100387, the square root of which is about 317. The range of all twelve observations was less than 100 and the standard deviation per unit can therefore hardly be as large as 317.

Exercise 5.1

In an experiment to compare the effects of four drugs, A, B, C and a placebo, or inactive substance, D on the lymphocyte counts in mice a randomized block design with four mice from each of five litters was used, the litters being regarded as blocks. The lymphocyte counts (thousands per mm^3 of blood) were:

Litters	1	2	3	4	5	Drug totals
Drugs						
A	7.1	6.1	6.9	5.6	6.4	32,1
B	6.7	5.1	5.9	5.1	5.8	28,6
C	7.1	5.8	6.2	5.0	6.2	30,3
D	6.7	5.4	5.7	5.2	5.3	28,3
Block totals	27,6	22,4	24,7	20,9	23,7	

Complete the following analysis of variance table and summarize the results of the analysis. The uncorrected total s.s. $= \Sigma(x^2)$ is 720.51.

Var. due to	d.f.	s.s.	m.s.	F
Block	4	6,403	1,6	30,15 **
Treatment	3	1,845	0,615	11,58 **
Error	12	0,637	0,053	
Total	19	8,885		

c.f. $= (119,3)^2/20 \doteq 711,624$

s.d. = 0,23 g.m. $= \dfrac{119,3}{20} = 5,965$ c. of v. = 3,86%

Strain means:

A	6,42
B	5,72
C	6,06
D	5,66

s.e. for comparing any two strain means (12 d.f.) = $\sqrt{\dfrac{2(0,053)}{5}} = 0,145$

5.3 Meaning of the error mean square

It is rather more difficult in the randomized block design than in the completely randomized design to see just what the quantity, s^2, is estimating. To get a better idea, we shall consider two further ways of obtaining the error mean square. The first is in the analysis of a simple randomized block experiment in which there are only two treatments. Such an experiment is called a *paired comparison experiment*. In a test of a particular treatment supposed to induce growth, twenty plants were grouped into ten pairs so that the two members of each pair were as alike as possible. One plant of each pair was chosen, randomly, and treated; the other was left as a control. The increases in height (in centimetres) of plants over a two-week period were as follows:

Pair (Block)	Treated plant	Control plant
1	7	4
2	10	6
3	9	10
4	8	8
5	7	5
6	6	3
7	8	10
8	9	8
9	12	8
10	13	10

To carry out the usual analysis of variance we calculate the block totals (11, 16, 19, 16, 12, 9, 18, 17, 20 and 23), the treatment totals (89 for the treated plants, and 72 for the control), the overall total (161) and the sum of squares of the original observations (1415). Hence we obtain the following analysis of variance:

Source of variation	s.s.	d.f.	m.s.
Pairs	84.45	9	
Treatments	14.45	1	14
Error	20.05	9	$s^2 = 2.23$
Total	118.95	19	
c.f.	1296		

The standard error of the difference between the two treatment means is $\sqrt{(2s^2/10)} = 0.67$ cm on 9 d.f., and the difference between the treatment means is $8.9 - 7.2 = 1.7$ cm.

An alternative way of looking at these results is to consider, for each pair, the difference between the growth of the treated plant and the growth of the control plant. Since we are interested only in whether the treatment does increase growth, these differences are reasonable quantities to consider. In fact, we have a sample of ten differences. Note that by considering differences we have eliminated differences between blocks – in this case pairs of plants. If we consider this sample of differences, we can calculate the sample mean and sample variance and obtain, from the sample variance, the standard error of the sample mean. Since the methods of this approach and the analysis of variance approach are essentially the same, i.e. eliminate block differences and estimate variation within blocks, it would be surprising if the sample mean and its standard error were not the same as before.

The actual differences are 3, 4, −1, 0, 2, 3, −2, 1, 4 and 3, giving a mean of 1.7 cm, a sample variance of 4.46 and the standard error of the mean $\sqrt{4.46/10} = 0.67$ cm as before. The sample variance is exactly double the previous error mean square because it is the variance of differences between two values (Section 3.4). Thus for the particular case with two treatments in a randomized block experiment the random variation, s^2, can be regarded as a measure of the consistency of the difference between the two treatments in the different blocks. When there are three or more treatments this interpretation of s^2 must be modified since with three treatments there are three possible pairs of treatments whose differences could be considered and these are not independent. As the number of treatments increases the problem becomes greater. However, it is still true that the error m.s., s^2, is a measure of the consistency of treatment effects from block to block.

