Hasse diagrams as a visual aid for linear models and analysis of variance

R. A. Bailey, School of Mathematics and Statistics,
University of St Andrews, St Andrews, Fife KY16 9SS, UK
and School of Mathematical Sciences (emerita),
Queen Mary University of London,
Mile End Road, London E1 4NS, UK

September 24, 2019

Abstract

The expectation part of a linear model is often presented as a single equation with unknown parameters, and the reader is supposed to know that this is shorthand for a whole family of expectation models (for example, is there interaction or not?). It is helpful to list the whole family of models separately and then represent them on a Hasse diagram. This shows which models are sub-models of others, which helps the user to respect marginality when choosing the most parsimonious model to explain the data.

Each row in an analysis-of-variance table corresponds to an edge in the Hasse diagram. In the scaled version of the Hasse diagram, the length of each edge is proportional to the appropriate mean square. This gives a visual display of the analysis of variance. For some people, this is easier to interpret than the standard analysis-of-variance table. Moreover, the scaled Hasse diagram makes clear the difficulties in model choice that can occur under non-orthogonality.

The ideas are illustrated using some familiar families of models defined by crossed and nested factors, possibly including polynomial terms for quantitative factors, as well as some more recently introduced families of models for experiments in biodiversity.

Keywords   Analysis of variance; Hasse diagram; Linear model; Marginality; Mean square; Scaled Hasse diagram.

1   Introduction

R. A. Fisher is deservedly famous for his wide-ranging pioneering work on designed experiments in the first half of the twentieth century: how to think about the factors, how to

design the experiment, how to randomize the layout, how to conduct the experiment, how to analyse the data. What is surprising to today's reader is that Fisher (1925, 1926, 1935) never wrote down an explicit equation to show his underlying assumptions. Later, Frank Yates, who was Fisher's colleague at Rothamsted Experimental Station before succeeding him as Head of Statistics there, wrote out the assumptions in a far more mathematical form, apparently with Fisher's blessing: see Yates (1935, 1964, 1965).

Nowadays, it is absolutely standard to use an equation for the assumed linear model for the response on an experimental unit in a designed experiment. There are two good reasons for this. One is to prevent misunderstanding between different people involved in the experiment or in the exposition of the results. The other is to assist in giving a single formula whose input into an appropriate statistical computer program gives all the information needed for the data analysis, without the need to modify the formula and run the program again. The second aspect was particularly important in the days when you had to wait overnight to get the output from a single run of the program.

However, there is a downside to having a single equation for this linear model. The single equation obscures the fact that the experimenter probably has several models in mind, and would like to choose the simplest one which explains the data adequately. Thus data analysis has two stages: analysis of variance gives the information to use for hypothesis tests, which are used to select this model; and then the estimated parameters for this model, together with their standard errors, are presented. The single equation obscures the process of model selection, and may make it difficult to give appropriate parameters for the selected model. John Nelder, who succeeded Frank Yates at Rothamsted, was well aware of this problem, which he explained in Nelder (1977, 1994, 1998a,b).

Hasse diagrams are often used by algebraists and combinatorialists to display partial orders. Brien (1989), Bailey (2004, 2008, 2015), Clarke (2008) and Kirwan et al. (2009) recommended using them to show the family of models being considered for the expectation of the response. Thirty years ago, such figures were discouraged because of the cost of including them in printed work, but that objection is no longer valid.

The rest of this paper is organized as follows. In Section 2 we use some common examples to show difficulties that can arise when a single equation is used for a whole family of expectation models. Section 3 concentrates on being explicit about the family of models, and showing them on a Hasse diagram. Section 4 relates the Hasse diagram to the data analysis. Section 5 proposes using the output from the analysis of variance (anova) to scale the lengths of the edges of the Hasse diagram. Not only do some people find this display easier to interpret than numbers in an anova table, but the diagram also makes it harder to ignore the possiblity of conflicting results from non-orthogonality.

Until this point, the collections of models are all fairly standard. Section 6 shows the potential for applying the ideas more widely by describing their use in some biodiversity experiments. Finally, Section 7 discusses potential further work.

2   Equations for linear models

Denote by Y_ω the response on experimental unit ω. In a linear model, Y_ω can be expressed as a sum of several items, some of which are constants and some of which are random variables with zero mean. Thus the sum of the former gives the expectation $\mathbb{E}(Y_\omega)$ of Y_ω. Denote by ε_ω the sum of the latter.

Let N be the number of experimental units. Denote by $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ the vectors $(Y_1, \ldots, Y_N)^\top$ and $(\varepsilon_1, \ldots, \varepsilon_N)^\top$ respectively. Then $\mathbb{E}(\mathbf{Y})$ is a vector in $\mathbb{R}^N$ and the $N \times N$ variance-covariance matrix $\mathrm{Cov}(\mathbf{Y})$ is equal to $\mathrm{Cov}(\boldsymbol{\varepsilon})$. At first we assume that $\mathrm{Cov}(\boldsymbol{\varepsilon})$ is a scalar matrix $\sigma^2 \mathbf{I}_N$, but we generalize this later. This section focuses on equations for $\mathbb{E}(\mathbf{Y})$. Three examples demonstrate some misunderstandings that can arise, and serve to introduce our approach.

Example 1 (Polynomial regression) Given a known vector $\mathbf{x}$ in $\mathbb{R}^N$, the expectation part of the linear model for the cubic regression of $\mathbf{Y}$ on $\mathbf{x}$ is

$$\mathbb{E}(Y_\omega) = \alpha_0 + \alpha_1 x_\omega + \alpha_2 x_\omega^2 + \alpha_3 x_\omega^3 \tag{1}$$

for unknown constants α_0, α_1, α_2 and α_3, where x_ω denotes the entry in $\mathbf{x}$ on experimental unit ω. The constant α_0 is called the *intercept*, because it gives $\mathbb{E}(Y_\omega)$ when $x_\omega = 0$.

When model (1) is fitted to some real data, it may be possible to conclude that the right-hand side of Equation (1) can be simplified to a quadratic expression in x_ω. In most such cases, the best-fitting quadratic model is not obtained by replacing α_0, α_1 and α_2 in Equation (1) by their estimated values $\hat{\alpha}_0$, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ and putting $\alpha_3 = 0$. Statisticians know this, but not everyone who uses statistical software to fit models to data does.

In fact, Equation (1) is shorthand for the family of four expectation models shown in Table 1, each with a corresponding *expectation subspace*. Under each linear model, the expectation subspace consists of all potential vectors of fitted values. In particular, V_0 is the one-dimensional subspace of constant vectors. The coefficients in different models in Table 1 are given in different notation to emphasize the fact that the estimated coefficients in a smaller model are not necessarily obtained by fitting a larger model and then equating some coefficients to zero.

Example 2 (One qualitative factor) Suppose that n different levels of a qualitative factor F are applied to the experimental units. Denote by $F(\omega)$ the level of F on experimental unit ω. The expectation part of the assumed linear model is often written as

$$\mathbb{E}(Y_\omega) = \mu + \alpha_{F(\omega)} \tag{2}$$

for unknown constants μ, α_1, $\ldots$, α_n, where μ is sometimes called the intercept, by analogy with Example 1, and sometimes called the *overall mean*. Why is the parameter μ included? It is to emphasize that Equation (2) is shorthand for the family of two expectation models shown in Table 2. Here V_F denotes the n-dimensional subspace of $\mathbb{R}^N$ consisting of vectors

Table 1: Family of expectation models in Example 1

Expectation model	Subspace	Dimension
$\mathbb{E}(Y_\omega) = \alpha_0 + \alpha_1 x_\omega + \alpha_2 x_\omega^2 + \alpha_3 x_\omega^3$	V_3	4
$\mathbb{E}(Y_\omega) = \beta_0 + \beta_1 x_\omega + \beta_2 x_\omega^2$	V_2	3
$\mathbb{E}(Y_\omega) = \gamma_0 + \gamma_1 x_\omega$	V_1	2
$\mathbb{E}(Y_\omega) = \delta_0$	V_0	1

Table 2: Family of expectation models in Example 2

Expectation model	Subspace	Dimension
$\mathbb{E}(Y_\omega) = \alpha_{F(\omega)}$	V_F	n
$\mathbb{E}(Y_\omega) = \mu$	V_0	1

which are constant on each level of F. It is spanned by the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$, where $\mathbf{u}_i$ is the vector whose entry in position ω is 1 if $F(\omega) = i$ and is 0 otherwise.

A problem with Equation (2) is that the parameters are not uniquely identifiable. For example, if we replace μ by $\mu - 1$ and α_i by $\alpha_i + 1$ for $i = 1, \ldots, n$ then we do not change $\mathbb{E}(Y_\omega)$. Most statistical software needs linear constraints imposed on the parameters in order to cope with such linear dependence. In this example, the constraint is typically one of the following, where r_i denotes the number of experimental units receiving level i of F.

(C1) $\sum_i \alpha_i = 0$.

(C2) $\sum_i r_i \alpha_i = 0$.

(C3) $\alpha_1 = 0$.

(C4) $\alpha_n = 0$.

Now there are two ways in which a naive user can misinterpret the output from the data analysis. The first is that the estimated value $\hat{\alpha}_i$ may not be the fitted value on those experimental units receiving level i of F. The second is that, unless constraint (**C2**) is used, $\hat{\mu}$ is not the constant fitted value in the smaller model: in particular, it is not the overall mean.

Table 3: Family of expectation models in Example 3

Expectation model		Subspace	Dimension
$\mathbb{E}(Y_\omega) = \gamma_{ij}$	if $F(\omega) = i$ and $G(\omega) = j$	$V_{F \wedge G}$	nm
$\mathbb{E}(Y_\omega) = \theta_i + \phi_j$	if $F(\omega) = i$ and $G(\omega) = j$	$V_F + V_G$	$n + m - 1$
$\mathbb{E}(Y_\omega) = \alpha_i$	if $F(\omega) = i$	V_F	n
$\mathbb{E}(Y_\omega) = \beta_j$	if $G(\omega) = j$	V_G	m
$\mathbb{E}(Y_\omega) = \mu$		V_0	1

Example 3 (Two qualitative factors) Now extend Example 2 by including a second qualitative factor G with m levels. Denote by $F \wedge G$ the factor whose levels are all combinations of a level of F with a level of G. If all combinations occur then $F \wedge G$ has nm levels and so $\dim(V_{F \wedge G}) = nm$. This also implies that $V_F \cap V_G = V_0$, and hence that

$$\dim(V_F + V_G) = \dim(V_F) + \dim(V_G) - \dim(V_F \cap V_G) = n + m - 1.$$

Here $V_F + V_G$ is the usual notation for the smallest subspace which contains both V_F and V_G. This consists of all vectors of the form $\mathbf{v} + \mathbf{w}$ with $\mathbf{v}$ in V_F and $\mathbf{w}$ in V_G. It is strictly larger than the union $V_F \cup V_G$ unless one of those subspaces is contained in the other. On the other hand, $V_F \cap V_G$ is the largest subspace that is contained in both V_F and V_G.

