Chapter 8

Small Units inside Large Units

8.1 Experimental units bigger than observational units

8.1.1 The context

Example 8.1 (Example 1.2 continued: Calf-feeding) Four feed treatments are compared on 80 calves. The calves are kept in eight pens, with 10 calves per pen. Each feed is applied to two whole pens, with calves feeding ad lib. Even though every calf is weighed individually, differences between feeds should be compared with differences between pens, not with differences between calves. So should we say that the replication is 20 or 2?

Example 1.8 is similar: treatments are applied to whole classes but individual children are measured. So is one version of Example 1.11: pullets are fed but eggs are weighed. So are Examples 7.7 and 7.8: treatments are applied to whole doctors' practices or to villages, but it is individual people that are measured.

In general, suppose that there are m experimental units, each of which consists of k observational units, and that there are t treatments, each of which is applied to s experimental units, so that $st = m$.

8.1.2 Construction and randomization

This is just like the completely randomized design in Section 2.1, using the experimental units instead of the plots.

8.1.3 Model and strata

For simplicity, I use the language of Example 8.1. If pens give fixed effects then we cannot estimate anything about treatments, so we assume that pens give random

effects. This gives a model like the random-effects model in Sections 4.4 and 4.6:

$$\mathbb{E}(Y_\omega) = \tau_{T(\omega)}$$

for each calf ω, and

$$\text{cov}(Y_\alpha, Y_\beta) = \begin{cases} \sigma^2 & \text{if } \alpha = \beta \\ \rho_1\sigma^2 & \text{if } \alpha \text{ and } \beta \text{ are different calves in the same pen} \\ \rho_2\sigma^2 & \text{if } \alpha \text{ and } \beta \text{ are in different pens.} \end{cases}$$

Usually $\rho_1 > \rho_2 > 0$ unless there is so little food that there is competition among calves in the same pen.

Put

$$V_P = \{\text{vectors in } V \text{ which take a constant value on each pen}\}$$

and

$$W_P = V_P \cap V_0^\perp.$$

From Section 4.6, we see that the eigenspaces of $\text{Cov}(\mathbf{Y})$ are

	V_0	W_P	$V_P^\perp$
with dimensions	1	$m-1$	$m(k-1)$
and eigenvalues	ξ_0	ξ_1	ξ_2,

where

$$\begin{aligned} \xi_0 &= \sigma^2(1-\rho_1) + \sigma^2 k(\rho_1 - \rho_2) + \sigma^2 bk\rho_2 \\ \xi_1 &= \sigma^2(1-\rho_1) + \sigma^2 k(\rho_1 - \rho_2) \\ \xi_2 &= \sigma^2(1-\rho_1). \end{aligned}$$

If $\rho_1 > \rho_2$ then $\xi_1 > \xi_2$; if $\rho_2 > 0$ then $\xi_0 > \xi_1$.

8.1.4 Analysis

Here I show how to build up the analysis-of-variance table in stages, so that it does not have to be memorized. The first step is to write out the list of strata and their degrees of freedom in a table, with one row for each stratum. This table is called the *null analysis of variance*. It is shown in Table 8.1.

The next two steps can be done in either order. The second step can be done before treatments are thought about. Simply take the rows of the null analysis-of-variance table, and show the calculations associated with each one. See Table 8.2. The first column contains sums of squares. As usual, the sum of squares for the mean is sum^2/N. The sum of squares for pens, SS(pens), is equal to CSS(pens) − SS(mean), where

$$\text{CSS(pens)} = \sum_{i=1}^{m} \frac{\text{sum}^2_{\text{pen}=i}}{k}.$$

Stratum	df
mean	1
pens	$m-1$
calves	$m(k-1)$
Total	$mk=N$

Table 8.1: Null analysis of variance

SS	EMS
sum^2/N	ξ_0
$\sum_i \mathsf{sum}^2_{\mathrm{pen}=i}/k - \mathrm{SS}(\mathrm{mean})$	ξ_1
$\sum y_\omega^2 - \mathrm{CSS}(\mathrm{pens})$	ξ_2
$\sum y_\omega^2$	

Table 8.2: Calculations ignoring treatments

The total sum of squares is

$$\sum_{\omega\in\Omega} y_\omega^2,$$

and the sum of squares for calves is obtained by subtraction, so that it is equal to $\sum y_\omega^2 - \mathrm{CSS}(\mathrm{pens})$.

The second column contains the expected mean squares if there are no treatments. These are just the appropriate eigenvalues of $\mathrm{Cov}(\mathbf{Y})$.

The third step is to take the null analysis of variance and expand it to show where treatments lie. This gives the *skeleton analysis of variance*, shown in Table 8.3. An extra column for source is inserted after the stratum column, and the degrees-of-freedom column is split into two. If there is any stratum that contains no treatment subspace (except for V_0), then its name is copied into the source column and its degrees of freedom are written on the right-hand side of the 'df' column. This rule gives the rows labelled 'mean' and 'calves' in Table 8.3. The next lemma shows why there are no treatment subspaces in the calves stratum.

Lemma 8.1 *If treatments are applied to whole pens then $W_T \subseteq W_P$.*

Proof Let $\mathbf{v}$ be in V_T. If α and β are in the same pen then $T(\alpha)=T(\beta)$ so $v_\alpha = v_\beta$ because $\mathbf{v}\in V_T$. Hence $\mathbf{v}\in V_P$. Therefore $V_T \subseteq V_P$, so $W_T = V_T \cap V_0^\perp \subseteq V_P \cap V_0^\perp = W_P$. ■

Since the pens stratum contains W_T, which is the subspace for differences between feeds, the row labelled 'pens' in the null analysis-of-variance table has to be split into three, with names shown in the source column. The first is W_T, labelled 'feed', whose degrees of freedom are shown on the left-hand side of the 'df' column. The third is labelled 'total'; it is really a copy of the 'pens' row in Table 8.1 so its degrees of freedom are shown on the right of the 'df' column. In between is the source labelled 'residual', which is the space $W_P \cap W_T^\perp$. Its degrees of freedom, shown on the left of the 'df' column, are calculated by subtraction as $(m-1)-(t-1)=m-t$.