Another way of examining the meaning of the error s.s. is to return to the original model for the observed yield of a unit and to consider the difference between the observed yield and that which would be obtained if the unit within a block were not variable. In other words we have to calculate the difference between

$$\begin{pmatrix}\text{observed}\\ \text{yield}\end{pmatrix} = \begin{pmatrix}\text{block}\\ \text{mean}\end{pmatrix} + \begin{pmatrix}\text{treatment}\\ \text{effect}\end{pmatrix} + \begin{pmatrix}\text{random unit}\\ \text{variation}\end{pmatrix}$$

and an estimate of it,

$$\begin{pmatrix}\text{expected}\\ \text{yield}\end{pmatrix} = \begin{pmatrix}\text{block}\\ \text{mean}\end{pmatrix} + \begin{pmatrix}\text{treatment}\\ \text{effect}\end{pmatrix}$$

Such differences are referred to as *residuals*. In estimating expected values for units it is easier to use the form

$$\begin{pmatrix}\text{expected}\\ \text{yield}\end{pmatrix} = \begin{pmatrix}\text{overall}\\ \text{mean}\end{pmatrix} + \begin{pmatrix}\text{block}\\ \text{effect}\end{pmatrix} + \begin{pmatrix}\text{treatment}\\ \text{effect}\end{pmatrix}$$

where block effects (and treatment effects) are deviations of the particular block or treatment mean from the overall mean. Thus we estimate the effect of block 1 by the difference between the mean for block 1 and the overall mean.

Table 5.1 Yields for a randomized block experiment (Example 5.1) with means and estimates of effects

Treatment	*Block* 1	2	3	4	*Treatment mean*	*Treatment effect*
O	330	288	295	313	306.5	−26.25
E	372	340	343	341	349.0	+16.25
F	359	337	373	302	342.75	+10.00
Block mean	353.67	321.67	337.0	318.67	Overall mean = 332.75	
Block effect	+20.92	−11.08	+4.25	−14.08		

Table 5.1 gives yields, means and effect estimates for the experiment of Example 5.1. The calculation of the treatment and block means is obvious. Thus the mean for treatment O is (330 + 288 + 295 + 313)/4 = 306.5, and similarly the overall mean is (330 + 288 + ... + 302)/12 = 332.75. The estimates of effects are obtained by subtracting the overall mean from the particular treatment of block mean. Thus the effect of block 2 is 321.67 − 332.75 = −11.08. In other words, on average the yields for units in block 2 are 11.08 eggs per pen less than the mean of all the units. The estimated expected yields, given in Table 5.2, are then obtained as

$$\text{overall mean} + \text{block effect} + \text{treatment effect}$$

so that for the unit receiving treatment F in block 3 the estimated expected yields is 332.75 + 4.25 + 10.00 = 347.00 compared with the observed yield of 373.

Table 5.2 Expected yields, assuming a simple additive model of treatment and block effects, for the data of Example 5.1

Treatment	*Block* 1	2	3	4
O	327.42	295.42	310.75	292.42
E	369.92	337.92	353.25	334.92
F	363.67	331.67	347.00	328.67

The residuals, the differences between observed and expected yields, are given in Table 5.3; the residual for the unit with treatment F in block 3 is 373 − 347.00 = +26; the residual for treatment F in block 4 is 302 − 328.67 = −26.67, etc.

Table 5.3 Residuals for the data of Example 5.1

Treatment	*Block* 1	2	3	4
O	+2.58	−7.42	−15.75	+20.58
E	+2.08	+2.08	−10.25	+6.08
F	−4.67	+5.33	+26.00	−26.67

Note that the sum of the residuals for a particular treatment or block is zero, apart from rounding-off errors. Why? The sum of the squares of the residuals in Table 5.3 gives 2321.70 which, apart from small rounding-off errors, is the same as the error s.s. in Example 5.1. Thus we can see that the error sum of squares in the analysis of variance is made up from the deviations of the actual observations from the set of yields predicted from a simple additive model of treatment effect + block effect. So we can think of the deviations, and therefore of the error sum of squares, as representing the consistency of the treatment effect pattern over the different blocks. We have already met this interpretation of the error sum of squares in our discussion earlier in this section of the comparison of just two treatments. It must be emphasized, however, that this is not the simplest way of calculating the error sum of squares; the analysis of variance provides that.

