

---

## Chapter 2

# Unstructured Experiments

---

### 2.1 Completely randomized designs

If there is no reason to group the plots into blocks then we say that  $\Omega$  is *unstructured*.

Suppose that treatment  $i$  is applied to  $r_i$  plots, in other words that  $i$  is replicated  $r_i$  times. Then

$$\sum_{i=1}^t r_i = |\Omega| = N.$$

Treatments should be allocated to the plots at random. Then the design is said to be *completely randomized*.

To construct and randomize the design, proceed as follows.

- (i) Number the plots  $1, 2, \dots, N$ .
- (ii) Apply treatment 1 to plots  $1, \dots, r_1$ ; apply treatment 2 to plots  $r_1 + 1, \dots, r_1 + r_2$ , and so on, to obtain a systematic design.
- (iii) Choose a random permutation of  $\{1, 2, \dots, N\}$  and apply it to the design.

**Example 2.1 (Fictitious)** Suppose that there are three treatments coded  $A, B$  and  $C$  with  $r_A = 5$  and  $r_B = r_C = 4$ . Then there are 13 plots. The systematic design is as follows.

plot	1	2	3	4	5	6	7	8	9	10	11	12	13
treatment	A	A	A	A	A	B	B	B	B	C	C	C	C

Suppose that the random permutation is

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ 6 & 2 & 8 & 11 & 13 & 1 & 12 & 5 & 7 & 4 & 9 & 3 & 10 \end{pmatrix},$$

where we are using the usual two-line way of displaying a permutation, which here indicates that 1 and 6 are interchanged, 2 does not move, 3 is moved to 8, and so on. In applying this permutation to the design we move the treatment *A* on plot 1 to plot 6, leave the treatment *A* on plot 2 still on plot 2, move the treatment *A* on plot 3 to plot 8, and so on. This gives the following plan.

plot	1	2	3	4	5	6	7	8	9	10	11	12	13
treatment	<i>B</i>	<i>A</i>	<i>C</i>	<i>C</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>C</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>A</i> .

## 2.2 Why and how to randomize

Why do we randomize? It is to avoid

**systematic bias** for example, doing all the tests on treatment *A* in January then all the tests on treatment *B* in March;

**selection bias** for example, choosing the most healthy patients for the treatment that you are trying to prove is best;

**accidental bias** for example, using the first rats that the animal handler takes out of the cage for one treatment and the last rats for the other;

**cheating** by the experimenter.

Cheating is not always badly intentioned. For example, an experimenter may decide to give the extra milk rations to those schoolchildren who are most undernourished or she may choose to put a patient in a trial if she thinks that the patient will particularly benefit from the new treatment. This sort of cheating is for the benefit of the (small) number of people in the experiment but, by biasing the results, may be to the disadvantage of the (large) number of people who could benefit in future from a treatment which has been demonstrated, without bias, to be superior. As another example, she may be secretly trying to remove bias by trying to balance numbers over some nuisance factor without troubling the statistician, but this too can produce false results unless this ‘balancing’ is taken into account in the analysis. Yet again, she may be trying to make life a little easier for the technician by telling him to do all of one treatment first.

Thus doing an objective randomization and presenting the experimenter with the plan has the added benefit that it may force her to tell you something which she had thought unnecessary, such as “We cannot do it that way because ...” or “that will put all replicates of treatment *A* in the shady part of the field”.

How do we choose a random permutation? The process must be objective, so that you have no chance to cheat either. Simply writing down the numbers 1, ..., *N* in an apparently haphazard order is not good enough.

One excellent way to randomize is to shuffle a pack of cards. A normal pack of 52 playing cards should be thoroughly randomized after seven riffle shuffles. In

Example 2.1 one can deal out the shuffled pack, noting the number (with Jack = 11, etc) but not the suit, and ignoring repeated numbers.

Another good method is to ask a computer to produce a (pseudo-)random order of the numbers 1, . . . ,  $N$ . Even a palm-top can do this. The permutation in Example 2.1 corresponds to the random order

6 2 8 11 13 1 12 5 7 4 9 3 10.

Two other methods of randomizing use sequences of random numbers, which can be generated by computers or calculators or found in books of tables. Both methods will be illustrated here for a random permutation of the numbers 1, . . . , 13.

My calculator produces random numbers uniformly distributed between 0 and 1. We need 13 different numbers, so record the random numbers to 2 decimal places. These are shown in the top row in Table 2.1. To turn these into the numbers 1–13 in a simple way, keeping the uniform distribution, multiply each random number by 100 and subtract multiples of 20 to leave a number in the range 1–20. Cross out numbers bigger than 13: these are marked X in the table. Remove any number that has already occurred: these are marked R in the table. Continue to produce random numbers until 12 different numbers have been listed. Then the final number must be the missing one so it can be written down without further ado.

0.72	0.34	0.65	0.99	0.01	0.23	0.30	0.57	0.17	0.63	0.14
12	14	5	19	1	3	10	17	17	3	14
12	X	5	X	1	3	10	X	X	R	X
1.00	0.94	0.29	0.05	0.38	0.86	0.59	0.51	0.22	0.30	0.05
20	14	9	5	18	6	19	11	2	10	5
X	X	9	R	X	6	X	11	2	R	R
0.24	0.78	0.63	0.35	0.13	0.12	0.54	0.96	0.67		
4	18	3	15	13	12	14	16	7		
4	X	R	X	13	R	X	X	7		

Table 2.1: Using a sequence of random numbers to generate a random permutation of 13 numbers

Thus the sequence of random numbers given in Table 2.1 gives us the random order

12 5 1 3 10 9 6 11 2 4 13 7 8.

Although this process is simple, it usually requires the generation of far more random numbers than the number of plots being permuted. In the second method we generate only a few more random numbers than the number of plots, because only the exact repeats are crossed out. Then place a 1 under the smallest number, a

0.71	0.66	0.31	0.71	0.78	0.11	0.11	0.41	0.91	0.24	0.99
10	8	5	X	11	2	X	6	12	4	13
0.20	0.53	0.68	0.24	0.03						
3	7	9	X	1						

Table 2.2: Second method of using a sequence of random numbers to generate a random permutation of 13 numbers

2 under the second smallest, and so on, replacing each random number by its rank. This process is shown in Table 2.2. It gives us the random order

10 8 5 11 2 6 12 4 13 3 7 9 1.

## 2.3 The treatment subspace

**Definition** The function  $T: \Omega \rightarrow \mathcal{T}$  is called the *treatment factor*.

There is an  $N$ -dimensional vector space associated with  $\Omega$ . It consists of  $N$ -tuples of real numbers, with each place in the  $N$ -tuple associated with one plot. Formally this vector space is  $\mathbb{R}^\Omega$ , but it will be called  $V$  for the remainder of this book. I shall be very cavalier about whether the vectors are row vectors or column vectors: in fact, if  $\Omega$  is a rectangular array of plots in a field then the vectors will also look like rectangles.

**Definition** The *treatment subspace* of  $V$  consists of those vectors in  $V$  which are constant on each treatment.

**Notation** Since the treatment factor is  $T$ , the treatment subspace will be denoted  $V_T$ .

**Example 2.1 revisited (Fictitious)** Figure 2.1 shows the set  $\Omega$ , the treatment factor  $T$ , a typical vector  $\mathbf{v}$  in  $V$  and the data vector  $\mathbf{y}$ . Beneath this are some vectors in the treatment subspace  $V_T$ . These include the special vectors  $\mathbf{u}_A$ ,  $\mathbf{u}_B$  and  $\mathbf{u}_C$ , which will be defined below.

Also in the treatment subspace is the vector  $\boldsymbol{\tau}$  of unknown treatment parameters. Its value on plot  $\omega$  is equal to  $\tau_{T(\omega)}$ . Under the linear model assumed in Section 2.5,  $\mathbb{E}(Y_\omega) = \tau_{T(\omega)}$ , so the *fitted value* on plot  $\omega$  is equal to the estimated value  $\hat{\tau}_{T(\omega)}$  of  $\tau_{T(\omega)}$ . Thus we have a vector  $\hat{\boldsymbol{\tau}}$  of fitted values, which is also in the treatment subspace.

