CHAPTER 3

Control of haphazard variation

3.1 General remarks

In the previous chapter the primary emphasis was on the elimination of systematic error. We now turn to the control of haphazard error, which may enter at any of the phases of an investigation. Sources of haphazard error include intrinsic variation in the experimental units, variation introduced in the intermediate phases of an investigation and measurement or sampling error in recording response.

It is important that measures to control the effect of such variation cover all the main sources of variation and some knowledge, even if rather qualitative, of the relative importance of the different sources is needed.

The ways in which the effect of haphazard variability can be reduced include the following approaches.

1. It may be possible to use more uniform material, improved measuring techniques and more internal replication, i.e. repeat observations on each unit.
2. It may be possible to use more experimental units.
3. The technique of blocking, discussed in detail below, is a widely applicable technique for improving precision.
4. Adjustment for baseline features by the techniques for bias removal discussed in Section 2.3 can be used.
5. Special models of error structure may be constructed, for example based on a time series or spatial model.

On the first two points we make here only incidental comments.

There will usually be limits to the increase in precision achievable by use of more uniform material and in technological experiments the wide applicability of the conclusions may be prejudiced if artificial uniformity is forced.

Illustration. In some contexts it may be possible to use pairs

of homozygotic twins as experimental units in the way set out in detail in Section 3.3. There may, however, be some doubt as to whether conclusions apply to a wider population of individuals. More broadly, in a study to elucidate some new phenomenon or suspected effect it will usually be best to begin with the circumstances under which that effect occurs in its most clear-cut form. In a study in which practical application is of fairly direct concern the representativeness of the experimental conditions merits more emphasis, especially if it is suspected that the treatment effects have different signs in different individuals.

In principle precision can always be improved by increasing the number of experimental units. The standard error of treatment comparisons is inversely proportional to the square root of the number of units, provided the residual standard deviation remains constant. In practice the investigator's control may be weaker in large investigations than in small so that the theoretical increase in the number of units needed to shorten the resulting confidence limits for treatment effects is often an underestimate.

3.2 Precision improvement by blocking

The central idea behind blocking is an entirely commonsense one of aiming to compare like with like. Using whatever prior knowledge is available about which baseline features of the units and other aspects of the experimental set-up are strongly associated with potential response, we group the units into blocks such that all the units in any one block are likely to give similar responses in the absence of treatment differences. Then, in the simplest case, by allocating one unit in each block to each treatment, treatments are compared on units within the same block.

The formation of blocks is usually, however, quite constrained in addition by the way in which the experiment is conducted. For example, in a laboratory experiment a block might correspond to the work that can be done in a day. In our initial discussion we regard the different blocks as merely convenient groupings without individual interpretation. Thus it makes no sense to try to interpret differences between blocks, except possibly as a guide for future experimentation to see whether the blocking has been effective in error control. Sometimes, however, some aspects of blocking do have a clear interpretation, and then the issues of Chapter 5 concerned

with factorial experiments apply. In such cases it is preferable to use the term *stratification* rather than blocking.

Illustrations. Typical ways of forming blocks are to group together neighbouring plots of ground, responses from one subject in one session of a psychological experiment under different conditions, batches of material produced on one machine, where several similar machines are producing nominally the same product, groups of genetically similar animals of the same gender and initial body weight, pairs of homozygotic twins, the two eyes of the same subject in an opthalmological experiment, and so on. Note, however, that if gender were a defining variable for blocks, i.e. strata, we would likely want not only to compare treatments but also to examine whether treatment differences are the same for males and females and this brings in aspects that we ignore in the present chapter.

3.3 Matched pairs

3.3.1 Model and analysis

Suppose that we have just two treatments, T and C, for comparison and that we can group the experimental units into pairs, so that in the absence of treatment differences similar responses are to be expected in the two units within the same pair or block.

It is now reasonable from many viewpoints to assign one member of the pair to T and one to C and, moreover, in the absence of additional structure, to randomize the allocation within each pair independently from pair to pair. This yields what we call the *matched pair* design.

Thus if we label the units

$$U_{11}, U_{21}; \quad U_{12}, U_{22}; \quad \ldots; \quad U_{1r}, U_{2r} \tag{3.1}$$

a possible design would be

$$T, C; \quad C, T; \quad \ldots; \quad T, C. \tag{3.2}$$

As in Chapter 2, a linear model that directly corresponds with randomization theory can be constructed. The broad principle in setting up such a physical linear model is that randomization constraints forced by the design are represented by parameters in the linear model. Writing Y_{Ts}, Y_{Cs} for the observations on treatment

and control for the sth pair, we have the model

$$Y_{Ts} = \mu + \beta_s + \delta + \epsilon_{Ts}, \quad Y_{Cs} = \mu + \beta_s - \delta + \epsilon_{Cs}, \tag{3.3}$$

where the ϵ are random variables of mean zero. As in Section 2.2, either the normal theory or the second moment assumption about the errors may be made; the normal theory assumption leads to distributional results and strong optimality properties.

Model (3.3) is overparameterized, but this is often convenient to achieve a symmetrical formulation. The redundancy could be avoided here by, for example, setting μ to any arbitrary known value, such as zero.

