Chapter 5

Factorial Treatment Structure

5.1 Treatment factors and their subspaces

In this section we consider experiments where the treatment structure is that treatments consist of all combinations of levels of two treatment factors.

Example 5.1 (Example 1.5 continued: Rye-grass) Here the twelve treatments are all combinations of the levels of the following two treatment factors.

factor	levels
Cultivar (C)	Cropper, Melle, Melba
Fertilizer (F)	0, 80, 160, 240 kg/ha

The treatments may be labelled $1, \ldots, 12$ according to the following table.

	F			
C	0	80	160	240
Cropper	1	2	3	4
Melle	5	6	7	8
Melba	9	10	11	12

Thus treatment 6 is the ordered pair $(\text{Melle}, 80)$. In particular,

$$T(\omega) = 6 = (\text{Melle}, 80) \iff \begin{cases} C(\omega) = \text{Melle} & \text{and} \\ F(\omega) = 80. \end{cases}$$

Notation If treatments are all combinations of levels of treatment factors F and G, write $T = F \wedge G$ to show that T is defined by the combinations of levels of F and G.

In Example 5.1, $T = C \wedge F$.

Notation If F is a treatment factor, write n_F for the number of levels of F.

In Example 5.1, $n_C = 3$ and $n_F = 4$.

Now we define subspaces V_F, W_F, V_G and W_G of V analogous to those defined for treatments in Sections 2.3 and 2.10 and for blocks in Section 4.2. Thus we put

$$\begin{aligned} V_F &= \{\text{vectors in } V \text{ which are constant on each level of } F\}, \\ W_F &= V_F \cap V_0^\perp, \\ V_G &= \{\text{vectors in } V \text{ which are constant on each level of } G\}, \\ W_G &= V_G \cap V_0^\perp. \end{aligned}$$

Then

$$\begin{aligned} \dim V_F &= n_F, \qquad \dim W_F = n_F - 1, \\ \dim V_G &= n_G, \qquad \dim W_G = n_G - 1. \end{aligned}$$

Theorem 5.1 *If every combination of level of factors F and G occurs on the same number of plots then $W_F \perp W_G$.*

Proof Similar to the proof of Theorem 4.1. ∎

Theorem 5.1 shows that

$$V_F + V_G = V_0 \oplus W_F \oplus W_G \tag{5.1}$$

orthogonally, and hence that

$$\dim(V_F + V_G) = 1 + (n_F - 1) + (n_G - 1).$$

If the treatment factor T is defined by all combinations of levels of F and G then $\dim V_T = t = n_F n_G$. Now, $V_F + V_G \subset V_T$, so put

$$W_{F \wedge G} = V_T \cap (V_F + V_G)^\perp. \tag{5.2}$$

Then $\dim W_{F \wedge G} = \dim V_T - \dim(V_F + V_G) = n_F n_G - 1 - (n_F - 1) - (n_G - 1) = (n_F - 1)(n_G - 1)$.

Figure 5.1 may be helpful. Note that this is not a Venn diagram. For example, V_0 and V_F are complementary *subspaces* of V_F, not complementary subsets.

5.2 Interaction

We continue to suppose that treatments consist of all combinations of levels of F and G. Table 5.1 shows some plausible models for the expectation of the response in this situation. The relationships among the subspaces considered as models are shown in Figure 5.2.

If $\mathbb{E}(\mathbf{Y}) \in V_F + V_G$ then the difference between $\mathbf{Y}$ values for different levels of F does not depend on the level of G, and vice versa. Thus plotting $\mathsf{mean}_{F=i}$ against i for each level of G gives approximately parallel curves.

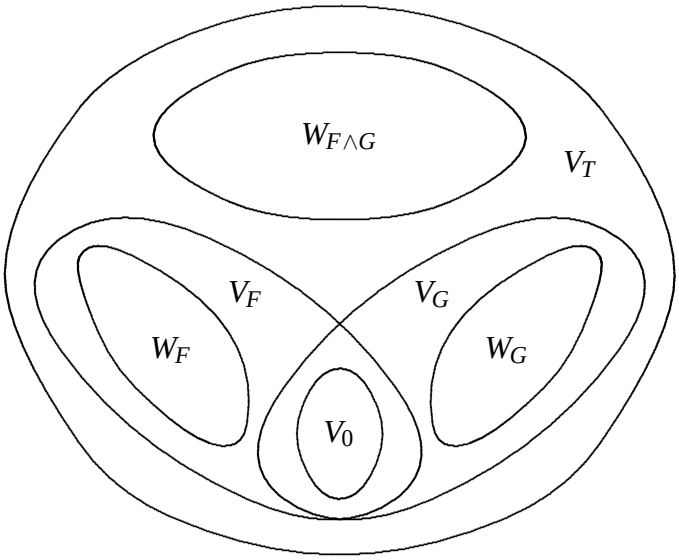

Figure 5.1: Orthogonal subspaces of the treatment space when treatments are all combinations of the levels of two treatment factors

coordinate (and parameters)	vector (and subspace)	model name
$\mathbb{E}(Y_\omega) = \kappa$	$\mathbb{E}(\mathbf{Y}) \in V_0$	null model
$\mathbb{E}(Y_\omega) = \lambda_{F(\omega)}$	$\mathbb{E}(\mathbf{Y}) \in V_F$	F only
$\mathbb{E}(Y_\omega) = \mu_{G(\omega)}$	$\mathbb{E}(\mathbf{Y}) \in V_G$	G only
$\mathbb{E}(Y_\omega) = \lambda_{F(\omega)} + \mu_{G(\omega)}$	$\mathbb{E}(\mathbf{Y}) \in V_F + V_G$	additive in F and G
$\mathbb{E}(Y_\omega) = \tau_{T(\omega)}$	$\mathbb{E}(\mathbf{Y}) \in V_T$	full treatment model

Table 5.1: Some models for $\mathbb{E}(Y_\omega)$ when treatments are all combinations of levels of F and G

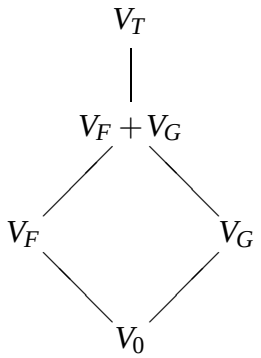

Figure 5.2: Relationships among the subspaces listed as possible models in Table 5.1

Example 5.1 revisited (Rye-grass) In this experiment the response on each plot was the percentage of water-soluble carbohydrate in the grain. If the twelve treatment means are as shown in Figure 5.3 then we have approximately parallel curves. Thus we can report, for example, that the percentage of water-soluble carbohydrate in the grain is 0.95 higher for Cropper than it is for Melle, irrespective of the amount of fertilizer.

If $\mathbb{E}(\mathbf{Y})$ is not in $V_F + V_G$ then we say that there is an *interaction* between F and G. In Example 5.1 there is no interaction. The most extreme form of interaction occurs when the separate curves actually cross over each other.

Example 5.2 (Cowpeas) In an experiment in South Africa, the treatments consisted of five varieties of cowpea in combination with three methods of cultivation. The mean yields, in tonnes/hectare, are shown in Figure 5.2. Here there is *crossover* interaction. It is important to report this interaction. Clearly cultivation method 1 is best for varieties *C* and *D* and worst for the other three, while cultivation method 2 is worst for varieties *C* and *D* and best for the other three. Cultivation method 3 is intermediate for every variety and so might be the safest one to use if a farmer is trying a new variety not among the five tested here.

Some people draw lines between successive points with the same symbol in diagrams like those in Figures 5.3 and 5.4. The lines certainly aid the human eye to see if the successions of points are parallel. However, they can also be misleading. In Figure 5.3 such a line would suggest a value for the percentage of water-soluble carbohydrate in the grain if nitrogen fertilizer were applied at 100 kg/ha, and the value might not be too far from the truth. In Figure 5.4 such a line would suggest a variety intermediate between varieties *A* and *B*, which may be nonsense.

If one treatment factor is quantitive and the other is qualitative, then people who draw lines would plot the levels of the quantitative factor along the *x*-axis, as in Figure 5.3, so that the intermediate values suggested by the lines might have some meaning. However, it is often easier to read the diagram if the factor with the higher

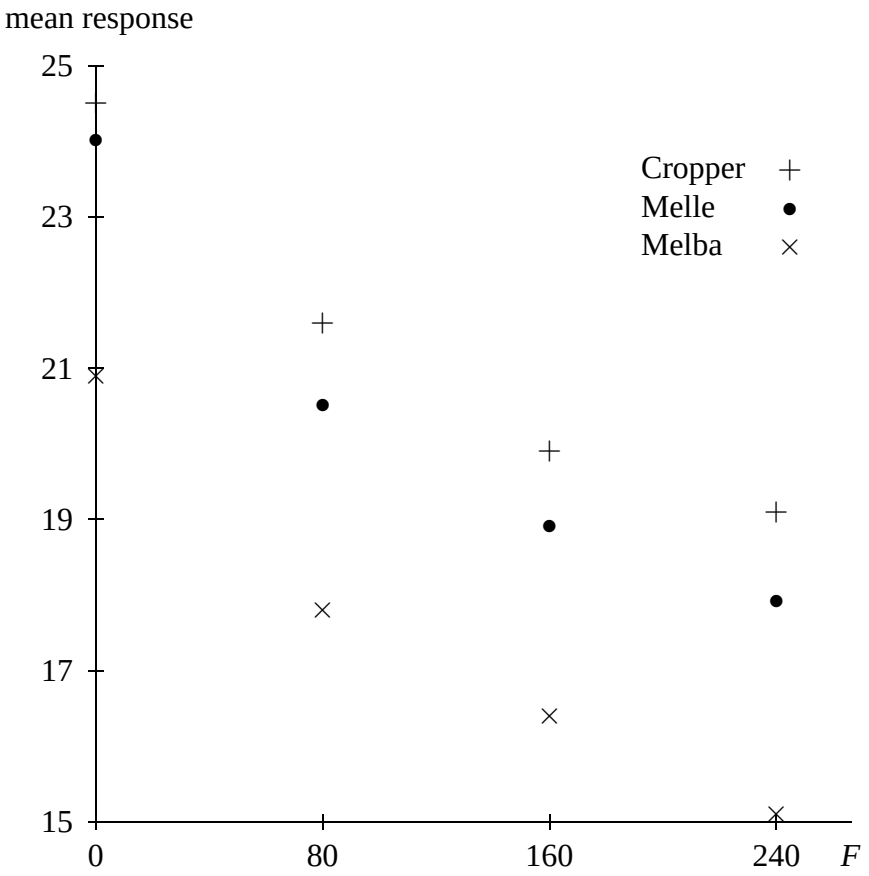

Figure 5.3: No interaction between Cultivar and Fertilizer

number of levels is plotted along the x-axis, and the levels of the other factor are shown by symbols, whether or not either of them is quantitative.