**Definition** A vector  $\mathbf{v}$  in  $V$  is a *treatment vector* if  $\mathbf{v} \in V_T$ ; it is a *treatment contrast* if  $\mathbf{v} \in V_T$  and  $\sum_{\omega \in \Omega} v_\omega = 0$ .

$\Omega$	1	2	3	4	5	6	7	8	9	10	11	12	13	
$T$	$B$	$A$	$C$	$C$	$B$	$A$	$B$	$A$	$C$	$C$	$A$	$B$	$A$	
typical $\mathbf{v}$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_9$	$v_{10}$	$v_{11}$	$v_{12}$	$v_{13}$	
data $\mathbf{y}$	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	$y_7$	$y_8$	$y_9$	$y_{10}$	$y_{11}$	$y_{12}$	$y_{13}$	
some vectors in $V_T$	1	0	3	3	1	0	1	0	3	3	0	1	0	$= \mathbf{u}_B + 3\mathbf{u}_C$
	0	$\frac{1}{5}$	0	0	0	$\frac{1}{5}$	0	$\frac{1}{5}$	0	0	$\frac{1}{5}$	0	$\frac{1}{5}$	$= (\frac{1}{5})\mathbf{u}_A$
	$-\frac{1}{4}$	$\frac{1}{5}$	0	0	$-\frac{1}{4}$	$\frac{1}{5}$	$-\frac{1}{4}$	$\frac{1}{5}$	0	0	$\frac{1}{5}$	$-\frac{1}{4}$	$\frac{1}{5}$	$= (\frac{1}{5})\mathbf{u}_A - (\frac{1}{4})\mathbf{u}_B$
orthogonal basis for $V_T$	0	1	0	0	0	1	0	1	0	0	1	0	1	$= \mathbf{u}_A$
	1	0	0	0	1	0	1	0	0	0	0	1	0	$= \mathbf{u}_B$
	0	0	1	1	0	0	0	0	1	1	0	0	0	$= \mathbf{u}_C$
unknown treatment parameters	$\tau_B$	$\tau_A$	$\tau_C$	$\tau_C$	$\tau_B$	$\tau_A$	$\tau_B$	$\tau_A$	$\tau_C$	$\tau_C$	$\tau_A$	$\tau_B$	$\tau_A$	$= \boldsymbol{\tau}$
fitted values	$\hat{\tau}_B$	$\hat{\tau}_A$	$\hat{\tau}_C$	$\hat{\tau}_C$	$\hat{\tau}_B$	$\hat{\tau}_A$	$\hat{\tau}_B$	$\hat{\tau}_A$	$\hat{\tau}_C$	$\hat{\tau}_C$	$\hat{\tau}_A$	$\hat{\tau}_B$	$\hat{\tau}_A$	$= \hat{\boldsymbol{\tau}}$

Figure 2.1: Some vectors in Example 2.1

The vector  $(1/5)\mathbf{u}_A - (1/4)\mathbf{u}_B$  in Figure 2.1 is a treatment contrast.

Recall the *scalar product*: for vectors  $\mathbf{v}$  and  $\mathbf{w}$  in  $V$ , the scalar product  $\mathbf{v} \cdot \mathbf{w}$  of  $\mathbf{v}$  and  $\mathbf{w}$  is defined by

$$\mathbf{v} \cdot \mathbf{w} = \sum_{\omega \in \Omega} v_{\omega} w_{\omega}.$$

In particular,  $\mathbf{v} \cdot \mathbf{v} = \sum_{\omega \in \Omega} v_{\omega}^2$ , which is sometimes called the *sum of squares* of  $\mathbf{v}$  and sometimes the *squared length* of  $\mathbf{v}$ ; it is also written as  $\|\mathbf{v}\|^2$ .

We say that  $\mathbf{v}$  is *orthogonal* to  $\mathbf{w}$  (written  $\mathbf{v} \perp \mathbf{w}$ ) if  $\mathbf{v} \cdot \mathbf{w} = 0$ .

**Proposition 2.1** For each treatment  $i$  let  $\mathbf{u}_i$  be the vector whose value on plot  $\omega$  is equal to

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

Then  $\{\mathbf{u}_i : i \in \mathcal{T}\}$  is an orthogonal basis for  $V_T$ .

**Corollary 2.2** If there are  $t$  treatments then  $\dim(V_T) = t$ .

## 2.4 Orthogonal projection

Orthogonality is important in statistics, partly because orthogonal vectors correspond to random variables with zero correlation. Many procedures in estimation and analysis of variance are nothing more than the decomposition of the data vector into orthogonal pieces.

**Definition** If  $W$  is a subspace of  $V$  then the *orthogonal complement* of  $W$  is  $\{\mathbf{v} \in V : \mathbf{v} \text{ is orthogonal to } W\}$ .

**Notation** The orthogonal complement of  $W$  is written  $W^\perp$ , which is pronounced “ $W$  perp”.

**Theorem 2.3** *Let  $W$  be a subspace of  $V$ . Then the following hold.*

- (i)  $W^\perp$  is also a subspace of  $V$ .
- (ii)  $(W^\perp)^\perp = W$ .
- (iii)  $\dim(W^\perp) = \dim V - \dim W$ .
- (iv)  $V$  is the internal direct sum  $W \oplus W^\perp$ ; this means that given any vector  $\mathbf{v}$  in  $V$  there is a unique vector  $\mathbf{x}$  in  $W$  and a unique vector  $\mathbf{z}$  in  $W^\perp$  such that  $\mathbf{v} = \mathbf{x} + \mathbf{z}$ . We call  $\mathbf{x}$  the orthogonal projection of  $\mathbf{v}$  onto  $W$ , and write  $\mathbf{x} = P_W \mathbf{v}$ . See Figure 2.2
- (v)  $P_{W^\perp} \mathbf{v} = \mathbf{z} = \mathbf{v} - \mathbf{x} = \mathbf{v} - P_W \mathbf{v}$ .
- (vi) For a fixed vector  $\mathbf{v}$  in  $V$  and vector  $\mathbf{w}$  in  $W$ ,  $\sum_{\omega \in \Omega} (v_\omega - w_\omega)^2 = \|\mathbf{v} - \mathbf{w}\|^2$ . As  $\mathbf{w}$  varies over  $W$ , this sum of squares of differences is minimized when  $\mathbf{w} = P_W \mathbf{v}$ .
- (vii) If  $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$  is an orthogonal basis for  $W$  then

$$P_W \mathbf{v} = \left( \frac{\mathbf{v} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1 + \left( \frac{\mathbf{v} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \right) \mathbf{u}_2 + \dots + \left( \frac{\mathbf{v} \cdot \mathbf{u}_n}{\mathbf{u}_n \cdot \mathbf{u}_n} \right) \mathbf{u}_n.$$

## 2.5 Linear model

For unstructured plots we assume that

$$\mathbf{Y} = \boldsymbol{\tau} + \mathbf{Z},$$

where  $\boldsymbol{\tau} \in V_T$ ,  $\mathbb{E}(\mathbf{Z}) = \mathbf{0}$ ,  $\text{Var}(Z_\omega) = \sigma^2$  for all  $\omega$  in  $\Omega$ , and  $\text{cov}(Z_\alpha, Z_\beta) = 0$  for different plots  $\alpha$  and  $\beta$ . In other words,  $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\tau}$ , which is an unknown vector in  $V_T$ , and  $\text{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}$ , where  $\mathbf{I}$  is the  $N \times N$  identity matrix.

Under these assumptions, standard linear model theory gives the following results.