The fourth step is to combine Tables 8.2 and 8.3 and insert extra information to give Table 8.4, which is the *full* analysis of variance. The sums of squares for the

Stratum	source	degrees of freedom
mean	mean	1
pens	feed	$t-1$
	residual	$m-t$
	total	$m-1$
calves	calves	$m(k-1)$
Total		N

Table 8.3: Skeleton analysis of variance

mean and calves strata are copied from Table 8.2 to Table 8.4. In the pens stratum, the sum of squares is copied from Table 8.2 to the 'total' line of the pens stratum in Table 8.4. The sum of squares for treatments is calculated in the usual way as

$$\mathrm{SS}(\text{feed}) = \sum_{i=1}^{t} \frac{\mathsf{sum}_{T=i}^2}{sk} - \frac{\mathsf{sum}^2}{N}.$$

The sum of squares for residual is obtained by subtraction:

$$\mathrm{SS}(\text{residual}) = \mathrm{SS}(\text{total}) - \mathrm{SS}(\text{feed}).$$

If there is space, it is useful to split the 'sum of squares' column into two just like the 'df' column.

Expected mean squares are copied from Table 8.2 to Table 8.4, being put in the right-hand side of the 'EMS' column. Thus ξ_0 is written for the mean, ξ_2 for calves, and ξ_1 for both of the non-total sources in the pens stratum. To these stratum variances must be added sums of squares of expected values. As in Section 4.5, $\mathbb{E}(P_{V_0}\mathbf{Y}) = \tau_0$ and $\mathbb{E}(P_{W_T}\mathbf{Y}) = \tau_T$, so Theorem 2.10(ii) shows that $\mathrm{EMS}(\text{mean}) = \|\tau_0\|^2 + \xi_0$,

$$\mathrm{EMS}(\text{feed}) = \frac{\|\tau_T\|^2}{t-1} + \xi_1,$$

$\mathrm{EMS}(\text{residual}) = \xi_1$ and $\mathrm{EMS}(\text{calves}) = \xi_2$.

The final column shows what is the appropriate variance ratio, if any. There is no independent estimate of ξ_0, so there is no test for whether τ_0 is zero. If $\tau_T = \mathbf{0}$ then $\mathrm{EMS}(\text{feed}) = \mathrm{EMS}(\text{residual})$ so the appropriate variance ratio to test for differences between feeds is

$$\frac{\mathrm{MS}(\text{feed})}{\mathrm{MS}(\text{residual})}.$$

8.1.5 Hypothesis testing

Test for difference between feeds by using the variance ratio

$$\frac{\mathrm{MS}(\text{feed})}{\mathrm{MS}(\text{residual})}.$$

Stratum	source	df	SS	EMS	VR
mean	mean	1	$\dfrac{\text{sum}^2}{N}$	$\|\tau_0\|^2+\xi_0$	–
pens	feed	$t-1$	$\dfrac{\sum \text{sum}_{T=i}^2}{sk} - \dfrac{\text{sum}^2}{N}$	$\dfrac{\|\tau_T\|^2}{t-1}+\xi_1$	$\dfrac{\text{MS(feed)}}{\text{MS(residual)}}$
	residual	$\leftarrow \ldots\ldots\ldots\ldots$	by subtraction $\ldots\ldots\ldots\ldots \rightarrow$	ξ_1	–
	total	$m-1$	$\dfrac{\sum P_i^2}{k} - \dfrac{\text{sum}^2}{N}$		
calves	calves	$m(k-1)$	$\sum y_\omega^2 - \dfrac{\sum P_i^2}{k}$	ξ_2	–
Total		N	$\sum_\omega y_\omega^2$		

Table 8.4: Full analysis of variance: here P_i denotes the total of the data on the i-th pen

The appropriate numbers of degrees of freedom for the F-test are $t-1$ and $m-t$.

If $s=1$ then $m=t$ and so there is no residual in the pens stratum and so no test can be done. In this case people sometimes wrongly compare MS(feed) to MS(calves). If $\rho_1 > \rho_2$ then $\xi_1 > \xi_2$ so this test is likely to detect differences that do not really exist. This is known as *false replication* or *pseudo-replication*, because the replication of calves within the pens is falsely taken as a substitute for the replication of pens within treatments.

Probably the single most common mistake in designing experiments is to allocate treatments to large units with $s=1$. This is the mistake in the ladybirds experiment in Example 1.1.

Even when $s \neq 1$, some people mistakenly combine the sum of squares for residual in the pens stratum with the calves sum of squares, to give a source whose expected mean square is equal to

$$\frac{(m-t)\xi_1 + m(k-1)\xi_2}{N-t}.$$

This is smaller than ξ_1 if $\rho_1 > \rho_2$, so again there is spurious precision. There are also more degrees of freedom than there should be, which leads to spurious power.

Note that increasing k does not increase the number of degrees of freedom for residual.

8.1.6 Decreasing variance

Theorem 2.10(iv) shows that the variance of the estimator of a simple treatment difference $\tau_i - \tau_j$ is equal to

$$\frac{2}{sk}\xi_1 = \frac{2}{s}\sigma^2\left[\frac{(1-\rho_1)}{k} + (\rho_1 - \rho_2)\right],$$

which is estimated by $(2/sk)\,\mathrm{MS(residual)}$. To decrease the variance, it is more fruitful to increase s than to increase k, even though the experimenter may find it simpler to increase k.