The expectation model for this situation is typically written as

$$\mathbb{E}(Y_\omega) = \mu + \alpha_{F(\omega)} + \beta_{G(\omega)} + \gamma_{F(\omega),G(\omega)} \tag{3}$$

for unknown constants μ, α_i, β_j and γ_{ij}, where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Table 3 shows the family of expectation models covertly indicated by Equation (3). The expectation subspace $V_F + V_G$ corresponds to the largest model in which the effects of F and G are additive.

Now there is a third problem, in addition to the two noted in the previous examples: the parameters in Equation (3) do not all have the same status. If $\hat{\gamma}_{ij} = 0$ for all i and j then the model can be simplified to the additive model with expectation subspace $V_F + V_G$. On the contrary, if $\hat{\beta}_j = 0$ for all j but the $\hat{\gamma}_{ij}$ are not all zero then the model does not simplify at all. This problem was pointed out by Nelder (1977, 1994, 1998a).

There is also a problem with notation. Some people use the notation $F * G$ for the factor $F \wedge G$; some use it for the interaction between F and G; some for the expectation model $V_{F \wedge G}$; and some for the whole family of models in Table 3. Others use $F.G$ or $F\colon G$ or $F \# G$ or $F \times G$ for one or more of these.

5

3 A collection of expectation models

3.1 Expectation subspaces

The problems shown in Examples 1–3 suggest that it would be better to replace the single equation for $\mathbb{E}(Y_\omega)$ by an explicit collection $\mathcal{E}$ of expectation models. As demonstrated in Examples 1–3, each expectation model can be shown first as an equation for $\mathbb{E}(Y_\omega)$ and then identified with its expectation subspace V, which contains all potential vectors of fitted values. Christensen (1987), Saville and Wood (1996) and Bailey (2008) show the utility of thinking about expectation models in terms of subspaces.

The usual inner product on $\mathbb{R}^N$ is defined by $\mathbf{v} \bullet \mathbf{w} = \mathbf{v}^\top \mathbf{w}$. Two vectors are orthogonal to each other if their inner product is zero. If $\mathbf{y}$ is the data vector, then the vector of fitted values for expectation model V is the orthogonal projection of $\mathbf{y}$ onto V, which is the unique vector $P_V \mathbf{y}$ in V such that $\mathbf{y} - P_V \mathbf{y}$ is orthogonal to all vectors in V: in other words $\mathbf{y} - P_V \mathbf{y}$ is in the orthogonal complement $V^\perp$ of V. Let us write this vector $P_V \mathbf{y}$ of fitted values as $\mathrm{Fit}(V)$. The sum of the squares of the fitted values is $(\mathrm{Fit}(V)) \bullet (\mathrm{Fit}(V))$: let us write this as $\mathrm{SSF}(V)$.

If V_1 and V_2 are both in $\mathcal{E}$, we write $V_1 \leq V_2$ to indicate that V_1 is a subspace of V_2, possibly with $V_1 = V_2$, and $V_1 < V_2$ to indicate that $V_1 \leq V_2$ but $V_1 \neq V_2$. If $V_1 < V_2$ then $\mathrm{SSF}(V_1) \leq \mathrm{SSF}(V_2)$, but if the difference is not very large then V_1 may be regarded as explaining the data just as well as V_2.

The relation "is a subspace of" is a *partial order* on $\mathcal{E}$, because it satisfies the following conditions.

(PO1) $V \leq V$ for all V in $\mathcal{E}$.

(PO2) If V_1 and V_2 are in $\mathcal{E}$, and $V_1 \leq V_2$ and $V_2 \leq V_1$, then $V_1 = V_2$.

(PO3) If V_1, V_2 and V_3 are in $\mathcal{E}$, and $V_1 \leq V_2$ and $V_2 \leq V_3$, then $V_1 \leq V_3$.

Thus $\mathcal{E}$ is called a *partially ordered set*, or *poset* for short. We need to understand this partial order so that we can understand the process of model selection.

3.2 Showing the subspaces on a Hasse diagram

The Hasse diagram is a well-established method of representing a partial order visually. The Hasse diagram for a collection $\mathcal{E}$ of expectation models has a labelled symbol (such as a dot or a square) for each model subspace V in $\mathcal{E}$. Lines are drawn between some of these symbols in such a way that $V_1 < V_2$ if and only if the symbol for V_1 is lower in the diagram than the symbol for V_2 and is joined to it by a line or series of lines that are traversed upwards. There are no horizontal lines. It is helpful to show $\dim(V)$ beside the symbol for V.

Figures 1 and 2 show the Hasse diagrams for the collections of models in Examples 1 and 3.

As with all partial orders, there are two opposite conventions for which way up to draw the Hasse diagram. Kirwan et al. (2009) use the opposite convention to the one in this paper.
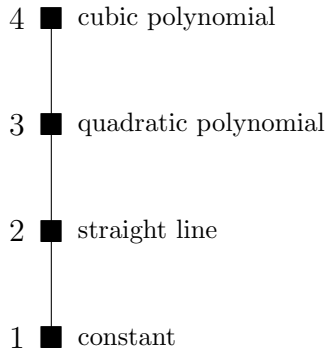
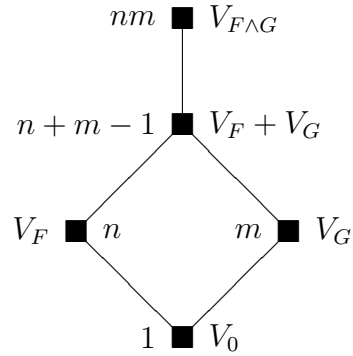Figure 1: Hasse diagram of expectation models in Example 1

Figure 2: Hasse diagram of expectation models in Example 3

3.3   Factorial effects

For factorial designs, a Hasse diagram like that in Figure 2 helps with understanding notions of main effects and interactions. In Example 3, the *interaction* between F and G is the extra fit for $V_{F \wedge G}$ over and above the fit for $V_F + V_G$. This is the difference $\text{Fit}(V_{F \wedge G}) - \text{Fit}(V_F + V_G)$, which can be associated with the edge between $V_{F \wedge G}$ and $V_F + V_G$ in the diagram. This interpretation of interaction is given by Yates (1937, page 13).

The main effects are slightly more complicated. There are unique vectors $\mathbf{v}_F$ in $V_F \cap V_0^{\perp}$ and $\mathbf{v}_G$ in $V_G \cap V_0^{\perp}$ such that

$$\text{Fit}(V_F + V_G) = \text{Fit}(V_0) + \mathbf{v}_F + \mathbf{v}_G.$$

These vectors $\mathbf{v}_F$ and $\mathbf{v}_G$ are the *main effects* of F and G respectively.

Tjur (1984) defined subspaces V_1 and V_2 to be *geometrically orthogonal* to each other if $V_1 \cap (V_1 \cap V_2)^{\perp}$ is orthogonal to $V_2 \cap (V_1 \cap V_2)^{\perp}$. In Example 3, we have $V_F \cap V_G = V_0$, and so V_F is geometrically orthogonal to V_G if and only if $V_F \cap V_0^{\perp}$ is orthogonal to $V_G \cap V_0^{\perp}$. If V_F is geometrically orthogonal to V_G then

$$\mathbf{v}_F = \text{Fit}(V_F) - \text{Fit}(V_0) = \text{Fit}(V_F + V_G) - \text{Fit}(V_G),$$

and so this main effect can be associated both with the edge joining V_0 to V_F and with the edge joining V_G to $V_F + V_G$. Likewise, in this case

$$\mathbf{v}_G = \text{Fit}(V_G) - \text{Fit}(V_0) = \text{Fit}(V_F + V_G) - \text{Fit}(V_F),$$

and the main effect of G can be associated with the edge joining V_0 to V_G as well as with the edge joining V_F to $V_F + V_G$.

3.4   Conditions on the collection of models

Section 3.3 leads us naturally to consider some conditions that we may expect $\mathcal{E}$ to satisfy. The following two are often mentioned.

7

(L1) If V_1 and V_2 are in $\mathcal{E}$, then $V_1 \cap V_2$ is in $\mathcal{E}$.

(L2) If V_1 and V_2 are in $\mathcal{E}$, then $V_1 + V_2$ is in $\mathcal{E}$.

If Condition (**L1**) is not satisfied then there can be ambiguity in model fitting. For example, suppose that V_1 and V_2 are in $\mathcal{E}$ but that $V_1 \cap V_2$ is not. It may happen that the data vector $\mathbf{y}$ is close to $V_1 \cap V_2$ but that the differences $\text{Fit}(V_1) - \text{Fit}(V_1 \cap V_2)$ and $\text{Fit}(V_2) - \text{Fit}(V_1 \cap V_2)$ are both very small. If we are not allowed to conclude that $\mathbb{E}(\mathbf{Y}) \in V_1 \cap V_2$ then we have no way of deciding whether to report that the fitted model is V_1 or V_2.

It seems to be fairly well agreed that Condition (**L1**) is necessary.

Bailey (2008, 2015) argues that Condition (**L2**) is also needed. For example, modify Example 3 by omitting the two largest models, so that $\mathcal{E} = \{V_0, V_F, V_G\}$. It can happen that $\text{Fit}(V_F) - \text{Fit}(V_0)$ and $\text{Fit}(V_G) - \text{Fit}(V_0)$ are both large relative to the appropriate residual mean square: see Section 4. If we are not allowed to conclude that $\mathbb{E}(\mathbf{Y}) \in V_F + V_G$ then we have to make an arbitrary choice between V_F and V_G.

Condition (**L2**) is not so universally accepted. For example, search designs are used in screening experiments where many factors are tested but it is assumed that only a small number of them will have non-zero effects: see Dean and Lewis (2006). Such experiments do not fit easily within the paradigm of this paper.

Condition (**L2**) has another advantage. It ensures that $\mathcal{E}$ has a single largest model $V_{\max}$, which contains all the others. This gives a unique starting point for most approaches to anova, as described in Section 4. Moreover, many authors, including Gilmour and Trinca (2012), argue that no part of $\text{Fit}(V_{\max})$ should ever be put into a residual sum of squares.