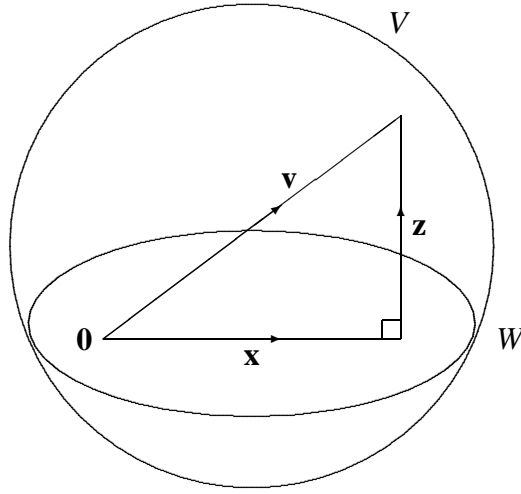


Figure 2.2: The vector  $\mathbf{x}$  is the orthogonal projection of the vector  $\mathbf{v}$  onto the subspace  $W$

**Theorem 2.4** Assume that  $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\tau}$  and that  $\text{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}$ . Let  $W$  be a  $d$ -dimensional subspace of  $V$ . Then

$$(i) \quad \mathbb{E}(P_W \mathbf{Y}) = P_W(\mathbb{E}(\mathbf{Y})) = P_W \boldsymbol{\tau};$$

$$(ii) \quad \mathbb{E}(\|P_W \mathbf{Y}\|^2) = \|P_W \boldsymbol{\tau}\|^2 + d\sigma^2.$$

**Theorem 2.5** Assume that  $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\tau} \in V_T$  and that  $\text{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}$ . Let  $\mathbf{x}$  and  $\mathbf{z}$  be any vectors in  $V_T$ . Then

(i) the best (that is, minimum variance) linear unbiased estimator of the scalar  $\mathbf{x} \cdot \boldsymbol{\tau}$  is  $\mathbf{x} \cdot \mathbf{Y}$ ;

(ii) the variance of the estimator  $\mathbf{x} \cdot \mathbf{Y}$  is  $\|\mathbf{x}\|^2 \sigma^2$ ;

(iii) the covariance of  $\mathbf{x} \cdot \mathbf{Y}$  and  $\mathbf{z} \cdot \mathbf{Y}$  is  $(\mathbf{x} \cdot \mathbf{z}) \sigma^2$ .

## 2.6 Estimation

Note that if the  $\mathbf{u}_i$  are the vectors defined in Proposition 2.1 then  $\mathbf{u}_i \cdot \mathbf{u}_i = r_i$ . Moreover, if  $\mathbf{v}$  is any other vector in  $V$  then  $\mathbf{u}_i \cdot \mathbf{v}$  is equal to the sum of the values of  $\mathbf{v}$  on those plots with treatment  $i$ . Write  $\text{SUM}_{T=i}$  for the sum of the values of  $\mathbf{Y}$  on the plots with treatment  $i$  and  $\text{sum}_{T=i}$  for the sum of the values of  $\mathbf{y}$  on the plots with treatment  $i$ . Then

$$\mathbf{u}_i \cdot \mathbf{Y} = \text{SUM}_{T=i} \quad \text{and} \quad \mathbf{u}_i \cdot \mathbf{y} = \text{sum}_{T=i}.$$

Moreover, write the means  $\text{SUM}_{T=i}/r_i$  and  $\text{sum}_{T=i}/r_i$  as  $\text{MEAN}_{T=i}$  and  $\text{mean}_{T=i}$  respectively.

Similarly, let  $\mathbf{u}_0$  be the all-1 vector; that is,  $\mathbf{u}_0 = \sum_{i=1}^t \mathbf{u}_i$ . For every vector  $\mathbf{v}$  in  $V$ , write

$$\bar{v} = \sum_{\omega \in \Omega} v_{\omega} / N.$$

Then  $\mathbf{u}_0 \cdot \mathbf{v} = \sum_{\omega \in \Omega} v_{\omega} = N\bar{v}$  for all  $\mathbf{v}$  in  $V$ . In particular,  $\mathbf{u}_0 \cdot \mathbf{u}_0 = N$ ,

$$\mathbf{u}_0 \cdot \mathbf{Y} = \text{SUM} = N\bar{Y} \quad \text{and} \quad \mathbf{u}_0 \cdot \mathbf{y} = \text{sum} = N\bar{y},$$

where  $\text{SUM}$  and  $\text{sum}$  are the grand totals  $\sum_{\omega \in \Omega} Y_{\omega}$  and  $\sum_{\omega \in \Omega} y_{\omega}$  respectively and  $\bar{Y}$  and  $\bar{y}$  are the grand means  $\text{SUM}/N$  and  $\text{sum}/N$  respectively.

To estimate the treatment parameter  $\tau_i$ , put  $\mathbf{x} = (1/r_i)\mathbf{u}_i$ . Then  $\mathbf{x} \cdot \boldsymbol{\tau} = \tau_i$  and  $\mathbf{x} \cdot \mathbf{Y} = \text{SUM}_{T=i}/r_i = \text{MEAN}_{T=i}$ . Therefore Theorem 2.5(i) shows that the best linear unbiased estimator of  $\tau_i$  is  $\text{MEAN}_{T=i}$ , with corresponding estimate  $\hat{\tau}_i$  equal to  $\text{mean}_{T=i}$ .

Similarly, to estimate a linear combination such as  $\sum_{i=1}^t \lambda_i \tau_i$ , put  $\mathbf{x} = \sum_{i=1}^t (\lambda_i/r_i)\mathbf{u}_i$ . Now

$$\mathbf{x} \cdot \boldsymbol{\tau} = \sum_{i=1}^t \lambda_i \left( \frac{1}{r_i} \mathbf{u}_i \cdot \boldsymbol{\tau} \right) = \sum_{i=1}^t \lambda_i \tau_i$$

and

$$\mathbf{x} \cdot \mathbf{Y} = \sum_{i=1}^t \lambda_i \left( \frac{1}{r_i} \mathbf{u}_i \cdot \mathbf{Y} \right) = \sum_{i=1}^t \lambda_i \text{MEAN}_{T=i},$$

so  $\sum \lambda_i \hat{\tau}_i$  is the best linear unbiased estimate of  $\sum \lambda_i \tau_i$ .

In particular, put  $\bar{\boldsymbol{\tau}} = \sum_{i=1}^t r_i \tau_i / N$ , which is the linear combination of  $\tau_1, \dots, \tau_t$  with  $\lambda_i = r_i/N$ . Then  $\mathbf{x} = (1/N)\mathbf{u}_0$ , so

$$\mathbf{x} \cdot \mathbf{Y} = \frac{\text{SUM}}{N} = \bar{Y},$$

and this is the best linear unbiased estimator of  $\bar{\boldsymbol{\tau}}$ .

Now we look at Theorem 2.4 with  $W = V_T$ . Since  $\boldsymbol{\tau} \in V_T$ , we have

$$P_{V_T} \boldsymbol{\tau} = \boldsymbol{\tau} = \sum_{i=1}^t \tau_i \mathbf{u}_i.$$

Proposition 2.1 and Theorem 2.3(vii) show that

$$P_{V_T} \mathbf{Y} = \sum_{i=1}^t \left( \frac{\mathbf{Y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} \right) \mathbf{u}_i = \sum_{i=1}^t \frac{\text{SUM}_{T=i}}{r_i} \mathbf{u}_i = \sum_{i=1}^t \text{MEAN}_{T=i} \mathbf{u}_i.$$

Theorem 2.4(i) confirms that this is an unbiased estimator of  $\boldsymbol{\tau}$ .



## 2.7 Sums of squares

**Definition** Let  $W$  be any subspace of  $V$ . The *sum of squares* for  $W$  means either  $\|P_W \mathbf{Y}\|^2$  or  $\|P_W \mathbf{y}\|^2$ . The *degrees of freedom* for  $W$  is another name for  $\dim W$ . The *mean square* for  $W$  is

$$\frac{\text{sum of squares for } W}{\text{degrees of freedom for } W},$$

for either sense of sum of squares. The *expected mean square* for  $W$ , written  $\text{EMS}(W)$ , is the expectation of the mean square in the random sense; that is

$$\text{EMS}(W) = \frac{\mathbb{E}(\|P_W \mathbf{Y}\|^2)}{\dim W}.$$

First we apply these ideas with  $W = V_T$ . Since

$$P_{V_T} \mathbf{Y} = \sum_{i=1}^t \frac{\text{SUM}_{T=i}}{r_i} \mathbf{u}_i,$$

the sum of squares for  $V_T$  is equal to

$$\left( \sum_{i=1}^t \frac{\text{SUM}_{T=i}}{r_i} \mathbf{u}_i \right) \cdot \left( \sum_{j=1}^t \frac{\text{SUM}_{T=j}}{r_j} \mathbf{u}_j \right).$$

Now,  $\mathbf{u}_i \cdot \mathbf{u}_j = 0$  whenever  $i \neq j$ , so this sum of squares

$$= \sum_{i=1}^t \frac{\text{SUM}_{T=i}^2}{r_i^2} \mathbf{u}_i \cdot \mathbf{u}_i = \sum_{i=1}^t \frac{\text{SUM}_{T=i}^2}{r_i}.$$

The quantity  $\sum_i (\text{sum}_{T=i}^2 / r_i)$  is called the *crude sum of squares for treatments*, which may be abbreviated to  $\text{CSS}(\text{treatments})$ .