Example 8.1 revisited (Calf-feeding) Plausible values for the two correlations are $\rho_1 = 0.3$ and $\rho_2 = 0$. Then the variance of the estimator of a difference is equal to

$$2\sigma^2\left[\frac{0.7}{sk} + \frac{0.3}{s}\right].$$

Some combinations of values of s and k give the following coefficients of $2\sigma^2$.

	$k=7$	$k=10$	$k=15$
$s=2$	0.200	0.185	0.173
$s=3$	0.133	0.123	0.116

The original design has variance $2\sigma^2 \times 0.185$. Increasing the number of calves per pen from 10 to 15 only decreases the variance to $2\sigma^2 \times 0.173$, even though 50%

more feed is used. On the other hand, replacing 8 pens of 10 calves by 12 pens of 7 calves decreases the variance substantially to $2\sigma^2 \times 0.133$ while using almost the same number of calves and almost the same amount of feed. However, the extra four pens may not be available, or they may be expensive to set up.

It is almost always easier for the experimenter to randomize treatments to a few large units than to many smaller units. Part of the statistician's job is to explain the loss of precision and power that will result from this.

8.2 Treatment factors in different strata

Example 8.1 revisited (Calf-feeding) Suppose that the four feeds consist of all combinations of two types of hay, which is put directly into the pen, with two types of cake, which is fed to calves individually. If all calves in the same pen have the same type of cake then the design and analysis are just as in Section 8.1, except that the treatments line in the analysis of variance is split into three, giving the skeleton analysis of variance in Table 8.5.

Stratum	source	degrees of freedom
mean	mean	1
pens	hay	1
	cake	1
	hay $\wedge$ cake	1
	residual	4
	total	7
calves	calves	72
Total		80

Table 8.5: Skeleton analysis of variance when types of hay and types of cake are both applied to whole pens

It might be better to give five calves in each pen one type of cake and the other five calves the other type of cake.

In general, suppose that there are m large units, each of which consists of k small units; and that there are n_H levels of factor H, each of which is applied to r_H large units (so that $n_H r_H = m$); and that there are n_C levels of treatment factor C, each of which is applied to r_C small units per large unit (so that $n_C r_C = k$). Now the number of treatments is $n_H n_C$ and they all have replication $r_H r_C$. Thus $N = mk = n_H n_C r_H r_C$.

The construction and randomization are as follows.

(i) Apply levels of H to large units just like a completely randomized design.

(ii) Within each large unit independently, apply levels of C just like a completely randomized design.

Let L be the factor for large units, so that $L(\omega)$ is the large unit which contains small unit ω. Define subspaces V_L and W_L like the subspaces V_P and W_P in Section 8.1.

Theorem 8.2 *If levels of H are applied to large units with equal replication and levels of C are applied to small units with equal replication within each large unit, then*

(i) $W_H \subseteq W_L$;

(ii) $W_C \subseteq V_L^{\perp}$;

(iii) $W_{H \wedge C} \subseteq V_L^{\perp}$.

Proof (i) Use Lemma 8.1.

(ii) and (iii) Let Δ be any large unit. Suppose that H has level h on Δ. Let $\mathbf{v}$ be any vector in W_C or $W_{H \wedge C}$. Then $\mathbf{v} \in V_T$ and $\mathbf{v} \in V_H^{\perp}$. The r_H large units on which H has level h all have the same treatments, so they all have the same values of $\mathbf{v}$, although probably in different orders. Therefore

$$\sum_{\omega \in \Gamma} v_\omega = r_H \sum_{\omega \in \Delta} v_\omega, \tag{8.1}$$

where Γ is the union of the large units on which H takes level h. Let $\mathbf{w}_\Gamma$ be the vector whose entry on plot ω is equal to

$$\begin{cases} 1 & \text{if } \omega \in \Gamma \\ 0 & \text{otherwise,} \end{cases}$$

and define the vector $\mathbf{w}_\Delta$ analogously. Equation (8.1) can be rewritten as

$$\mathbf{v} \cdot \mathbf{w}_\Gamma = r_H \mathbf{v} \cdot \mathbf{w}_\Delta.$$

Now, $\mathbf{w}_\Gamma$ is in V_H and $\mathbf{v}$ is orthogonal to V_H, so $\mathbf{v} \cdot \mathbf{w}_\Gamma = 0$ so $\mathbf{v} \cdot \mathbf{w}_\Delta = 0$. This is true for all large units Δ, and V_L is spanned by the vectors $\mathbf{w}_\Delta$ as Δ runs over the large units, so $\mathbf{v}$ is orthogonal to V_L. ■

The analysis-of-variance table is built up as in Section 8.1. The null analysis of variance is shown in Table 8.6 and the calculations ignoring treatments in Table 8.7. Theorem 8.2 shows us how to expand Table 8.6 to the skeleton analysis of variance in Table 8.8. The row for the large units stratum must be split to give a line for H, a residual line and a total line. The total line is the same as in the null analysis of variance, and the residual is the difference between total and H. The row for the small units stratum must be split even further: it has to show two treatment lines, the main effect of C and the H-by-C interaction; the residual is obtained by subtracting both of these from the total. Note that we now have two separate lines called ‘residual’, one in the large units stratum and one in the small units stratum.

