 1/(1/\text{var}(\widehat{t}_W) + 1/\text{var}(\widehat{t}_B)).$$

For our data,

$$K = 1/(1/4.45 + 1/43.06) = 4.03,$$

giving

$$\widehat{t}_{RE} = 4.03 \times (4.25/4.45 + 5/43.06) = 4.03 \times 1.07 = 4.32.$$

To calculate $\text{var}(\widehat{t}_{RE})$, we use the property $\text{var}(nx) = n^2\text{var}(x)$, so that

$$\text{var}(\widehat{t}_{RE}) = K^2 \times [\text{var}(\widehat{t}_W)/(\text{var}(\widehat{t}_W))^2 + \text{var}(\widehat{t}_B)/(\text{var}(\widehat{t}_B))^2],$$

giving

$$\begin{aligned}\text{var}(\widehat{t}_{RE}) &= K^2 \times (1/\text{var}(\widehat{t}_W) + 1/\text{var}(\widehat{t}_B)) \\ &= K.\end{aligned}$$

Thus, for our data:

$$\text{var}(\widehat{t}_{RE}) = 4.03,$$

and

$$\text{SE}(\widehat{t}_{RE}) = 2.01.$$

These results are identical to those obtained initially using `PROC MIXED`. However, it is not usually quite so simple to combine estimates manually from different error strata. A general formula for calculating fixed effects estimates for all types of mixed model will be given in Section 2.2.2.

The point that we hope has been made clear by the example is the way in which the random effects model has used the information from patients 5 and 6, which would have been lost in a fixed effects analysis.

1.2.4 Estimation (or prediction) of random effects

In the previous model, the patient terms were regarded as random effects. That is, they were defined as realisations of samples from a normal distribution, with mean equal to zero and with variance σ_p^2. Thus, their expected values are zero. We know, however, that patients may differ from one another, and the idea that all have the same expected value is counterintuitive. We resolve this paradox by attempting to determine for each individual patient a *prediction* of the location within the normal distribution from which that patient's observations have arisen. This prediction will be affected by that for all other patients and will differ from the corresponding estimate in the fixed effects model. The predictions will be less widely spread than the fixed effects estimates, and because of this, they are described as *shrunken*. The extent of this shrinkage depends on the relative sizes of the patient and residual variance components. In the extreme case where the estimate of the patient variance component is zero, all patients will have equal predictions. Shrinkage will also be relatively greater when there are fewer observations per patient. It occurs for both balanced and unbalanced data, and the relevant formula is given in Section 2.2.3. Although, on technical grounds, it is more accurate to refer to *predictions* of *random effects categories* (e.g. of individual patients), in this book, we will use the more colloquial form of expression and refer to estimates of patient effects.

In our example, using the complete trial data, the random effects estimates can be obtained computationally using `PROC MIXED`. They are listed as follows along with the fixed effects patient means.

Patient number	**1**	**2**	**3**	**4**	**5**	**6**
Fixed patients	16.0	25.0	16.5	25.0	21.5	20.5
Random patients	17.2	23.9	17.6	23.9	21.3	20.6

We observe that the mean estimates are indeed 'shrunken' towards the grand mean of 20.8. Shrinkage has occurred because patients are treated as a sample from the overall patient population.

1.3 A multi-centre hypertension trial

We now introduce a more complex 'real-life' clinical trial. Measurements from this trial will be used to provide data for several examples in future chapters. Although

it is by no means the only example we will be presenting, by the repeated use of this trial, we hope that the reader will identify more readily with the analyses.

The trial was a randomised, double blind comparison of three treatments for hypertension and has been reported by Hall *et al.* (1991). One treatment was a new drug (A), and the other two (B and C) were standard drugs for controlling hypertension (A = Carvedilol, B = Nifedipine, C = Atenolol). Twenty-nine centres participated in the trial, and patients were randomised in the order of entry. Two pre-treatment and four post-treatment visits were made as follows:

- Visit 1 (week 0): measurements were made to determine whether patients met the eligibility criteria for the trial. Patients who did so received a placebo treatment for 1 week, after which they returned for a second visit.
- Visit 2 (week 1): measurements were repeated, and patients who still satisfied the eligibility criteria were entered into the study and randomised to receive one of the three treatments.
- Visits 3–6 (weeks 3, 5, 7 and 9): measurements were repeated at four post-treatment visits, which occurred at 2-weekly intervals.
- Three hundred and eleven patients were assessed for entry into the study. Of these, 288 patients were suitable and were randomised to receive one of the three treatments. Thirty patients dropped out of the study prior to Visit 6.
- Measurements on cardiac function, laboratory values and adverse events were recorded at each visit. Diastolic blood pressure (DBP) was the primary endpoint, and we will consider its analysis in this section.
- The frequencies of patients attending at least one post-treatment visit at each of the 29 centres are shown in Table 1.1.

1.3.1 Modelling the data

The main purpose of this trial was to assess the effect of the three treatments on the primary endpoint, DBP recorded at the final visit. As in the previous example, we can do this by forming a statistical model. We will now describe several possible models. A simple model (Model A) to assess just the effects of treatment could be expressed as

$$\mathrm{DBP}_i = \mu + t_k + e_i,$$

where

$\mathrm{DBP}_i =$ diastolic blood pressure at final visit for patient i,
$\mu =$ intercept,
$t_k =$ kth treatment effect (where patient i has received treatment k),
$e_i =$ error term (residual) for the ith patient.

Before the model is fitted, we should be certain that we have the most relevant dataset for our objectives. In this trial, 30 patients dropped out of the study before their final visit. If treatments have influenced whether patients dropped out,

Table 1.1 Number of patients included in analyses of final visits by treatment and centre.