Example 5.3 (Modern cereals) Modern cereals have been bred to produce high yields, but they often need large amounts of fertilizer and other additives in order to do so. For such a cereal, the yield rises rather steeply in response to extra fertilizer. By comparison, a traditional old variety will not be capable of producing the highest yields, but it may have a rather mild response to fertilizer, so that it does better than the new variety if not much fertilizer is applied. This is another example of crossover interaction. A farmer with low resources, or with poor soil, might do better to grow the traditional variety.

Example 5.4 (Enzyme in blood) Two chemical preparations were injected into mice to see if they affected the quantity of some enzyme in their blood. Thus there were two treatment factors, F and G, whose levels were absence and presence of each of the chemicals respectively. The mean responses are shown in Figure 5.5. It is clear that both chemicals are needed in order to increase the quantity of enzyme. Thus their effects are not additive, and there is interaction between them.

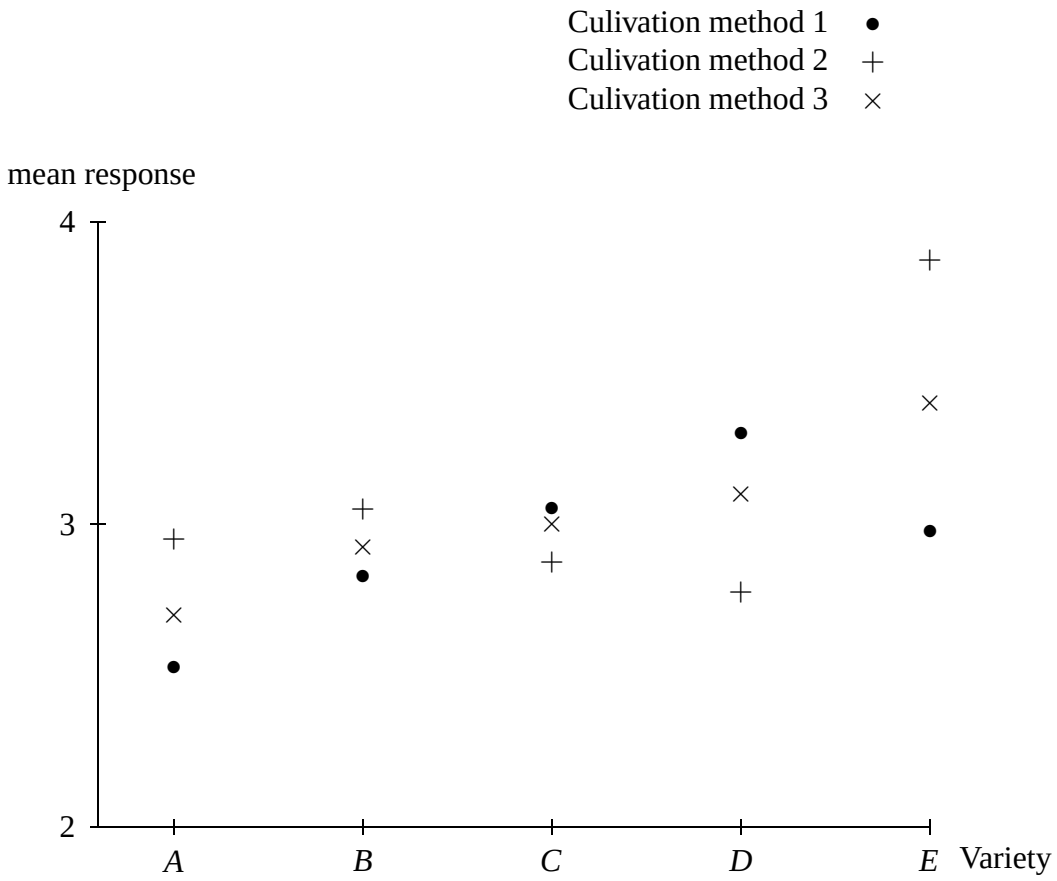

Figure 5.4: Crossover interaction between Variety and Cultivation method in Example 5.2

Example 5.4 shows *threshold* interaction: each treatment factor needs to be present at a certain level before the other can take effect. In more complicated cases the treatment factors may have more than two levels, and it may be that only one of them acts as a threshold for the other.

Example 5.5 (Saplings) At a forestry research station, saplings were planted in an experiment to compare treatments for getting the young trees established. Four types of collar were put around the saplings to prevent them from predators and other damage: these were combined with ?? levels of ??. It turned out that one type of collar excluded so much light that all the plants died, so the response on those saplings was zero irrespective of the level of ?? This is an extreme form of threshold interaction. Unless the qualitative factor Collar-type is at a level that permits the plants to grow, no other treatment factor can have any effect.

The opposite of a threshold interaction is a *trigger* interaction. Here at least one treatment factor needs to have the correct level to change the response but there is no further gain from having both at the correct level.

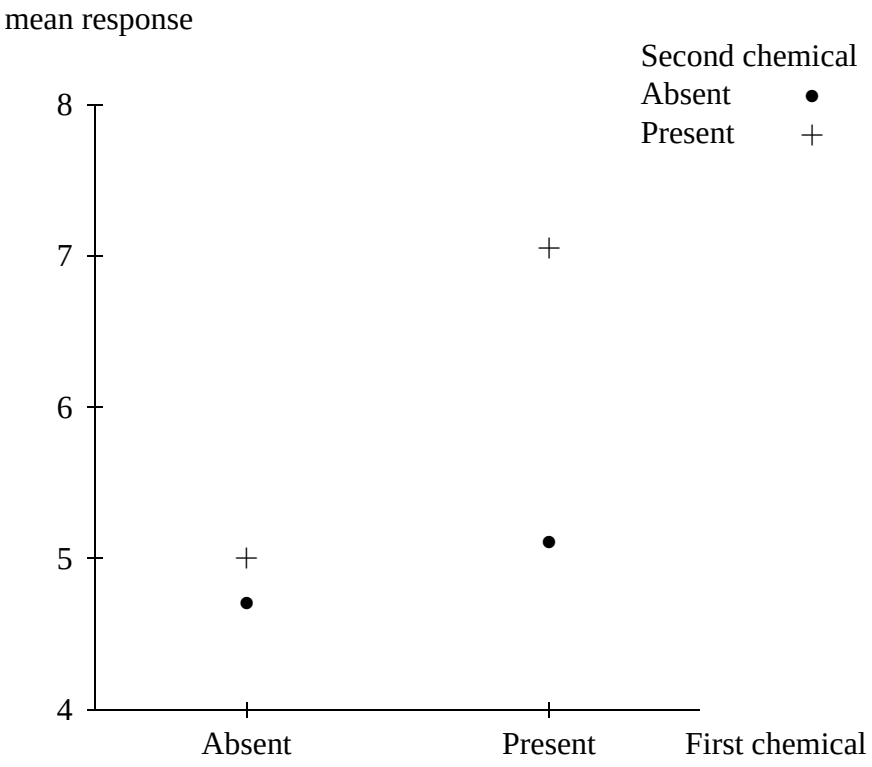

Figure 5.5: Threshold interaction in Example 5.4

Example 5.6 (Catalysts in a chemical reaction) A chemical process to produce an industrial chemical has several stages. At each stage a catalyst may be present to improve the reaction. Figure 5.6 shows the effect of having a catalyst present or absent at each of two stages. The response is a measure of the quality of the chemical produced. Relative to the residual mean square, the three reponses with at least one catalyst present were judged to be the same.

Two further types of interaction are worth mentioning. If there is no crossover then the distances between the curves may increase as the general response increases. This can be an indication that we are measuring on the wrong scale. The basic assumption (1.1) is that treatment i adds a constant τ_i to the response on each plot where it is applied. Similarly, we assume in Equation (4.1) that each block adds a fixed constant.

Example 5.7 (Counts of bacteria) If the purpose of the treatments is to reduce the number of bacteria in milk then it is much more likely that the effect of each treatment is to *multiply* the initial number of bacteria by a positive constant (which we hope is less than 1). Thus we would expect Equation (1.1) to apply to the logarithms of the counts rather than to the counts themselves.