The number of degrees of freedom for  $V_T$  is simply the dimension of  $V_T$ , which is equal to  $t$ .

The mean square for  $V_T$  is equal to

$$\sum_{i=1}^t \frac{\text{SUM}_{T=i}^2}{r_i} / t.$$

Theorem 2.4(ii) shows that

$$\mathbb{E}(\|P_{V_T} \mathbf{Y}\|^2) = \|P_{V_T} \boldsymbol{\tau}\|^2 + t\sigma^2 = \sum_{i=1}^t r_i \tau_i^2 + t\sigma^2,$$

because  $P_{V_T} \boldsymbol{\tau} = \sum_{i=1}^t \tau_i \mathbf{u}_i$ . Hence the expected mean square for  $V_T$  is equal to  $\sum r_i \tau_i^2 / t + \sigma^2$ .

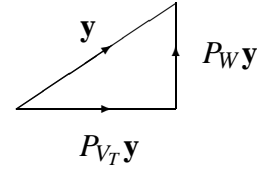
Secondly we apply the ideas with  $W = V_T^\perp$ . By Theorem 2.3(v),

$$\begin{aligned} P_W \mathbf{y} &= \mathbf{y} - P_{V_T} \mathbf{y} \\ &= \mathbf{y} - \sum_{i=1}^t \hat{\tau}_i \mathbf{u}_i \\ &= \text{data vector} - \text{vector of fitted values} \\ &= \text{residual vector,} \end{aligned}$$

so  $\|P_W \mathbf{y}\|^2$  is equal to the sum of the squares of the residuals. For this reason, all the quantities associated with  $W$  are named ‘residual’. (The word ‘error’ is sometimes used, but this can be confusing to non-statisticians, who tend to interpret it as ‘mistake’.)

Now,  $\mathbf{y}$  is the sum of the orthogonal vectors  $P_{V_T} \mathbf{y}$  and  $P_W \mathbf{y}$ , so Pythagoras’s Theorem shows that

$$\sum_{\omega \in \Omega} y_\omega^2 = \|\mathbf{y}\|^2 = \|P_{V_T} \mathbf{y}\|^2 + \|P_W \mathbf{y}\|^2.$$



The quantity  $\sum_{\omega \in \Omega} y_\omega^2$  is just the *total sum of squares*, so the sum of squares for residual is equal to the difference between the total sum of squares and the crude sum of squares for treatments; indeed, this is usually the easiest way to calculate it.

The number of degrees of freedom for residual is equal to the dimension of  $W$ , which is  $N - t$ , by Theorem 2.3(iii). Hence the mean square for residual is equal to

$$\frac{\text{sum of squares for residual}}{N - t}.$$

This will be denoted MS(residual).

We know that  $P_W \boldsymbol{\tau} = \mathbf{0}$  because  $\boldsymbol{\tau} \in V_T$ . Thus Theorem 2.4(ii) shows that  $\mathbb{E}(\|P_W \mathbf{Y}\|^2) = (N - t)\sigma^2$  and hence

$$\text{EMS}(\text{residual}) = \sigma^2. \quad (2.1)$$

## 2.8 Variance

Section 2.6 showed that the best linear unbiased estimator of the linear combination  $\sum \lambda_i \tau_i$  is  $\mathbf{x} \cdot \mathbf{Y}$ , where  $\mathbf{x} = \sum (\lambda_i / r_i) \mathbf{u}_i$ . By Theorem 2.5(ii), the variance of this estimator is equal to  $\|\mathbf{x}\|^2 \sigma^2$ . Now

$$\begin{aligned} \|\mathbf{x}\|^2 \sigma^2 &= \left( \sum_{i=1}^t \frac{\lambda_i}{r_i} \mathbf{u}_i \right) \cdot \left( \sum_{j=1}^t \frac{\lambda_j}{r_j} \mathbf{u}_j \right) \sigma^2 \\ &= \sum_{i=1}^t \frac{\lambda_i^2}{r_i^2} (\mathbf{u}_i \cdot \mathbf{u}_i) \sigma^2 \\ &= \left( \sum_{i=1}^t \frac{\lambda_i^2}{r_i} \right) \sigma^2. \end{aligned}$$

Two cases are particularly important. To estimate the treatment parameter  $\tau_i$ , for a fixed treatment  $i$ , put  $\lambda_i = 1$  and  $\lambda_j = 0$  for  $j \neq i$ . Then the variance is  $\sigma^2/r_i$ . To estimate the simple difference  $\tau_i - \tau_j$  for fixed different  $i$  and  $j$ , put  $\lambda_i = 1$ ,  $\lambda_j = -1$  and  $\lambda_k = 0$  if  $k \neq i$  and  $k \neq j$ . Then the variance is equal to

$$\sigma^2 \left( \frac{1}{r_i} + \frac{1}{r_j} \right). \quad (2.2)$$

Equation (2.1) shows that MS(residual) is an unbiased estimator of  $\sigma^2$ , so

$$\sum_{i=1}^t \frac{\lambda_i^2}{r_i} \times \text{MS}(\text{residual}) \quad (2.3)$$

is an unbiased estimator of the variance of the estimator of  $\sum \lambda_i \tau_i$ .

**Definition** The *standard error* for the estimate  $\sum \lambda_i \hat{\tau}_i$  is the square root of the estimate of the variance given by (2.3); that is, the standard error is equal to

$$\sqrt{\sum_{i=1}^t \frac{\lambda_i^2}{r_i} \times \text{MS}(\text{residual})}.$$

Thus the standard error for  $\hat{\tau}_i$  is  $\sqrt{\text{MS}(\text{residual})/r_i}$ . This is called the *standard error of a mean*, which may be abbreviated to s. e. m. Similarly, the standard error for  $\hat{\tau}_i - \hat{\tau}_j$  is

$$\sqrt{\text{MS}(\text{residual}) \left( \frac{1}{r_i} + \frac{1}{r_j} \right)},$$

which is called the *standard error of a difference* and abbreviated to s. e. d.

**Example 2.1 revisited (Fictitious)** Here the estimate  $\hat{\tau}_A$  of  $\tau_A$  is equal to  $\text{sum}_A/5$ , with variance  $\sigma^2/5$ . The simple difference  $\tau_A - \tau_B$  is estimated by  $\text{sum}_A/5 - \text{sum}_B/4$ , with variance  $9\sigma^2/20$ .

The estimators  $\hat{\tau}_A$ ,  $\hat{\tau}_B$  and  $\hat{\tau}_C$  are mutually uncorrelated, but

$$\text{Cov}(\hat{\tau}_A - \hat{\tau}_B, \hat{\tau}_A - \hat{\tau}_C) = \sigma^2/5.$$

Furthermore, the estimators  $\hat{\tau}_B - \hat{\tau}_C$  and  $\tau_A - (\hat{\tau}_B + \hat{\tau}_C)/2$  are uncorrelated.

The theory from Theorem 2.4 onwards has made no assumptions about the distribution of  $\mathbf{Y}$  about from its expectation and covariance. Assuming multivariate normality enables us to say more about the distributions of some statistics.

**Theorem 2.6** Suppose that the distribution of  $\mathbf{Y}$  is multivariate normal, that  $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\tau} \in V_T$  and that  $\text{Cov}(\mathbf{Y})$  is a scalar matrix. Then the following hold.