	Treatment			
Centre	**A**	**B**	**C**	**Total**
1	13	14	12	39
2	3	4	3	10
3	3	3	2	8
4	4	4	4	12
5	4	5	2	11
6	2	1	2	5
7	6	6	6	18
8	2	2	2	6
9	0	0	1	1
11	4	4	4	12
12	4	3	4	11
13	1	1	2	4
14	8	8	8	24
15	4	4	3	11
18	2	2	2	6
23	1	0	2	3
24	0	0	1	1
25	3	2	2	7
26	3	4	3	10
27	0	1	1	2
29	1	0	2	3
30	1	2	2	5
31	12	12	12	36
32	2	1	1	4
35	2	1	1	4
36	9	6	8	23
37	3	1	2	6
40	1	1	0	2
41	2	1	1	4
Total	100	91	94	288

Note: Several additional centres were numbered but did not eventually participate in the study.

omitting these patients from the analysis could give rise to biased estimates of treatment effects. We therefore adopt a 'last value carried forward' approach and substitute the last recorded value for the final visit values in these patients (the issue of how to deal with missing data will be considered again in Section 2.4.7.)

1.3.2 Including a baseline covariate (Model B)

Model A was a very simple model for assessing the effect of treatment on DBP. It is usually reasonable to assume that there may be some relationship between pre-treatment and post-treatment values on individual patients. Patients with relatively high DBP before treatment are likely to have higher values after treatment and likewise for patients with relatively low DBPs. We can utilise this information in the model by fitting the baseline (pre-treatment) DBP as an additional effect in Model A:

$$\text{DBP}_i = \mu + b \cdot pre + t_k + e_i,$$

where

$b =$ baseline covariate effect,
$pre =$ baseline (pre-treatment) DBP.

In this case, we will take the values recorded at visit 2 as the baseline values. We could, of course, have considered using either the visit 1 value or the average of the visit 1 and visit 2 values, instead. The visit 2 value was chosen because it measured the DBP immediately prior to randomisation, after 1 week, during which all patients received the same placebo medication. The baseline DBP is measured on a quantitative scale (unlike treatments). Such quantitative variables are commonly described as *covariate effects*, and an analysis based on the above model is often referred to as *analysis of covariance*. The term b is a constant that has to be estimated from our data. There is an implicit assumption in our model that the relationship between the final DBP and the baseline value is linear; Additionally, that within each treatment group, an increase of 1 unit in the baseline DBP is associated with an average increase of b units in the final DBP. Figure 1.1 shows the results from fitting this model to the data (only a sample of data points is shown, for clarity).

This demonstrates that performing an analysis of covariance is equivalent to fitting separate parallel lines for each treatment to the relationship between post-treatment DBP and baseline DBP. The separation between the lines represents the magnitude of the treatment effects. The analysis will be considered in much greater detail in Section 2.5, but we note for now that two of the treatments appear to be similar to one another, while the lowest post-treatment blood pressures occur with treatment C.

The use of a baseline covariate will usually improve the precision of the estimates of the treatment effects. It will also compensate for any differences between the mean levels of the covariate in the treatment groups prior to treatment being received. Of course, our assumption that there is a linear relationship between pre-treatment and post-treatment values may not be true. If this were the case, fitting a baseline covariate could lead to less precise results. However, in practice,

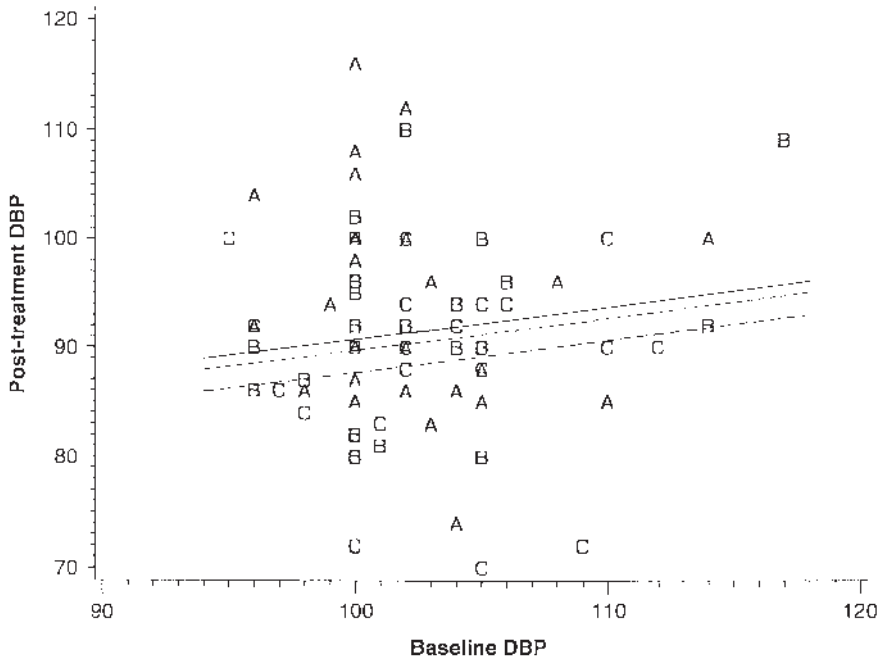

Figure 1.1 Plot to illustrate the analysis of covariance. Treatment: ———A; --------- B; – – – – – C.

the assumption is very frequently justified in medicine, and it has become almost standard to take baseline values into account in the model if they are available.

An alternative way of using baseline values (which we do not recommend) is to analyse the differences between pre-treatment and post-treatment values. However, this generally leads to less accurate results than the 'covariate' approach, particularly when the relationship between pre-treatment and post-treatment values is weak.

1.3.3 Modelling centre effects (Model C)

So far, the model has taken no account of the fact that the data are recorded at different centres. It is possible that values in some centres may tend to be higher than those in other centres. Such differences could be due, for example, to differences in the techniques of personnel across centres. It is also possible that some centres/clinics may recruit patients with differing degrees of severity of hypertension (within the bounds of the study entry criteria) who could, on average, have higher or lower values of DBP. We can allow for these possibilities by adding centre effects to Model B:

$$\mathrm{DBP}_i = \mu + b \cdot pre + t_k + c_j + e_i,$$

where

c_j = the jth centre effect.