A least squares analysis of this model can be done in several ways. The simplest, for this very special case, is to transform the Y_{Ts}, Y_{Cs} to sums, B_s and differences, D_s. Because this is proportional to an orthogonal transformation, the transformed observations are also uncorrelated and have constant variance. Further in the linear model for the new variables we have

$$E(B_s) = 2(\mu + \beta_s), \quad E(D_s) = 2\delta = \Delta. \tag{3.4}$$

It follows that, so long as the β_s are regarded as unknown parameters unconnected with Δ, the least squares estimate of Δ depends only on the differences D_s and is in fact the mean of the differences,

$$\hat{\Delta} = \bar{D}_\cdot = \bar{Y}_{T\cdot} - \bar{Y}_{C\cdot}, \tag{3.5}$$

with

$$\text{var}(\hat{\Delta}) = \text{var}(D_s)/r = 2\sigma^2/r, \tag{3.6}$$

where σ^2 is the variance of ϵ. Finally σ^2 is estimated as

$$s^2 = \Sigma(D_s - \bar{D}_\cdot)^2/\{2(r-1)\}, \tag{3.7}$$

so that

$$\text{evar}(\hat{\Delta}) = 2s^2/r. \tag{3.8}$$

In line with the discussion in Section 2.2.4 we now show that the properties just established under the linear model and the second moment assumption also follow from the randomization used in allocating treatments to units, under the unit-treatment additivity assumption. This assumption specifies the response on the sth pair to be $(\xi_{1s} + \delta, \xi_{2s} - \delta)$ if the first unit in that pair is randomized to treatment and $(\xi_{1s} - \delta, \xi_{2s} + \delta)$ if it is randomized to control. We

then have

$$E_R(\hat{\Delta}) = \Delta, \quad E_R\{\text{evar}(\hat{\Delta})\} = \text{var}_R(\hat{\Delta}). \tag{3.9}$$

To prove the second result we note that both sides of the equation do not depend on Δ and are quadratic functions of the ξ_{js}. They are invariant under permutations of the numbering of the pairs $1, \ldots, r$, and under permutations of the two units in any pair. Both sides are zero if $\xi_{1s} = \xi_{2s}, s = 1, \ldots, r$. It follows that both sides of the equation are constant multiples of

$$\Sigma(\xi_{1s} - \xi_{2s})^2 \tag{3.10}$$

and consistency with the least squares analysis requires that the constants of proportionality are equal. In fact, for example,

$$E_R(s^2) = \Sigma(\xi_{1s} - \xi_{2s})^2/(2r). \tag{3.11}$$

Although not necessary for the discussion of the matched pair design, it is helpful for later discussion to set out the relation with analysis of variance. In terms of the original responses Y the estimation of μ, β_s is orthogonal to the estimation of Δ and the analysis of variance arises from the following decompositions.

First there is a representation of the originating random observations in the form

$$\begin{aligned} Y_{Ts} &= \bar{Y}_{\cdot\cdot} + (\bar{Y}_{T\cdot} - \bar{Y}_{\cdot\cdot}) + (\bar{Y}_{\cdot s} - \bar{Y}_{\cdot\cdot}) \\ &\qquad + (\bar{Y}_{Ts} - \bar{Y}_{T\cdot} - \bar{Y}_{\cdot s} + \bar{Y}_{\cdot\cdot}), \qquad (3.12) \\ Y_{Cs} &= \bar{Y}_{\cdot\cdot} + (\bar{Y}_{C\cdot} - \bar{Y}_{\cdot\cdot}) + (\bar{Y}_{\cdot s} - \bar{Y}_{\cdot\cdot}) \\ &\qquad + (\bar{Y}_{Cs} - \bar{Y}_{C\cdot} - \bar{Y}_{\cdot s} + \bar{Y}_{\cdot\cdot}). \qquad (3.13) \end{aligned}$$

Regarded as a decomposition of the full vector of observations, this has orthogonal components.

Secondly because of that orthogonality the squared norms of the components add to give

$$\Sigma Y_{js}^2 = \Sigma \bar{Y}_{\cdot\cdot}^2 + \Sigma(\bar{Y}_{j\cdot} - \bar{Y}_{\cdot\cdot})^2 + \Sigma(\bar{Y}_{\cdot s} - \bar{Y}_{\cdot\cdot})^2 + \Sigma(Y_{js} - \bar{Y}_{j\cdot} - \bar{Y}_{\cdot s} + \bar{Y}_{\cdot\cdot})^2 : \tag{3.14}$$

note that Σ represents a sum over all observations so that, for example, $\Sigma \bar{Y}_{\cdot\cdot}^2 = 2r\bar{Y}_{\cdot\cdot}^2$. In this particular case the sums of squares can be expressed in simpler forms. For example the last term is $\Sigma(D_s - \bar{D}_\cdot)^2/2$. The squared norms on the right-hand side are conventionally called respectively sums of squares for general mean, for treatments, for pairs and for residual or error.

Thirdly the dimensions of the spaces spanned by the component vectors, as the vector of observations lies in the full space of dimension $2r$, also are additive:

$$2r = 1 + 1 + (r-1) + (r-1). \tag{3.15}$$

These are conventionally called *degrees of freedom* and *mean squares* are defined for each term as the sum of squares divided by the degrees of freedom. Finally, under the physical linear model (3.3) the residual mean square has expectation σ^2.

3.3.2 A modified matched pair design

In some matched pairs experiments we might wish to include some pairs of units both of which receive the same treatment. Cost considerations might sometimes suggest this as a preferable design, although in that case redefinition of an experimental unit as a pair of original units would be called for and the use of a mixture of designs would not be entirely natural. If, however, there is some suspicion that the two units in a pair do not react independently, i.e. there is doubt about one of the fundamental assumptions of unit-treatment additivity, then a mixture of matched pairs and pairs both treated the same might be appropriate.

Illustration. An opthalmological use of matched pairs might involve using left and right eyes as distinct units, assigning different treatments to the two eyes. This would not be a good design unless there were firm *a priori* grounds for considering that the treatment applied to one eye had negligible influence on the response in the other eye. Nevertheless as a check it might be decided for some patients to assign the same treatment to both eyes, in effect to see whether the treatment difference is the same in both environments. Such checks are, however, often of low sensitivity.