Let Y_ω be the number of bacteria per ml in sample ω. Suppose that treatment factors F and G each act multiplicatively on the counts, so that $Y_\omega = Z_\omega \times \lambda_{F(\omega)}$ if only factor F is applied and $Y_\omega = Z_\omega \times \mu_{G(\omega)}$ if only factor G is applied. If

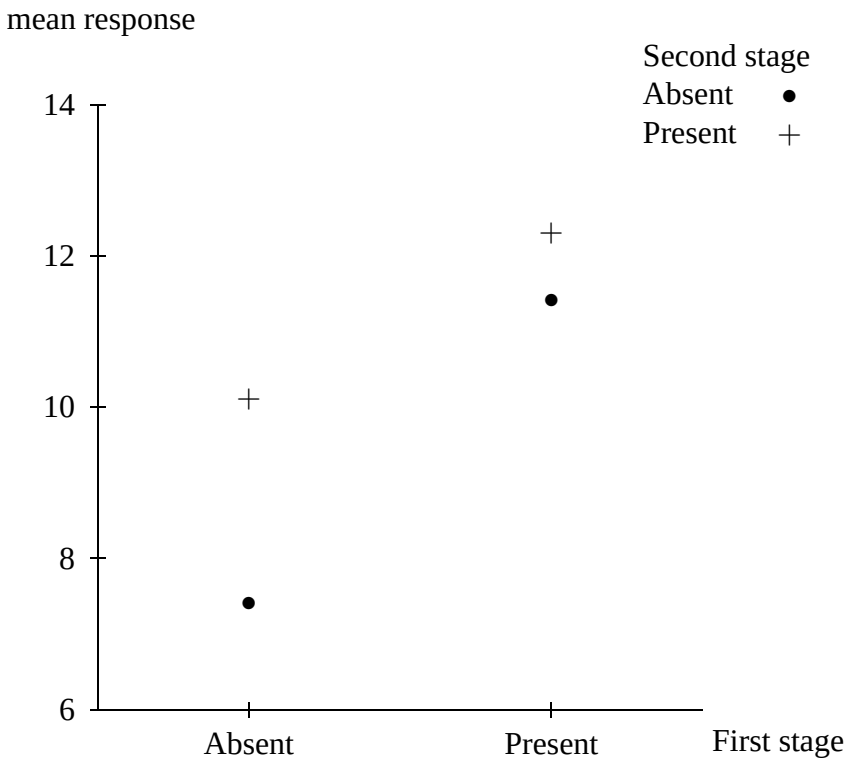

Figure 5.6: Trigger interaction in Example 5.6

these factors continue to act multiplicatively in the presence of each other, then $Y_\omega = Z_\omega \times \lambda_{F(\omega)} \times \mu_{G(\omega)}$, so that $\log(Y_\omega) = \log(Z_\omega) + \log(\lambda_{F(\omega)}) + \log(\mu_{G(\omega)})$: in other words, the model is additive in F and G on the log scale but not on the original count scale.

Similarly, we may measure volume but expect the linear model (1.1) to apply to the linear measurement, so that we need to take cube roots before analysing the data.

Suppose that f is a nonlinear monotonic function such that $f(Y)$ satisfies the linear model (1.1). If there is no interaction between treatment factors F and G when we consider the transformed data $f(y)$, then there *will* be interaction between them when we consider the untransformed data y. This is rather hard to spot graphically. If f is nearly linear over the range of the data, zero interaction on the transformed scale looks like zero interaction on the original scale. If f is not nearly linear over this range then calculating means of the data on the wrong scale gives a seriously misleading impression.

Finally, the distances between the curves may increase as the levels of one of the treatment factors increases. There may be a simple explanation for this.

Example 5.8 (Vetch and oats) An experiment on forage crops compared five seed mixtures in the presence and absence of nitrogen fertilizer. Figure 5.7 shows the mean responses in tonnes/ha. The seed mixtures range from all oats, no vetch to no

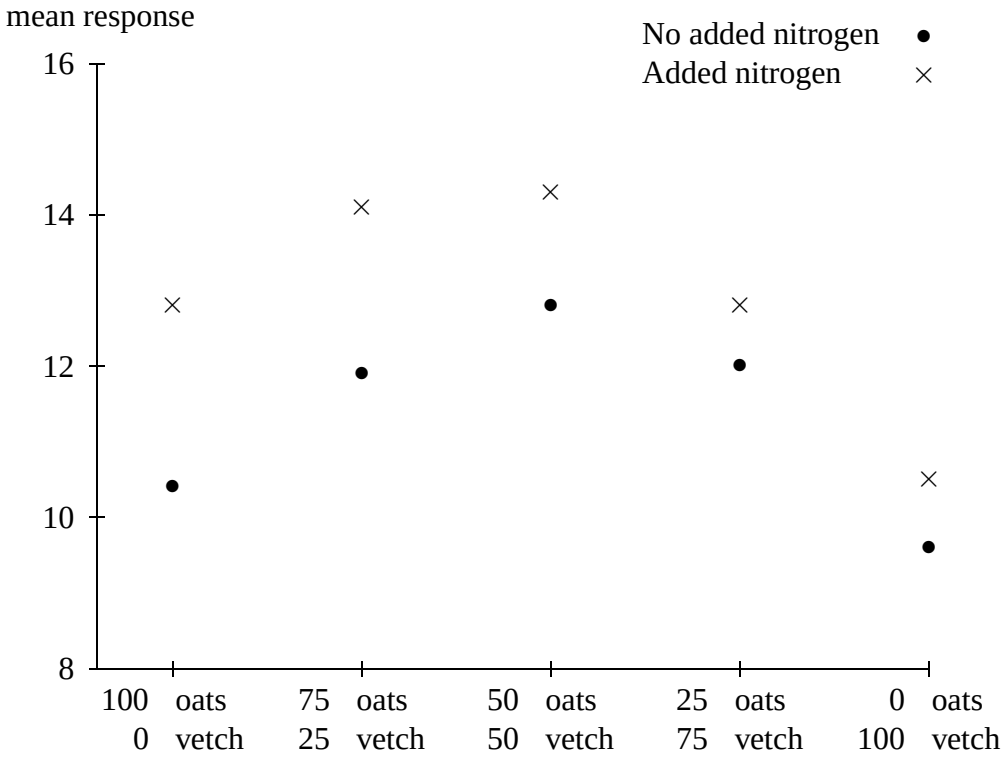

Figure 5.7: Improvement due to added nitrogen declines as the proportion of vetch increases: see Example 5.8

oats, all vetch. The yield is highest for the 50 : 50 mixture. Added nitrogen improves the yield of all of the mixtures, but the improvement declines as the proportion of vetch increases. This makes perfect sense. Cereal crops, such as oats, need to take nitrogen from the soil if they can, while legumes, such as vetch, actually fix nitrogen in the soil.

Examples 5.4–5.8 all exhibit interaction even though there is a simple explanation of the non-additivity in some cases.

If there are no fixed block effects then $\mathbb{E}(\mathbf{Y}) = \tau$. Using similar arguments to those in Section 4.5, we can put

$$\tau_0 = \bar{\tau}\mathbf{u}_0 = P_{V_0}\tau,$$

$$\tau_F = P_{W_F}\tau = P_{V_F}\tau - \tau_0,$$

and

$$\tau_G = P_{W_G}\tau = P_{V_G}\tau - \tau_0.$$

The vector τ_F is called the *main effect* of treatment factor F. Sometimes the entries in τ_F, which must sum to zero, are called the *effects* of F. Similarly, τ_G is the main effect of treatment factor G.

Write τ_{FG} for the projection of τ onto $W_{F\wedge G}$. This vector τ_{FG} (or its entries) is called the F-by-G *interaction*. Thus the F-by-G interaction is zero if and only

if $\tau \in V_F + V_G$. If $W_F \perp W_G$ then Equation (5.1) shows that $\tau_0 + \tau_F + \tau_G$ is the projection of τ onto $V_F + V_G$, so Equation (5.2) shows that $\tau_{FG} = \tau - \tau_0 - \tau_F - \tau_G$.

Of course, what we mean by "approximately parallel" in graphs such as Figure 5.3 depends on the size of the variance. The estimate of the interaction is a measure of the departure of the fit in the full model V_T from the fit in the submodel $V_F + V_G$. The significance of this departure is assessed by comparing its size (divided by its degrees of freedom) with the residual mean square, as we show in Section 5.5.

5.3 Principles of expectation models

It it time to come clean over an issue that I have been fudging until now. In Chapter 1 I suggested that there was a clear dichotomy between estimation and testing. That may be true when the treatments are unstructured. However, we have now met several cases where we are interested in many different models for the expectation: see Section 2.10, Chapter 3 and Table 5.1. In these circumstances we usually do hypothesis tests to select the smallest model supported by the data, and then estimate the parameters of that model.

A collection of subspaces of V which are to serve as expectation models cannot be arbitrary, but should obey the following principles.

Principle 5.2 (Intersection Principle) If V_1 and V_2 are both expectation models then $V_1 \cap V_2$ should also be an expectation model.

The intersection principle is there to avoid ambiguity in model-fitting. If $\mathbf{y}$ is in (or is close to) $V_1 \cap V_2$, then our fitted model should be V_1 (or a subspace of it) and it should also be V_2 (or a subspace of it): if the intersection principle is not satisfied then we cannot fit the model $V_1 \cap V_2$ but there is no way of deciding between the models V_1 and V_2.

Principle 5.3 (Sum Principle) If V_1 and V_2 are both expectation models then $V_1 + V_2$ should also be an expectation model.

There are three reasons for the sum principle. First, it is a feature of all linear models that if $\mathbf{v}$ and $\mathbf{w}$ are allowable as vectors of fitted values then so should $\mathbf{v}+\mathbf{w}$ be: apply this with $\mathbf{v}$ in V_1 and $\mathbf{w}$ in V_2.