Thus, part of the residual term in Model B may now be explained by the centre effects, c_j. If there are differences between the centres, this model will have a smaller residual variance than Model B (i.e. a smaller σ^2). This in turn allows treatment effects to be calculated with greater accuracy.

1.3.4 Including centre-by-treatment interaction effects (Model D)

In Model C, we took account of the fact that there may be an underlying difference in DBP between the centres. We did so in such a way that the effect of a patient being in a particular centre would be additive to the effect of treatment. Another possibility is that the response of patients to treatments may vary between the centres. That is, the effects of centre and treatment are non-additive or that there is an *interaction*. For example, in any multi-centre trial, if some centres tended to have more severely ill patients, it is plausible that the reaction of these patients to the treatments would differ from that of patients at other centres who are less severely ill. We can take this possibility into account in the model by allowing the treatment effects to vary between the centres. This is achieved by adding a centre·treatment *interaction* to Model C. It causes a separate set of treatment effects to be fitted for each centre.

$$\mathrm{DBP}_i = \mu + b \cdot pre + t_k + c_j + (ct)_{jk} + e_i,$$

where

$(ct)_{jk}$ = the kth treatment effect at the jth centre.

Throughout this book, we will refer to such interactions using the notation 'centre·treatment'. When Model D is fitted, the first question of interest is whether the centre·treatment effect is statistically significant. If the interaction term is significant, then we have evidence that the treatment effect differs between the centres. It will then be inadvisable to report the overall treatment effect across the centres. Results will need to be reported for each centre. If the interaction is not significant, centre·treatment may be removed from the model and the results from Model C reported. Further discussion on centre·treatment interactions appears in Chapter 5.

As we will see in more detail in Section 2.5, the centre·treatment effect is non-significant for our data ($p = 0.19$), and the results of Model C can be presented. Centre effects are statistically significant in Model C ($p = 0.004$), and so this model will be preferred to Model B.

From our data, b is estimated to be 0.22, with a SE of 0.11. Thus, if the baseline DBPs of two patients receiving the same treatment differ by 10 mm Hg, we can expect that their final DBPs will differ by only 2.2 mm Hg (0.22×10), as illustrated in Figure 1.1. The relationship is therefore weak, and hence we can anticipate

that the analysis of covariance approach will be preferable to a simple analysis of change in DBP. In fact, the statistical significance of the treatment differences is $p = 0.054$ using the analysis of covariance compared with $p = 0.072$ for the analysis of change.

1.3.5 Modelling centre and centre·treatment effects as random (Model E)

Models A–D can all be described as fixed effects models, and only the residual term is assumed to have a distribution. Alternatively, we could assume that the centre and centre·treatment effects also arose from a distribution. We again write the model as:

$$\mathrm{DBP}_i = \mu + b + t_k + c_j + (ct)_{jk} + e_i,$$

but now we assume that the residual, centre and centre·treatment effects are all realisations of separate distributions, all with zero means:

$$e_i \sim \mathrm{N}(0, \sigma^2),$$
$$c_j \sim \mathrm{N}(0, \sigma_{\mathrm{c}}^2),$$
$$(ct)_{jk} \sim \mathrm{N}(0, \sigma_{\mathrm{ct}}^2).$$

Hence, c_j and $(ct)_{jk}$ are now random effects, and b and t_k are fixed effects. This random effects model can be described as *hierarchical* since treatment effects are *contained* within the random centre·treatment effects. The concept of containment will be picked up again in Section 1.6.

Since we have assumed that centre·treatment effects have a distribution, that is that differences between treatments vary randomly across the centres, we can relate our results to the population of potential centres. This is in contrast to Model D, where treatment effects are assumed to be specific to the centres observed.

There are no hard and fast rules about whether effects should be modelled as fixed or random (or indeed whether some effects should be fitted at all). In this case, various approaches are acceptable, but they offer us different interpretations of the results. These various approaches will be discussed in much greater detail in Section 2.5, but for now, we pick up on just one point: the precision with which treatment effects are estimated. We have seen previously that fitting centre and centre·treatment effects as random enables our inferences to apply to a 'population' of centres. There is a price to be paid, however. The SEs of the treatment estimates will be inflated because we allow the treatment effects to vary randomly across centres. Thus, the mean difference in final DBP between treatments A and C is estimated as 2.93 mm Hg, with a SE of 1.41 mm Hg. In contrast, using Model C, the corresponding estimate is 2.99 mm Hg, with a smaller SE of 1.23 mm Hg. Arguments in favour of the random effects model are

the wider scope of the inferences and perhaps a more appropriate modelling of the data. In some circumstances, however, it is adequate to establish treatment differences in a specific set of centres. Statisticians in the pharmaceutical industry, for example, may prefer to avoid the penalty of less precise treatment estimates, with a corresponding reduction in *power* (the probability of obtaining statistically significant treatment differences when treatments do differ in their effect) and will often use a fixed effects model. This discussion point will be taken up again in Chapter 5.

1.4 Repeated measures data

There were four post-treatment visits in the multi-centre hypertension trial introduced in the previous section. However, so far in this chapter, we have chosen only to model measurements made at the final visit, which were of primary interest. An alternative strategy would be to include measurements from all four post-treatment visits in the model. Since measurements are made repeatedly on the same patients, we can describe these types of data as *repeated measures* data. For illustrative purposes, we now assume that the centre has no effect at all on the results and consider which models are appropriate for analysing repeated measures data. The mean levels for the three treatments at all time points are shown in Figure 1.2.