For the rest of this paper, we assume that $\mathcal{E}$ satisfies Conditions (**L1**) and (**L2**). Condition (**L1**) ensures that whenever V_1 and V_2 are in $\mathcal{E}$ then there is a largest subspace in $\mathcal{E}$ among those which are contained in both V_1 and V_2. Condition (**L2**) ensures the dual result: whenever V_1 and V_2 are in $\mathcal{E}$ then there is a smallest subspace in $\mathcal{E}$ among those which contain both V_1 and V_2. In the language of poset theory, this means that Conditions (**L1**) and (**L2**) ensure that the poset $(\mathcal{E}, \leq)$ is a *lattice*. Slightly confusingly, Brien (1989) and Clarke (2008) use the word "lattice" in place of "Hasse diagram".

Conditions (**L1**) and (**L2**) give part of the answer to the question "What expectation models should we include in $\mathcal{E}$?" Now we name a further condition, already discussed in Section 3.3, which helps in the analysis of data.

(L3) If V_1 and V_2 are in $\mathcal{E}$, then V_1 is geometrically orthogonal to V_2.

Condition (**L3**) covers most statisticians' ideas of orthogonality. As discussed in Section 4, if it is satisfied then there is no ambiguity in choosing the model to fit to the data. Hence experiments are often designed in such a way that Condition (**L3**) is satisfied. We do not always assume it here. However, when $\mathcal{E}$ satisfies all three Conditions (**L1**), (**L2**) and (**L3**) then various topics covered later in this paper are simplified.

Table 4: Family of expectation models in Example 4

Expectation model	Subspace	Dimension
$\mathbb{E}(Y_\omega) = \beta_{B(\omega)} + \alpha_{F(\omega)}$	$V_B + V_F$	$b + n - 1$
$\mathbb{E}(Y_\omega) = \gamma_{B(\omega)}$	V_B	b

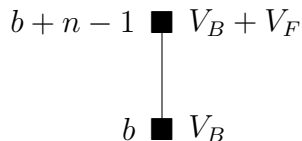$$b + n - 1 \;\blacksquare\; V_B + V_F$$

$$b \;\blacksquare\; V_B$$

Figure 3: Hasse diagram of expectation models in Example 4

3.5 Nuisance factors

In designed experiments, there are often inherent nuisance factors such as blocks, rows or columns. There is usually no interest in estimating their effects, or testing whether these are non-zero, but their presence is taken notice of in the design of the experiment, in its conduct, and in the ensuing data analysis. In such cases, if all of the nuisance factors have fixed effects then it makes sense to include them all in every expectation model in $\mathcal{E}$.

Example 4 (Randomized complete-block design) Suppose that there are b blocks, each consisting of n plots, and that the levels of factor F from Example 2 are applied to plots in a randomized complete-block design. Then the collection of expectation models in Table 2 should be replaced by the collection in Table 4, where B denotes the block factor. Figure 3 gives the corresponding Hasse diagram.

Nuisance factors with random effects are covered in Section 4.2.

3.6 Marginality

Nelder (1977, 1994, 1998a,b) argued that no fitted model should contain an interaction effect if it did not include all the corresponding main effects. For example, if the fitted model in Example 3 includes any part of $V_{F \wedge G} \cap (V_F + V_G)^\perp$ then it should include the whole of $V_F + V_G$. He described the two main effects as *marginal* to their interaction.

Presumably he was thinking of an $n \times m$ two-way table of treatment means, with row means and columns means in its margins. This choice of terminology is slightly unfortunate: it can be argued that the F-by-G interaction is the part of the fit in $V_{F \wedge G}$ that is outside the

9

subspace $V_F + V_G$, so that it is the interaction that is marginal. This wording has certainly caused some misunderstandings.

More generally, the *marginality principle* means that the model for $\mathbb{E}(\mathbf{Y})$ chosen to explain the data vector $\mathbf{y}$ should be one of those in the specified list $\mathcal{E}$, rather than something obtained from $V_{\max}$ by setting an arbitrary subset of the parameters equal to zero.

More recently, this concept of marginality has been re-invented as *strong effect heredity*: see Chipman (1996). A weaker version was introduced by Hamada and Wu (1992). Nelder (1998b) criticized the weak version. See Wu and Hamada (2000) and Mukerjee and Wu (2006) for the related concept of *effect hierarchy*. The term *hierarchy principle* is used by Montgomery (2012) to mean exactly the same thing as marginality principle. However, as Nelder & Lane (1995) and Wu (2015) point out, the concept is essentially present in Yates (1935, 1937, 1965). It certainly seems to have been core to statistical thinking at Rothamsted from the 1930s onwards.

Using a single equation, such as Equation (3), to express all the possibilities for $\mathbb{E}(\mathbf{Y})$ encourages users to fit models in which arbitrary subsets of the parameters are zero. Some software tries to prevent this: for example, when the command `aov` in R (R Core Team, 2012) is used, all terms involving a single factor or variable are fitted before all other terms, even if one of those single factors is an all-encompassing one such as `Treatment`.

In their excellent explanation of the marginality principle for the polynomial model in Example 1, Grafen and Hails (2002, Chapter 10) point out that some software can give a test for fitting x after x^2 and x^3: in other words, it is implicitly allowing the possibility that $\mathbb{E}(Y_\omega) = \theta_0 + \theta_2 x_\omega^2 + \theta_3 x_\omega^3$, even though this does not appear in the list in Table 1.

When the collection of expectation subspaces is ordered in a single vertical line, as in Figures 1 and 3, it is relatively easy to observe the marginality principle. In other cases, viewing the collection of expectation subspaces on a Hasse diagram really helps the user both to understand the marginality principle and to observe it.

3.7   Improving the visual effect of the diagram

Figure 2 is all very well for a generic experiment with two treatment factors, but it may not be appealing to the scientists involved. Perkins et al. (2015) suggested replacing each labelled black square by a small box containing a brief description of the model in words and/or symbols. In this case, the dimension can also be shown inside the box, in parentheses.

Define a subspace V in $\mathcal{E}$ to be *irreducible* if there are no smaller subspaces V_1 and V_2 in $\mathcal{E}$ such that $V_1 + V_2 = V$. If $\mathcal{E}$ satisfies conditions (**L1**) and (**L2**) then V is irreducible if either V is the smallest model $V_{\min}$ in $\mathcal{E}$ or there is a single edge coming down from V. The standard way of writing anova tables encourages readers to ignore models that are not irreducible.

Perkins et al. (2015) suggested using a bold line for the outline of the box for an irreducible model, and a dashed line for other models.

Example 5 (Chicken Feed) Bailey (2008, Chapter 5) gives a small subset of data taken from a larger set given by Carpenter and Duckworth (1941). The experiment is about
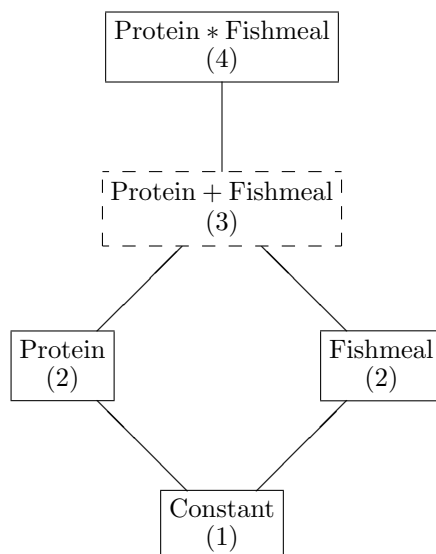
Figure 4: Hasse diagram of expectation models in Example 5

comparing diets for newly-hatched chickens. The diets are the combinations of two two-level factors:

$$\begin{array}{ll} \text{Protein:} & \text{groundnut or soya bean;} \\ \text{Fishmeal:} & \text{added or not.} \end{array}$$

The list of expectation models is similar to that in Table 3, with Protein, Fishmeal, 2 and 2 in place of F, G, n and m. Figure 4 shows the Hasse diagram using the conventions of Perkins et al. (2015), except that we use a normal solid line for an irreducible model and a dashed line for other models.

3.8   Dummy treatments

As Bailey and Łacka (2015) report, it is not uncommon to include a "do nothing" control treatment in experiments, especially in factorial experiments where one of the factors is quantitative. Yates (1937) called this a *dummy* treatment. Since there may be a substantial difference between the dummy treatment and the other treatments, it is a good idea to include a two-level factor which distinguishes the dummy from the rest. The next two examples include such a factor.

Example 6 (Fungicides) Dagnelie (1997) reports an experiment on growing potatoes to compare four fungicides both with each other and with a dummy treatment (called a "witness" in French) consisting of no fungicide. Denote by T the five-level treatment factor, and by F the two-level factor distinguishing actual fungicide from no fungicide. Thus $V_F < V_T$.

Table 5: Family of expectation models in Example 6

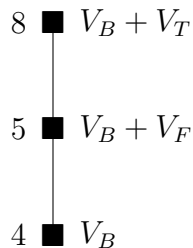| Expectation model	Subspace	Dimension
$\mathbb{E}(Y_\omega) = \beta_{B(\omega)} + \gamma_{T(\omega)}$	$V_B + V_T$	8
$\mathbb{E}(Y_\omega) = \alpha_{B(\omega)} + \delta_{F(\omega)}$	$V_B + V_F$	5
$\mathbb{E}(Y_\omega) = \lambda_{B(\omega)}$	V_B	4

$$8 \quad \blacksquare \quad V_B + V_T$$

$$5 \quad \blacksquare \quad V_B + V_F$$

$$4 \quad \blacksquare \quad V_B$$

Figure 5: Hasse diagram of expectation models in Example 6

The experiment was conducted in four complete blocks, each consisting of five plots. Denoting the nuisance factor for blocks by B, we obtain the collection of expectation models in Table 5 and the Hasse diagram in Figure 5.

Example 7 (Factorial plus dummy) An experiment on the control of eelworms is reported in full in the 1935 Rothamsted Experimental Station Annual Report, and used as an example by Cochran and Cox (1957) and Bailey (2008). There were nine treatments. Eight of these comprised all combinations of two factors:

Type (T):   type of chemical, with four levels;
Dose (D):   single or double.

For the ninth treatment, no chemical was applied. Thus we need an extra two-level factor:

Fumigant (F):   chemical applied or not.

Both T and D need an extra level ("irrelevant" and zero respectively) to cover the ninth treatment. Then $V_F = V_T \cap V_D$.