(i) If  $\mathbf{x} = \sum_{i=1}^t (\lambda_i/r_i) \mathbf{u}_i$  then

$$\frac{\mathbf{x} \cdot \mathbf{Y} - \sum \lambda_i \tau_i}{\sqrt{\left(\sum \frac{\lambda_i^2}{r_i}\right) \times \text{MS}(\text{residual})}}$$

has a  $t$ -distribution on  $N - t$  degrees of freedom.

(ii) If  $\mathbf{x}$  and  $\mathbf{z}$  are in  $V$  and  $\mathbf{x} \cdot \mathbf{z} = 0$  then  $\mathbf{x} \cdot \mathbf{Y}$  and  $\mathbf{z} \cdot \mathbf{Y}$  are independent estimators.

## 2.9 Replication: equal or unequal

If the treatments are unstructured we assume that all estimates of simple treatment differences are equally important. Thus the variances of all these estimators should be as small as possible. Equation (2.2) shows that the average variance of these estimators is equal to

$$\frac{1}{t(t-1)} \sum_{i=1}^t \sum_{j \neq i} \left( \frac{1}{r_i} + \frac{1}{r_j} \right) = \frac{1}{t} \sum_{i=1}^t \frac{1}{r_i}.$$

**Proposition 2.7** *If positive numbers  $r_1, \dots, r_t$  have a fixed sum  $R$  then  $\sum(1/r_i)$  is minimized when  $r_1 = r_2 = \dots = r_t = R/t$ .*

This proposition shows that the average variance of the estimators of simple differences is minimized when the replications  $r_1, \dots, r_t$  are as equal as possible. That is why most designs are equi-replicate.

Suppose that there is one treatment which is not sufficiently available for it to have replication  $N/t$ . If there are only two treatments this is often not a problem, because replacing replications of  $N/2$  and  $N/2$  by  $N/3$  and  $2N/3$  increases the variance from  $4\sigma^2/N$  to  $9\sigma^2/2N$ , an increase of only 12.5%. You may even be able to increase the number of plots slightly, to maintain the variance, if there is an unlimited supply of the second treatment.

**Example 2.2 (Limited availability)** Suppose that there are two treatments and 16 plots. Ideally, each treatment is applied to 8 plots, in which case the variance of the difference is  $\sigma^2/4$ . Suppose that there is only enough of the first treatment for 6 plots but that there are unlimited supplies of the second treatment. Keeping 16 plots we can have replications 6 and 10, giving a variance of  $4\sigma^2/15$ . If we can use two more plots then we can increase the replication of the second treatment to 12, in which case the variance returns to  $\sigma^2/4$ .

Life is not so simple when there are more than two treatments. Here the best that you can do is use the maximum amount of any treatment(s) whose availability is less than  $N/t$ , and make the remaining replications as equal as possible. The number of plots is rarely specified exactly in advance, so it will usually be possible to make all the remaining replications equal by including a few extra plots.

## 2.10 Allowing for the overall mean

In Section 2.7 we saw that

$$\text{EMS}(V_T) = \frac{\sum_{i=1}^t r_i \tau_i^2}{t} + \sigma^2 = \frac{\sum_{i=1}^t r_i \tau_i^2}{t} + \text{EMS}(\text{residual}).$$

Thus the difference between  $\text{EMS}(V_T)$  and  $\text{EMS}(\text{residual})$  is nonnegative, and is equal to zero if and only if  $\tau_1 = \tau_2 = \dots = \tau_t = 0$ .

Usually we do not measure on a scale that makes it plausible that  $\tau_1 = \tau_2 = \dots = \tau_t = 0$ . However, it is often plausible that  $\tau_1 = \tau_2 = \dots = \tau_t$ . This is called the *null model*, in which

$$\mathbb{E}(Y_\omega) = \kappa \quad \text{for all } \omega \text{ in } \Omega,$$

where  $\kappa$  is an unknown constant. In other words,  $\mathbb{E}(\mathbf{Y})$  is a scalar multiple of  $\mathbf{u}_0$ .

Let  $V_0$  be the subspace of  $V$  which consists of scalar multiples of  $\mathbf{u}_0$ . Then  $\{\mathbf{u}_0\}$  is a basis for  $V_0$  and  $\dim V_0 = 1$ . Theorem 2.3(vii) gives the following.

**Proposition 2.8** *If  $\mathbf{v} \in V$  then*

$$P_{V_0} \mathbf{v} = \left( \frac{\mathbf{v} \cdot \mathbf{u}_0}{\mathbf{u}_0 \cdot \mathbf{u}_0} \right) \mathbf{u}_0 = \left( \frac{\text{grand total of } \mathbf{v}}{N} \right) \mathbf{u}_0 = \bar{v} \mathbf{u}_0$$

and

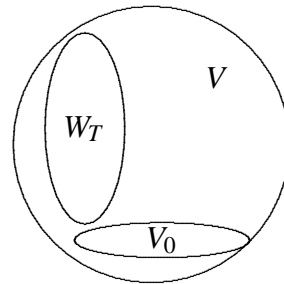
$$\begin{aligned} \|P_{V_0} \mathbf{v}\|^2 &= \left( \frac{\text{grand total of } \mathbf{v}}{N} \right)^2 \mathbf{u}_0 \cdot \mathbf{u}_0 \\ &= \frac{(\text{grand total of } \mathbf{v})^2}{N} \\ &= N \bar{v}^2. \end{aligned}$$

Now,  $\mathbf{u}_0 = \mathbf{u}_1 + \dots + \mathbf{u}_t$ , which is in  $V_T$ , and so  $V_0$  is a subspace of  $V_T$ . We can apply the ideas of Section 2.4 with  $V_0$  in place of  $W$  and  $V_T$  in place of  $V$ . Thus we define

$$\begin{aligned} W_T &= \{ \mathbf{v} \in V_T : \mathbf{v} \text{ is orthogonal to } V_0 \} \\ &= V_T \cap V_0^\perp, \end{aligned}$$

and find that

- (i)  $\dim W_T = \dim V_T - \dim V_0 = t - 1$ ;
- (ii)  $P_{W_T} \mathbf{v} = P_{V_T} \mathbf{v} - P_{V_0} \mathbf{v}$  for all  $\mathbf{v}$  in  $V$ ;
- (iii)  $\|P_{W_T} \mathbf{v}\|^2 + \|P_{V_0} \mathbf{v}\|^2 = \|P_{V_T} \mathbf{v}\|^2$  for all  $\mathbf{v}$  in  $V$ .



Applying (ii) and (iii) with  $\mathbf{v} = \boldsymbol{\tau}$  gives

$$P_{W_T}\boldsymbol{\tau} = \boldsymbol{\tau} - P_{V_0}\boldsymbol{\tau} = \boldsymbol{\tau} - \bar{\boldsymbol{\tau}}\mathbf{u}_0 = \sum_{i=1}^t (\tau_i - \bar{\tau})\mathbf{u}_i \quad (2.4)$$

and

$$\sum_{i=1}^t r_i (\tau_i - \bar{\tau})^2 = \|P_{W_T}\boldsymbol{\tau}\|^2 = \|\boldsymbol{\tau}\|^2 - \|\bar{\boldsymbol{\tau}}\mathbf{u}_0\|^2 = \sum_{i=1}^t r_i \tau_i^2 - N\bar{\tau}^2,$$

which is zero if and only if all the  $\tau_i$  are equal.

Applying (ii) with  $\mathbf{v} = \mathbf{y}$  gives

$$\begin{aligned} P_{W_T}\mathbf{y} &= \sum_{i=1}^t \left( \frac{\text{sum}_{T=i}}{r_i} \right) \mathbf{u}_i - \frac{\text{sum}}{N} \mathbf{u}_0 \\ &= \text{fitted values for treatments} - \text{fit for null model} \\ &= \sum_{i=1}^t (\hat{\tau}_i - \bar{y}) \mathbf{u}_i. \end{aligned}$$

The coefficients  $\hat{\tau}_i - \bar{y}$  are called *treatment effects*. Taking sums of squares gives

$$\|P_{W_T}\mathbf{y}\|^2 = \sum_{i=1}^t \frac{\text{sum}_{T=i}^2}{r_i} - \frac{\text{sum}^2}{N}.$$

The sum of squares for  $W_T$  is called the *sum of squares for treatments*, which may be abbreviated to SS(treatments), while the sum of squares for  $V_0$  is called the *crude sum of squares for the mean*, or just the *sum of square for the mean*, so we have

$$\begin{aligned} \text{sum of squares for treatments} &= \text{crude sum of squares for treatments} \\ &\quad - \text{crude sum of squares for the mean.} \end{aligned}$$

Correspondingly, the *mean square for treatments*, MS(treatments), and the *mean square for the mean*, MS(mean), are given by

$$\text{MS(treatments)} = \frac{\text{SS(treatments)}}{t-1},$$

$$\text{MS(mean)} = \frac{\text{SS(mean)}}{1} = \frac{\text{sum}^2}{N}.$$

The sum of squares for the mean is sometimes called the *correction for the mean*, which suggests, rather misleadingly, that there is something incorrect about the full data.