1.4.1 Covariance pattern models

Again, our primary objective is to assess the effect of the treatments on DBP, and we might again consider models which fit treatment and baseline DBP as in Model B in Section 1.3. The models will, of necessity, be more complicated, as we now have four observations per patient. In addition, it is possible that there is an underlying change in DBP over the four post-randomisation visits, and we can allow for this in the model by including a time effect, which we will denote by m. It is also possible that treatment effects may differ across time points, and to allow for this, we can also include a treatment-by-time interaction, (tm). Thus, the jth observation on patient i can be modelled as:

$$\text{DBP}_{ij} = \mu + b \cdot pre + t_k + m_j + (tm)_{jk} + e_{ij},$$

where

$m_j =$ time effect at the jth post-treatment visit,
$(tm)_{jk} =$ the kth treatment effect at the jth post-treatment visit,
$e_{ij} =$ residual term for the ith patient at the jth post-treatment visit.

So far, in developing this model, we have taken no account of the fact that post-treatment measurements taken on the same patient may not be independent

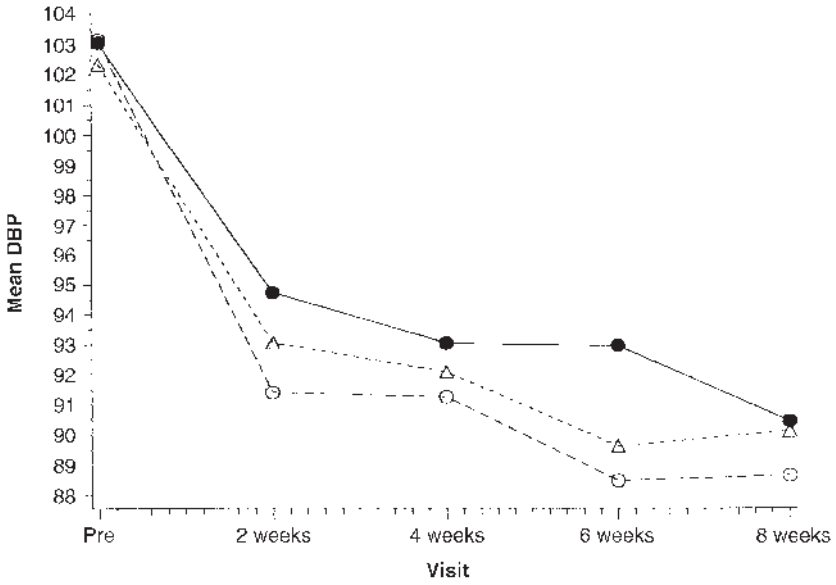

Figure 1.2 Plot of mean DBP by treatment group and visit. Treatment: ———A; ---------- B; – – – – – C.

of one another. A straightforward way to do this would be to assume that there is a constant correlation for all pairs of measurements on the same patient. Then, we could write the correlation between the residuals as

$$\text{corr}(e_{ij}, e_{ij'}) = \rho, \quad j \neq j'.$$

Alternatively, it is possible that the correlation between pairs of measurements decays as they become more widely separated in time. We could then write

$$\text{corr}(e_{ij}, e_{ij'}) = \rho^{|j'-j|}, \quad j \neq j'.$$

In the extreme, we can set a separate correlation for each pair of visits and may write

$$\text{corr}(e_{ij}, e_{ij'}) = \rho_{j,j'}, \quad j \neq j'.$$

A *covariance pattern model* can be used to fit any of these covariance (or correlation) patterns. This type of model forms another class of mixed models. Fitting covariance patterns leads to a more appropriate analysis than occurs when the fact that the repeated observations are correlated is ignored. The covariance parameter estimates may also uncover additional information about the data. They are considered in more detail in Section 6.2, and the analysis of this example is presented in Section 6.3.

1.4.2 Random coefficients models

In the previous section, the pattern of covariance between the repeated observations was modelled. An alternative approach to modelling repeated measures data would be to devise a model that explained arithmetically the relationship between DBP and time. A very simple way to do this would be to include a quantitative time effect (e.g. in measured weeks) as a covariate in the model.

$$\text{DBP}_{ij} = \mu + b \cdot pre + t_k + m \cdot time_{ij} + e_{ij},$$

where

$time_{ij} =$ time of observation j for patient i (weeks),
$m =$ constant representing the change in DBP for unit time (week).

Thus, we obtain a time slope with gradient m, which defines a linear relationship between DBP and time. It is also possible (and indeed likely) that the relationship between DBP and time would vary between patients. To allow for this, we could model a separate regression of DBP on time for each patient. To do this, we fit patient effects to provide the intercept terms for each patient and a patient·time interaction to provide the slopes for each patient.

$$\text{DBP}_{ij} = \mu + b \cdot pre + t_k + p_i + m \cdot time_{ij} + (pm)_i \cdot time_{ij} + e_{ij},$$

where

$(pm)_i =$ difference in slope for the ith patient from the average slope,
$p_i =$ difference from average in the intercept term for the ith patient.

It would seem reasonable to regard the values of patient effects and their slopes against time as arising from a distribution. Thus, patient and patient·time effects can both be fitted as random effects. However, the statistical properties of a model where some of the random effects involve covariate terms (time in this example) differ from ordinary random effects models (where the random effects do not involve any covariates). For this reason, we distinguish these models from ordinary random effects models and refer to them as *random coefficients models*. They form a third class of mixed models.