Secondly, it avoids ambiguity, just like the intersection principle. Suppose that $\mathbf{y} = \mathbf{v} + \mathbf{w}$, where $\mathbf{v}$ and $\mathbf{w}$ are non-zero vectors in V_1 and V_2 respectively. Unless $V_1 + V_2$ is an expection model, we shall be forced to make an arbitrary choice between V_1 and V_2.

Thirdly, it implies that all the expectation models are contained in a single maximal model. For most of this book the maximal model is V_T. In Chapters 9 and 12 we shall take a different maximal model when we can assume that some interactions are zero. The maximal model is the starting point for testing hypotheses about

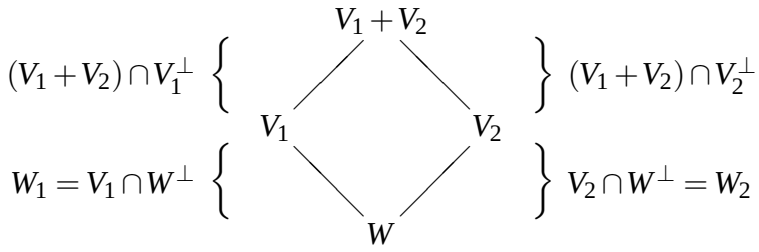

Figure 5.8: Fitting submodels of V_1+V_2

models. It also defines the residual: nothing in the maximal model is *ever* put into residual, even when we fit a smaller model.

The intersection and sum principles are both automatically satsified if the collection of expectation models forms a chain, as in Figures 2.3, 3.1 and 3.2. Moreover, neither is affected by the numbers of replications of the treatments.

The third principle can be affected by the numbers of replications. It is also more controversial.

Principle 5.4 (Orthogonality Principle) If V_1 and V_2 are both expectation models and if $W = V_1 \cap V_2$ then $V_1 \cap W^\perp$ should be orthogonal to $V_2 \cap W^\perp$.

Put $W_1 = V_1 \cap W^\perp$ and $W_2 = V_2 \cap W^\perp$. Then $V_1 = W + W_1$ and $V_2 = W + W_2$, so $V_1 + V_2 = W + W_1 + W_2 = V_1 + W_2$. If W_2 is orthogonal to W_1 then W_2 is orthogonal to V_1 and so $(V_1+V_2)\cap V_1^\perp = W_2$; similarly $(V_1+V_2)\cap V_2^\perp = W_1$. It can be shown that the converse is also true: if either of these equations holds then W_1 is orthogonal to W_2.

Suppose that we have accepted the hypothesis that $\mathbb{E}(\mathbf{Y}) \in V_1 + V_2$ and want to see if we can reduce the expectation model to V_1. To do this, we examine the size of $P_{V_1+V_2}\mathbf{y} - P_{V_1}\mathbf{y}$ in relation to $d_1\,\mathrm{MS}(\text{residual})$, where $d_1 = \dim(V_1+V_2) - \dim V_1 = \dim V_2 - \dim(V_1 \cap V_2) = \dim V_2 - \dim W$. Now, $P_{V_1+V_2}\mathbf{y} - P_{V_1}\mathbf{y}$ is just the projection of $\mathbf{y}$ onto $(V_1+V_2)\cap V_1^\perp$, which is $P_{W_2}\mathbf{y}$ if W_1 is orthogonal to W_2.

If we accept the hypothesis that $\mathbb{E}(\mathbf{Y}) \in V_1$ then we go on to test whether we can reduce the expectation model to W. We do this by examining the size of $P_{V_1}\mathbf{y} - P_W\mathbf{y}$, which is $P_{W_1}\mathbf{y}$.

The models we are discussing are shown in Figure 5.8. Starting at V_1+V_2, there are two routes down to W, and there is no good reason to choose one rather than the other. The advantage of orthogonality is that both routes give the same result, because in both cases we are examining the sizes of $P_{W_1}\mathbf{y}$ and $P_{W_2}\mathbf{y}$. In other words, the test for reducing model V_1+V_2 to V_1 is exactly the same as the test for reducing V_2 to W, and similarly with V_1 and V_2 interchanged.

If W_1 is not orthogonal to W_2, then there are some values of the data vector $\mathbf{y}$ which give contradictory results. Inference can be difficult in these circumstances. Of course, there is unlikely to be a problem if W_1 and W_2 are 'nearly' orthogonal.

In some practical circumstances we are forced to deal with non-orthogonal models, especially when the data are observational. Nonetheless, in this book we limit ourselves to collections of expectation models that conform to all three principles.

The proof of Theorem 4.1 shows that, in general, whether or not the subspaces defined by two treatment factors satisfy the orthogonality principle depends on the numbers of replications of the combinations of levels. Thus the orthogonality principle has implications for the design of experiments.

Example 5.9 (Example 3.3 continued: Drugs at different stages of development) Figure 3.3 shows that the only pair of subspaces that we need worry about is V_A and V_B. Now, $\mathbf{v}$ is in V_A if and only if v_ω is the same for all ω which receive a dose of the new formulation, while $\mathbf{v}$ is in V_B if and only if v_ω is the same for all ω which receive a dose of the old formulation. Hence $V_A \cap V_B = V_F$. Moreover, $V_A + V_B \subset V_T$ and $\dim(V_A + V_B) = \dim(V_A) + \dim(V_B) - \dim(V_A \cap V_B) = 4 + 4 - \dim(V_F) = 8 - 2 = 6 = \dim(V_T)$ so $V_A + V_B = V_T$. Finally, if $\mathbf{v} \in V_A \cap V_F^\perp$ then $v_\omega = 0$ whenever $T(\omega)$ is dose of the new formulation, while if $\mathbf{w} \in V_B \cap V_F^\perp$ then $w_\omega = 0$ whenever $T(\omega)$ is dose of the old formulation, and so the spaces $V_A \cap V_F^\perp$ and $V_B \cap V_F^\perp$ are orthogonal to each other. Therefore the collection of expectation models in Figure 3.3 satisfies the three principles.

Similar arguments about intersection and orthogonality apply to all the models in Chapter 3. To satisfy the sum principle, we have to explicitly ensure that sums of models are included. For example, Figure 5.9 shows the collection of expectation models for Example 3.4. Here V_C is the model with only two treatment parameters, one for the control treatment and one for the rest. The space V_Q has five treatment parameters, one for the control treatment, one for each treatment of the 'quantity' type, and one for the rest. The spaces V_R and V_S are defined similarly by the 'roughage' and 'time' types. Thus the model $V_Q + V_R$ forces all treatments of the 'time type' to have the same parameter, but otherwise allows for different treatment parameters.

Given a collection of expectation models that satisfies the three principles, we test submodels by starting at the maximal model and working downwards. At each stage we test the next submodel by examining the difference between the sums of squares for the fit in the current model and for the fit in the submodel, divided by the difference between their dimensions. This mean square is always compared to the original residual mean square. If we accept the submodel, we move down the diagram to it and continue from there. If at any stage we have rejected all submodels immediately below the current model, we decide that the current one is the smallest that is supported by the data. Because of orthogonality, it does not matter in what order we test submodels when there is a choice.

Once we have decided on the smallest model, we estimate its parameters, which are usually shown in a table of means, along with their standard errors of differences.

Fortunately, the calculations needed for estimation and testing are virtually the same. For estimating the parameters of expectation model V_i we need $P_{V_i}\mathbf{y}$. If

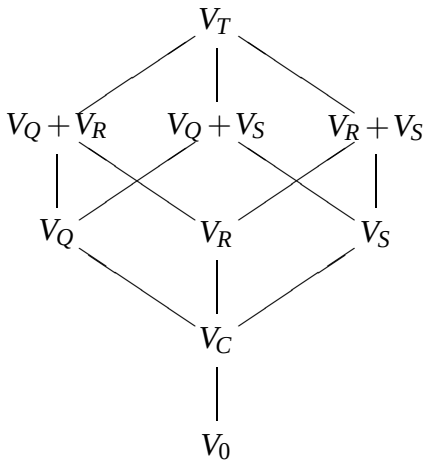

Figure 5.9: Collection of expectation models in Example 3.4

submodel V_j is immediately below V_i then testing for V_j needs the sum of squares $\left\|P_{V_i}\mathbf{y} - P_{V_j}\mathbf{y}\right\|^2$, which is equal to $\|P_{V_i}\mathbf{y}\|^2 - \left\|P_{V_j}\mathbf{y}\right\|^2$. These sums of squares are displayed in an anova table. Convention dictates the opposite order for the anova table to that used in the diagram of submodels, so model testing proceeds by starting at the bottom of the anova table and working upwards. In the next two sections this is described in detail for the case where treatments consist of all combinations of two treatment factors.

5.4 Decomposing the treatment subspace

We return to the case where treatments consist of all combinations of factors F and G. We need to show that the collection of models in Table 5.1 satsifies the Intersection Principle, the Sum Principle and the Orthogonality Principle. Figure 5.2 shows that the only pair of subspaces that we need to check is V_F and V_G. The collection of models contains the additive model $V_F + V_G$, so the Sum Principle is satisfied.

Suppose that $\mathbf{v} \in V_F \cap V_G$ and that α is a plot for which $F(\alpha) = G(\alpha) = 1$. Let β be any other plot and suppose that $G(\beta) = j$. All combinations of the levels of F and G occur, so there is a plot γ for which $F(\gamma) = 1$ and $G(\gamma) = j$. Now, $\mathbf{v} \in V_F$ so $v_\alpha = v_\gamma$. Also $\mathbf{v} \in V_G$, so $v_\gamma = v_\beta$. Hence $v_\alpha = v_\beta$ for all β and so $\mathbf{v} \in V_0$. This shows that $V_F \cap V_G = V_0$, and so the Intersection Principle is satisfied.