The experiment was carried out in four blocks, each of which had four plots with no chemical and one for each other treatment. The number of eelworm cysts in a 400 gm soil sample from each plot was recorded before treatments were applied and again after harvest. Since we would expect the latter number to be proportional to the former, Bailey (2008)

Table 6: Family of expectation models in Example 7

Expectation model	Subspace	Dimension
$\mathbb{E}(Y_\omega) = \beta_{B(\omega)} + \gamma_{D(\omega),T(\omega)}$	$V_B + V_{D \wedge T}$	12
$\mathbb{E}(Y_\omega) = \alpha_{B(\omega)} + \phi_{D(\omega)} + \psi_{T(\omega)}$	$V_B + V_D + V_T$	9
$\mathbb{E}(Y_\omega) = \delta_{B(\omega)} + \eta_{T(\omega)}$	$V_B + V_T$	8
$\mathbb{E}(Y_\omega) = \theta_{B(\omega)} + \zeta_{D(\omega)}$	$V_B + V_D$	6
$\mathbb{E}(Y_\omega) = \rho_{B(\omega)} + \tau_{F(\omega)}$	$V_B + V_F$	5
$\mathbb{E}(Y_\omega) = \lambda_{B(\omega)}$	V_B	4

argued that the appropriate response Y_ω is the difference between the logarithm of the latter and the logarithm of the former. For definiteness, logarithms were taken to base e. This gives the collection of expectation models in Table 6, where B denotes the block factor. The corresponding Hasse diagram is in Figure 6.

3.9 Quantitative factors

As discussed in Example 1, when dealing with a quantitative factor with generic value x it usually does not make sense to include any power of x in a polynomial expectation model unless all lower powers are included. If F and G are quantitative factors with n and m levels respectively, generically denoted by x and y, this idea generalizes to the following conditions.

(M1) If the term $x^a y^b$ is in the multinomial expression for $\mathbb{E}(Y_\omega)$ and $c \leq a$ and $d \leq b$ then the term $x^c y^d$ is also included.

(M2) If the term $x^a y^b$ is in the equation for the expectation model then $a < n$ and $b < m$.

When $n = m$, condition **(M2)** is sometimes replaced by the following stronger condition.

(M2′) If the term $x^a y^b$ is in the equation for the expectation model then $a + b < n$.

These generalize to three or more quantitative factors in the obvious way. Families of polynomial expectation models which satisfy condition **(M1)** are called *well-formulated* by Peixoto (1987, 1990) and *well-formed* by Nelder (2000).

In situations, such as response-surface methodology, where all non-nuisance factors are quantitative, it is routine to take $\mathcal{E}$ to be the collection of multinomial expectation models satisfying conditions **(M1)** and either **(M2)** or **(M2′)**. If everybody involved is aware of this, then there is no need to list the models in $\mathcal{E}$ explicitly. However, as noted in Section 3.6, it is still possible to ask your software to fit models which do not satisfy these conditions.
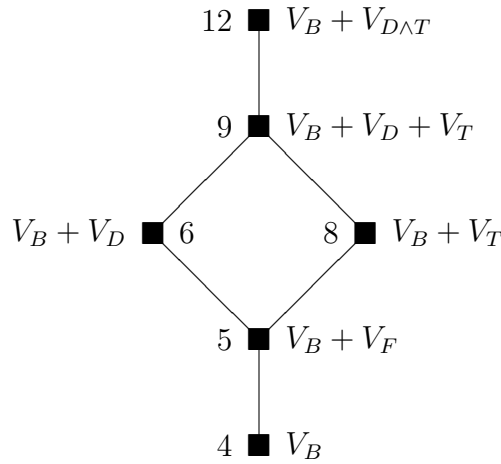
Figure 6: Hasse diagram of expectation models in Example 7

Example 8 (Two quantitative factors) Suppose that $n = m = 3$. Then we may take $\mathcal{E}$ to consist of all multinomials in x and y satisfying conditions (**M1**) and (**M2′**), and also include the full expectation model for $F \wedge G$, which has nine parameters. This gives the Hasse diagram in Figure 7, which is taken from Bailey (2012) but redrawn using the conventions in Section 3.7.

Example 9 (Mixture factorial) Yates (1937) described an experiment on forage crops conducted in five complete blocks. Each plot had a mixture of oats and vetch, to which nitrogen fertilizer was either added or not. The two treatment factors can be summarized as follows.

Mixture (M):   percentage x of vetch (levels 0, 25, 50, 75 and 100);
Fertilizer (F):   added or not (levels 2 and 1 respectively).

Yates (1937) gave the standard anova table for the crop yields in this experiment, using a collection of expectation models like that in Figure 2 but with V_B added to each expectation subspace. This analysis suggests that there is no interaction between F and M. However, Bailey (2008) noticed that the mean improvement from adding fertilizer declines essentially linearly as x increases, and that this accords with well-known differences between cereal crops and legumes. Figure 8 shows this decrease. This figure is consistent with Figure 5.8 of Bailey (2008), which shows all ten treatment means.

This observation suggests including another expectation model, intermediate between $V_B + V_F + V_M$ and $V_B + V_{F \wedge M}$, which allows the mean improvement to be a linear function of x. It is convenient to write this linear term as $(-1)^{F(\omega)} \lambda x_\omega$ for some constant λ. Define the vector $\mathbf{z}$ by $z_\omega = (-1)^{F(\omega)} x_\omega = (2F(\omega) - 3)x_\omega$, and let Z be the one-dimensional subspace spanned by $\mathbf{z}$. This gives the enhanced collection of expectation models in Table 7. Figure 9 gives the corresponding Hasse diagram.
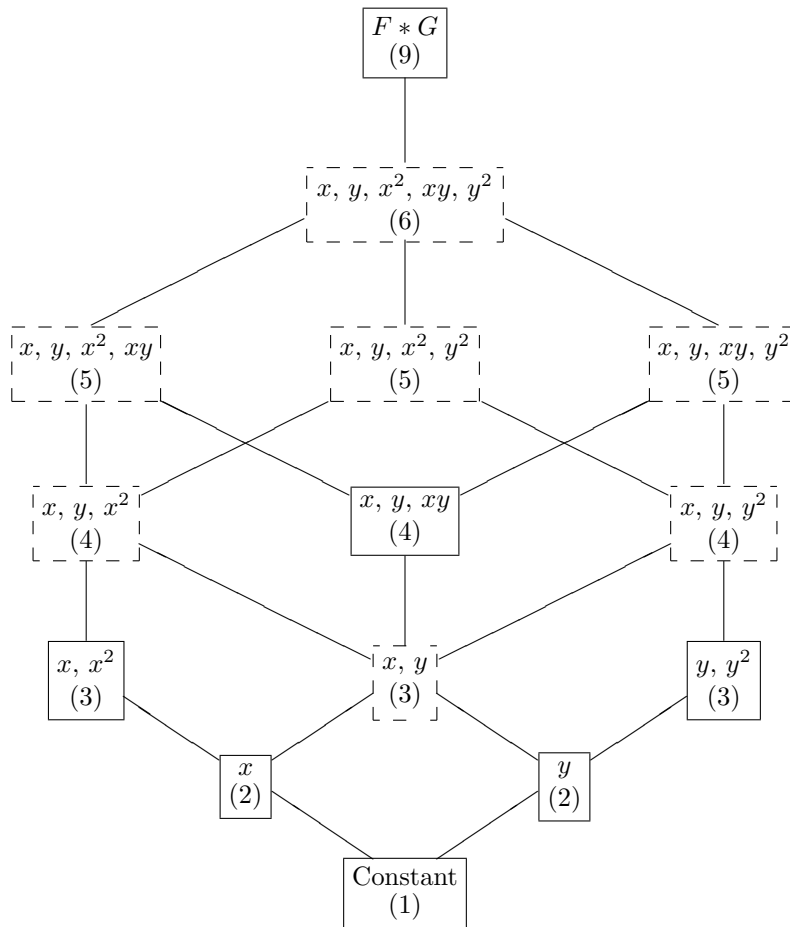
14

$F * G$
(9)

x, y, x^2, xy, y^2
(6)

x, y, x^2, xy x, y, x^2, y^2 x, y, xy, y^2
(5)   (5)   (5)

x, y, x^2 x, y, xy x, y, y^2
(4)   (4)   (4)

x, x^2 x, y y, y^2
(3)   (3)   (3)

x y
(2)   (2)

Constant
(1)

Figure 7: Hasse diagram of expectation models in Example 8

3.10   Neighbour effects or carry-over effects

Sometimes the treatment applied to one experimental unit may have a neighbour effect, or carry-over effect, on other units that are nearby in either space or time. At first sight this appears to give a factorial model, but it is unlikely that there are neighbour effects in the absence of direct effects.

Example 10 (Cross-over trial) Consider a cross-over trial using 16 subjects in several periods to compare four drugs to alleviate symptoms of a long-term health condition. In each period, four subjects are allocated to each drug. A drug taken in one period may have a carry-over effect on that subject in the next period. For simplicity here, we assume that there is a pre-period, so that all periods in the trial have carry-over effects; and that there are 16 periods, so that each subject can receive each ordered pair of drugs in one pair of consecutive periods. Let P, S, D and C be the factors for periods, subjects, drugs and

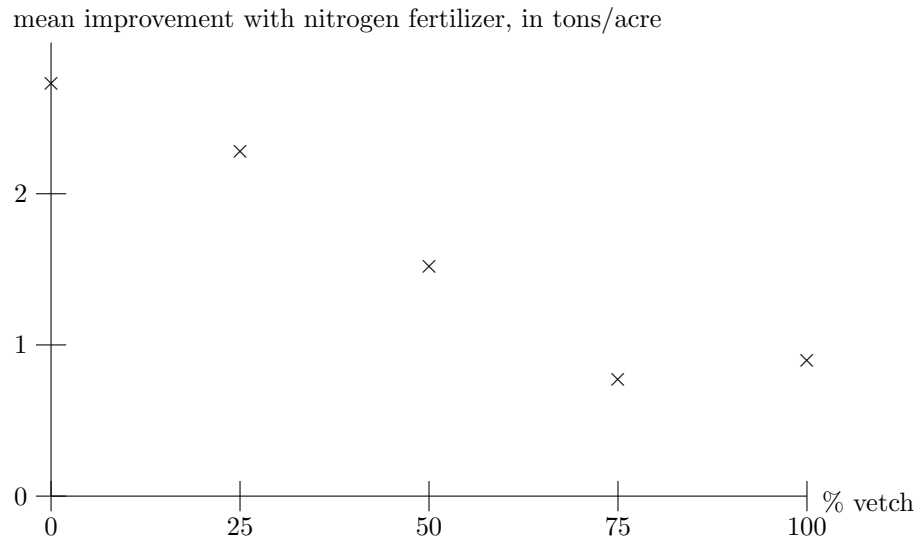mean improvement with nitrogen fertilizer, in tons/acre