The statistical properties of random coefficients models are similar in many respects to random effects models. The residuals again are assumed to be independent and to have a normal distribution, with zero mean:

$$\text{var}(e_{ij}) = \sigma^2.$$

The main statistical difference from ordinary random effects models arises from the fact that when we fit a straight line, the estimates of the slope and the intercept are not independent. Thus, the patient effects (intercepts) and patient·time effects (slopes) are correlated within each patient. We therefore need to extend the approach met earlier, where separate normal distributions were used for

each random effect. We do this by use of the bivariate normal distribution. As well as terms for the means of both effects (which, as usual, are zero) and the variance components σ^2_{p} and σ^2_{pm} for patients and patient·time, this incorporates a covariance parameter $\sigma_{\mathrm{p,pm}}$. We denote the bivariate normal distribution as

$$\begin{pmatrix} p_i \\ pm_i \end{pmatrix} \sim \mathrm{N}(\mathbf{0}, \mathbf{G}),$$

where

$$\mathbf{G} = \begin{pmatrix} \sigma^2_{\mathrm{p}} & \sigma_{p,pm} \\ \sigma_{p,pm} & \sigma^2_{pm} \end{pmatrix}.$$

Thus, repeated measures data can be modelled using two alternative types of mixed model. Either the pattern of covariance between the repeated observations is modelled using a covariance pattern model or the relationship with time can be modelled using a random coefficients model. The latter approach is usually more appropriate if the repeated measurements do not occur at fixed intervals or when the relationship with time is of particular interest.

1.5 More about mixed models

In Sections 1.2–1.4, we used examples to introduce various concepts and types of mixed models. In this section, we pull together some of the ideas introduced earlier and define them more concisely. We also discuss some general points about mixed models. Finally, we present a perspective of mixed models, giving an outline of the history of their development.

1.5.1 What is a mixed model?

We have already met a number of models that have been described as mixed models, but it may not be clear what unites them. The key distinguishing feature of mixed models compared with fixed effects models is that they are able to model data in which the observations are not independent. To express this more positively, we say that mixed models are able to model the covariance structure of the data.

A simple type of mixed model is the *random effects model*, which was introduced in Sections 1.2 and 1.3. Here, certain effects in the model are assumed to have arisen from a distribution and thus give rise to another source of random variation in addition to the residual variation. These effects are referred to as *random effects*. For example, when patient effects were fitted in the trial introduced in Section 1.2, random variation occurred both between patients and as residual variation. Any number of random effects can be specified in a model; for example, in a multi-centre trial (as in Section 1.3), both centre and centre·treatment effects can be fitted as random, giving rise to two additional sources of variation.

In *random coefficients models*, a covariate effect is allowed to vary randomly. For example, in the repeated measures hypertension data considered in Section 1.4, interest might centre on the rate of change of DBP measured over the four treatment visits in the three arms of the trial. The random coefficients model allows this rate of change (or slope) to vary randomly between patients. This is achieved technically by fitting patients and the patient·slope interaction as random, and these effects are referred to as *random coefficients*.

The *covariance pattern model*, introduced in Section 1.4, is a third type of mixed model that directly models a pattern of correlations between observations. For example, in repeated measures trials, interest is focused on several observations of the response variable made over a period, and we can allow for the correlations (or, equivalently, covariances) between these observations. Suitable mixed models lead to more appropriate estimates of fixed effects and can investigate the nature of these covariances.

Random effects models, random coefficients models and covariance pattern models form three categories of mixed models. Mixed models can also be defined with combinations of random effects, random coefficient effects and covariance patterns. The choice will depend on the application and the objectives of the analysis.

1.5.2 Why use mixed models?

To stimulate further interest, we now mention some potential advantages that can be gained by using a mixed model. In some situations, a mixed model may simply be the most plausible model for a given data structure. For example, it is clearly desirable to take account of correlations between measurements in repeated measures data. In other circumstances, the choice is less obvious between a fixed effects model and a mixed model. Factors influencing the decision will depend partly on the structure of the data. For example, in a multi-centre trial (as in Section 1.3), the decision depends mainly on the interpretation to be put on the results. When centre and centre·treatment effects are fitted as fixed, inference can only formally be applied to the centres observed, but if they are fitted as random, inference can be applied with more confidence to a wider population of centres.

Some potential advantages that can be gained by using a mixed model are as follows:

- *Fitting covariance pattern models leads to more appropriate fixed effects estimates and SEs*. This type of model is of particular use for analysing repeated measures data. An important advantage is that the presence of missing data does not pose the major problems for analysis that can occur with a traditional analysis. The covariance parameter estimates may also uncover additional information about the data.
- *Results from a mixed model may be more appropriate to the required inference when the data structure is hierarchical*. For example, by fitting centre·treatment effects

as random in a multi-centre trial analysis (as in Section 1.3), treatment effects are allowed to vary randomly across centres, and the treatment SE increases to allow for this. Inference can then be applied to the full population of centres. However, if centre and centre·treatment effects were fitted as fixed, treatment effects would be specific to the centres observed, and inference should only be applied to these centres.

- *In a cross-over trial, estimates of treatment effects can become more accurate in datasets where there are missing data* (as in Section 1.2). The degree of benefit from using a mixed model in this situation will depend on the amount of missing data. If the original trial design was balanced and only occasional values were missing, there would be little to be gained. However, if several values were missing, treatment estimates could become notably more accurate.
- *In a random effects model, estimates of random effects are 'shrunken' compared with their fixed effects counterparts*. That is, their mean values are closer to the overall mean than if they were fitted as fixed. This helps to avoid the potential problem of extreme parameter estimates occurring due to chance when the estimates are based on small numbers. For example, in Section 1.1, we introduced an example on surgical audit. If failure rates from a particular type of operation were measured at several hospitals, a model fitting hospitals as fixed would produce unreliable failure rates for hospitals performing a small number of operations. Sometimes, these would appear as outliers compared with other hospitals, purely due to chance variation. A model fitting hospitals as random would estimate failure rates that were shrunken towards the overall failure rate. The shrinkage is greatest for hospitals performing fewer operations because less is known about them, and so misleading outliers are avoided.
- *Different variances can be fitted in a mixed model for each treatment group*. Such different variances for the treatment groups often arise in clinical trials comparing active treatments with a placebo, but they are rarely accounted for in fixed effects analyses.
- *Problems caused by missing data when fitting fixed effects models do not arise in mixed models*, provided that missing data can be assumed missing at random. This applies particularly in repeated measures trials, as noted previously, and in cross-over trials.