Theorem 5.1 shows that the Orthogonality Principle is satisfied if all combinations of levels of F and G have the same replication. That is why we insist, in this book, that all combinations of treatment factors must occur equally often. (The apparent exceptions in Chapter 3 will be explained in Chapter 10.)

We now have

$$V_T = V_0 \oplus W_F \oplus W_G \oplus W_{F \wedge G}$$

orthogonally. As in Section 4.5, this decomposition leads to decompositions of the

symbol	0	F	G	$F \wedge G$
V-subspace	V_0	V_F	V_G	$V_{F\wedge G}$
dimension	1	n_F	n_G	$n_F n_G$
fit = vector of fitted values	$P_{V_0}\mathbf{y}$	$P_{V_F}\mathbf{y}$	$P_{V_G}\mathbf{y}$	$P_{V_{F\wedge G}}\mathbf{y}$
coordinate in fit	mean	$\mathsf{mean}_{F=i}$	$\mathsf{mean}_{G=j}$	$\mathsf{mean}_{F=i,G=j}$
$\|\text{fit}\|^2 = \text{CSS}$	$\frac{\mathsf{sum}^2}{n_F n_G r}$	$\sum_i \frac{\mathsf{sum}^2_{F=i}}{n_G r}$	$\sum_j \frac{\mathsf{sum}^2_{G=j}}{n_F r}$	$\sum_{i,j} \frac{\mathsf{sum}^2_{F=i,G=j}}{r}$

Table 5.2: Quantities associated with the V-subspaces

dimension of V_T and of the the vector of fitted values for the full treatment model, as follows.

$$\begin{array}{lrcccccc} & V_T & = & V_0 & \oplus & W_F & \oplus & W_G & \oplus & W_{F\wedge G} \\ \text{dimension} & n_F n_G & = & 1 & + (n_F-1) & + (n_G-1) & + (n_F-1)(n_G-1) & & & \\ \text{vector} & P_{V_T}\mathbf{y} & = & P_{V_0}\mathbf{y} & + & P_{W_F}\mathbf{y} & + & P_{W_G}\mathbf{y} & + & P_{W_{F\wedge G}}\mathbf{y} \end{array}$$

Here $P_{V_T}\mathbf{y}$ is the fit for the full treatment model, $P_{V_0}\mathbf{y}$ is the fit for the null model, $P_{W_F}\mathbf{y}$ is the estimate of the main effect of F, $P_{W_G}\mathbf{y}$ is the estimate of the main effect of G, and $P_{W_{F\wedge G}}\mathbf{y}$ is the estimate of the F-by-G interaction.

We need to calculate various quantities associated with the W-subspaces. The easiest way of doing this is to calculate them by subtraction from the quantities associated with the V-subspaces, which are shown in Table 5.2. Here we write V_T as $V_{F\wedge G}$, because $T = F \wedge G$.

From these we calculate the quantities associated with the W-subspaces by successively subtracting W-quantities from V-quantities, as shown in Table 5.3.

Example 5.10 (Protein in feed for chickens) Eight newly-hatched chicks took part in a feeding experiment. Four different feeds (A, B, C and D) were made available to two chicks each. The protein in feeds A and B was groundnuts, while the protein in feeds C and D was soya bean. Moreover, feeds C and D contained added fishmeal. Thus the treatments consisted of all combinations of levels of two treatments factors: P (protein) with levels g and s; and M (fishmeal) with levels $+$ and $-$.

The top line of Table 5.4 shows the weights of the chicks (in grams) at the end of six weeks.

The first four lines of Table 5.4 below the description of the treatments are the fits for four expectation models: V_0, V_P, V_M and V_T. All the coordinates are obtained

symbol	0	F	G	$F \wedge G$
W-subspace	$W_0 = V_0$	W_F	W_G	$W_{F\wedge G}$
dimension	1	$n_F - 1$	$n_G - 1$	$(n_F - 1)(n_G - 1)$
effect = extra fit	$P_{V_0}\mathbf{y}$ $= P_{W_0}\mathbf{y}$	$P_{V_F}\mathbf{y} - P_{V_0}\mathbf{y}$ $= P_{W_F}\mathbf{y}$	$P_{V_G}\mathbf{y} - P_{V_0}\mathbf{y}$ $= P_{W_G}\mathbf{y}$	$P_{V_{F\wedge G}}\mathbf{y} - P_{W_0}\mathbf{y} - P_{W_F}\mathbf{y} - P_{W_G}\mathbf{y}$ $= P_{W_{F\wedge G}}\mathbf{y}$
coordinate in effect	mean	$\text{mean}_{F=i} - \text{mean}$	$\text{mean}_{G=j} - \text{mean}$	$\text{mean}_{F=i,G=j} - \text{mean}_{F=i} - \text{mean}_{G=j} + \text{mean}$
$\lVert\text{effect}\rVert^2$ = sum of squares	$\text{CSS}(\text{mean})$	$\text{CSS}(F) - \text{SS}(\text{mean})$	$\text{CSS}(G) - \text{SS}(\text{mean})$	$\text{CSS}(F \wedge G) - \text{SS}(\text{mean}) - \text{SS}(F) - \text{SS}(G)$

Table 5.3: Quantities associated with the W-subspaces: here "effect" is short for "estimated effect"

as simple averages of entries from the 'weight' vector. The squared length of each of these vectors can be calculated either as the sum of the squares of all its entries or by using the appropiate formula for a crude sum of squares.

The line for the estimate of the main effect of P is obtained by subtracting the line for the fit for the null model from the line for the fit for P. The squared length of this vector can be calculated either as the sum of the squares of all its entries or by using the appropriate formula for a sum of squares, that is, by subtracting the squared length of the fit for the null model from the squared length of the fit for P. The line for the estimate of the main effect of M is similar.

The line for the fit for $P+M$ is obtained by adding the lines for the fit for the null model, the main effect of P and the main effect of M. The squared length is equal to the sum of the squares of all its entries; it is also the sum of the squared lengths of three vectors which have been added to obtain it.

The line for the estimate of the P-by-M interaction is obtained by subtracting the line for the fit for $P+M$ from the line for the fit for T. The squared length may be obtained by (a) calculating sum of the squares of all its entries, (b) taking the difference between the squared length of the fit for T and the squared length of the fit for $P+M$, (c) using the formula for the sum of squares for interaction shown in Table 5.3.

Finally, the line for the residual is the difference between the original 'weight' vector and the fit for T. Its squared length is equal to the sum of the squares of all its entries; it is also the difference between the total sum of squares (squared length of the 'weight' vector') and the crude sum of squares for treatments (squared length of the fit for T), as described in Section 2.7.

5.5 Analysis

If the treatments are all combinations of two treatment factors F and G then we replace the treatments line in the anova table by lines for the two main effects and one for the interaction. Theorem 2.4(ii) shows that, if $\mathrm{Cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$, then the expected mean squares for these lines are those shown in Table 5.5.

First we use

$$\frac{\mathrm{MS(interaction)}}{\mathrm{MS(residual)}}$$

to test for interaction. If we cannot assume that the interaction is zero, then report that we cannot use the simpler, additive model, give a table of the treatment means and standard errors of their differences, and stop.

If we can assume that the interaction is zero, report this clearly (in Example 5.10, we can report that added fishmeal increases weight by 39.5gm irrespective of type of protein). Then test separately for each main effect. If either main effect is nonzero, give its table of means and the standard errors of their differences.

Here is a warning about vocabulary. Many scientists say that two factors "interact" to mean that they "act together" in the sense that you can add their separate

									squared length
weight	410	393	442	424	443	441	500	479	1567860.0
feed (T)	A	A	B	B	C	C	D	D	
protein (P)	g	g	g	g	s	s	s	s	
fishmeal (M)	$-$	$-$	$+$	$+$	$-$	$-$	$+$	$+$	
fit for null	441.50	441.50	441.50	441.50	441.50	441.50	441.50	441.50	1559378.0
fit for P	417.25	417.25	417.25	417.25	465.75	465.75	465.75	465.75	1564082.5
fit for M	421.75	421.75	461.25	461.25	421.75	421.75	461.25	461.25	1562498.5
fit for T	401.50	401.50	433.00	433.0	442.00	442.00	489.50	489.50	1567331.0
main effect of P	-24.25	-24.25	-24.25	-24.25	24.25	24.25	24.25	24.25	4704.5
main effect of M	-19.75	-19.75	19.75	19.75	-19.75	-19.75	19.75	19.75	3120.5
fit for $P+M$	397.50	397.50	437.00	437.00	446.00	446.00	485.50	485.50	1567203.0
P-by-M interaction	4.00	4.00	-4.00	-4.00	-4.00	-4.00	4.00	4.00	128.0
residual	8.50	-8.50	9.00	$9-.00$	1.00	-1.00	10.50	-10.50	529.0

Table 5.4: Calculating fits, effects and sums of squares in Example 5.10

source	sum of squares	degrees of freedom	mean square	EMS	variance ratio
F	$\mathrm{SS}(F)$	n_F-1	$\dfrac{\mathrm{SS}(F)}{n_F-1}$	$\dfrac{\|\tau_F\|^2}{n_F-1}+\sigma^2$	$\dfrac{\mathrm{MS}(F)}{\mathrm{MS}(\text{residual})}$
G	$\mathrm{SS}(G)$	n_G-1	$\dfrac{\mathrm{SS}(G)}{n_G-1}$	$\dfrac{\|\tau_G\|^2}{n_G-1}+\sigma^2$	$\dfrac{\mathrm{MS}(G)}{\mathrm{MS}(\text{residual})}$
F-by-G	$\mathrm{SS}(F\wedge G)$	d_{FG}	$\dfrac{\mathrm{SS}(F\wedge G)}{d_{FG}}$	$\dfrac{\|\tau_{FG}\|^2}{d_{FG}}+\sigma^2$	$\dfrac{\mathrm{MS}(F\wedge G)}{\mathrm{MS}(\text{residual})}$

Table 5.5: Treatment lines in the anova table for a factorial experiment with two treatment factors: $d_{FG}=(n_F-1)(n_G-1)$

main effects. This is precisely what statisticians call "zero interaction". Thus it is always a good idea to report the presence or absence of interaction by pointing out what this means in the particular case.