Consider a design in which the r matched pairs are augmented by m pairs in which both units receive the same treatment, m_T pairs receiving T and m_C receiving C, with $m_T + m_C = m$. So long as the parameters β_s in the matched pairs model describing inter-pair differences are arbitrary the additional observations give no information about the treatment effect. In particular a comparison of the means of the m_T and the m_C complete pairs estimates Δ plus a contrast of totally unknown β's.

Suppose, however, that the pairs are randomized between com-

plete and incomplete assignments. Then under randomization analysis the β's can be regarded in effect as random variables. In terms of a corresponding physical model we write for each observation

$$Y_{js} = \mu \pm \delta + \beta_s + \epsilon_{js}, \tag{3.16}$$

where the sign of δ depends on the treatment involved, the β_s are now zero mean random variables of variance σ_B^2 and the ϵ_{js} are, as before, zero mean random variables of variance now denoted by σ_W^2. All random variables are mutually uncorrelated or, in the normal theory version, independently normally distributed.

It is again convenient to replace the individual observations by sums and differences. An outline of the analysis is as follows. Let Δ_{MP} and Δ_{UM} denote treatment effects in the matched pairs and the unmatched data respectively. These are estimated by the previous estimate, now denoted by $\bar{Y}_{\mathrm{MP}T} - \bar{Y}_{\mathrm{MP}C}$, with variance $2\sigma_W^2/r$ and by $\bar{Y}_{\mathrm{UM}T} - \bar{Y}_{\mathrm{UM}C}$ with variance

$$(\sigma_B^2 + \sigma_W^2/2)(1/m_T + 1/m_C). \tag{3.17}$$

If, as might quite often be the case, σ_B^2 is large compared with σ_W^2, the between block comparison may be of such low precision as to be virtually useless.

If the variance components are known we can thus test the hypothesis that the treatment effect is, as anticipated *a priori*, the same in the two parts of the experiment and subject to homogeneity find a weighted mean as an estimate of the common Δ. Estimation of the two variance components is based on the sum of squares within pairs adjusting for treatment differences in the matched pair portion and on the sum of squares between pair totals adjusting for treatment differences in the unmatched pair portion.

Under normal theory assumptions a preferable analysis for a common Δ is summarized in Exercise 3.3. There are five sufficient statistics, two sums of squares and three means, and four unknown parameters. The log likelihood of these statistics can be found and a profile log likelihood for Δ calculated.

The procedure of combining information from within and between pair comparisons can be regarded as the simplest special case of the recovery of between-block information. More general cases are discussed in Section 4.2.

3.4 Randomized block design

3.4.1 Model and analysis

Suppose now that we have more than two treatments and that they are regarded as unstructured and on an equal footing and therefore to be equally replicated. The discussion extends in a fairly direct way when some treatments receive additional replication. With v treatments, or varieties in the plant breeding context, we aim to produce blocks of v units. As with matched pairs we try, subject to administrative constraints on the experiment, to arrange that in the absence of treatment effects, very similar responses are to be anticipated on the units within any one block. We allocate treatments independently from block to block and at random within each block, subject to the constraint that each treatment occurs once in each block.

Illustration. Typical ways of forming blocks include compact arrangements of plots in a field chosen in the light of any knowledge about fertility gradients, batches of material that can be produced in one day or production period, and animals grouped on the basis of gender and initial body weight.

Let Y_{js} denote the observation on treatment T_j in block s. Note that because of the randomization this observation may be on any one of the units in block s in their original listing. In accordance with the general principle that constraints on the randomization are represented by parameters in the associated linear model, we represent Y_{js} in the form

$$Y_{js} = \mu + \tau_j + \beta_s + \epsilon_{js}, \tag{3.18}$$

where $j = 1, \ldots, v$; $s = 1, \ldots, r$ and ϵ_{js} are zero mean random variables satisfying the second moment or normal theory assumptions. The least squares estimates of the parameters are determined by the row and column means and in particular under the summation constraints $\Sigma\tau_j = 0$, $\Sigma\beta_s = 0$, we have $\hat{\tau}_j = \bar{Y}_{j.} - \bar{Y}_{..}$ and $\hat{\beta}_s = \bar{Y}_{.s} - \bar{Y}_{..}$. The contrast $L_\tau = \Sigma l_j \tau_j$ is estimated by $\hat{L}_\tau = \Sigma l_j \bar{Y}_{j.}$.

The decomposition of the observations, the sums of squares and the degrees of freedom are as follows:

1. For the observations we write

$$\begin{aligned} Y_{js} &= \bar{Y}_{..} + (\bar{Y}_{j.} - \bar{Y}_{..}) + (\bar{Y}_{.s} - \bar{Y}_{..}) \\ &\quad + (Y_{js} - \bar{Y}_{j.} - \bar{Y}_{.s} + \bar{Y}_{..}), \end{aligned} \tag{3.19}$$

a decomposition into orthogonal components.

2. For the sums of squares we therefore have

$$\begin{aligned}\Sigma Y_{js}^2 &= \Sigma \bar{Y}_{..}^2 + \Sigma(\bar{Y}_{j.} - \bar{Y}_{..})^2 + \Sigma(\bar{Y}_{.s} - \bar{Y}_{..})^2 \\ &\quad + \Sigma(Y_{js} - \bar{Y}_{j.} - \bar{Y}_{.s} + \bar{Y}_{..})^2, \quad (3.20)\end{aligned}$$

where the summation is always over both suffices.