## 2.11 Hypothesis Testing

The previous section shows how to decompose the vector space  $V$  into the sum of three orthogonal pieces:

$$V = V_T \oplus V_T^\perp = V_0 \oplus W_T \oplus V_T^\perp.$$

Correspondingly, the overall dimension  $N$ , the data vector  $\mathbf{y}$  and its sum of squares can all be shown as the sum of three pieces. The sums of squares have their corresponding mean squares and expected mean squares, although there is no longer any sense in adding the three pieces.

	$V$	$=$	$V_0$	$\oplus$	$W_T$	$\oplus$	$V_T^\perp$
dimension	$N$	$=$	$1$	$+$	$(t-1)$	$+$	$(N-t)$
data	$\mathbf{y}$	$=$	$\bar{y}\mathbf{u}_0$	$+$	$\left(\sum_i \text{mean}_{T=i}\mathbf{u}_i - \bar{y}\mathbf{u}_0\right)$	$+$	$\left(\mathbf{y} - \sum_i \text{mean}_{T=i}\mathbf{u}_i\right)$
					treatment effects		residual
sum of squares	$\sum_{\omega \in \Omega} Y_\omega^2$	$=$	$\frac{\text{SUM}^2}{N}$	$+$	$\text{SS}(\text{treatments})$	$+$	$\text{SS}(\text{residual})$
mean square			$\frac{\text{SUM}^2}{N}$		$\frac{\text{SS}(\text{treatments})}{t-1}$		$\frac{\text{SS}(\text{residual})}{N-t}$
expected mean square			$N\bar{\tau}^2 + \sigma^2$		$\frac{\sum_i r_i \tau_i^2 - N\bar{\tau}^2}{t-1} + \sigma^2$		$\sigma^2$

To test the null hypothesis

$$H_0 : \bar{\tau} = 0$$

against the alternative hypothesis

$$H_1 : \bar{\tau} \neq 0,$$

look at MS(mean). If  $\text{MS}(\text{mean}) \approx \text{MS}(\text{residual})$  then we can conclude that  $\bar{\tau}$  may well be zero. However, we are not usually interested in this.

To test the null hypothesis

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_t$$

against the alternative hypothesis

$$H_1 : \tau \text{ is not a constant vector}$$

look at MS(treatments). If  $\text{MS}(\text{treatments}) \approx \text{MS}(\text{residual})$  then we can conclude that  $\tau$  may well be constant, in other words that there are no treatment differences.

source	sum of squares	degrees of freedom	mean square	variance ratio
mean	$\frac{\text{sum}^2}{N}$	1	SS(mean)	$\frac{\text{MS}(\text{mean})}{\text{MS}(\text{residual})}$
treatments	$\sum_i \frac{\text{sum}_{T=i}^2}{r_i} - \frac{\text{sum}^2}{N}$	$t - 1$	$\frac{\text{SS}(\text{treatments})}{t - 1}$	$\frac{\text{MS}(\text{treatments})}{\text{MS}(\text{residual})}$
residual	← . . . . . by subtraction . . . . . →		$\frac{\text{SS}(\text{residual})}{\text{df}(\text{residual})}$	—
Total	$\sum_{\omega} y_{\omega}^2$	$N$		

Table 2.3: Anova table for unstructured plots and unstructured treatments under the simple model

Note that both of the above tests are one-sided, because  $\text{EMS}(\text{mean}) - \sigma^2$  and  $\text{EMS}(\text{treatments}) - \sigma^2$  are both nonnegative.

The calculations are shown in an analysis-of-variance table (usually abbreviated to ‘anova table’). There is one row for each ‘source’; that is, subspace. The quantity in the column headed ‘variance ratio’ is the ratio of two mean squares whose expectations are equal under some null hypothesis to be tested: the numerator is the mean square for the current row, while the denominator is another mean square. Table 2.3 is the anova table for unstructured plots and unstructured treatments.

Comparing the size of a mean square with the mean square for residual gives an indication that some parameter of interest is non-zero. However, a proper significance test cannot be done without knowing the distribution of the variance ratio under the null hypothesis.

**Theorem 2.9** *Suppose that the distribution of  $\mathbf{Y}$  is multivariate normal. Let  $W_1$  and  $W_2$  be subspaces of  $V$  with dimensions  $d_1$  and  $d_2$ . Then the following hold.*

- (i) *If  $P_{W_1}\boldsymbol{\tau} = \mathbf{0}$  then  $\text{SS}(W_1)/\sigma^2$  has a  $\chi^2$ -distribution with  $d_1$  degrees of freedom.*
- (ii) *If  $W_1$  is orthogonal to  $W_2$  and  $P_{W_1}\boldsymbol{\tau} = P_{W_2}\boldsymbol{\tau} = \mathbf{0}$  then  $\text{MS}(W_1)/\text{MS}(W_2)$  has an F-distribution with  $d_1$  and  $d_2$  degrees of freedom.*

There are many experiments where the response variable is manifestly not normally distributed: for example, the observations may be counts (Example 1.1), percentages (Example 1.3) or ordinal scores (Example 1.9). Nonetheless, an F-test will usually give the qualitatively correct conclusion.

Most textbooks and computer output do not show the line for the mean in the anova table, as I have done in Table 2.3. There are three reasons why I do so. In the first place, I think that the calculations are more transparent when the ‘sum of



squares' and 'degrees of freedom' columns add up to the genuine totals rather than to adjusted totals.

In the second place, fitting  $V_0$  as a submodel of  $V_T$  is a first taste of what we shall do many times with structured treatments: fit submodels and see what is left over. Figure 2.3 shows the chain of vector subspaces  $\{\mathbf{0}\} \subseteq V_0 \subseteq V_T$  according to the usual convention that a smaller space is shown below a larger space, with a vertical line to indicate that the smaller space is contained in the larger. This is the opposite convention to the one used in the anova table, where the subspaces at the bottom of the figure are shown at the top of the table. Each hypothesis test compares the size of the extra fit in one subspace (compared to the fit in the subspace below) against the residual mean square. I see no benefit in treating  $V_0$  differently from any other submodel.

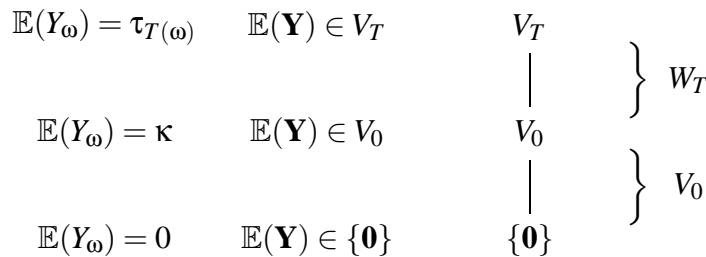


Figure 2.3: Three models for the expectation of  $\mathbf{Y}$ .

The third reason will become clear in Section 2.13, where we shall see that retaining a line for the mean points to the need to split up the rows of the anova table by strata even in the simplest case.

## 2.12 Sufficient replication for power

Suppose that there are two treatments and that  $\tau_1 - \tau_2 = \delta$ . Let  $\Delta$  be the estimator for  $\delta$  given in Section 2.6. Then  $\Delta - \delta$  has expectation 0 and variance  $\sigma^2 \nu$ , where  $\nu = 1/r_1 + 1/r_2$ , from Equation (2.2). Let  $\Gamma$  be the estimator for  $\sigma^2$  given in Section 2.8. Put  $X = (\Delta - \delta) / \sqrt{\nu\Gamma}$ . Under normality, Theorem 2.6 shows that  $X$  has a t-distribution on  $d$  degrees of freedom, where  $d$  is the number of residual degrees of freedom.