The phrase "main effect" can also be misinterpreted. In Example 5.10, a chicken breeder who thinks that groundnuts are the obvious source of protein may say that the "main effect" of fishmeal is

$$\hat{\tau}_{g,+}-\hat{\tau}_{g,-}=433.00-401.50=31.50.$$

5.6 Three treatment factors

Now suppose that we have three treatment factors F, G and H whose sets of levels are L_F, L_G and L_F respectively. Then $n_F=|L_F|$, $n_G=|L_G|$ and $n_H=|L_H|$. We assume that the treatment set consists of all combinations of levels of F, G and H, so that $\mathcal{T}=L_F\times L_G\times L_H$.

Now the factor $F\wedge G$ is no longer the same as the treatment factor T. In fact, $F\wedge G$ is the function from Ω to $L_F\times L_G$ defined by

$$(F\wedge G)(\omega)=(F(\omega),G(\omega)).$$

In other words, $F\wedge G$ tells us what combinations of levels of F and G is on plot ω, ignoring all other factors.

Definition A *class* of $F\wedge G$ consists of all plots having the same level of F and the same level of G.

Thus $F\wedge G$ has n_Fn_G levels and n_Fn_G classes. The vector subspace $V_{F\wedge G}$ consists of all vectors in V which are constant on each class of $F\wedge G$, so $V_{F\wedge G}$ has dimension n_Fn_G.

Factors $F \wedge H$ and $G \wedge H$ are defined similarly. They have $n_F n_H$ and $n_G n_H$ levels respectively.

As in Section 5.1, we put

$$W_{F\wedge G} = V_{F\wedge G} \cap (V_F + V_G)^{\perp},$$

so that

$$\dim W_{F\wedge G} = (n_F - 1)(n_G - 1).$$

As in Section 5.4, the fit for the expectation model $V_{F\wedge G}$, called the fit for $F \wedge G$, is the projection of the data vector $\mathbf{y}$ onto $V_{F\wedge G}$, whose coordinate on plot ω is equal to the mean of $\mathbf{y}$ on the $F \wedge G$-class containing ω. The crude sum of squares for $F \wedge G$, written $\mathrm{CSS}(F \wedge G)$, is the sum of the squares of the entries in the fit, which is equal to

$$\sum_{F\wedge G\text{-classes}} \frac{(\text{class total})^2}{\text{class size}}.$$

Also as in Section 5.4, the effect of the F-by-G interaction is defined to be the difference between the projection of τ onto $V_{F\wedge G}$ and the projection of τ onto $V_F + V_G$, which is estimated by

$$\text{fit in } V_{F\wedge G} - \text{fit in } (V_F + V_G).$$

The sum of squares for the F-by-G interaction, written $\mathrm{SS}(F \wedge G)$, is the sum of the squares of the entries in the estimate of the effect, so that

$$\begin{aligned} \mathrm{SS}(F\wedge G) &= \mathrm{CSS}(F\wedge G) - \mathrm{SS}(\text{mean}) - \mathrm{SS}(F) - \mathrm{SS}(G) \\ &= \mathrm{CSS}(F\wedge G) - \mathrm{CSS}(F) - \mathrm{CSS}(G) + \mathrm{CSS}(\text{mean}). \end{aligned}$$

The subspaces $V_{F\wedge H}$, $V_{G\wedge H}$, $W_{F\wedge G}$ and $W_{G\wedge H}$, together with the associated fits, crude sums of squares, effects and sums of squares, are defined analogously.

With three treatment factors, we have $T = F \wedge G \wedge H$, and the treatment space V_T contains the seven subspaces shown in Table 5.6. If all treatments have the same replication then every pair of these spaces is orthogonal to each other. Thereom 5.1 shows that W_F is orthogonal to W_G and to W_H; by construction, W_F is orthogonal to $W_{F\wedge G}$ and to $W_{F\wedge H}$; the proof that $W_{G\wedge H}$ is orthogonal to W_F and to $W_{F\wedge G}$ will be given in Chapter 10.

Now,

$$\begin{aligned} V_{F\wedge G} + V_{F\wedge H} + V_{G\wedge H} &= (W_0 + W_F + W_G + W_{F\wedge G}) \\ &\quad + (W_0 + W_F + W_H + W_{F\wedge H}) \\ &\quad + (W_0 + W_G + W_H + W_{G\wedge H}) \\ &= W_0 + W_F + W_G + W_H + W_{F\wedge G} + W_{F\wedge H} + W_{G\wedge H}. \end{aligned}$$

These W-subspaces are mutually orthogonal, so

$$\begin{aligned} \dim(V_{F\wedge G} + V_{F\wedge H} + V_{G\wedge H}) &= 1 + d_F + d_G + d_H + d_{FG} + d_{FH} + d_{GH} \\ &= 1 + d_F + d_G + d_H + d_F d_G + d_F d_H + d_G d_H \\ &= (d_F + 1)(d_G + 1)(d_H + 1) - d_F d_G d_H \\ &= n_F n_G n_H - (n_F - 1)(n_G - 1)(n_H - 1). \end{aligned}$$

factor	subspace	dimension
mean	$W_0 = V_0$	1
F	W_F	$d_F = n_F - 1$
G	W_G	$d_G = n_G - 1$
H	W_H	$d_H = n_H - 1$
$F \wedge G$	$W_{F\wedge G}$	$d_{FG} = (n_F - 1)(n_G - 1)$
$F \wedge H$	$W_{F\wedge H}$	$d_{FH} = (n_F - 1)(n_H - 1)$
$G \wedge H$	$W_{G\wedge H}$	$d_{GH} = (n_G - 1)(n_H - 1)$

Table 5.6: Seven of the W-subspaces for three treatment factors

However, $V_{F\wedge G} + V_{F\wedge H} + V_{G\wedge H} \subset V_T$, and $V_T = V_{F\wedge G\wedge H}$, so we may define a new subspace $W_{F\wedge G\wedge H}$ by

$$W_{F\wedge G\wedge H} = V_{F\wedge G\wedge H} \cap (V_{F\wedge G} + V_{F\wedge H} + V_{G\wedge H})^{\perp}.$$

Put $d_{FGH} = \dim W_{F\wedge G\wedge H}$. Then

$$\begin{aligned} d_{FGH} &= \dim V_{F\wedge G\wedge H} - \dim(V_{F\wedge G} + V_{F\wedge H} + V_{G\wedge H}) \\ &= (n_F - 1)(n_G - 1)(n_H - 1). \end{aligned}$$

The F-by-G-by-H interaction is defined to be the difference between the projection of τ onto $V_{F\wedge G\wedge H}$ and the projection of τ onto $V_{F\wedge G} + V_{F\wedge H} + V_{G\wedge H}$. This is an example of a *three-factor interaction*; by contrast, the previous interactions were *two-factor interactions*. It is estimated by

$$(\text{fit in } V_{F\wedge G\wedge H}) - (\text{fit in } (V_{F\wedge G} + V_{F\wedge H} + V_{G\wedge H})),$$

which is the projection of $\mathbf{y}$ onto $W_{F\wedge G\wedge H}$. The crude sum of squares for $F \wedge G \wedge H$, written $\mathrm{CSS}(F \wedge G \wedge H)$, is given by

$$\mathrm{CSS}(F \wedge G \wedge H) = \sum_{F\wedge G\wedge H\text{-classes}} \frac{(\text{class total})^2}{\text{class size}}.$$

The sum of squares for $F \wedge G \wedge H$, written $\mathrm{SS}(F \wedge G \wedge H)$, which is equal to the sum of the squares of the entries of the projection of $\mathbf{y}$ onto $W_{F\wedge G\wedge H}$, is given by

$$\begin{aligned} \mathrm{SS}(F \wedge G \wedge H) &= \mathrm{CSS}(F \wedge G \wedge H) - \mathrm{SS}(\text{mean}) - \mathrm{SS}(F) - \mathrm{SS}(G) - \mathrm{SS}(H) \\ &\quad - \mathrm{SS}(F \wedge G) - \mathrm{SS}(F \wedge H) - \mathrm{SS}(G \wedge H). \end{aligned}$$

Figure 5.10 shows the collection of expectation model subspaces for three treatment factors. In the analysis, first test for the three-factor interaction. If it is nonzero, report this, give the table of means and standard errors of differences for the factor $F \wedge G \wedge H$, and stop.