5.4 Missing observations in a randomized block design

One result of the more rigid form of the randomized block design is that, unlike the completely randomized design, the analysis cannot deal simply with the situation where the yields for some units are missing. There are many reasons why yields may be missing. Accidents may occur during the experiment such as the death of an animal, damage to a crop from an external source; records may be lost, illegible or even ridiculous. There are situations where it may be decided that some observations cannot be recorded. For example, in a trial on methods of propagating cauliflower seedlings, one of the propagation treatments may not give sufficient seedlings to plant out the required number of plots for that treatment.

The absence of one or more observations changes the form of the design since the balanced nature of the relationship between treatments and blocks is destroyed. Various methods of dealing with this situation are available but most of these are statistically complex. In practice therefore an approximate method is used in which estimated values are substituted for the missing observations and the analysis of variance calculated almost as before.

The simplest situation is when one observation is missing. Suppose, for simplicity, that this is the observation on the unit receiving treatment 1 in block 1. Assume also that the randomized block design has b blocks and t treatments. Then to calculate the estimated value for this unit we require the total of the non-missing observations in block 1, B_1', the total of the non-missing observations for treatment 1, T_1', and the grand total of all recorded observations, G'. The estimated value for the missing unit is then

$$\frac{bB_1' + tT_1' - G'}{(b-1)(t-1)}$$

This form of estimate is derived using the principle that the estimate shall be such as to make the error sum of squares as small as possible, and this is in fact

achieved by making the residual for the missing unit zero or, equivalently, by substituting for the missing value an estimate as consistent as possible with the rest of the data. Using this estimated value we complete the analysis of variance as usual, except that, because our estimate is derived from the remaining values, the degrees of freedom for both total variation and error variation are one less than usual. This is reasonable for we have one less piece of information than in the full experiment but there are still t treatment means and b block means so that the degrees of freedom for treatments and blocks must still be $(t-1)$ and $(b-1)$ respectively. In calculating the standard errors of treatment means we must obviously take account of the fact that the mean for the treatment with a missing unit is based on only $(b-1)$ actual observations. This leads to a revised standard error of a treatment mean, when the treatment has a missing unit,

$$\text{s.e. of treatment mean} = \sqrt{\left\{\frac{s^2}{b}\left[1 + \frac{t}{(b-1)(t-1)}\right]\right\}}$$

and the corresponding standard error of the difference between this treatment mean and any other treatment mean,

$$\text{s.e. of treatment mean difference} = \sqrt{\left\{\frac{s^2}{b}\left[2 + \frac{t}{(b-1)(t-1)}\right]\right\}}$$

These standard errors are, of course, based on the degrees of freedom of s^2, i.e. $(b-1)(t-1)-1$.

The standard errors of treatment means and of differences between treatment means not involving the missing unit are in the usual form, i.e. $\sqrt{(s^2/b)}$ and $\sqrt{(2s^2/b)}$, again based on $(b-1)(t-1)-1$ degrees of freedom.

Example 5.2 A single missing unit in a randomized block analysis of variance
Suppose that in the experiment of Example 5.1 the record of the number of eggs for treatment E in block 3 had been destroyed accidentally, the records then being:

	Block				
Treatments	1	2	3	4	Total
O	330	288	295	313	1226
E	372	340		341	
F	359	337	373	302	1371
Total	1061	965		956	

The block total for recorded yields in block 3 is $B_3' = 668$, the treatment total for recorded yields for treatment E is $T_E' = 1053$, and the overall total of recorded yields is $G' = 3650$. The estimated yield for the missing unit is then

$$\frac{4 \times 668 + 3 \times 1053 - 3650}{3 \times 2} = 363.5$$