3. For the degrees of freedom we have

$$rv = 1 + (v-1) + (r-1) + (r-1)(v-1). \quad (3.21)$$

The residual mean square provides an unbiased estimate of the variance. Let

$$s^2 = \Sigma(Y_{js} - \bar{Y}_{j.} - \bar{Y}_{.s} + \bar{Y}_{..})^2/\{(r-1)(v-1)\}. \quad (3.22)$$

We now indicate how to establish the result $E(s^2) = \sigma^2$ under the second moment assumptions. In the linear model the residual sum of squares depends only on $\{\epsilon_{js}\}$, and not on the fixed parameters μ, $\{\tau_j\}$ and $\{\beta_s\}$. Thus for the purpose of computing the expected value of (3.22) we can set these parameters to zero. All sums of squares in (3.20) other than the residual have simple expectations: for example

$$\begin{aligned}E\{\Sigma_{j,s}(\bar{Y}_{j.} - \bar{Y}_{..})^2\} &= rE\{\Sigma_j(\bar{\epsilon}_{j.} - \bar{\epsilon}_{..})^2\} \quad (3.23) \\ &= r(v-1)\mathrm{var}(\bar{\epsilon}_{j.}) = (v-1)\sigma^2. \quad (3.24)\end{aligned}$$

Similarly $E\{\Sigma_{j,s}(\bar{Y}_{.s} - \bar{Y}_{..})^2\} = (r-1)\sigma^2$, $E(\Sigma_{j,s}\bar{Y}_{..}^2) = \sigma^2$, and that for the residual sum of squares follows by subtraction. Thus the unbiased estimate of the variance of $\hat{L}_\tau$ is

$$\mathrm{evar}(\hat{L}_\tau) = \Sigma_j l_j^2 s^2/r. \quad (3.25)$$

The partition of the sums of squares given by (3.20) is often set out in an analysis of variance table, as for example Table 3.2 below. This table has one line for each component of the sum of squares, with the usual convention that the sums of squares due to the overall mean, $n\bar{Y}_{..}^2$, is not displayed, and the total sum of squares is thus a corrected total $\Sigma(Y_{js} - \bar{Y}_{..})^2$.

The simple decomposition of the data vector and sum of squares depend crucially on the balance of the design. If, for example, some treatments were missing in some blocks not merely would the orthogonality of the component vectors be lost but the contrasts of treatment means would not be independent of differences between blocks and *vice versa*. To extend the discussion to such cases more

elaborate methods based on a least squares analysis are needed. It becomes crucial to distinguish, for example, between the sum of squares for treatments ignoring blocks and the sum of squares for treatments adjusting for blocks, the latter measuring the effect of introducing treatment effects after first allowing for block differences.

The randomization model for the randomized block design uses the assumption of unit-treatment additivity, as in the matched pairs design. We label the units

$$U_{11},\ldots,U_{v1};\quad U_{12},\ldots,U_{v2};\quad \ldots;\quad U_{1r},\ldots,U_{vr}. \tag{3.26}$$

The response on the unit in the sth block that is randomized to treatment T_j is

$$\xi_{T_j s} + \tau_j \tag{3.27}$$

where $\xi_{T_j s}$ is the response of that unit in block s in the absence of treatment.

Under randomization theory properties such as

$$E_R\{\text{evar}(\hat{L}_\tau)\} = \text{var}_R(\hat{L}_\tau) \tag{3.28}$$

are established by first showing that both sides are multiples of

$$\Sigma(\xi_{js} - \bar{\xi}_{.s})^2. \tag{3.29}$$

3.4.2 Example

This example is taken from Cochran and Cox (1958, Chapter 3), and is based on an agricultural field trial. In such trials blocks are naturally formed from large sections of field, sometimes roughly square; the shape of individual plots and their arrangement into plots is usually settled by a mixture of technological convenience, for example ease of harvesting, and special knowledge of the particular area.

This experiment tested the effects of five levels of application of potash on the strength of cotton fibres. A single sample of cotton was taken from each plot, and four measurements of strength were made on each sample. The data in Table 3.1 are the means of these four measurements.

The marginal means are given in Table 3.1, and seem to indicate decreasing strength with increasing amount of potash, with perhaps some curvature in the response, since the mean strength

Table 3.1 *Strength index of cotton, from Cochran and Cox (1958), with marginal means.*

		Pounds of potash per acre					
		36	54	72	108	144	Mean
Block	I	7.62	8.14	7.76	7.17	7.46	7.63
	II	8.00	8.15	7.73	7.57	7.68	7.83
	III	7.93	7.87	7.74	7.80	7.21	7.71
Mean		7.85	8.05	7.74	7.51	7.45	7.72

Table 3.2 *Analysis of variance for strength index of cotton.*

Source	Sums of squares	Degrees of freedom	Mean square
Treatment	0.7324	4	0.1831
Blocks	0.0971	2	0.0486
Residual	0.3495	8	0.0437

at 36 pounds is less than that at 54 pounds, where the maximum is reached.

The analysis of variance outlined in Section 3.4.1 is given in Table 3.2. The main use of the analysis of variance table is to provide an estimate of the standard error for assessing the precision of contrasts of the treatment means. The mean square residual is an unbiased estimate of the variance of an individual observation, so the standard error for example for comparing two treatment means is $\sqrt{(2 \times 0.0437/3)} = 0.17$, which suggests that the observed decrease in strength over the levels of potash used is a real effect, but the observed initial increase is not.

It is possible to construct more formal tests for the shape of the response, by partitioning the sums of squares for treatments, and this is considered further in Section 3.5 below.

The S-PLUS code for carrying out the analysis of variance in this and the following examples is given in Appendix C. As with many

other statistical packages, the emphasis in the basic commands is on the analysis of variance table and the associated F-tests, which in nearly all cases are not the most useful summary information.