Figure 8: The improvement in yield due to added fertilizer in Example 9

Table 7: Family of expectation models in Example 9

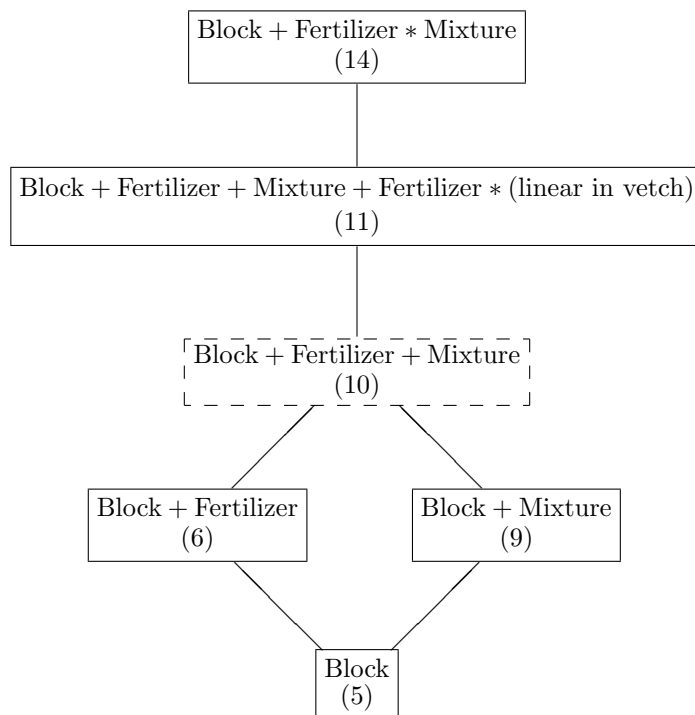| Expectation model	Subspace	Dimension
$\mathbb{E}(Y_\omega) = \alpha_{B(\omega)} + \gamma_{F(\omega),M(\omega)}$	$V_B + V_{F \wedge M}$	14
$\mathbb{E}(Y_\omega) = \beta_{B(\omega)} + \phi_{F(\omega)} + \psi_{M(\omega)} + (-1)^{F(\omega)}\lambda x_\omega$	$V_B + V_F + V_M + Z$	11
$\mathbb{E}(Y_\omega) = \delta_{B(\omega)} + \eta_{F(\omega)} + \zeta_{M(\omega)}$	$V_B + V_F + V_M$	10
$\mathbb{E}(Y_\omega) = \theta_{B(\omega)} + \xi_{F(\omega)}$	$V_B + V_F$	6
$\mathbb{E}(Y_\omega) = \rho_{B(\omega)} + \tau_{M(\omega)}$	$V_B + V_M$	9
$\mathbb{E}(Y_\omega) = \chi_{B(\omega)}$	V_B	5

16

Figure 9: Hasse diagram of expectation models in Example 9

carry-over effects of drugs. For experimental unit ω, denote by ω' the experimental unit which is the same subject in the previous period. Thus $C(\omega) = D(\omega')$.

The most usual single expectation model for this situation is

$$\mathbb{E}(Y_\omega) = \alpha_{P(\omega)} + \beta_{S(\omega)} + \gamma_{D(\omega)} + \delta_{C(\omega)}. \tag{4}$$

Although the treatments appear to be factorial, it does not make sense to assume that C has an effect when D does not. Sometimes researchers allow for the possibility of an interaction between C and D.

Since P and S are nuisance factors, we have the collection of expectation models in Table 8 and the Hasse diagram in Figure 10.

3.11 Recommendation

The overall recommendation of this section is to replace the single equation for the expectation model, such as Equation (1), (2) or (3), by two things.

The first is the explicit list of equations for all individual expectation models, as in Tables 1–8. For a knowledgeable readership, only the irreducible models need to be shown, so long as it is made clear that Condition (**L2**) is assumed. This list is needed for precision about details, assumptions and notation.

Table 8: Family of expectation models in Example 10

Expectation model	Subspace	Dimension
$\mathbb{E}(Y_\omega) = \theta_{P(\omega)} + \phi_{S(\omega)} + \psi_{D(\omega),C(\omega)}$	$V_P + V_S + V_{D \wedge C}$	46
$\mathbb{E}(Y_\omega) = \alpha_{P(\omega)} + \beta_{S(\omega)} + \gamma_{D(\omega)} + \delta_{C(\omega)}$	$V_P + V_S + V_D + V_C$	37
$\mathbb{E}(Y_\omega) = \chi_{P(\omega)} + \eta_{S(\omega)} + \zeta_{D(\omega)}$	$V_P + V_S + V_D$	34
$\mathbb{E}(Y_\omega) = \rho_{P(\omega)} + \tau_{S(\omega)}$	$V_P + V_S$	31

Period + Subject + Drug * Carry-over
(46)

Period + Subject + Drug + Carry-over
(37)

Period + Subject + Drug
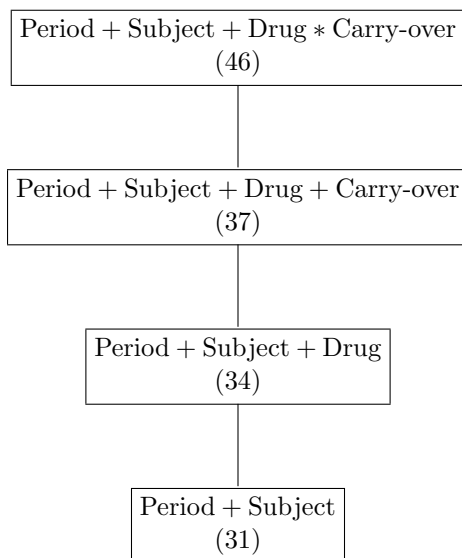(34)

Period + Subject
(31)

Figure 10: Hasse diagram of expectation models in Example 10

The second is the Hasse diagram of expectation models, as in Figures 1–7, 9 and 10. This is needed to show the relationships between the models. It reinforces the ideas of main effects, interactions, and extra fit in one model compared to a smaller one.

In my experience, the inclusion of both of these does increase the readers' understanding, particularly in cross-disciplinary papers. Some people absorb the necessary information better from mathematical notation; some from words; and others from diagrams.

4 Data analysis

4.1 Model choice

Given a non-standard collection $\mathcal{E}$ of expectation models, a bare-hands approach to data analysis begins by fitting each model M in $\mathcal{E}$ separately, then calculating $\mathrm{Fit}(M)$ and $\mathrm{SSF}(M)$.

First assume that $\mathrm{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}_N$. Then the single residual sum of squares is $\mathbf{y}^\top \mathbf{y} - \mathrm{SSF}(V_{\max})$. The corresponding degrees of freedom are $N - \dim(V_{\max})$, and so the residual mean square RMS is

$$\frac{\mathbf{y}^\top \mathbf{y} - \mathrm{SSF}(V_{\max})}{N - \dim(V_{\max})}.$$

As Bailey (2004) and Bowman (2019) point out, the other mean squares that appear in an anova table are associated not with individual models for $\mathbb{E}(\mathbf{Y})$ but with edges between them in the Hasse diagram. If V_1 and V_2 are in $\mathcal{E}$, and $V_1 < V_2$, and V_1 is joined to V_2 by an edge in the Hasse diagram, let us denote that edge by (V_2, V_1). The sum of squares $\mathrm{SS}(V_2, V_1)$, the degrees of freedom $\mathrm{df}(V_2, V_1)$ and the mean square $\mathrm{MS}(V_2, V_1)$ are defined by

$$\begin{aligned}
\mathrm{SS}(V_2, V_1) &= \mathrm{SSF}(V_2) - \mathrm{SSF}(V_1), \\
\mathrm{df}(V_2, V_1) &= \dim(V_2) - \dim(V_1) \quad \text{and} \\
\mathrm{MS}(V_2, V_1) &= \frac{\mathrm{SS}(V_2, V_1)}{\mathrm{df}(V_2, V_1)}.
\end{aligned}$$

Under normality, if we already know that $\mathbb{E}(\mathbf{Y}) \in V_2$ then we can perform an F-test on the ratio $\mathrm{MS}(V_2, V_1) / \mathrm{RMS}$ to decide whether we can simplify the expectation model to V_1.

If (V_2, V_1) is an edge in the Hasse diagram then there is a unique irreducible expectation subspace W such that $V_2 = V_1 + W$. If V_2 is itself irreducible then $V_2 = W$. Define the *label* $\lambda(V_2, V_1)$ of the edge (V_2, V_1) by $\lambda(V_2, V_1) = W$. When Condition (**L3**) is satisfied, all edges with the same label have the same mean square. In Example 3, $\lambda(V_F, V_0) = \lambda(V_F + V_G, V_G) = V_F$ and $\mathrm{MS}(V_F, V_0) = \mathrm{MS}(V_F + V_G, V_G)$.

The usual convention in anova tables is to use this label as the name for the mean square associated with one such edge, if the marginality principle is respected. See Section 4.5.

When $\mathbb{E}(Y_\omega)$ is specified by a single equation with linearly independent parameters, the set of parameters which occur when $\mathbb{E}(\mathbf{Y})$ is in the irreducible expectation subspace W but do not occur for any smaller expectation subspace is often called the *term* for W. A term is not the same thing as an expectation subspace: in Example 3 there are four terms but five expectation subspaces.

4.2 Random effects

So far, we have assumed that $\mathrm{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}_N$. If there are random effects in the model, then $\mathrm{Cov}(\mathbf{Y})$ becomes more complicated. We shall assume that the relationship between the fixed effects and the random effects is orthogonal in the sense that the following two conditions are satisfied.

(O1) $\mathrm{Cov}(\mathbf{Y})$ is an *orthogonal block structure* in the sense defined by Houtman and Speed (1983). This means that the eigenspaces of $\mathrm{Cov}(\mathbf{Y})$ are determined by the pattern of the entries in that matrix, and do not depend on the values of those entries; furthermore, there are no linear constraints on the eigenvalues. These eigenspaces are often called *strata*, and are known in advance of collecting the data.

(O2) Whenever V_1 and V_2 are expectation subspaces in $\mathcal{E}$ and (V_2, V_1) is an edge, the subspace $V_2 \cap V_1^\perp$ is contained in a single stratum. This is always true if every expectation subspace in $\mathcal{E}$ is geometrically orthogonal to every stratum.

The original orthogonal block structures introduced by Nelder (1965) satisfy condition **(O1)**.