Stratum	df
mean	1
large units	$m-1$
small units	$m(k-1)$
Total	$mk=N$

Table 8.6: Null analysis of variance

SS	EMS
sum^2/N	ξ_0
$\sum_i \mathsf{sum}^2_{L=i}/k - \mathrm{SS}(\mathrm{mean})$	ξ_1
$\sum y_\omega^2 - \mathrm{CSS}(\text{large units})$	ξ_2
$\sum y_\omega^2$	

Table 8.7: Calculations ignoring treatments

Stratum	source	degrees of freedom
mean	mean	1
large units	H	n_H-1
	residual	$m-n_H$
	total	$m-1$
small units	C	n_C-1
	$H\wedge C$	$(n_H-1)(n_C-1)$
	residual	$m(k-1)-n_H(n_C-1)$
	total	$m(k-1)$
Total		N

Table 8.8: Skeleton analysis of variance

Where necessary, I distinguish these by writing, for example, 'large units residual' in full.

Tables 8.7 and 8.8 are combined to give the full analysis of variance in Table 8.9. Here $\tau_0 = P_{V_0}\tau$, $\tau_H = P_{W_H}\tau$, $\tau_C = P_{W_C}\tau$, $\tau_{HC} = P_{W_{H\wedge C}}\tau$,

$$\begin{aligned}
\mathrm{CSS}(L) &= \sum_{l=1}^{m} \frac{\mathsf{sum}^2_{L=l}}{k}, \\
\mathrm{SS}(H) &= \sum_{i=1}^{n_H} \frac{\mathsf{sum}^2_{H=i}}{r_H k} - \frac{\mathsf{sum}^2}{N}, \\
\mathrm{SS}(C) &= \sum_{j=1}^{n_C} \frac{\mathsf{sum}^2_{C=j}}{r_C m} - \frac{\mathsf{sum}^2}{N}
\end{aligned}$$

and

$$\mathrm{SS}(H \wedge C) = \sum_{i=1}^{n_H} \sum_{j=1}^{n_C} \frac{\mathsf{sum}^2_{H=i,C=j}}{r_H r_C} - \frac{\mathsf{sum}^2}{N} - \mathrm{SS}(H) - \mathrm{SS}(C).$$

Now differences between levels of H are assessed against the variability of large units. Testing uses

$$\frac{\mathrm{MS}(H)}{\mathrm{MS}(\text{large units residual})}.$$

The standard error of a difference between levels of H is

$$\sqrt{\frac{2}{r_H k}\,\mathrm{MS}(\text{large units residual})},$$

and the variance of the estimator of such a difference is $2\xi_1/r_H k$.

The main effect of C and the H-by-C interaction are assessed against the variability of small units within large units. The tests use the variance ratios

$$\frac{\mathrm{MS}(C)}{\mathrm{MS}(\text{small units residual})}$$

and

$$\frac{\mathrm{MS}(H \wedge C)}{\mathrm{MS}(\text{small units residual})}$$

respectively.

The variance of the estimator of the difference between two levels of C is $2\xi_2/r_C m$, so the standard error of a such a difference is

$$\sqrt{\frac{2}{r_C m}\,\mathrm{MS}(\text{small units residual})}.$$

Standard errors of differences between indiviual treatments require a little more care. For $i = 1, \ldots, n_H$ and $j = 1, \ldots, n_C$, let $\mathbf{u}_{ij}$ be the vector in V_T whose coordinate on plot ω is equal to

$$\begin{cases} 1 & \text{if } H(\omega) = i \text{ and } C(\omega) = j \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, let $\mathbf{v}_i$ be the vector whose coordinate on ω is equal to

$$\begin{cases} 1 & \text{if } H(\omega) = i \\ 0 & \text{otherwise} \end{cases}$$

and $\mathbf{w}_j$ the vector whose coordinate on ω is equal to

$$\begin{cases} 1 & \text{if } C(\omega) = j \\ 0 & \text{otherwise.} \end{cases}$$

Fix levels i, i' of H and j, j' of C, with (i, j) different from (i', j'). Put $\mathbf{u} = (\mathbf{u}_{ij} - \mathbf{u}_{i'j'})/r$, where $r = r_H r_C$. Then $\mathbf{u} \in V_0^{\perp}$. Let $\mathbf{v}$ be the projection of $\mathbf{u}$ onto

Stratum	source	df	SS	EMS	VR
mean	mean	1	$\frac{\text{sum}^2}{N}$	$\|\tau_0\|^2+\xi_0$	–
large units	H	n_H-1	$\text{SS}(H)$	$\frac{\|\tau_H\|^2}{n_H-1}+\xi_1$	$\frac{\text{MS}(H)}{\text{MS(large units residual)}}$
	residual	←…… by subtraction…… →		ξ_1	–
	total	$m-1$	$\text{CSS}(L)-\frac{\text{sum}^2}{N}$		
small units	C	n_C-1	$\text{SS}(C)$	$\frac{\|\tau_C\|^2}{n_C-1}+\xi_2$	$\frac{\text{MS}(C)}{\text{MS(small units residual)}}$
	$H\wedge C$	d_{HC}	$\text{SS}(H\wedge C)$	$\frac{\|\tau_{HC}\|^2}{d_{HC}}\xi_2$	$\frac{\text{MS}(H\wedge C)}{\text{MS(small units residual)}}$
	residual	←…… by subtraction…… →		ξ_2	–
	total	$m(k-1)$	$\Sigma y_\omega^2-\text{CSS}(L)$		
Total		N	$\Sigma_\omega y_\omega^2$		

Table 8.9: Full analysis of variance: here $d_{HC}=(n_H-1)(n_C-1)$

V_H. Theorem 2.3(vii) shows that $\mathbf{v} = (\mathbf{v}_i - \mathbf{v}_{i'})/rn_C$, which is in W_H, so $\mathbf{v} = P_{W_H}\mathbf{u}$. Similarly, if $\mathbf{w} = P_{W_C}\mathbf{u}$ then $\mathbf{w} = (\mathbf{w}_j - \mathbf{w}_{j'})/rn_H$. Therefore

$$\mathbf{u} = \mathbf{v} + \mathbf{w} + \mathbf{x}, \tag{8.2}$$

where $\mathbf{x} \in W_{H \wedge C}$.