Although we have listed several advantages to mixed models, there is a potential disadvantage. This is that more distributional assumptions are made, and approximations are used to estimate certain model parameters. Consequently, the conclusions are dependent on more assumptions being valid, and there will be some circumstances where parameter estimates are biased. These difficulties are addressed in Section 2.4.

1.5.3 Communicating results

Statistical methods have been defined as those which elucidate data affected by a multiplicity of causes. A problem with methods of increasing complexity can be

difficulty in communicating the results of the analysis to the practitioner. There is the danger of obfuscating rather than elucidating. Estimation methods for mixed models are more complex than those used for fixed effects models, and results can therefore be more difficult to justify to non-statistical colleagues. It is not usually realistic to describe the exact methodology. However, a satisfactory explanation can often be given by emphasising the key point that mixed models take account of the covariance structure or interdependence of the data, whereas more conventional fixed effects methods assume that all observations are independent. Mixed models may therefore provide results that are more appropriate to the study design. A (hypothetical) statistical methods section in a medical journal might read:

> *The trial was analysed using a mixed model (see Brown and Prescott, 2015) with centres and the centre·treatment interaction fitted as random, so that possible differences in the size of the treatment effect across centres could be assessed.*

1.5.4 Mixed models in medicine

Frequently, there are advantages to be gained from using mixed models in medical applications. Data in medical studies are often clustered; for example, data may be recorded at several centres, hospitals or general practices. This design can be described as hierarchical, and wider inferences can be made by fitting the clustering effect as random. Repeated measures designs are also often used in medicine, and it is not uncommon for some of the observations to be missing. There are then advantages to be gained from using a mixed models analysis, which makes allowance for the missing data. Another consideration is that it is ethically desirable to use as few patients as possible, and therefore any improvements in the accuracy of treatment estimates gained by using a mixed model are particularly important. Although several examples of using mixed models in medicine have appeared in the literature for some time (e.g. Brown and Kempton, 1994), their use is still in the process of becoming routine.

1.5.5 Mixed models in perspective

It is interesting to see the application of mixed models in its historical context. In doing so, we will have to use occasional technical terms that have not yet been introduced in this book. They will, however, be met later on, and readers for whom some of the terms are unfamiliar may wish to return to this section after reading subsequent chapters.

The idea of attributing random variation to different sources by fitting random effects is not new. Fisher (1925), in his book *Statistical Methods for Research Workers*, outlined the basic method for estimating variance components by equating

the mean squares from an ANOVA table to their expected values (as described in Section 1.2). However, this method was only appropriate for balanced data. Yates (1940) and Henderson (1953) showed how Fisher's technique could be extended to unbalanced data, but their method did not always lead to unique variance components estimates. Hartley and Rao (1967) showed that unique estimates could be obtained using the method of maximum likelihood (see Section 2.2.1 for details on maximum likelihood). However, the estimates of the variance components are generally biased downwards because the method assumes that the fixed effects are known, rather than being estimated from the data. This problem of bias was overcome by Patterson and Thompson (1971) who proposed a method known as residual maximum likelihood (REML) (see Section 2.2.1), which automatically adjusted for the degrees of freedom corresponding to estimated fixed effects, as does ANOVA for balanced data. Many of the methods we describe in this book will be based on the REML method. Likelihood-based methods have only been adopted slowly because they are computationally intensive, and this has limited their use until recently.

In the past 30 years, there have been developments in parallel, in the theory and practice of using the different types of mixed model that we described earlier. Random coefficients problems have sometimes in the past been handled in two stages: first, by estimating time slopes for each patient and then by performing an analysis of the time slopes (e.g. Rowell and Walters, 1976). An early theoretical article describing the fitting of a random coefficients model in a single stage, as we will do in this book, is by Laird and Ware (1982). We consider random coefficients models again in Section 6.5.

Covariance pattern models have developed largely from time series models. Jennrich and Schluchter (1986) described the use of different covariance pattern models for analysing repeated measures data and gave some indication of how to choose between them. These models are considered more in detail in Section 6.2.

Random effects models have been frequently applied in agriculture. They have been used extensively in animal breeding to estimate heritabilities and predict genetic gain from breeding programmes (Meyer, 1986; Thompson, 1977). They have also been used for analysing crop variety trials. For example, Talbot (1984) used random effects models to estimate variance components for variety trailing systems carried out across several centres and years for different crops and was thus able to compare their general precision and effectiveness. The adoption of these models in medicine has been much slower, and a review of applications in clinical trials was given by Brown and Kempton (1994). Since then, there has been an increasing acceptability of these methods, not only by medical statisticians, but also by the regulatory authorities. The Food and Drug Administration (FDA) website contains, for example, recommended code using `SAS` to fit mixed models to multi-period cross-over trials to establish bioequivalence (www.fda.gov). Analyses of such designs are considered in Section 8.15, and other cross-over designs are considered in Chapter 7.

More recently, mixed models have become popular in the social sciences. However, they are usually described as multi-level or hierarchical models, and the terminology used for defining the models differs from that used in this book. This reflects parallel developments in different areas of application. However, the basic concept of allowing the data to have a covariance structure is the same. Two books published in this area are *Multilevel Statistical Models, Fourth Edition* by Goldstein (2010) and *Random Coefficients Models* by Longford (1993).