Condition **(O2)** is a generalization of the condition given in Section 10.12.2 of Bailey (2008) for the case where all expectation subspaces, as well as the form of $\mathrm{Cov}(\mathbf{Y})$, are defined by factors.

4.3 Model choice under random effects

The choice of expectation model is slightly more complicated when $\mathrm{Cov}(\mathbf{Y})$ is not a scalar matrix. However, when conditons **(O1)** and **(O2)** hold then $V_{\max}^\perp$ is a direct sum of orthogonal spaces, each of which is contained in a single stratum. Suppose that $V_{\max}^\perp = W_1 \oplus \cdots \oplus W_s$, where W_i is the intersection of $V_{\max}^\perp$ with the i-th stratum. The residual mean square $\mathrm{RMS}(i)$ for this stratum is given by

$$\mathrm{RMS}(i) = \frac{(P_{W_i}\mathbf{y}) \bullet (P_{W_i}\mathbf{y})}{\dim(W_i)};$$

its expectation is the eigenvalue for that stratum.

Condition **(O2)** associates a unique stratum with each edge. If (V_2, V_1) is associated with stratum i, we say that $\mathrm{RMS}(i)$ is its *appropriate* residual mean square. In this case $\mathrm{MS}(V_2, V_1)/\mathrm{RMS}(i)$ gives the test statistic for deciding whether the expectation model can be simplified from V_2 to V_1.

It can happen that a stratum is contained in $V_{\max}$, and so has no residual mean square. For example, this occurs in a fractional factorial main-effects-only design for $n + 1$ mutually orthogonal treatment factors each with n levels in an experiment with n^2 experimental units. It also occurs under pseudo-replication: see Hurlbert (1984). If this happens for the stratum associated with edge (V_2, V_1), then there is no statistical test for simplifying $\mathbb{E}(\mathbf{Y})$ from V_2 to V_1.

4.4 An algorithm

Once all the mean squares have been calculated, the Hasse diagram gives a guide to the following algorithm for selecting the model for $\mathbb{E}(\mathbf{Y})$.

1. Start at the top point $V_{\max}$ of the Hasse diagram.

2. At point V, assume that $\mathbb{E}(\mathbf{Y})$ is in, or close to, V. Then

   2.1. Choose a unused downwards edge from point V.

   2.2. Suppose that the point at the bottom of the edge is W.

   2.3. Perform a test of the hypothesis that $P_V(\mathbb{E}(\mathbf{Y})) - P_W(\mathbb{E}(\mathbf{Y})) = 0$, using $\mathrm{MS}(V, W)$ and the residual mean square in the appropriate stratum.

   2.4. If the hypothesis is not rejected then

      2.4.1. conclude that $\mathbb{E}(\mathbf{Y})$ is close enough to W for our purposes;

      2.4.2. do not change the residual mean square;

      2.4.3. move down to point W, and repeat from Step 2.

   2.5. Otherwise, return to Step 2.1, if possible.

   2.6. If there are no unused downwards edges from V then

      2.6.1. report that the model cannot be simplified from V;

      2.6.2. report the vector of fitted values in V;

      2.6.3. if there is more than one edge downwards from V, then the fitted model is additive in some smaller models, so it is equivalent (and helpful) to report the vectors of fitted values for the endpoints of all these edges;

      2.6.4. use appropriate residual mean squares to report standard errors of differences between these fitted values;

      2.6.5. stop.

When the orthogonality condition (**L3**) is satisfied, all routes down the diagram lead to the same conclusion, and so the order of choosing in Step 2.1 is immaterial.

Step 2.4.2 is slightly controversial. Some people like to put $\mathrm{SS}(V, W)$ into the residual sum of squares in this case, and add $\mathrm{df}(V, W)$ to the residual degrees of freedom. However, this has been shown to introduce biases: see Wolde-Tsadik and Afifi (1980), Draper and Smith (1998), Janky (2000) and Gilmour and Trinca (2012). Changing the residual sum of squares in this way can also give different results from different orders of choice in Step 2.1, even when Condition (**L3**) is satisfied.

4.5 Comparison of different approaches

The methodology described in this section explicitly begins by finding $\mathrm{Fit}(V)$ for every expectation subspace in $\mathcal{E}$.

If the models in $\mathcal{E}$ are defined by factors, with crossing and/or nesting relationships, then execution of the command `anova` in Genstat (VSN International, 2017) finds Fit(V) for every irreducible model in $\mathcal{E}$. The output includes these as tables of means, together with their standard errors, which are calculated from the appropriate residual mean squares. The output also gives the anova table for a single route from top to bottom of the Hasse diagram. The statistician or scientist can use that table to select the model, then report the parameters for that model using the appropriate table(s) of means. This is equivalent to Steps 2.6.3 and 2.6.4.

Most software gives this single-route anova table, and maybe some others as well. However, rather than the tables of means, most software other than Genstat gives estimates of all the parameters in the single equation for $\mathbb{E}(\mathbf{Y})$. In order to report meaningful parameters for the selected model, the user has to know precisely which linear constraints were used. Sometimes a user reports a single parameter for a two-level factor without saying which level it is attached to nor which linear constraints have been assumed.

Some software allows the creation of anova tables that do not correspond to a route through the Hasse diagram. Each sum of squares is associated with a term rather than with an edge. As explained by Gardiner and Gettinby (1998), Dean and Voss (1999) and Montgomery (2012), there are several types of sums of squares. Type I corresponds to adjoining the terms in some given order. Grafen and Hails (2002) call these *sequential*. If the order does not respect the marginality principle, then some of the sums of squares do not correspond to edges in the Hasse diagram. Some of the other types do respect marginality, and can be the same when condition (**L3**) is satisfied. For example, Type II gives the sum of squares for the highest edge labelled by any particular irreducible subspace. Type III is the sum of squares for including this term after all of the rest, which does not always correspond to an edge. Grafen and Hails (2002) call this *adjusted*.

5 Scaling the Hasse diagram

5.1 The main idea

Given an anova table, many applied scientists concentrate on the column displaying P-values. Maybe this is because they are not really comfortable with the numerical information in the rest of the table. In particular, they may not absorb the relative sizes of mean squares as quickly as statisticians do.

Bailey and Reiss (2014) suggested overcoming this problem by presenting a new version of the Hasse diagram once the anova calculations have been made. Now each edge (V_2, V_1) is scaled to have length proportional to its mean square MS(V_2, V_1). For comparison, another line is shown at the side of the diagram: its length is proportional to the residual mean square RMS.

The scaled Hasse diagram has many advantages. Most users can see very quickly that the lengths of some edges are much smaller than the RMS, while some others are much larger. The diagram forces the reader to respect the marginality principle during model selection.

It reinforces the message that each mean square is associated with an edge between models, not with a single model. Unlike a single anova table, the scaled Hasse diagram does not restrict the user to a single route from top to bottom.

Of course, the scaled Hasse diagram is an addtion to the anova table, not a replacement for it. When the length of an edge is a similar order of magnitude to RMS, the usual F-test is needed, and this depends on the degrees of freedom. However, with online publication, and the addition of supplementary material in many papers, there should be no objection to including both.

There is a question about what should be considered as the length of the edge (V_2, V_1) in the diagram. If the subspaces are considered as the points where the edges meet, the answer is clear. If the symbol for each subspace is a uniform-sized dot or small square, the length is probably still considered to be the length between their central points. However, if the subspaces are shown in the format used in Figures 4, 7, 9 and 10, the viewer may subconsciously interpret "length" as the length of the line between the edges of the boxes. Making this concept of length proportional to $MS(V_2, V_1)$ is different from making the point-to-point length proportional to it. Moreover, boxes may have unequal heights or widths. Suppose that $MS(V_2, V_1) = MS(V_4, V_3)$. A scaled Hasse diagram which shows the point-to-point lengths of the edges (V_2, V_1) and (V_4, V_3) as equal may, therefore, not show their box-to-box lengths as equal.

Here we avoid this problem by using uniform symbols and point-to-point lengths, but sometimes retain the less concise way of naming the model, deliberately using various different conventions for this. In Section 5.2 we show scaled versions of some of the Hasse diagrams in Section 3, all of which satisfy Condition (**L3**) and have a single residual mean square. We discuss some other potential problems, as well as advantages, as we go along. Section 5.3 shows how to generalize the method to cover multiple residual mean squares. Section 5.4 shows what happens under non-orthogonality in the sense that Condition (**L3**) is violated.

5.2   Examples

In each of these examples, mean squares are rounded versions of those in the cited source but retain enough precision for the scaled Hasse diagram.

Example 5 continued   The anova table given by Bailey (2008, Section 5.5) shows that $MS(V_{P \wedge F}, V_P + V_F) = 128$, $MS(V_P + V_F, V_F) = MS(V_P, V_0) = 4705$, $MS(V_P + V_F, V_P) = MS(V_F, V_0) = 3120$ and $RMS = 132$. These give the scaled Hasse diagram in Figure 11.

As this diagram shows, sometimes the length of an edge can be so small that the edge would be entirely hidden if any symbols were shown for the subspaces at each end. In this case, the symbols for $V_{F \wedge P}$ and $V_F + V_P$ are both omitted, but the positioning of the names of those subspaces should make things clear.

This scaled Hasse diagram shows that, for this example, we have no need to look in tables of critical values of the F-distribution. The length of $(V_{P \wedge F}, V_P + V_F)$ is almost the same as RMS. This means that there is no evidence of interaction, and so we can simplify to the
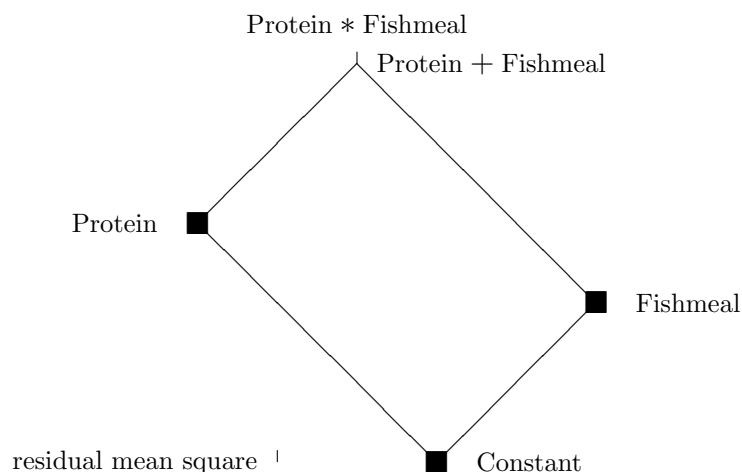
Figure 11: Scaled Hasse diagram in Example 5

additive model $V_P + V_F$. The two edges down from this point are both visibly more than ten times as long as RMS, so that we conclude that neither main effect is zero and we cannot simplify the model any further.