3.4.3 Efficiency of blocking

As noted above the differences between blocks are regarded as of no intrinsic interest, so long as no relevant baseline information is available about them. Sometimes, however, it may be useful to ask how much gain in efficiency there has been as compared with complete randomization. The randomization model provides a means of assessing how effective the blocking has been in improving precision. In terms of randomization theory the variance of the difference between two treatment means in a completely randomized experiment is determined by

$$\frac{2}{r}\Sigma(\xi_{js} - \bar{\xi}_{..})^2/(vr-1), \tag{3.30}$$

whereas in the randomized block experiment it is

$$\frac{2}{r}\Sigma(\xi_{js} - \bar{\xi}_{.s})^2/\{r(v-1)\}. \tag{3.31}$$

Also in the randomization model the mean square between blocks is constant with value

$$v\Sigma(\bar{\xi}_{.s} - \bar{\xi}_{..})^2/(r-1). \tag{3.32}$$

As a result the relative efficiency for comparing two treatment means in the two designs is estimated by

$$\frac{2}{r}\frac{\mathrm{SS}_B + r(v-1)\mathrm{MS}_R}{(vr-1)\mathrm{MS}_R}. \tag{3.33}$$

Here SS_B and MS_R are respectively the sum of squares for blocks and the residual mean square in the original randomized block analysis.

To produce from the original analysis of variance table for the randomized block design an estimate of the effective residual variance for the completely randomized design we may therefore produce a new formal analysis of variance table as follows. Replace the treatment mean square by the residual mean square, add the sums of squares for modified treatments, blocks and residual and divide by the degrees of freedom, namely $vr-1$. The ratio of the two residual mean squares, the one in the analysis of the randomized

block experiment to the notional one just reconstructed, measures the reduction in effective variance induced by blocking.

There is a further aspect, however; if confidence limits for Δ are found from normal theory using the Student t distribution, the degrees of freedom are $(v-1)(r-1)$ and $v(r-1)$ respectively in the randomized block and completely randomized designs, showing some advantage to the latter if the error variances remain the same. Except in very small experiments, however, this aspect is relatively minor.

3.5 Partitioning sums of squares

3.5.1 General remarks

We have in this chapter emphasized that the objective of the analysis is the estimation of comparisons between the treatments. In the context of analysis of variance the sum of squares for treatments is a summary measure of the variation between treatments and could be the basis of a test of the overall null hypothesis that all treatments have identical effect, i.e. that the response obtained on any unit is unaffected by the particular treatment assigned to it. Such a null hypothesis is, however, very rarely of concern and therefore the sum of squares for treatments is of importance primarily in connection with the computation of the residual sum of squares, the basis for estimating the error variance.

It is, however, important to note that the treatment sum of squares can be decomposed into components corresponding to comparisons of the individual effects and this we now develop.

3.5.2 Contrasts

Recall from Section 2.2.6 that if the treatment parameters are denoted by $\tau_1, \ldots, \tau_v$ a linear combination $L_\tau = \Sigma l_j \tau_j$ is called a treatment contrast if $\Sigma l_j = 0$. The contrast L_τ is estimated in the randomized block design by

$$\hat{L}_\tau = \Sigma_j l_j \bar{Y}_{j.}, \tag{3.34}$$

where $\bar{Y}_{j.}$ is the mean response on the jth treatment, averaged over blocks. Equivalently we can write

$$\hat{L}_\tau = \Sigma_{j,s} l_j Y_{js}/r, \tag{3.35}$$

where the sum is over individual observations and r is the number of replications of each treatment.

Under the linear model (3.18) and the second moment assumption,

$$E(\hat{L}_\tau) = L_\tau, \quad \text{var}(\hat{L}_\tau) = \sigma^2 \Sigma_j l_j^2 / r. \tag{3.36}$$

We now define the sum of squares with one degree of freedom associated with L_τ to be

$$\text{SS}_L = r\hat{L}_\tau^2 / \Sigma l_j^2. \tag{3.37}$$

This definition is in some ways most easily recalled by noting that $\hat{L}_\tau$ is a linear combination of responses, and hence SS_L is the squared length of the orthogonal projection of the observation vector onto the vector whose components are determined by l.

The following properties are derived directly from the definitions:

1. $E(\hat{L}_\tau) = L_\tau$ and is zero if and only if the population contrast is zero.
2. $E(\text{SS}_L) = \sigma^2 + rL_\tau^2 / \Sigma l_j^2$.
3. Under the normal theory assumption SS_L is proportional to a noncentral chi-squared random variable with one degree of freedom reducing to the central chi-squared form if and only if $L_\tau = 0$.
4. The square of the Student t statistic for testing the null hypothesis $L_\tau = 0$ is the analysis of variance F statistic for comparing SS_L with the residual mean square.

In applications the Student t form is to be preferred to its square, partly because it preserves the information in the sign and more importantly because it leads to the determination of confidence limits.

3.5.3 Mutually orthogonal contrasts

Several contrasts $L_\tau^{(1)}, L_\tau^{(2)}, \ldots$ are called mutually orthogonal if for all $p \neq q$

$$\Sigma l_j^{(p)} l_j^{(q)} = 0. \tag{3.38}$$

Note that under the normal theory assumption the estimates of orthogonal contrasts are independent. The corresponding Student t statistics are not quite independent because of the use of a common estimate of σ^2, although this is a minor effect unless the residual degrees of freedom are very small.

Now suppose that there is a complete set of $v-1$ mutually orthogonal contrasts. Then by forming an orthogonal transformation of $\bar{Y}_{1.},\ldots,\bar{Y}_{v.}$ from $(1/\sqrt{v},\ldots,1/\sqrt{v})$ and the normalized contrast vectors, it follows that

$$r\Sigma_{js}(\bar{Y}_{j.}-\bar{Y}_{..})^2 = \mathrm{SS}_{L_\tau^{(1)}}+\ldots+\mathrm{SS}_{L_\tau^{(v)}}, \tag{3.39}$$

that is the treatment sum of squares has been decomposed into single degrees of freedom.