Let  $a$  be the 0.975 point of the t-distribution on  $d$  degrees of freedom. Some values of  $a$  are shown in Table 2.4. If a two-sided t-test with significance level 0.05 is performed on the data then we will conclude that  $\delta \neq 0$  if  $|\Delta| / \sqrt{\nu\Gamma} > a$ . Thus the probability  $p$  of not finding enough evidence to conclude that  $\delta \neq 0$  is given by

$$p = \Pr \left[ -a\sqrt{\nu\Gamma} < \Delta < a\sqrt{\nu\Gamma} \right] = \Pr \left[ -a - \frac{\delta}{\sqrt{\nu\Gamma}} < X < a - \frac{\delta}{\sqrt{\nu\Gamma}} \right].$$

The probability density function of each t-distribution takes larger values between 0 and  $a$  than it does between  $-2a$  and  $-a$ , so

$$\Pr[-a - x < X < a - x]$$

$d$	2	4	6	10	15	20	30	60
$a$	4.30	2.78	2.45	2.23	2.13	2.09	2.04	2.00
$\Pr[X > 2a]$	< 0.01	< 0.003	0.001	0.0005	0.0003	0.0002	0.0001	0.0001

Table 2.4: Values of  $a$  such that  $\Pr[X \leq a] = 0.975$ , where  $X$  has a t-distribution on  $d$  degrees of freedom

decreases from  $x = 0$  to  $x = a$  and also decreases from  $x = 0$  to  $x = -a$ . Thus if  $-a < \delta/\sqrt{v\Gamma} < a$  then  $p > \Pr[0 < X < 2a] = 0.5 - \Pr[X > 2a]$ . Table 2.4 shows that  $\Pr[X > 2a]$  is negligibly small, so  $p$  is unacceptably high.

If  $\delta$  is bigger than  $a\sqrt{v\Gamma}$  then

$$\Pr[X < -a - \delta/\sqrt{v\Gamma}] < \Pr[X < -2a] = \Pr[X > 2a],$$

and so  $p \approx \Pr[X < a - \delta/\sqrt{v\Gamma}]$ . If we want  $p$  to be at most 0.1 then we need to have  $a - \delta/\sqrt{v\Gamma} < -b$ , where  $-b$  is the 0.1 point of the t-distribution with  $d$  degrees of freedom (and so  $b$  is the 0.9 point). Thus  $a + b < \delta/\sqrt{v\Gamma}$ . If  $\delta$  is less than  $-a\sqrt{v\Gamma}$  then a similar argument shows that we must have  $\delta/\sqrt{v\Gamma} < -(a+b)$ . Thus we need

$$(a+b)^2 v\Gamma < \delta^2.$$

Replacing  $\Gamma$  by its expectation  $\sigma^2$  gives

$$(a+b)^2 v < (\delta/\sigma)^2. \quad (2.5)$$

Consider the ingredients in Equation (2.5). We assume that  $|\delta|$  is a known quantity, the size of the smallest difference that we want to detect. The variance  $\sigma^2$  is assumed unknown, but previous experiments on similar material may give a rough estimate for its value. In more complicated experiments (see Sections 2.13 and 4.6 and Chapter 8), we shall need to replace  $\sigma^2$  by the appropriate stratum variance. In many experiments the variance  $v\sigma^2$  of the estimator of  $\delta$  is given by  $v = 1/r_1 + 1/r_2$ , but Chapter 11 shows that a more complicated formula is needed in non-orthogonal designs. The values  $a$  and  $b$  depend partly on the number of degrees of freedom, which depends on the design. They also depend on some other choices:  $a$  on the significance level of the t-test;  $b$  on the upper limit of acceptability for  $p$ .

If we have a even a rough idea of the size of  $|\delta|/\sigma$ , and have set the value of  $a$  by choosing a significance level for the t-test, then Equation (2.5) gives an inequality to be satisfied by  $v$  and  $b$ . There are two ways in which this can be used.

In some areas, such as agricultural research, it is typical to propose the number of treatments and their replications first, according to resources available. This gives a value for  $v$ , from which Equation (2.5) gives an upper bound for  $b$ , from which t-tables give a value for  $p$ . If this value is acceptably low the experiment proceeds. If not, a modified experiment is proposed with a smaller value of  $v$ , usually by increasing resources or omitting less interesting treatments.

In other areas, such as clinical trials, it is more common to set both the significance level and the power (which is equal to  $1 - p$ ) in advance. Assuming equal replication  $r$ , Equation (2.5) is used to update values of  $r$  and  $d$  alternately until convergence is achieved.

**Example 2.3 (Calculation of replication)** Suppose that there are two treatments with equal replication  $r$  and that  $|\delta|/\sigma = 3$ . Then  $v = 2/r$  and Equation (2.5) gives

$$r > \frac{2}{9}(a+b)^2.$$

Start with  $d = \infty$ , for which the t-distribution is standard normal and so  $a = 1.960$  and  $b = 1.282$ . Then

$$r > 2(1.960 + 1.282)^2/9 \approx 2.3.$$

Take  $r$  to be the smallest value that satisfies this inequality, namely  $r = 3$ . Then  $N = 6$  and  $d = 4$ .

Repeat the cycle. Now that  $d = 4$  we have  $a = 2.776$  and  $b = 1.533$ . Hence

$$r > 2(2.776 + 1.533)^2/9 \approx 4.1$$

so put  $r = 5$ . Then  $d = 8$ .

This new value of  $d$  gives  $a = 2.306$  and  $b = 1.397$ , so

$$r > 2(2.306 + 1.397)^2/9 \approx 3.04.$$

Thus we put  $r = 4$ . Then  $d = 6$ .

Now  $a = 2.447$  and  $b = 1.440$  so

$$r > 2(2.447 + 1.440)^2/9 \approx 3.4.$$

This is satisfied by the current value of  $r$ , and we know that the value immediately below does not satisfy Equation (2.5), so we stop. We conclude that eight experimental units should suffice.

Note that power can be increased by including extra treatments (because this increases  $d$ ) but that this does not alter the variance of the estimator of a difference between two treatments.

## 2.13 A more general model

The chapter concludes with a slightly more general model than the one in Section 2.5.

As before, we assume that

$$\mathbb{E}(\mathbf{Y}) = \boldsymbol{\tau} \in V_T.$$

However, we change the assumption about covariance to

$$\text{cov}(Z_\alpha, Z_\beta) = \begin{cases} \sigma^2 & \text{if } \alpha = \beta \\ \rho\sigma^2 & \text{if } \alpha \neq \beta. \end{cases}$$

In other words, the correlation between responses on pairs of different plots is  $\rho$ , which may not be zero. Thus

$$\text{Cov}(\mathbf{Y}) = \sigma^2\mathbf{I} + \rho\sigma^2(\mathbf{J} - \mathbf{I}) = \sigma^2[(1 - \rho)\mathbf{I} + \rho\mathbf{J}],$$

where  $\mathbf{J}$  is the  $N \times N$  all-1 matrix.

Now,  $\mathbf{J}\mathbf{u}_0 = \mathbf{u}_0$ , and it is easily checked that  $\mathbf{J}\mathbf{u}_0 = N\mathbf{u}_0$ . Therefore

$$\text{Cov}(\mathbf{Y})\mathbf{u}_0 = \sigma^2(1 - \rho + N\rho)\mathbf{u}_0,$$

so  $\mathbf{u}_0$  is an eigenvector of  $\text{Cov}(\mathbf{Y})$  with eigenvalue  $\sigma^2(1 - \rho + N\rho)$ .