Example 6 continued The anova of the data in Dagnelie (1997, Section 2.2) give $\text{MS}(V_B + V_T, V_B + V_F) = 8$, $\text{MS}(V_B + V_F, V_B) = 387$ and $\text{RMS} = 12$. Figure 12 shows the scaled Hasse diagram. Because the two edges at the point $V_B + V_F$ are part of a single vertical line through that point in the Hasse diagram, we cannot completely omit the symbol for that point. The compromise in Figure 12 is to use a short horizontal line as the symbol.

This scaled Hasse diagram makes it immediately clear that there is no evidence of any differences between the four genuine fungicides but there is definitely a difference between applying fungicide and not applying any. Thus the expectation model simplifies to $V_B + V_F$ but no further.

Example 7 continued The anova of the logcount data given in Bailey (2008, Chapter 4) gives $\text{MS}(V_B + V_{D \wedge T}, V_B + V_D + V_T) = 0.15$, $\text{MS}(V_B + V_D + V_T, V_B + V_T) = \text{MS}(V_B + V_D, V_B + V_F) = 0.11$, $\text{MS}(V_B + V_D + V_T, V_B + V_D) = \text{MS}(V_B + V_T, V_B + V_F) = 1.46$, $\text{MS}(V_B + V_F, V_B) = 1.33$ and $\text{RMS} = 0.17$. This gives the scaled Hasse diagram in Figure 13.

In contrast with Figures 11 and 12, this shows every model subspace as a black square, with the result that three very short edges are completely obscured. This should not matter if the reader can see the unscaled Figure 6 at the same time.

It is clear from Figure 13 that the full model can be simplified first to the additive model $V_B + V_D + V_T$ and then to $V_B + V_T$ but no further. In other words, there is no evidence for interaction between Dose and Type, or for any difference between single and double doses. Thus the differences between the five types of chemical, including the dummy, explain all
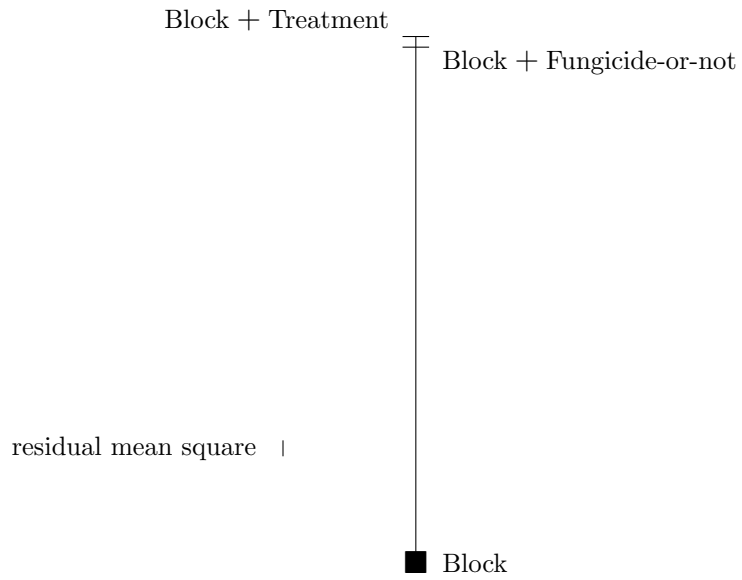
Block + Treatment

Block + Fungicide-or-not

residual mean square |

■ Block

Figure 12: Scaled Hasse diagram in Example 6

$V_{\text{Block}} + V_{\text{Dose}\wedge\text{Type}}$

$V_{\text{Block}} + V_{\text{Dose}} + V_{\text{Type}}$

$V_{\text{Block}} + V_{\text{Type}}$

$V_{\text{Block}} + V_{\text{Dose}}$

$V_{\text{Block}} + V_{\text{Fumigant}}$

residual mean square |

■ V_{Block}
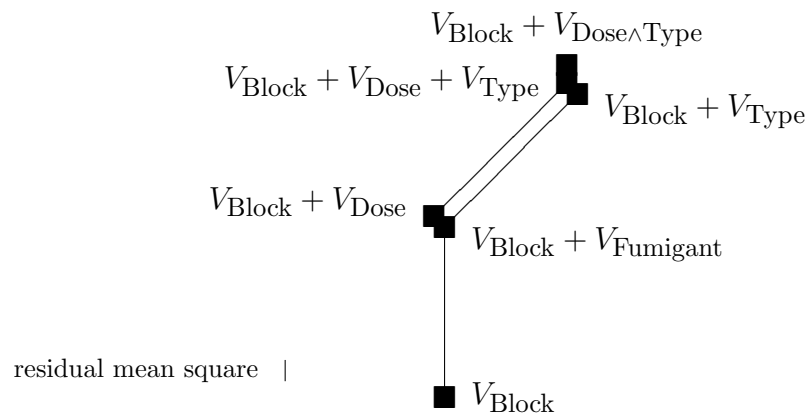
Figure 13: Scaled Hasse diagram in Example 7
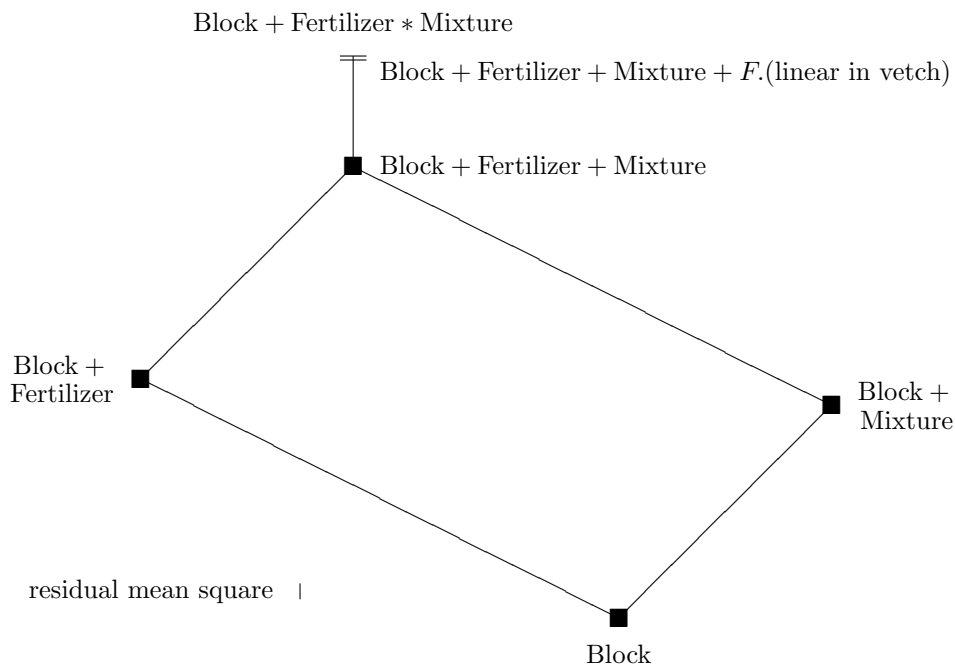
Figure 14: Scaled Hasse diagram in Example 9

the differences between treatments. However, there are more differences among the types than just that between the dummy and the rest.

Example 9 continued Yates (1937) reported the yields in pounds per (1/80)-th of an acre. Bailey (2008) converted the data to tons per acre, showing these numbers correctly in her Figure 5.8. Unfortunately, the treatment totals reported in her Table 5.8 are actually treatment means, with with the consequence that all sums of squares in her Table 5.9 should be multiplied by by 25. This gives $\mathrm{MS}(V_B + V_{F \wedge M}, V_B + V_F + V_M + Z) = 0.20$, $\mathrm{MS}(V_B + V_F + V_M + Z, V_B + V_F + V_M) = 6.68$, $\mathrm{MS}(V_B + V_F + V_M, V_B + V_F) = \mathrm{MS}(V_B + V_M, V_B) = 19.23$, $\mathrm{MS}(V_B + V_F + V_M, V_B + V_M) = \mathrm{MS}(V_B + V_F, V_B) = 33.56$ and RMS $= 1.01$. Hence we obtain the scaled Hasse diagram in Figure 14. This makes it very clear that the full model

$$\text{Block} + \text{Fertilizer} * \text{Mixture}$$

can be simplified to the one in which the effect of F is linear in the proportion of vetch, but no further.

5.3   More than one residual mean square

The approach shown in the previous examples can be extended to situations where there is more than one relevant residual mean square. Show the length of each residual mean square using a different type of line, such as solid, dashed or dotted. In the Hasse diagram itself, use the same type of line for each edge as is used for its appropriate residual mean square.

Example 11 (Falling apples) Bailey (2008, Chapter 8) describes an experiment on cider-apple trees reported in the 1941 Annual Report for Long Ashton Research Station. The 15 treatments were all combinations of five pruning methods (factor P) with three quantities of naphthalene spray (factor S).

It is not practicable to spray individual trees without affecting other trees nearby. Thus the experiment took place on six blocks of five apple trees each, all of the same variety. Each pruning method was used on one tree per block, while each quantity of spray was applied to two whole blocks.

The block effects were regarded as random. Thus the Hasse diagram of expectation models is like the one in Figure 2.

The aim of the experiment was to find an effective way of stopping apples falling to the ground before picking time. There was an unusually strong gale in October that year, so on the following day the experimenters counted how many apples had fallen from each tree, and recorded y_ω as the percentage of apples on tree ω that had fallen.

The anova table in Bailey (2008, Section 8.2) gives $\mathrm{MS}(V_{P\wedge S}, V_P + V_S) = 35$, $\mathrm{MS}(V_P + V_S, V_S) = \mathrm{MS}(V_P, V_0) = 459$ and $\mathrm{MS}(V_P + V_S, V_P) = \mathrm{MS}(V_S, V_0) = 558$. The first two of these must be compared with the residual mean square for trees, which is 74, while the last must be compared with the residual mean square for blocks, which is 239.

Figure 15 shows the scaled Hasse diagram. Edges whose mean squares should be compared with the mean square for trees are shown as solid lines, while those whose mean square should be compared with the mean square for blocks are shown as dotted lines.