Further if there is a smaller set of $v_1 < v-1$ mutually orthogonal contrasts, then the treatment sum of squares can be decomposed into

Selected individual contrasts	v_1
Remainder	$v-1-v_1$
Total for treatments	$v-1$

In this analysis comparison of the mean square for the remainder term with the residual mean square tests the hypothesis that all treatment effects are accounted for within the space of the v_1 identified contrasts. Thus with six treatments and the single degree of freedom contrasts identified by

$$L_\tau^{(1)} = (\tau_1+\tau_2)/2-\tau_3, \tag{3.40}$$

$$L_\tau^{(2)} = (\tau_1+\tau_2+\tau_3)/3-(\tau_4+\tau_5+\tau_6)/3, \tag{3.41}$$

we have the partition

$L_\tau^{(1)}$	1
$L_\tau^{(2)}$	1
Remainder	3
Total for treatments	5

The remainder term could be divided further, perhaps most naturally initially into a contrast of τ_1 with τ_2 and a comparison with two degrees of freedom among the last three treatments.

The orthogonality of the contrasts is required for the simple decomposition of the sum of squares. Subject-matter relevance of the comparisons of course overrides mathematical simplicity and it may be unavoidable to look at nonorthogonal comparisons.

We have in this section used notation appropriate to partitioning the treatment sums of squares in a randomized block design,

but the same ideas apply directly to more general settings, with $\bar{Y}_{j.}$ above replaced by the average of all observations on the jth treatment, and r replaced by the number of replications of each treatment. When in Chapter 5 we consider more complex treatments defined by factors exactly the same analysis can be applied to interactions.

3.5.4 Equally spaced treatment levels

A particularly important special case arises when treatments are defined by levels of a quantitative variable, often indeed by equally spaced values of that variable. For example a dose might be set at four levels defined by log dose $= 0, 1, 2, 3$ on some suitable scale, or a temperature might have three levels defined by temperatures of 30, 40, 50 degrees Celsius, and so on.

We now discuss the partitioning of the sums of squares for such a quantitative treatment in orthogonal components, corresponding to regression on that variable. It is usual, and sensible, with quantitative factors at equally spaced levels, to use contrasts representing linear, quadratic, cubic, ... dependence of the response on the underlying variable determining the factor levels. Tables of these contrasts are widely available and are easily constructed from first principles via orthogonal polynomials, i.e. via Gram-Schmidt orthogonalization of $\{1, x, x^2, \ldots\}$. For a factor with three equally spaced levels, the linear and quadratic contrasts are

$$\begin{array}{rrr} -1 & 0 & 1 \\ 1 & -2 & 1 \end{array}$$

and for one with four equally spaced levels, the linear, quadratic and cubic contrasts are

$$\begin{array}{rrrr} -3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & 3 & 1 \end{array}$$

The sums of squares associated with these can be compared with the appropriate residual sum of squares. In this way some notion of the shape of the dependence of the response on the variable defining the factor can be obtained.

3.5.5 Example 3.4 continued

In this example the treatments were defined by increasing levels of potash, in pounds per acre. The levels used were 36, 54, 72, 108 and 144. Of interest is the shape of the dependence of strength on level of potash; there is some indication in Table 3.1 of a levelling off or decrease of response at the highest level of potash.

These levels are not equally spaced, so the orthogonal polynomials of the previous subsection are not exactly correct for extracting linear, quadratic, and other components. The most accurate way of partitioning the sums of squares for treatments is to use regression methods or equivalently to construct the appropriate orthogonal polynomials from first principles. We will illustrate here the use of the usual contrasts, as the results are much the same.

The coefficients for the linear contrast with five treatment levels are $(-2, -1, 0, 1, 2)$, and the sum of squares associated with this contrast is $SS_{lin} = 3(-1.34)^2/10 = 0.5387$. The nonlinear contribution to the treatment sum of squares is thus just 0.1938 on three degrees of freedom, which indicates that the suggestion of nonlinearity in the response is not significant. The quadratic component, defined by the contrast $(2, -1, -2, -1, 2)$ has an associated sum of squares of 0.0440.

If we use the contrast exactly appropriate for a linear regression, which has entries proportional to

$$(-2, -1.23, -0.46, 1.08, 2.61),$$

we obtain the same conclusion.

With more extensive similar data, or with various sets of similar data, it would probably be best to fit a nonlinear model consistent with general subject-matter knowledge, for example an exponential model rising to an asymptote. Fitting such a model across various sets of data should be helpful for the comparison and synthesis of different studies.

3.6 Retrospective adjustment for improving precision

In Section 3.1 we reviewed various ways of improving precision and in Sections 3.2 and 3.3 developed the theme of comparing like with like via blocking the experimental units into relatively homogeneous sets, using baseline information. We now turn to a second use of baseline information. Suppose that on each experimental

unit there is a vector z of variables, either quantitative or indicators of qualitative groupings and that this information has either not been used in forming blocks or at least has been only partly used.

There are three rather different situations. The importance of z may have been realized only retrospectively, for example by an investigator different from the one involved in design. It may have been more important to block on features other than z; this is especially relevant when a large number of baseline features is available. Thirdly, any use of z to form blocks is qualitative and it may be that quantitative use of z instead of, or as well as, its use to form blocks may add sensitivity.

Illustrations. In many clinical trials there will be a large number of baseline features available at the start of a trial and the practicalities of randomization may restrict blocking to one or two key features such as gender and age or gender and initial severity. In an animal experiment comparing diets, blocks could be formed from animals of the same gender and roughly the same initial body weight but, especially in small experiments, appreciable variation in initial body weight might remain within blocks.

Values of z can be used to test aspects of unit-treatment additivity, in effect via tests of parallelism, but here we concentrate on precision improvement. The formal statistical procedures of introducing regression on z into a model have appeared in slightly different guise in Section 2.3 as techniques for retrospective bias removal and will not be repeated. In fact what from a design perspective is random error can become bias at the stage of analysis, when conditioning on relevant baseline features is appropriate. It is therefore not surprising that the same statistical technique reappears.