The vectors on the right-hand side of Equation (8.2) are mutually orthogonal, and so

$$\|\mathbf{x}\|^2 = \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 - \|\mathbf{w}\|^2.$$

Now, $\|\mathbf{u}\|^2 = 2/r$; $\|\mathbf{v}\|^2$ is equal to $2/rn_C$ if $i \neq i'$ and to zero otherwise; and $\|\mathbf{w}\|^2$ is equal to $2/rn_H$ if $j \neq j'$ and to zero otherwise. Therefore

$$\|\mathbf{x}\|^2 = \begin{cases} 2(n_H n_C - n_H - n_C)/N & \text{if } i \neq i' \text{ and } j \neq j' \\ 2n_C(n_H - 1)/N & \text{if } i = i' \text{ and } j \neq j' \\ 2n_H(n_C - 1)/N & \text{if } i \neq i' \text{ and } j = j'. \end{cases}$$

Parts (iii) and (iv) of Theorem 2.10 show that $\mathbf{v} \cdot \mathbf{Y}$ is the best linear unbiased estimator of $\mathbf{v} \cdot \tau$ and that it has variance $\|\mathbf{v}\|^2\xi_1$. Similarly, $\mathbf{w} \cdot \mathbf{Y}$ is the best linear unbiased estimator of $\mathbf{w} \cdot \tau$ and it has variance $\|\mathbf{w}\|^2\xi_2$, and $\mathbf{x} \cdot \mathbf{Y}$ is the best linear unbiased estimator of $\mathbf{x} \cdot \tau$ and it has variance $\|\mathbf{x}\|^2\xi_2$.

Hence $\mathbf{u} \cdot \mathbf{Y}$ is the best linear unbiased estimator of $\mathbf{u} \cdot \tau$, which is equal to $\tau_{ij} - \tau_{i'j'}$. Theorem 2.10(v) shows that three estimators are uncorrelated, and so

$$\mathrm{Var}(\mathbf{u} \cdot \mathbf{Y}) = \|\mathbf{v}\|^2\xi_1 + \|\mathbf{w}\|^2\xi_2 + \|\mathbf{x}\|^2\xi_2.$$

If $i = i'$ while $j \neq j'$ then

$$\begin{aligned} \mathrm{Var}(\mathbf{u} \cdot \mathbf{Y}) &= \left[\frac{2}{rn_H} + \frac{2n_C(n_H - 1)}{N}\right]\xi_2 \\ &= \frac{2n_C}{N}\left[1 + (n_H - 1)\right]\xi_2 \\ &= \frac{2n_C n_H}{N}\xi_2 \\ &= \frac{2}{r}\xi_2. \end{aligned}$$

If $i \neq i'$ but $j = j'$ then

$$\begin{aligned} \mathrm{Var}(\mathbf{u} \cdot \mathbf{Y}) &= \frac{2\xi_1}{rn_C} + \frac{2n_H(n_C - 1)\xi_2}{N} \\ &= \frac{2}{N}\left[n_H\xi_1 + n_H(n_C - 1)\xi_2\right] \\ &= \frac{2}{rn_C}\left[\xi_1 + (n_C - 1)\xi_2\right]. \end{aligned}$$

If $i \neq i'$ and $j \neq j'$ then

$$\begin{aligned}\mathrm{Var}(\mathbf{u}\cdot\mathbf{Y}) &= \frac{2\xi_1}{rn_C}+\frac{2\xi_2}{rn_H}+\frac{2(n_Hn_C-n_H-n_C)\xi_2}{N}\\ &= \frac{2}{N}\left[n_H\xi_1+n_H(n_C-1)\xi_2\right]\\ &= \frac{2}{rn_C}\left[\xi_1+(n_C-1)\xi_2\right].\end{aligned}$$

Thus the standard error of a difference is equal to

$$\sqrt{\frac{2}{r}\,\mathrm{MS}(\text{small units residual})}$$

if the levels of H are the same; otherwise it is equal to

$$\sqrt{\frac{2}{rn_C}[\mathrm{MS}(\text{large units residual})+(n_C-1)\,\mathrm{MS}(\text{small units residual})]}.$$

Since we expect ξ_2 to be smaller than ξ_1, this design gives more precise estimates of differences between levels of C than the design in Section 8.1. It also gives more precise estimates of differences between any two levels of $H \wedge C$. There is increased power for testing the main effect of C and the H-by-C interaction, for not only is ξ_2 smaller than ξ_1 but the number of residual degrees of freedom in the small units stratum is substantially more than the number of residual degrees of freedom in the large units stratum in the design in Section 8.1. There is even a slight increase in power for testing the main effect of H, because the number of residual degrees of freedom in the large units stratum increases from $m-n_Hn_C$ to $m-n_H$.

Example 8.1 revisited (Calf-feeding) If each type of hay is applied to four whole pens and each type of cake is given to five calves per pen, then we obtain the skeleton analysis of variance in Table 8.10. The improved precision for cake and for hay $\wedge$ cake is immediately evident.

To see the improved power for testing for the main effect of hay, we argue as we did in Sections 2.12 and 4.8. There is one degree of freedom for hay, and an F-test with one degree of freedom for the numerator is equivalent to a two-sided t-test on the square root of the statistic. The vector τ_H has just two distinct entries, one for each type of hay. Let δ be the modulus of their difference.