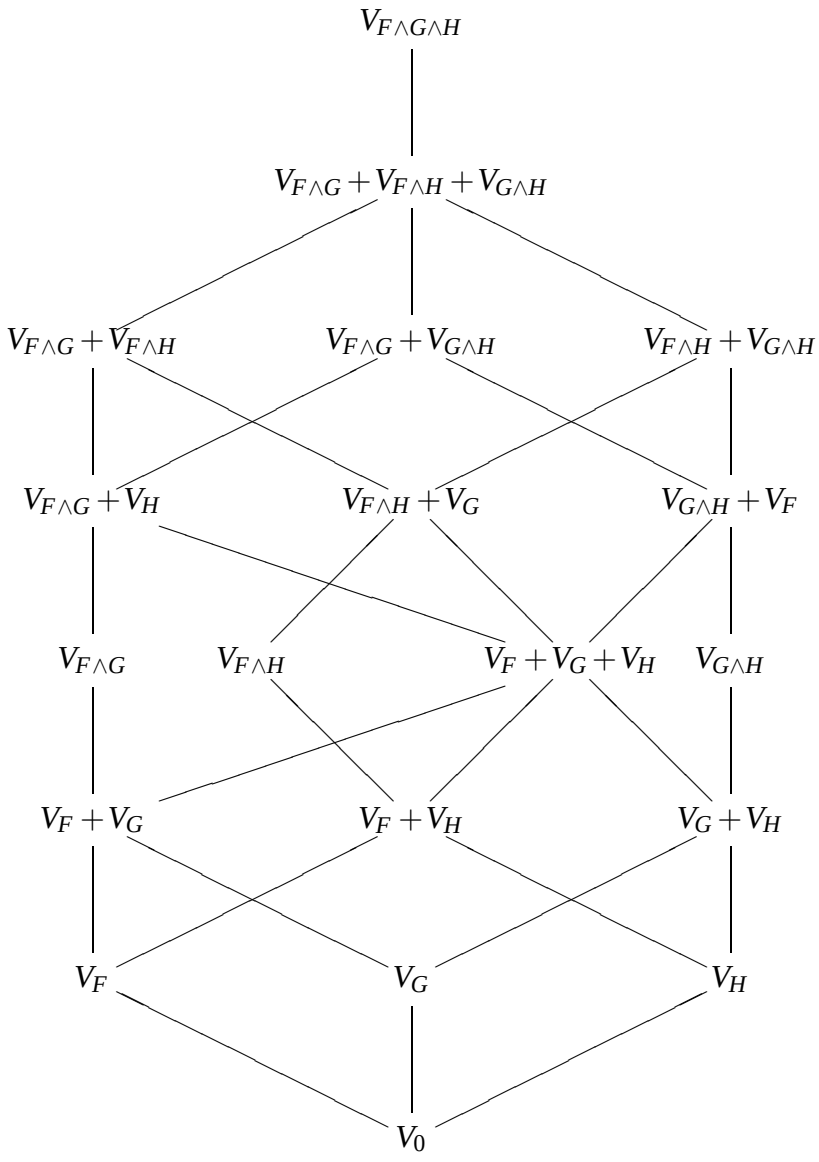

Figure 5.10: Collection of model subspaces for three treatment factors

If the three-factor interaction is zero, then test all of the two-factor interactions. If none of them is zero then the fitted model is $V_{F\wedge G}+V_{F\wedge H}+V_{G\wedge H}$. This is rather hard to report concisely. Give the table of means and seds for each of the factors $F\wedge G$, $F\wedge H$ and $G\wedge H$. It is hard to work out the overall fitted values from these tables, in spite of the additivity, so it is a good idea to also give a three-way table of fitted values and their seds.

If one of the two-factor interactions is zero, suppose that this one is the G-by-H-interaction. Then the fitted model is $V_{F\wedge G}+V_{F\wedge H}$. This should be reported and summarised in a similar way to the previous model.

If two of the two-factor interactions are zero, suppose that these are the F-by-H and the G-by-H interactions. Report that H does not interact with either of the other factors. Now the model is reduced to $V_{F\wedge G}+V_H$, and we can test for the main effect of H. If the main effect of H is not zero then report this and give the tables of means and seds both for $F\wedge G$ and for H. If the main effect of H is zero, then report this and give the tables of means and seds for $F\wedge G$ alone.

If all three of the two-factor interactions are zero, then the model reduces to $V_F+V_G+V_H$. Now test separately for all the main effects. Give the table of means and seds for each factor whose main effect is nonzero.

The extension of these methods to four or more treatment factors continues in the same way.

5.7 Factorial experiments

An experiment is said to be *factorial* if $\mathcal{T}=L_F\times L_G\times L_H\times\cdots$ for treatment factors, F, G, H, ...; that is, if the treatments are all combinations of levels of two or more factors.

Factorial experiments are better than experiments in which only one factors is changed at a time. This is because:

(i) they enable the best combination to be found if the interactions are nonzero;

(ii) they enable testing for the presence of interactions;

(iii) they achieve higher replication for the individual treatment factors.

Example 5.2 revisited (Cowpeas) Suppose that the three methods of cultivation are compared in an experiment in which only the variety C is sown. Figure 5.4 shows that the first method of cultivation would be chosen as the best. A second experiment to compare the five varieties, all with the first method of cultivation, would conclude that variety D is best. The best combination, which is variety E with the second method of cultivation, would not be found by this pair of experiments changing one factor at a time.

Example 5.1 revisited (Rye-grass) Compare the factorial design in Figure 1.1 with the pair of experiments that would be needed to test each factor separately. In the

first year, the three cultivars are tested at zero nitrogen, on two strips each, to find the best cultivar. Two strips each is the minimum to achieve any replication, so this (first) experiment takes as much space as the one in Figure 1.1. At the end of the first experiment the data are analysed to find the bext cultivar. In the second year, the best cultivar is used to test the four quantities of nitrogen. The quantities of nitrogen could be applied to six plots each, as in Figure 1.1. Three or four plots might suffice, but this is not a great saving compared to the overhead costs of the experiment.

This pair of experiments

- takes twice as long as the factorial experiment in Figure 1.1;
- costs almost twice as much;
- gives a larger variance for estimators of differences between levels of nitrogen (unless 24 plots are used in the second year);
- does not allow any detection of whether there is a nonzero cultivar-by-nitrogen interaction;
- may not find the best combination of cultivar with quantity of nitrogen.

Example 5.11 (Park Grass) J. B. Lawes founded Rothamsted Experimental Station in 1843 to investigate the effects of fertilizer on agricultural crops. He and J. H. Gilbert established many long-term experiments that continue to this day. One of these is known as *Park Grass*.

This experiment was started in 1856 in a field that had been in continuous grass for at least one hundred years. The field was divided into 20 plots, and a certain treatment regime allocated to each plot in perpetuity. Every summer the grass is cut for hay, which is made in situ, so that seeds from grass and wild flowers returns to the same plot. What is being observed is species diversity, as well as the yield of hay. Visitors to Rothamsted Experimental Station may visit Park Grass. The treatment allocated to each plot is clearly shown on a post at the end of the plot (there is no *blinding*: see Chapter 7). The plots have no fences or other physical separators between them, but their boundaries are not in doubt, because the mixture of species (clover, cow parsley, fescue etc.) visibly changes along a sharp straight line between each adjacent pair of plots.

Initially there were twenty plots and eighteen treatments, only two of which were replicated. The importance of replication was not appreciated in 1856. Four of the treatments were organic fertilizers, such as bullocks' manure. The remaining fourteen were combinations of different quantities and types of nitrogen with various combinations of other chemicals, as shown in Table 5.7. This table shows the influence of the 'change one factor at a time' dictum (factorial designs were not advocated until the 1920s). Two of the levels of nitrogen occur with four out of the five combinations of the chemicals Na, Mg, P and K. All but one of the levels of nitrogen occur with the combination PKNaMg.

	other chemicals				
	NaMg			–	
nitrogen	PK	P	K	P	–
none	√	√		√	√
0.4 cwt/acre as sulphate of ammonia					√
0.4 cwt/acre as nitrate of soda	√				√
0.8 cwt/acre as sulphate of ammonia	√	√	√	√	
0.8 cwt/acre as nitrate of soda	√				
1.2 cwt/acre as sulphate of ammonia	√				
1.2 cwt/acre as sulphate of ammonia plus silicate of soda	√				

Table 5.7: The fourteen inorganic fertilizer treatments in the Park Grass experiment

5.8 Construction and randomization of factorial designs

In simple cases, such as unstructured plots, orthogonal block designs, or row-column designs (Chapter 6), simply ignore the factorial structure on the treatments while the design is constructed and analysed.

Example 5.12 (Factorial design in unstructure plots) Suppose that there are eighteen plots, with no structure, and that there are two treatment factors: C with two levels c_1 and c_2, and D with three levels d_1, d_2 and d_3. For an equireplicate design, each treatment is assigned to three plots, giving the following systematic design.

plot	$\underbrace{1\ \ 2\ \ 3}$	$\underbrace{4\ \ 5\ \ 6}$	$\underbrace{7\ \ 8\ \ 9}$	$\underbrace{10\ \ 11\ \ 12}$	$\underbrace{13\ \ 14\ \ 15}$	$\underbrace{16\ \ 17\ \ 18}$
treatment	c_1d_1	c_1d_2	c_1d_3	c_2d_1	c_2d_2	c_2d_3

Supopse that the random permutation of the eighteen plots is

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \\ 8 & 5 & 14 & 10 & 1 & 15 & 3 & 16 & 9 & 4 & 12 & 11 & 13 & 7 & 2 & 18 & 6 & 17 \end{pmatrix}.$$

This gives the following plan.

plot	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
C	c_1	c_2	c_1	c_2	c_1	c_2	c_2	c_1	c_1	c_1	c_2	c_2	c_2	c_1	c_1	c_1	c_2	c_2
D	d_2	d_2	d_3	d_1	d_1	d_3	d_2	d_1	d_3	d_2	d_1	d_1	d_2	d_1	d_2	d_3	d_3	d_3

5.9 Factorial treatments plus control

Sometimes the treatments consist of a control treatment in addition to all combinations of two or more treatment factors: see Example 1.14 and Questions 3.2 and 4.1. There are no new principles involved.