Illustration. A group of animals with roughly equal numbers of males and females is randomized between two treatments T and C regardless of gender. It is then realized that there are substantially more males than females in T. From an initial design perspective this is a random fluctuation: it would not persist in a similar large study. On the other hand once the imbalance is observed, unless it can be dismissed as irrelevant or unimportant it is a potential source of bias and is to be removed by rerandomizing or, if it is too late for that, by appropriate analysis. This aspect is connected with

some difficult conceptual issues about randomization; see Section 2.4.

This discussion raises at least two theoretical issues. The first concerns the possible gains from using a single quantitative baseline variable both as a basis for blocking and after that also as a basis for an adjustment. It can be shown that only when the correlation between baseline feature and response is very high is this double use of it likely to lead to a substantial improvement in final precision.

Suppose now that there are baseline features that cannot be reasonably controlled by blocking and that they are controlled by a regression adjustment. Is there any penalty associated with adjusting unnecessarily?

To study this consider first an experiment to compare two treatments, with r replicates of each. After adjustment for the $q \times 1$ vector of baseline variables, z, the variance of the estimated difference between the treatments is

$$\mathrm{var}(\hat{\tau}_T - \hat{\tau}_C) = \sigma^2\{2/r + (\bar{z}_{T.} - \bar{z}_{C.})^T R_{zz}^{-1}(\bar{z}_{T.} - \bar{z}_{C.})\}, \quad (3.42)$$

where σ^2 is the variance per observation residual to regression on z and to any blocking system used, $\bar{z}_{T.}, \bar{z}_{C.}$ are the treatment mean vectors and R_{zz} is the matrix of sums of squares and cross-products of z within treatments again eliminating any block effects.

Now if treatment assignment is randomized

$$E_R(R_{zz}/d_w) = \Omega_{zz}, \quad (3.43)$$

where d_w is the degrees of freedom of the residual sum of squares in the analysis of variance table, and Ω_{zz} is a finite population covariance matrix of the unit constants within blocks. With $v = 2$ we have

$$E_R(\bar{z}_T - \bar{z}_C) = 0, \quad E_R\{(\bar{z}_T - \bar{z}_C)(\bar{z}_T - \bar{z}_C)^T\} = 2\Omega_{zz}/r. \quad (3.44)$$

Now

$$\frac{1}{2}r(\bar{z}_T - \bar{z}_C)^T\Omega_{zz}^{-1}(\bar{z}_T - \bar{z}_C) = \frac{1}{2}r||\bar{z}_T - \bar{z}_C||^2_{\Omega_{zz}}, \quad (3.45)$$

say, has expectation q and approximately a chi-squared distribution with q degrees of freedom.

That is, approximately

$$\mathrm{var}(\hat{\tau}_T - \hat{\tau}_C) = \frac{2\sigma^2}{r}(1 + W_q/d_w), \quad (3.46)$$

where W_q denotes a random variable depending on the outcome of

the randomization and having approximately a chi-squared distribution with q degrees of freedom.

More generally if there are v treatments each replicated r times

$$\text{ave}_{j\neq l}\text{var}(\hat{\tau}_j - \hat{\tau}_l) = \sigma^2[2/r + 2/\{r(v-1)\}\text{tr}(B_{zz}R_{zz}^{-1})], \quad (3.47)$$

where B_{zz} is the matrix of sums of squares and products between treatments and $\text{tr}(A)$ denotes the trace of the matrix A, i.e. the sum of the diagonal elements.

The simplest interpretation of this is obtained by replacing W_q by its expectation, and by supposing that the number of units n is large compared with the number of treatments and blocks, so that $d_w \sim n$. Then the variance of an estimated treatment difference is approximately

$$\frac{2\sigma^2}{r}(1 + \frac{q}{n}). \quad (3.48)$$

The inflation factor relative to the randomized block design is approximately $n/(n-q)$ leading to the conclusion that every unnecessary parameter fitted, i.e. adjustment made without reduction in the effective error variance per unit, σ^2, is equivalent to the loss of one experimental unit.

This conclusion is in some ways oversimplified, however, not only because of the various approximations in its derivation. First, in a situation such as a clinical trial with a potentially large value of q, adjustments would be made selectively in a way depending on the apparent reduction of error variance achieved. This makes assessment more difficult but the inflation would probably be rather more than that based on q_0, the dimension of the z actually used, this being potentially much less than q, the number of baseline features available.

The second point is that the variance inflation, which arises because of the nonorthogonality of treatments and regression analyses in the least squares formulation, is a random variable depending on the degree of imbalance in the configuration actually used. Now if this imbalance can be controlled by design, for example by rerandomizing until the value of W_q is appreciably smaller than its expectation, the consequences for variance inflation are reduced and possibly but not necessarily the need to adjust obviated. If, however, such control at the design stage is not possible, the average inflation may be a poor guide. It is unlikely though that the

inflation will be more for small ϵ than

$$(1 + W_{q,\epsilon}/n), \tag{3.49}$$

where $W_{q,\epsilon}$ is the upper ϵ point of the randomization distribution of W_q, approximately a chi-squared distribution with q degrees of freedom.

For example, with $\epsilon = 0.01$ and $q = 10$ it will be unlikely that there is more than a 10 per cent inflation if $n > 230$ as compared with $n > 100$ suggested by the analysis based on properties averaged over the randomization distribution. Note that when the unadjusted and adjusted effects differ immaterially simplicity of presentation may favour the former.

A final point concerns the possible justification of the adjusted analysis based on randomization and the assumption of unit treatment additivity. Such a justification is usually only approximate but can be based on an approximate conditional distribution regarding, in the simplest case of just two treatments, $\bar{z}_T - \bar{z}_C$ as fixed.