The design summarized in Table 8.5 has 4 residual degrees of freedom in the pens stratum. If we test at the 5% significance level then we need the 0.975 point of the t-distribution on 4 degrees of freedom, which is 2.776. The 0.90 point is 1.553. Thus the argument of Section 2.12 shows that to have probability at least 0.9 of detecting that the main effect of hay is non-zero we need

$$\delta > (2.776+1.533)\times\sqrt{\frac{2}{40}\xi_1} = 4.309\sqrt{\frac{\xi_1}{20}}.$$

Stratum	source	degrees of freedom
mean	mean	1
pens	hay	1
	residual	6
	total	7
calves	cake	1
	hay $\wedge$ cake	1
	residual	70
	total	72
Total		80

Table 8.10: Skeleton analysis of variance when types of hay are applied to whole pens and each type of cake is given to five calves per pen

On the other hand, the design summarized in Table 8.10 has 6 residual degrees of freedom in the pens stratum. The 0.975 and 0.90 points of the t-distribution on 6 degrees of freedom are 2.447 and 1.440 respectively. In order to have probability at least 0.9 of detecting that the main effect of hay is non-zero, we need

$$\delta > (2.447 + 1.440) \times \sqrt{\frac{2}{40}\xi_1} = 3.887\sqrt{\frac{\xi_1}{20}}.$$

Thus the second design can detect a difference between the types of hay which is 10% smaller than the difference detectable by the first design.

8.3 Split-plot designs

8.3.1 Blocking the large units

In the situtation described in either of Sections 8.1–8.2, the large units may themselves be grouped into blocks, as in Chapter 4. Then it is convenient to think of the large units as *small blocks* and the blocks as *large blocks*. Thus we have b large blocks each of which contains s small blocks each of which contains k plots.

Such large and small blocks often occur naturally.

Example 8.2 (Example 1.5 continued: Rye-grass) Here the large blocks are the fields and the small blocks are the strips. Thus $b = 2$, $s = 3$ and $k = 4$.

Example 8.3 (Animal-breeding) In animal breeding it is typical to mate each *sire* (male parent) with several *dams* (female parents), each of which may then produce several offspring. Then the large blocks are the sires, the small blocks are the dams, and the plots are the offspring.

Confusingly, the small blocks are sometimes called *whole plots* or *main plots* while the plots are called *subplots*. This is a historical accident based on a famous experiment at Rothamsted Experimental Station.

Example 8.4 (Example 5.11 continued: Park Grass) After the Park Grass experiment had been running for many decades, the soil became over-acidic. In 1903 each plot was split into two subplots. Thereafter, lime was applied to one subplot per plot in every fourth year. Although the subplot to be limed was not chosen randomly in the way described below, as far as I know this is the origin of the idea of splitting plots into subplots for the application of levels of a later treatment factor.

For simplicity, I shall describe only the classic *split-plot* design. This is like the second design in Section 8.2 except that

- the large units (small blocks) are grouped into b large blocks of size s;
- each level of H is applied to one small block per large block (so $n_H = s$ and $r_H = b$);
- each level of C is applied to one plot per small block (so $n_C = k$ and $r_C = 1$).

Example 8.2 is a classic split-plot design.

8.3.2 Construction and randomization

(i) Apply levels of H to small blocks just as in a complete-block design.

(ii) Within each small block independently, apply levels of C just as in a completely randomized design.

8.3.3 Model and strata

We assume that large blocks and small blocks both give random effects. Thus

$$\mathbb{E}(Y_\omega) = \tau_{H(\omega)C(\omega)}$$

for each plot ω, and

$$\operatorname{cov}(Y_\alpha, Y_\beta) = \begin{cases} \sigma^2 & \text{if } \alpha = \beta \\ \rho_1\sigma^2 & \text{if } \alpha \text{ and } \beta \text{ are different plots in the same small block} \\ \rho_2\sigma^2 & \text{if } \alpha \text{ and } \beta \text{ are in different small blocks in the same large block} \\ \rho_3\sigma^2 & \text{if } \alpha \text{ and } \beta \text{ are in different large blocks.} \end{cases} \tag{8.3}$$

Usually $\rho_1 > \rho_2 > \rho_3$.

Define V_B as in Chapter 4 and put $W_B = V_B \cap V_0^\perp$, so that $\dim V_B = b$ and $\dim W_B = b - 1$. Further, put

$$V_S = \{\text{vectors in } V \text{ which take a constant value on each small block}\}.$$

Then $\dim V_S = bs$. The argument in the proof of Lemma 8.1 shows that $V_B \subseteq V_S$, so it is natural to put $W_S = V_S \cap V_B^\perp$. Then $\dim W_S = \dim V_S - \dim V_B = bs - b = b(s-1)$. Finally, $\dim V_S^\perp = N - bs = bsk - bs = bs(k-1)$.

The proof of the following proposition is an extended version of the argument that finds the eigenspaces and eigenvectors of $\mathrm{Cov}(\mathbf{Y})$ in Section 4.6.

Proposition 8.3 *Under assumption (8.3), the eigenspaces of* $\mathrm{Cov}(\mathbf{Y})$ *are* W_0, W_B, W_S *and* $V_S^\perp$, *with eigenvalues* ξ_0, ξ_B, ξ_S *and* ξ *respectively, where*

$$\begin{aligned}
\xi_0 &= \sigma^2[(1-\rho_1) + k(\rho_1-\rho_2) + ks(\rho_2-\rho_3) + bks\rho_3] \\
\xi_B &= \sigma^2[(1-\rho_1) + k(\rho_1-\rho_2) + ks(\rho_2-\rho_3)] \\
\xi_S &= \sigma^2[(1-\rho_1) + k(\rho_1-\rho_2)] \\
\xi &= \sigma^2(1-\rho_1).
\end{aligned}$$