If  $\mathbf{x} \in V$  and  $\mathbf{x} \perp \mathbf{u}_0$  then  $\mathbf{J}\mathbf{x} = \mathbf{0}$ , because every entry in  $\mathbf{J}\mathbf{x}$  is equal to  $\mathbf{u}_0 \cdot \mathbf{x}$ . However,  $\mathbf{I}\mathbf{x} = \mathbf{x}$ , so

$$\text{Cov}(\mathbf{Y})\mathbf{x} = \sigma^2(1 - \rho)\mathbf{x},$$

and therefore  $\mathbf{x}$  is an eigenvector of  $\text{Cov}(\mathbf{Y})$  with eigenvalue  $\sigma^2(1 - \rho)$ .

The results from Theorem 2.4 onwards have assumed that  $\text{Cov}(\mathbf{Y})$  is a scalar matrix. Changing this assumption makes no difference to expectations of linear functions of  $\mathbf{Y}$ , but it does change the expectation of quadratic functions of  $\mathbf{Y}$ , that is, sums of squares. If  $\text{Cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$  then all formulas for variance or expected mean square involve  $\sigma^2$ . If  $\mathbf{x}$  is an eigenvector of  $\text{Cov}(\mathbf{Y})$  with eigenvalue  $\xi$  then  $\text{Cov}(\mathbf{Y})$  acts on  $\mathbf{x}$  just like  $\xi\mathbf{I}$ . Thus careful replacement of  $\sigma^2$  by the relevant eigenvalue gives the correct results. There is one possible difficulty in generalizing Theorem 2.5(iii) to the case when  $\mathbf{x}$  and  $\mathbf{z}$  are eigenvectors of  $\text{Cov}(\mathbf{Y})$  with different eigenvalues. However,  $\text{Cov}(\mathbf{Y})$  is a symmetric matrix, so eigenvectors with different eigenvalues are orthogonal to each other, so that  $\mathbf{x} \cdot \mathbf{z} = 0$  in this case. The other places where different eigenvalues might occur are the generalizations of Theorem 2.6(i) and Theorem 2.9(ii): we deal with both of these by restricting the results to eigenvectors with the same eigenvalue.

**Theorem 2.10** *Suppose that  $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\tau} \in V_T$  and  $\text{Cov}(\mathbf{Y}) = \mathbf{C}$ . Then the following hold.*

- (i) *If  $W$  is any subspace of  $V$  then  $\mathbb{E}(P_W(\mathbf{Y})) = P_W\boldsymbol{\tau}$ .*
- (ii) *If  $W$  consists entirely of eigenvectors of  $\mathbf{C}$  with eigenvalue  $\xi$ , and if  $\dim W = d$ , then*

$$\mathbb{E}(\|P_W(\mathbf{Y})\|^2) = \|P_W(\boldsymbol{\tau})\|^2 + d\xi.$$

- (iii) *If  $\mathbf{x} \in V_T$  and  $\mathbf{x}$  is an eigenvector of  $\mathbf{C}$  then the best linear unbiased estimator of  $\mathbf{x} \cdot \boldsymbol{\tau}$  is  $\mathbf{x} \cdot \mathbf{Y}$ .*

- (iv) If  $\mathbf{x}$  is an eigenvector of  $\mathbf{C}$  with eigenvalue  $\xi$  then the variance of  $\mathbf{x} \cdot \mathbf{Y}$  is  $\|\mathbf{x}\|^2 \xi$ .
- (v) Suppose that  $\mathbf{x}$  and  $\mathbf{z}$  are eigenvectors of  $\mathbf{C}$  with eigenvalues  $\xi$  and  $\eta$  respectively. If  $\xi = \eta$  then  $\text{cov}(\mathbf{x} \cdot \mathbf{Y}, \mathbf{z} \cdot \mathbf{Y}) = (\mathbf{x} \cdot \mathbf{z}) \xi$ ; if  $\xi \neq \eta$  then  $\text{cov}(\mathbf{x} \cdot \mathbf{Y}, \mathbf{z} \cdot \mathbf{Y}) = 0$ .
- (vi) Suppose that  $\mathbf{x}$  is an eigenvector of  $\mathbf{C}$  with eigenvalue  $\xi$ , that  $\mathbf{x} = \sum_i (\lambda_i / r_i) \mathbf{u}_i$ , that  $W$  is a  $d$ -dimensional subspace consisting of eigenvectors of  $\xi$  orthogonal to  $V_T$ . If  $\mathbf{Y}$  has a multivariate normal distribution then

$$\frac{\mathbf{x} \cdot \mathbf{Y} - \sum \lambda_i \tau_i}{\sqrt{\left(\sum \frac{\lambda_i^2}{r_i}\right) \times \text{MS}(W)}}$$

has a  $t$ -distribution on  $d$  degrees of freedom and  $\text{SS}(W)/\xi$  has a  $\chi^2$ -distribution on  $d$  degrees of freedom.

- (vii) If  $W_1$  and  $W_2$  are subspaces with dimensions  $d_1$  and  $d_2$ , both consisting of eigenvectors of  $\mathbf{C}$  with eigenvalue  $\xi$ , orthogonal to each other, with  $P_{W_1} \boldsymbol{\tau} = P_{W_2} \boldsymbol{\tau} = \mathbf{0}$ , and if  $\mathbf{Y}$  has a multivariate normal distribution then  $\text{MS}(W_1)/\text{MS}(W_2)$  has an  $F$ -distribution on  $d_1$  and  $d_2$  degrees of freedom.

**Definition** A *stratum* is an eigenspace of  $\text{Cov}(\mathbf{Y})$  (note that this is not the same as a stratum in sampling).

The analysis of variance proceeds just as before, except that we first decompose  $V$  into the different strata. Under the assumptions of this section,  $V_0$  is one stratum, with dimension 1 and eigenvalue  $\sigma^2(1 - \rho + N\rho)$ , while  $V_0^\perp$  is the other stratum, with dimension  $N - 1$  and eigenvalue  $\sigma^2(1 - \rho)$ . Call these eigenvalues  $\xi_0$  and  $\xi_1$  respectively. We then obtain the anova table shown in Table 2.5.

stratum	source	df	EMS	
$V_0$	'mean'	mean	1	$N\bar{\tau}^2 + \xi_0$
$V_0^\perp$	'plots'	treatments	$t - 1$	$\frac{\sum_i r_i (\tau_i - \bar{\tau})^2}{t - 1} + \xi_1$
		residual	$N - t$	$\xi_1$
Total			$N$	

Table 2.5: Anova table for unstructured plots and unstructured treatments under the more general model

Now we calculate the variance ratio only for terms with the same eigenvalue.

There is no way of estimating  $\xi_0$ , and hence no way of assessing whether  $\bar{\tau}$  is (statistically significantly) different from zero, and no way of estimating the variance of the estimator of any treatment parameter  $\tau_i$ . However, all treatment contrasts  $\mathbf{x}$  are in  $V_0^\perp$  so their linear combinations and their variances may be estimated just as before. Experiments in which we are interested only in treatment contrasts are called *comparative experiments*.

## Questions for Discussion

**2.1** A completely randomized experiment was conducted to compare seven treatments for their effectiveness in reducing scab disease in potatoes. The field plan is shown below.

2	1	6	4	6	7	5	3
9	12	18	10	24	17	30	16
1	5	4	3	5	1	1	6
10	7	4	10	21	24	29	12
2	7	3	1	3	7	2	4
9	7	18	30	18	16	16	4
5	1	7	6	1	4	1	2
9	18	17	19	32	5	26	4

The upper figure in each plot denotes the treatment, coded 1–7. The lower figure denotes an index of scabbiness of potatoes in that plot: 100 potatoes were randomly sampled from the plot, for each one the percentage of the surface area infected with scabs was assessed by eye and recorded, and the average of these 100 percentages was calculated to give the scabbiness index.

- Give the analysis-of-variance table for these data.
- Is there any evidence that the mean scabbiness is different according to different treatments? Justify your answer.
- Estimate the mean scabbiness produced by each treatment.
- What is the standard error of the above estimates?
- What is the standard error of the differences between means?

**2.2** A technician has to measure the acidity of four soils. You give him three samples of each soil and ask him to make the twelve measurements in random order. He says that a random order will confuse him and that it will be better if he measures the acidity of all three samples of soil A, then all three samples of soil B, and so on. Make notes on arguments you will use to persuade him that a random order is better.