Perhaps, the biggest change in the use of mixed models in recent years has been the increasing use of Bayesian methods. Historically, the dual problems of computational power and available software have been a factor in restricting the use of the Bayesian approach to analysis. While this approach is based on a different philosophy, it will often lead to superficially similar results to a conventional random effects model when used with uninformative priors. The increasing availability of good software to implement the Bayesian approach and, in particular, the implementation in `SAS` of `PROC MCMC` will undoubtedly lead to its wider use in future. There has also been a shift in terminology to make the methods more acceptable to statisticians who may distrust Bayesian methods by referring to them as simulation methods. Indeed, with flat priors, you are obtaining a simulation of the full likelihood. The Bayesian approach to modelling is considered in Section 2.3.

The expansion of interest in mixed models is illustrated by its wider coverage in undergraduate and postgraduate courses in statistics and the accompanying increase in books on the topic. These include *Linear Mixed Models for Longitudinal Data by* Verbeke and Molenberghs (2000), *Generalized, Linear, and Mixed Models* by McCulloch *et al.* (2008), *Linear Mixed Models: A Practical Guide Using Statistical Software* by West *et al.* (2006), and *Mixed Models: Theory and Applications with R* by Demidenko (2013).

1.6 Some useful definitions

We conclude this introductory chapter with some definitions. The terms we are introducing in this chapter will recur frequently within subsequent chapters, and the understanding of these definitions and their relevance should increase as their applications are seen in greater detail. The terms we will introduce are containment, balance and error strata. In the analyses we will be presenting, we usually wish to concentrate on estimates of treatment effects. With the help of the definitions we are introducing, we will be able to distinguish between situations where the treatment estimates are identical whether fixed effects models or mixed models are fitted. We will also be able to identify the situations where the treatment estimates will coincide with the simple average calculated from all observations involving that treatment. The first term we need to define is containment.

1.6.1 Containment

Containment occurs in two situations. First, consider the repeated measures data encountered in Section 1.4. In that hypertension trial, DBP was recorded at four visits after treatment had been started. In the analysis of that study, the residual variance will reflect variation *within* patients at individual visits. However, in this trial, the patients receive the same treatment throughout, and so all the observations on a patient will reflect the effect of that one treatment on the patient. It can therefore perhaps be appreciated intuitively that it is the variation in response *between* patients, which is appropriate for assessing the accuracy of the estimates of treatment effects rather than the residual or 'within-patient' variation. We can see this more dramatically with a totally artificial set of data which might have arisen from this trial.

		Post-treatment visits			
Patient	**Treatment**	**1**	**2**	**3**	**4**
1	A	80	80	80	80
2	B	85	85	85	85
3	B	85	85	85	85
4	A	91	91	91	91

In this situation, there is no within-patient variation, and the residual variance is zero. Thus, if the residual variance were used in the determination of the precision of treatment estimates, we would conclude that these data showed convincingly that treatment B produced lower DBPs than treatment A. Common sense tells us that this conclusion is ridiculous with these data and that between-patient variation must form the basis for any comparison.

Here, we say that treatment effects are contained within-patient effects.

The second situation where we can meet containment can also be illustrated with data from the hypertension trial, this time concentrating on the multi-centre aspect of the design. In Section 1.3, we actually met containment for the first time when dealing with Model E, and both centre effects and centre·treatment effects were fitted as random. We say in this context that the treatment effects are contained within centre·treatment effects. In fact, there is no requirement for the centre·treatment effects to be random for the definition of containment to hold. Thus, similarly, in Model D, where the centre·treatment effects were regarded as fixed, we can still refer to the treatment effects as being contained within centre·treatment effects. It applies in general to any data with a hierarchical

structure in which the fixed effects (treatment) appears in interaction terms with other effects.

1.6.2 Balance

In many statistical textbooks that discuss the concept of balance, it is never defined but, rather, left to the intuitive feel of the reader to determine whether an experimental design is balanced. Some authors (e.g. Searle *et al.*, 1992) have defined balance as occurring when there are equal numbers of observations per *cell*. Cells are formed by all possible combinations of the levels of all the effects in the model, otherwise known as the crossing between all effects fitted in the model. For example, if we fit centre effects and treatment effects in the analysis of a multi-centre trial, and we suppose that there are four centres and two treatments, then each of the eight combinations of centre and treatment requires the same number of patients to achieve balance.

When there is balance according to this definition, the estimate of a fixed effects mean will equal the mean of all the observations at that fixed effects level. To make this clearer, if we call one of the treatments A, then the estimate of the mean response to treatment A will simply be the average of all of the observations for all patients who received treatment A. In general, this will not happen when there is imbalance. Consider the dataset illustrated in the following section. If all of the observations are present, then the estimated means for treatments A and B are 85.0 and 95.0, respectively, corresponding to their means.

Centre	Treatment A	Treatment B
1	90	100
	80	90
2	90	100
	80	(90)

If the figure in brackets is missing, however, so that there is no longer balance, then the mean treatment estimates will be 85.0 and 97.0 compared with their means of 85.0 and 96.7.

Although the condition of equal numbers in all cells is sufficient for the fixed effects mean estimates to equal their 'raw' means, it is not a necessary condition. In the multi-centre trial, for example, as long as we do not fit centre·treatment effects, it does not matter if the numbers differ across centres, provided the

treatments are allocated evenly within the centres. The following dataset produces treatment mean estimates that equal their raw means.

Centre	Treatment A	Treatment B
1	90	100
	80	
2	85	95
	85	90
	80	
	80	

Another anomaly is the cross-over trial, which is always unbalanced by the Searle *et al.* definition if period effects are fitted, as well as patient and treatment effects. This leads to empty cells because we cannot have both treatments given in the same period to any patient. Nevertheless, in a simple two-period, cross-over trial, if every patient receives every treatment, equal numbers of patients receive each sequence of treatments, and no covariates are fitted, the treatment mean estimates will equal their raw means.

We suggest, therefore, an alternative definition of balance, whereby the fixed effects means will equal their raw means whenever data are balanced but not (in general) when they are unbalanced. Balance occurs for a fixed effect when both of the following conditions are met:

- Within each category of the fixed effect (e.g. treatment), observations occur in equal proportions among categories of every other effect, which is fitted at the same containment level (see the previous section).
- If the fixed effect (e.g. treatment) is contained within a random effect (e.g. centre·treatment), then an equal number of observations are required in each category of the containing effect.