8.3.4 Analysis

Stratum		df
W_0	mean	1
W_B	large blocks	$b-1$
W_S	small blocks	$b(s-1)$
$V_S^\perp$	plots	$bs(k-1)$
	Total	bsk

Table 8.11: Null analysis of variance for the classic split-plot design

SS	EMS
sum^2/N	ξ_0
$\sum_i \mathsf{sum}^2_{B=i}/sk - \mathrm{SS}(\mathrm{mean})$	ξ_B
$\sum_j \mathsf{sum}^2_{S=j}/k - \mathrm{CSS}(B)$	ξ_S
$\sum y_\omega^2 - \mathrm{CSS}(S)$	ξ
$\sum y_\omega^2$	

Table 8.12: Calculations ignoring treatments for the classic split-plot design

Proposition 8.3 shows that the null analysis of variance is as shown in Table 8.11. The calculations ignoring treatments are shown in Table 8.12, where S is the factor for small blocks,

$$\mathrm{CSS}(B) = \mathrm{CSS}(\text{large blocks}) = \sum_{i=1}^{b} \frac{\mathsf{sum}^2_{B=i}}{sk},$$

and

$$\mathrm{CSS}(S) = \mathrm{CSS}(\text{small blocks}) = \sum_{j=1}^{bs} \frac{\mathsf{sum}^2_{S=j}}{k}.$$

To work out the skeleton analysis of variance, we have to decide which stratum contains each treatment subspace. Every treatment occurs once in each block, so Theorem 4.1 shows that W_T is orthogonal to V_B. The argument of Lemma 8.1 shows that $W_H \subseteq V_S$. Since $W_H \subset W_T$, we obtain $W_H \subseteq V_S \cap V_B^\perp = W_S$. Theorem 8.2 shows that W_C and $W_{H\wedge C}$ are both contained in $V_S^\perp$. Hence we obtain the skeleton analysis of variance in Table 8.13.

Putting all the parts together gives the full analysis of variance in Table 8.14.

Stratum	source	degrees of freedom
mean	mean	1
large blocks	large blocks	$b-1$
small blocks	H	$s-1$
	residual	$(b-1)(s-1)$
	total	$b(s-1)$
plots	C	$k-1$
	$H \wedge C$	$(s-1)(k-1)$
	residual	$(b-1)s(k-1)$
	total	$bs(k-1)$
Total		bsk

Table 8.13: Skeleton analysis of variance for the classic split-plot design

8.3.5 Evaluation

Just as in Section 8.2, C and $H \wedge C$ are estimated more precisely than H, and with more power. So why would anyone deliberately choose a split-plot design? There are three good reasons.

(i) Treatment factor H is already applied to small blocks in a long-term experiment. At a later stage, it is decided to include treatment factor C, and it is practicable to apply levels of C to smaller units than the small blocks. This is what happened in Example 8.4.

(ii) It is not feasible to apply levels of H to such small units as is feasible for levels of C. This is the case in Example 8.2. Unless sowing is done by hand, varieties typically must be sown on relatively large areas while levels of other agricultural treatment factors, such as fertilizer, fungicide or pesticide, can be applied to smaller areas.

(iii) The differences between levels of H are already known and the main purpose of the experiment is to investigate C and the H-by-C interaction.

If none of those three conditions applies, think twice before recommending a split-plot design.

8.4 The split-plot principle

The split-plot idea is so simple to describe and to implement that there is a temptation to overuse it. Suppose that there are treatment factors F, G, H, ... Some scientists like to design experiments with repeated splitting as follows.

Stratum	source	df	SS	EMS	VR
mean	mean	1	$\frac{\text{sum}^2}{bsk}$	$\|\tau_0\|^2+\xi_0$	–
large blocks	large blocks	$b-1$	$\text{CSS}(B)-\text{SS}(\text{mean})$	ξ_B	–
small blocks	H	$s-1$	$\text{SS}(H)$	$\frac{\|\tau_H\|^2}{s-1}+\xi_S$	$\frac{\text{MS}(H)}{\text{MS}(\text{small blocks residual})}$
	residual	$(b-1)(s-1)$	by subtraction	ξ_S	–
	total	$b(s-1)$	$\text{CSS}(S)-\text{CSS}(B)$		
plots	C	$k-1$	$\text{SS}(C)$	$\frac{\|\tau_C\|^2}{k-1}+\xi$	$\frac{\text{MS}(C)}{\text{MS}(\text{plots residual})}$
	$H\wedge C$	$(s-1)(k-1)$	$\text{SS}(H\wedge C)$	$\frac{\|\tau_{HC}\|^2}{(s-1)(k-1)}+\xi$	$\frac{\text{MS}(H\wedge C)}{\text{MS}(\text{plots residual})}$
	residual	$(b-1)s(k-1)$	by subtraction	ξ	–
	total	$bs(k-1)$	$\sum y_\omega^2-\text{CSS}(S)$		
Total		bsk	$\sum_\omega y_\omega^2$		

Table 8.14: Full analysis of variance for the classic split-plot design

(i) Construct a complete-block design for F, or an equireplicate completely randomized design for F.

(ii) Split each plot into n_G subplots and apply levels of G randomly to subplots in each plot, independently within each plot.

(iii) Split each subplot into n_H sub-subplots and apply levels of H randomly to sub-subplots within each subplot, independently within each subplot.

(iv) And so on.

This is simple to understand, to construct and to randomize. It makes the experiment easy to conduct. There may be practical reasons why levels of F have to be applied to larger units than levels of G, and so on. If not, this is a bad design, because F has higher variance and fewer residual degrees of freedom than G and $F \wedge G$, which in turn have higher variance and fewer residual degrees of freedom than H, $F \wedge H$, $G \wedge H$ and $F \wedge G \wedge H$, and so on. This matters less if the main purpose of the experiment is to test for interactions.