3.7 Special models of error variation

In this chapter we have emphasized methods of error control by blocking which, combined with randomization, aim to increase the precision of estimated treatment contrasts without strong special assumptions about error structure. That is, while the effectiveness of the methods in improving precision depends on the way in which the blocks are formed, and hence on prior knowledge, the validity of the designs and the associated standard errors does not do so.

Sometimes, however, especially in relatively small experiments in which the experimental units are ordered in time or systematically arrayed in space a special stochastic model may reasonably be used to represent the error variation. Then there is the possibility of using a design that exploits that model structure. However, usually the associated method of analysis based on that model will not have a randomization justification and we will have to rely more strongly on the assumed model than for the designs discussed in this chapter.

When the experimental units are arranged in time the two main types of variation are a trend in time supplemented by totally random variation and a stationary time series representation. The latter is most simply formulated via a low order autoregression.

For spatial problems there are similar rather more complex representations. Because the methods of design and analysis associated with these models are more specialized we defer their discussion to Chapter 8.

3.8 Bibliographic notes

The central notions of blocking and of adjustment for baseline variables are part of the pioneering contributions of Fisher (1935), although the qualitative ideas especially of the former have a long history. The relation between the adjustment process and randomization theory was discussed by Cox (1982). See also the Bibliographic notes to Chapter 2. For the relative advantages of blocking and adjustment via a baseline variable, see Cox (1957).

The example in Section 3.4 is from Cochran and Cox (1958, Chapter 3), and the partitioning of the treatment sum of squares follows closely their discussion. The analysis of matched pairs and randomized blocks from the linear model is given in most books on design and analysis; see, for example, Montgomery (1997, Chapters 2 and 5) and Dean and Voss (1999, Chapter 10). The randomization analysis is given in detail in Hinkelmann and Kempthorne (1994, Chapter 9), as is the estimation of the efficiency of the randomized block design, following an argument attributed to Yates (1937).

3.9 Further results and exercises

1. Under what circumstances would it be reasonable to have a randomized block experiment in which each treatment occurred more than once, say, for example, twice, in each block, i.e. in which the number of units per block is twice the number of treatments? Set out the analysis of variance table for such a design and discuss what information is available that cannot be examined in a standard randomized block design.

2. Suppose in a matched pair design the responses are binary. Construct the randomization test for the null hypothesis of no treatment difference. Compare this with the test based on that for the binomial model, where Δ is the log odds-ratio. Carry out a similar comparison for responses which are counts of numbers of occurrences of point events modelled by the Poisson distribution.

3. Consider the likelihood analysis under the normal theory assumptions of the modified matched pair design of Section 3.3.2. There are r matched pairs, m_T pairs in which both units receive T and m_C pairs in which both units receive C; we assume a common treatment difference applies throughout. We transform the original pairs of responses to sums and differences as in Section 3.3.1.

   (a) Show that r of the differences have mean Δ, and that $m_T + m_C$ of them have mean zero, all differences being independently normally distributed with variance τ_D, say.

   (b) Show that independently of the differences the sums are independently normally distributed with variance τ_S, say, with r having mean ν, say, m_T having mean $\nu+\delta$ and m_C having mean $\nu-\delta$, where $\Delta = 2\delta$.

   (c) Hence show that minimal sufficient statistics are (i) the least squares estimate of ν from the sums; (ii) the least squares estimate $\hat{\Delta}_S$ of Δ from the unmatched pairs, i.e. the difference of the means of m_T and m_C pairs; (iii) the estimate $\hat{\Delta}_D$ from the matched pairs; (iv) a mean square MS_D with $d_D = r - 1 + m_T + m_C$ degrees of freedom estimating τ_D and (v) a mean square MS_S with $d_S = r - 2 + m_T + m_C$ degrees of freedom estimating τ_S. This shows that the system is a $(5, 4)$ curved exponential family.

   (d) Without developing a formal connection with randomization theory note that complete randomization of pairs to the three groups would give some justification to the strong homogeneity assumptions involved in the above. How would such homogeneity be examined from the data?

   (e) Show that a log likelihood function obtained by ignoring (i) and using the known densities of the four remaining statistics is

$$
\begin{aligned}
- & \frac{1}{2}\log\tau_S - \tilde{m}(\hat{\Delta}_S - \Delta)^2/(2\tau_S) \\
- & \frac{1}{2}\log\tau_D - r(\hat{\Delta}_D - \Delta)^2/(2\tau_D) \\
- & \frac{1}{2}d_D\log\tau_D - \frac{1}{2}d_D\mathrm{MS}_D/\tau_D \\
- & \frac{1}{2}d_S\log\tau_S - \frac{1}{2}d_S\mathrm{MS}_S/\tau_S,
\end{aligned}
$$

where $1/\tilde{m} = 1/m_D + 1/m_S$.

(f) Hence show, possibly via some simulated data, that only in quite small samples will the profile likelihood for Δ differ appreciably from that corresponding to a weighted combination of the two estimates of Δ replacing the variances and theoretically optimal weights by sample estimates and calculating confidence limits via the Student t distribution with effective degrees of freedom

$$\tilde{d} = (r\mathrm{MS}_S + \tilde{m}\mathrm{MS}_D)^2(r^2\mathrm{MS}_S^2/d_D + \tilde{m}^2\mathrm{MS}_D^2/d_S)^{-1}.$$

For somewhat related calculations, see Cox (1984b).

4. Suppose that n experimental units are arranged in sequence in time and that there is prior evidence that the errors are likely to be independent and identically distributed initially with mean zero except that at some as yet unknown point there is likely to be a shift in mean error. What design would be appropriate for the comparison of v treatments? After the experiment is completed and the responses obtained it is found that the discontinuity has indeed occurred. Under the usual linear assumptions what analysis would be suitable if

(a) the position of the discontinuity can be determined without error from supplementary information

(b) the position of the discontinuity is regarded as an unknown parameter.