Balance across random effects

It is of importance in this book to identify the situations in which the fixed effects means (usually treatments) will differ depending on whether a fixed effects model or a mixed model is used. When balance, as defined previously, is achieved, then the fixed effects mean estimates will equal the raw means, whether a fixed effects model or a mixed model has been applied. There are other situations when the fixed effects mean estimates will not equal their raw means, but the same estimates will be obtained whether the fixed effects approach or mixed models approach is followed. This occurs when both of the following conditions apply, and we have a situation that we define as balance across random effects:

- Within each category of the specific effect (e.g. treatment), observations are allocated in equal proportions among categories of every random effect (e.g. patient), which is fitted at the same containment level.
- If the effect (e.g. treatment) is contained within a random effect (e.g. centre·treatment), then an equal number of observations are required in each category of the containing effect.

An example of the subtle distinction between these two definitions is provided by the cross-over trial example. If there were an equal number of patients on the AB and BA sequence of treatments, with no missing values, then our definition of balance would be satisfied, as described earlier. If there were no missing values, but the numbers differed between the AB and BA sequences, then there would be balance over random effects. This is true because the only random effect is patients and within each category of the containing effect (i.e. within individual patients), each treatment occurs once, and hence the definition is satisfied. Thus, the treatment estimates will be identical whether the patient effect is fitted as fixed or random, but these estimates will (in general) differ from the raw means.

This definition has been applied in the context of one particular type of mixed model; namely, the random effects model. In random coefficients models, the random coefficient blocking effect (usually patients) can be substituted for 'random effect' in the definition. In covariance pattern models, the blocking effect within which the covariance pattern is defined (again usually patients) can be substituted for 'random effect'.

Assessing balance

It can sometimes be difficult to gain an immediate feel for when balance is achieved from these definitions. The three following common situations are easily classified:

- If any observations are missing, then imbalance across random effects occurs (except for simple parallel group situations).
- If a continuous effect is fitted, then imbalance will occur (unless identical means for the effect happen to occur within each fixed effects category). However, balance across the random effects may still be achieved.
- If an equal number of observations occur in every cell and no continuous covariate is fitted, then all fixed effects will be balanced.

1.6.3 Error strata

In the random effects model, an error stratum or error level is defined by each random effect and by the residual. For example, if patients are fitted as random in a cross-over trial, there are error strata corresponding to the patients and to the residual. The *containment stratum* for a particular fixed effect is defined by the

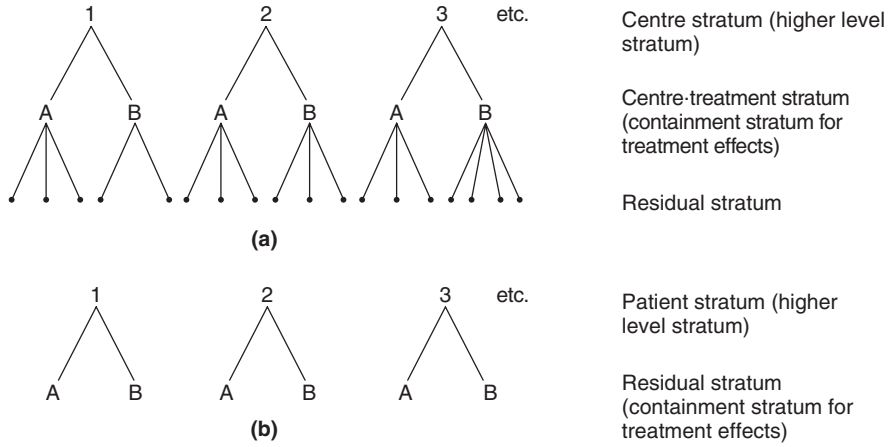

Figure 1.3 (a) Error strata for a multi-centre trial analysis fitting centre and centre·treatment effects as random; (b) error strata for a cross-over trial analysis fitting patient effects as random. A = treatment A; B = treatment B.

residual stratum, unless the effect is contained within a random effect in a random effects model or a blocking effect (see Section 6.2) in a random coefficients or covariance pattern model, in which case it is that of the containing effect. For example, in a repeated measures study, treatments are contained within patients, and thus the patient error stratum forms the containment stratum for treatments. Usually, an effect has only one containment stratum, and examples in this book will be restricted to this more usual situation. However, situations could be conceived where this is not the case. For example, if clinics and GPs were recorded in a trial and GP·treatment and clinic·treatment effects were fitted as random, then both of these effects would form containing strata for the treatment effect.

Higher level strata are defined by any random effects that are contained within the containment stratum. For example, in a multi-centre trial in which centre and centre·treatment effects are fitted as random, the centre·treatment stratum forms the containment stratum for treatment effects, and the centre stratum forms a higher level stratum (see Figure 1.3(a)). In a cross-over trial, the containment stratum for treatment effects is the residual stratum, and the patient stratum is a higher level stratum (see Figure 1.3(b)). Whenever higher level strata are present and data are not balanced across random effects, a fixed effect will be estimated using information from these strata, as well as from the containment stratum (i.e. information is *recovered* from the higher level strata).

Thus, in a cross-over trial with missing values, information is recovered from the patient level, as we saw in Section 1.2. The same occurs with missing values in a repeated measures trial where a covariance pattern is fitted. In random coefficients models, information is recovered from the patient level except in highly unusual

circumstances of equal numbers of observations at the same set of time points for all patients.

In random coefficients and covariance pattern models, error strata are not defined quite as easily because correlations occur between the random coefficients or residuals. However, random coefficients and blocking effects have a similar role to error strata, although their properties are not quite the same.