Example 8.5 (Insecticides on grasshoppers) An experiment was conducted on a prairie in Western Canada to find out if insecticides used to control grasshoppers affected the weight of young chicks of ring-necked pheasants, either by affecting the grass around the chicks or by affecting the grasshoppers eaten by the chicks. Three insecticides were used, at low and high doses. The low dose was the highest dose recommended by the department of agriculture; the high dose was four times as much as the recommended dose, to assess the effects of mistakes.

The experimental procedure took place in each of three consecutive weeks. On the first day of each week a number of newly hatched female pheasant chicks were placed in a brooder pen. On the third day, the chicks were randomly divided into twelve groups of six chicks each. Each chick was given an identification tape and weighed.

On the fourth day, a portion of the field was divided into three strips, each of which was divided into two swathes. The two swathes within each strip were sprayed with the two doses of the same insecticide. Two pens were erected on each strip, and one group of pheasant chicks was put into each pen.

For the next 48 hours, the chicks were fed with grasshoppers which had been collected locally. Half the grasshoppers were anaesthestized and sprayed with insecticide; the other half were also anaesthestized and handled in every way like the first half except that they were not sprayed. All grasshoppers were frozen. The experimenters maintained a supply of frozen grasshoppers to each pen, putting them on small platforms so that they would not absorb further insecticide from the grass. In each swath, one pen had unsprayed grasshoppers while the other had grasshoppers sprayed by the insecticide which had been applied to that swath. At the end of the 48 hours, the chicks were weighed again indiviudally.

Table 8.15 shows the skeleton analysis of variance.

Stratum	source	degrees of freedom
mean	mean	1
weeks	weeks	2
strips	insecticide	2
	residual	4
	total	6
swathes	dose	1
	insecticide ∧ dose	2
	residual	6
	total	9
pens	food	1
	insecticide ∧ food	2
	dose ∧ food	1
	insecticide ∧ dose ∧ food	2
	residual	12
	total	18
chicks	chicks	180
Total		216

Table 8.15: Skeleton analysis of variance for the grasshopper experiment

Questions for Discussion

8.1 Read the first five pages of the paper 'Effects of Rodeo® and Garlon® 3A on nontarget wetland species in central Washington', written by Susan C. Gardner and Christian E. Grue and published in the journal *Environmental Toxicology and Chemistry* in 1996 (volume 15, pages 441–451).

(a) Describe the experimental design used, and comment on it.

(b) Suggest a better design.

(c) Comment on the statistical analysis used.

8.2 The following extract is taken from *Manual of Crop Experimentation* by Pearce, Clarke, Dyke and Kempson. Calculate the analysis-of-variance table, the means for the two main effects, the standard error of a difference between two varieties, and the standard error of the difference between two cutting schemes.

> An example from *Statistical Methods* by G. W. Snedecor and W. G. Cochran (1967, p. 370) used three varieties of alfalfa (lucerne) on the main-plots in six randomized blocks. Each of the main plots was divided into four sub-plots, the sub-plots treatments being four cutting schemes. All sub-plots were cut twice: the second cut took place on 27th July. (The data of the first are not given.) Some of the plots received a further cut as follows: B, 1st September; C, 20th September; D, 7th October, but A

was not cut further. The yields in tons per acre for the following year are set out below.

		Block					
Variety	Cutting	I	II	III	IV	V	VI
Ladak	A	2.17	1.88	1.62	2.34	1.58	1.66
	B	1.58	1.26	1.22	1.59	1.25	0.94
	C	2.29	1.60	1.67	1.91	1.39	1.12
	D	2.23	2.01	1.82	2.10	1.66	1.10
Cossack	A	2.33	2.01	1.70	1.78	1.42	1.35
	B	1.38	1.30	1.85	1.09	1.13	1.06
	C	1.86	1.70	1.81	1.54	1.67	0.88
	D	2.27	1.81	2.01	1.40	1.31	1.06
Ranger	A	1.75	1.95	2.13	1.78	1.31	1.30
	B	1.52	1.47	1.80	1.37	1.01	1.31
	C	1.55	1.61	1.82	1.56	1.23	1.13
	D	1.56	1.72	1.99	1.55	1.51	1.33

8.3 On Monday 3rd March 1997 a surgeon sent a message to *Allstat*, the email list for British statisticians, asking their advice. He said that he was going to use 9 animals in an experiment to compare three methods of grafting skin. He would use 3 animals for each method. After the graft was complete he would take a sample of new skin from each animal. He would cut each sample into 20 (tiny!) pieces and use a precision instrument to measure the thickness of each piece. He wanted to know how to analyse the data.

(a) What are the observational units, and how many are there?

(b) What are the experimental units, and how many are there?

(c) What are the treatments, and how many are there?

(d) Construct the skeleton analysis-of-variance table.

8.4 A cattle breeder wants to find a way of protecting his cattle against a particular stomach disease. He wants to compare the effects of:

S: spraying the grass in the paddock with a special chemical
N: no spray

He also wants to compare the effect of

$+$: injecting each animal with a special vaccine
$-$: no injection

He has several paddocks. Each paddock contains 20 animals, labelled 1, ..., 20 with ear-tags. He wants to apply the treatments once. A month later he will assess the amount of stomach disease in each animal by counting the number of a certain type of bacterium in a sample from the stomach contents. The logarithm of this number will be analysed.

He wants to use two paddocks to find out the effect of the spray, using S on one paddock and N on the other. He wants to use another two paddocks to find out the effect of the injection, using $+$ on one paddock and $-$ on the other. He asks you:

> Since N and $-$ are both 'no treatment', can I just use three paddocks, one for S, one for $+$ and one for 'no treatment'?

How do you answer?