As explained in Section 5.3, the non-control treatments should all have the same replication. The relative replication of the control can be decided as in Section 3.1.

For an orthogonal block design, the replication of the control should be an integer multiple of the replication of the other treatments, so that each block can contain all the non-control treatments once and the control on a constant number of plots.

The design is then constructed and randomized ignoring the factorial treatment structure, as in Section 5.8.

In the analysis, the sums of squares for the main effects and interactions of the factorial treatments can be calculated by ignoring the plots with the control treatment. There is a further degree of freedom for comparing the control treatment with the mean of the rest: its sum of squares is calculated as described in Section 3.2.

Questions for Discussion

5.1 These data are based on a genuine agricultural experiment, but have been slightly modified for confidentiality. A completely randomized experiment was conducted with three treatment factors:

type of nitrogen fertilizer — Dutch or English;

method of fertilizer application — applied as a single dressing, or split into two halves and applied half at the normal time and half one month later;

quantity of nitrogen applied — 80 to 280 kg/ha, in increments of 40 kg/ha.

The following sheets show the data, the analysis-of-variance table and the tables of means. Interpret this analysis for the benefit of an agronomist who may not understand statistical jargon.

plot	1	2	3	4	5	6	7
method	split	split	split	single	single	single	single
type	Dutch	Dutch	Dutch	English	English	English	Dutch
nitrogen	280	120	160	160	120	280	80
yield	6.6	4.8	5.3	4.7	4.8	4.6	3.3

plot	8	9	10	11	12	13	14
method	single	single	split	split	split	single	split
type	Dutch	Dutch	English	English	English	English	Dutch
nitrogen	200	240	240	200	80	80	240
yield	3.7	4.5	5.9	6.1	4.8	3.9	5.4

plot	15	16	17	18	19	20	21
method	single	split	split	single	single	single	split
type	English	English	English	Dutch	Dutch	English	Dutch
nitrogen	240	160	280	120	280	200	200
yield	4.4	5.8	5.8	4.1	3.8	4.1	5.2

plot	22	23	24	25	26	27	28
method	split	single	split	split	split	single	split
type	Dutch	Dutch	English	Dutch	Dutch	Dutch	English
nitrogen	80	160	120	120	280	80	80
yield	4.3	3.9	4.9	4.7	5.2	3.9	3.8

plot	29	30	31	32	33	34	35
method	single	single	split	single	split	split	single
type	Dutch	English	Dutch	English	English	English	Dutch
nitrogen	200	160	160	280	200	240	240
yield	4.2	4.4	5.1	4.6	5.7	5.8	4.4

plot	36	37	38	39	40	41	42
method	single	single	split	single	split	single	split
type	English	Dutch	Dutch	Dutch	Dutch	English	English
nitrogen	120	160	240	280	80	240	120
yield	3.5	4.2	5.0	4.6	4.4	4.1	4.6

plot	43	44	45	46	47	48
method	split	single	single	split	single	split
type	English	English	Dutch	English	English	Dutch
nitrogen	160	80	120	280	200	200
yield	4.9	3.7	3.7	5.5	5.0	5.4

```
***** Analysis of variance *****

Variate: yield

Source of variation     d.f.       s.s.       m.s.    v.r.  F pr.
method                     1    12.9169    12.9169   72.01  <.001
type                       1     0.6769     0.6769    3.77  0.064
nitrogen                   5     6.6760     1.3352    7.44  <.001
method.type                1     0.0352     0.0352    0.20  0.662
method.nitrogen            5     1.0044     0.2009    1.12  0.376
type.nitrogen              5     0.4094     0.0819    0.46  0.804
method.type.nitrogen       5     0.6610     0.1322    0.74  0.603
Residual                  24     4.3050     0.1794
Total                     47    26.6848


***** Tables of means *****

Variate: yield

Grand mean  4.690

   method   single    split
             4.171    5.208

     type    Dutch  English
             4.571    4.808

 nitrogen       80      120      160      200      240      280
            4.012    4.387    4.787    4.925    4.937    5.087

   method     type    Dutch  English
   single             4.025    4.317
    split             5.117    5.300

   method nitrogen       80      120      160      200      240      280
   single             3.700    4.025    4.300    4.250    4.350    4.400
    split             4.325    4.750    5.275    5.600    5.525    5.775

     type nitrogen       80      120      160      200      240      280
    Dutch             3.975    4.325    4.625    4.625    4.825    5.050
  English             4.050    4.450    4.950    5.225    5.050    5.125

   method     type nitrogen       80      120      160      200      240
   single    Dutch            3.600    3.900    4.050    3.950    4.450
           English            3.800    4.150    4.550    4.550    4.250
    split    Dutch            4.350    4.750    5.200    5.300    5.200
```

```
          English             4.300    4.750    5.350    5.900    5.850

   method     type nitrogen      280
   single    Dutch             4.200
           English             4.600
    split    Dutch             5.900
           English             5.650


*** Standard errors of differences of means ***

Table                method        type    nitrogen      method
                                                             type
rep.                    24          24           8          12
d.f.                    24          24          24          24
s.e.d.                0.1223      0.1223      0.2118      0.1729

Table                method        type      method
                    nitrogen    nitrogen        type
                                                nitrogen
rep.                    4           4           2
d.f.                   24          24          24
s.e.d.                0.2995      0.2995      0.4235
```

5.2 A food company was interested in the precision of measuring the amount of the vitamin called niacin in bran products. Thirty-six samples of bran flakes were used. Twelve were left alone; twelve were enriched with 4 milligrams of niacin; and twelve were enriched with 8 milligrams of niacin.

Four laboratories were asked to take part in the study. The company was interested in differences between the laboratories, and also wanted to know if the laboratories were consistent in their estimation of differences in the amounts of niacin.

Three samples of each type were sent to each laboratory. Each laboratory put its nine samples into a random order, and then measured the amount of niacin in each sample according to instructions sent by the company. The measurements are shown below. The data have been reordered for ease of manual calculation.

	Niacin enrichment		
	+0 mg	+4 mg	+8 mg
Laboratory 1	8.03	11.50	15.10
	7.35	10.10	14.80
	7.66	11.70	15.70
Laboratory 2	8.50	11.75	16.20
	8.46	12.88	16.16
	8.53	12.64	16.48
Laboratory 3	7.31	11.11	15.00
	7.85	11.00	17.00
	7.92	11.67	15.50
Laboratory 4	8.82	12.90	17.30
	8.76	12.00	17.60
	8.52	13.50	18.40

(a) Calculate the analysis-of-variance table for these data.

(b) Are the laboratories consistent in their measurement of the differences in the amounts of niacin?

(c) Calculate the two tables of means for main effects.

(d) Give the standard error of the difference between two laboratories and the standard error of the difference between two enrichment amounts.

(e) Test the hypothesis that the method of measurement correctly gives the difference between the amount of niacin in the "+0 mg" samples and the "+4 mg" samples.

5.3 A factorial experiment has two treatment factors: C, which has three levels, and D, which has two levels. The design has four complete blocks. The systematic design is shown below.

Block	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4
Plot	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
C	1	1	2	2	3	3	1	1	2	2	3	3	1	1	2	2	3	3	1	1	2	2	3	3
D	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2

Which of the following methods of randomizing is correct? What is wrong with the other methods?

(a) Choose a random permutation of 24 objects and apply it to levels of both C and D at the same time.

(b) Choose a random permutation of 24 objects and apply it to levels of C; then choose another random permutation of 24 objects and apply it to levels of D.

(c) Within each block independently, choose a random permutation of 6 objects and apply it to levels of both *C* and *D* at the same time.

(d) Within each block independently, choose one random permutation of 6 objects and apply it to levels of *C*; then choose another random permutation of 6 objects and apply it to levels of *D*.

(e) Within each block and each level of *C*, toss a coin to decide the order of the two levels of *D*.

(f) Within each block and each level of *D*, choose a random permutation of three objects and apply it to the levels of *C*.

(g) Choose a random permutation of three objects and use it to relabel the levels of *C*; then do a similar thing for *D*.

5.4 A completely randomized factorial experiment is described by Davies in *Statistical Methods in Research and Production*, 1947. The plots were 36 samples of cement. One treatment factor was the gauger used to mix the cement and water and work the mixture. The second treatment factor was the breaker used to test the compressive strength of the mixture after it had set. There were three gaugers and three breakers; the replication was 4. The data, arranged in an order to help manual calculation, are shown below.

Breaker	Gauger		
	1	2	3
1	58.0	44.2	53.6
	47.6	52.8	55.0
	52.8	55.8	61.6
	55.2	49.0	43.2
2	50.2	53.4	55.6
	43.4	62.0	57.2
	62.0	49.6	47.6
	44.0	48.8	56.2
3	53.2	41.8	56.0
	46.0	44.8	44.6
	51.8	48.0	46.8
	41.6	46.0	49.3

Calculate the analysis-of-variance table, the two tables of means for main effects, the standard error of the difference between two breakers, and the standard error of the difference between two gaugers. What do you conclude?

5.5 A group of people researching ways to reduce the amount of cholesterol in the blood are planning their next experiments. One says:

> We know that aspirin reduces cholesterol. Let's experiment with the quantity of aspirin. We could enrol about 150 healthy men into the trial, give 50 of them one aspirin tablet per day for a year, another 50 one and a half aspirin tablets a day, and the final 50 will get two aspirin tablets per day.
>
> When we have decided which quantity is best, we can run another trial to find out if there is any difference between taking the aspirin after breakfast or after dinner.

How do you reply?