From the very short edge at the top of the diagram, it is clear that there is no evidence of interaction and so the full model $V_{P\wedge S}$ can be be simplified to the additive model $V_P + V_S$. The solid edge down from that point is more than five times as long as the residual mean square for trees, so the main effect of pruning is not zero and the model cannot be simplified down that edge. The dotted edge down from that point is the first one that we have seen in examples where we need to consult F-tables. The ratio with the mean square for blocks is 2.33, while the 90% point of F on 2 and 3 degrees of freedom is 5.46, so there is not sufficient evidence to reject a null hypothesis of no difference between sprays. However, this does not imply that there really is no difference. It often happens with designs like this (effectively "split-plot" designs) that the degrees of freedom are too small to draw satisfactory conclusions about the effect of the factor applied to blocks.

As we noted in Section 4.3, sometimes there may be edges in the Hasse diagram for which there is no appropriate residual mean square. In such cases, a special type of line should be used in the scaled Hasse diagram to warn the reader that no test of significance can be performed. See Example 15.

5.4 Non-orthogonality

Because the examples in Sections 5.2–5.3 all satisfy Condition (**L3**), certain subsets of edges are forced to have the same mean square, with the consequence that all routes downward
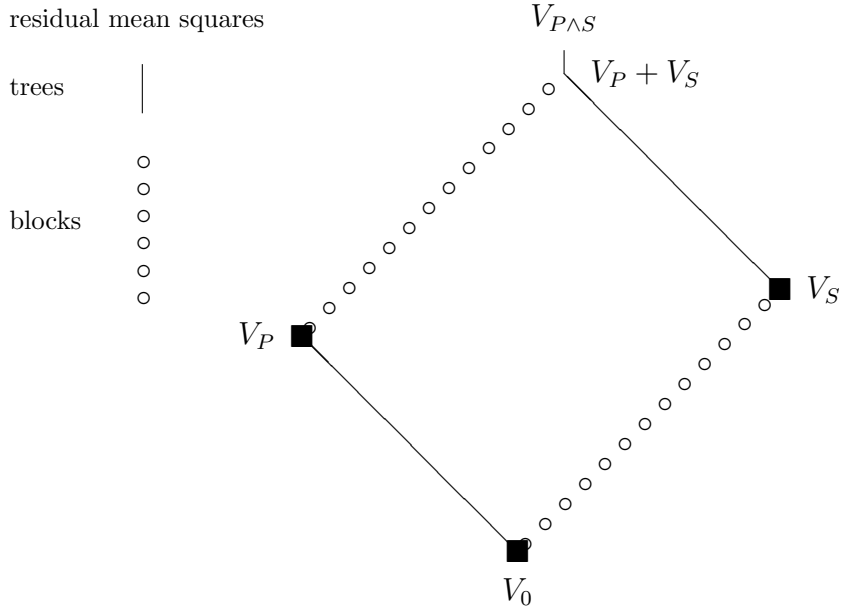
Figure 15: Scaled Hasse diagram in Example 11.

Table 9: Artificial data in Example 12

F	0	1	1
G	1	0	1
data1	6	6	12
data2	$2\sqrt{3}$	$-2\sqrt{3}$	0

from the top of the Hasse diagram lead to the same conclusion. Here we give a small artificial example to show what can happen under non-orthogonality.

Example 12 (Non-orthogonal factors) As Gerami and Lewis (1992) point out, the levels of many two-level factors are "absent" (level 0) and "present" (level 1). When all treatment factors are like this, there may be operational or ethical reasons for omitting the treatment combination in which all factors have level 0.

Suppose that there are just two such factors, F and G. Then $\dim(V_{F \wedge G}) = 3 = \dim(V_F + V_G)$, and so these two model subspaces are the same. The treatment combinations used are shown in Table 9. Suppose that each such combination has replication r and that there are no nuisance factors.

The treatment means of artificial data shown as **data1** in Table 9 give $\mathrm{MS}(V_F, V_0) = \mathrm{MS}(V_G, V_0) = 6r$ and $\mathrm{MS}(V_F + V_G, V_F) = \mathrm{MS}(V_F + V_G, V_G) = 18r$. This gives the scaled

28

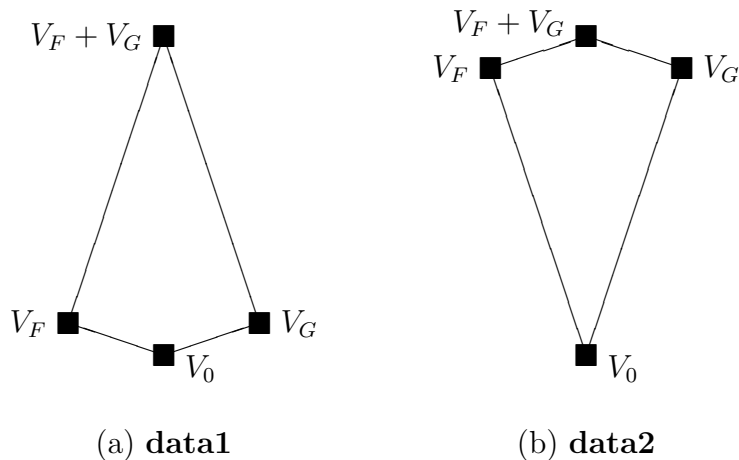(a) **data1**　　　　　　　　(b) **data2**

Figure 16: Scaled Hasse diagrams in Example 12

Hasse diagram in Figure 16(a). If RMS is approximately equal to $6r$ then the model for $\mathbb{E}(\mathbf{Y})$ cannot be simplified from $V_F + V_G$ even though neither V_F nor V_G seems to explain the data significantly better than V_0.

On the other hand, **data2** give $\mathrm{MS}(V_F, V_0) = \mathrm{MS}(V_G, V_0) = 18r$ and $\mathrm{MS}(V_F + V_G, V_F) = \mathrm{MS}(V_F + V_G, V_G) = 6r$. This gives the scaled Hasse diagram in Figure 16(b) and leads to the conclusion that the model for $\mathbb{E}(\mathbf{Y})$ does not need to include both V_F and V_G but that either single factor alone will do.

It is common to present results from data analysis in a single anova table even if there is non-orthogonality among the treatment model subspaces. Including a single scaled Hasse diagram like one of those in Figure 16 should make it clear to both the writer and the reader that it may be problematic to choose between models for $\mathbb{E}(\mathbf{Y})$, even though real data sets often produce less extreme results than these.

6 Biodiversity experiments

In experiments in biodiversity, sometimes the question of interest is whether the measured outcome is better with a more diverse collection of species. The expectation models are not those commonly encountered in statistically designed experiments, as they involve quantitative factors in the setting of mixture experiments where the total quantity is fixed, together with qualitative factors, as well as potential interactions between these. This section gives a few examples to show how Hasse diagrams can illuminate the process.

Example 13 (Compositions of individuals) Reiss et al. (2011) describe an experiment with six types of invertebrate detrivore, labelled A–F. Experimental units were glass jars, each containing 12 individual detrivores. In each monoculture, all individuals were of the

Table 10: Compositions in Example 13

C		Composition	R	x_1	x_2	x_3	x_4	x_5	x_6
1	A	12 of type A	1	12	0	0	0	0	0
$\vdots$			$\vdots$						
6	F	12 of type F	1	0	0	0	0	0	12
7	AB	6 of A, 6 of B	2	6	6	0	0	0	0
$\vdots$			$\vdots$						
21	EF	6 of E, 6 of F	2	0	0	0	0	6	6
22	ABC	4 of A, 4 of B, 4 of C	3	4	4	4	0	0	0
$\vdots$			$\vdots$						
41	DEF	4 of D, 4 of E, 4 of F	3	0	0	0	4	4	4

same type; each duoculture had six of each of two types; while each triculture had four of each of three types. This gave the 41 compositions shown in Table 10. Here x_1 denotes the number of individuals of type A, and so on, up to x_6 denoting the number of individuals of type F. There should be no confusion in writing $C(\omega)$ for the composition in jar ω, even though one of the types is labelled C. There is also a factor Richness (R), whose three levels are the number of different types present.

In the first experiment described by Reiss et al. (2011), each jar was filled with a standard quantity of leaf litter and submerged in a tank of water before twelve individual detrivores were introduced. After 28 days, the quantity of leaf material remaining was measured. The amount eaten was recorded as the response y_ω on jar ω. There was one jar per treatment in each of four blocks. Block effects were regarded as random.

Table 11 shows the family of expectation models that were considered. The symbol T is used to denote models that include the quantities x_1, ..., x_6 of each Type of detrivore. The model Richness says that only the level of R makes any difference. The model Type says that there are constants β_i such that each individual of type i ate amount β_i. In the model Richness * Type, the different types still do not influence each other, but the parameter β_i can change with each level of R.

The corresponding Hasse diagram is in Figure 17. Versions of Table 11 and Figure 17 are in Reiss et al. (2011) and Bailey and Reiss (2014).

The anova in Reiss et al. (2011) gives the scaled Hasse diagram in Figure 18, which is taken from Bailey and Reiss (2014). Here the edges $(V_R + V_T, V_R)$ and (V_T, V_0) are so much longer than the others that the whole figure has to be rather large. Moreover, a very small blob is used as the symbol for each expectation model. A multiple of the residual mean square is shown so that the reader can see it more clearly.

It is evident from Figure 18 that the model Richness does not explain the data at all. The model Type explains it well. There is no evidence that any larger model does any better.

Table 11: Family of expectation models in Example 13

Name	Expectation model	Subspace	Dimension
Composition	$\mathbb{E}(Y_\omega) = \theta_{C(\omega)}$	V_C	41
Richness $*$ Type	$\mathbb{E}(Y_\omega) = \sum_i \gamma_{i,R(\omega)} x_i(\omega)$	V_{R*T}	18
Richness $+$ Type	$\mathbb{E}(Y_\omega) = \phi_{R(\omega)} + \sum_i \psi_i x_i(\omega)$	$V_R + V_T$	8
Richness	$\mathbb{E}(Y_\omega) = \alpha_{R(\omega)}$	V_R	3
Type	$\mathbb{E}(Y_\omega) = \sum_i \beta_i x_i(\omega)$	V_T	6
Constant	$\mathbb{E}(Y_\omega) = \mu$	V_0	1
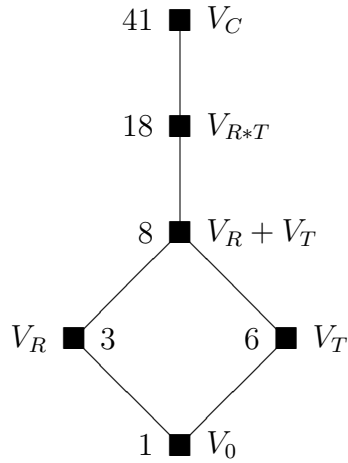
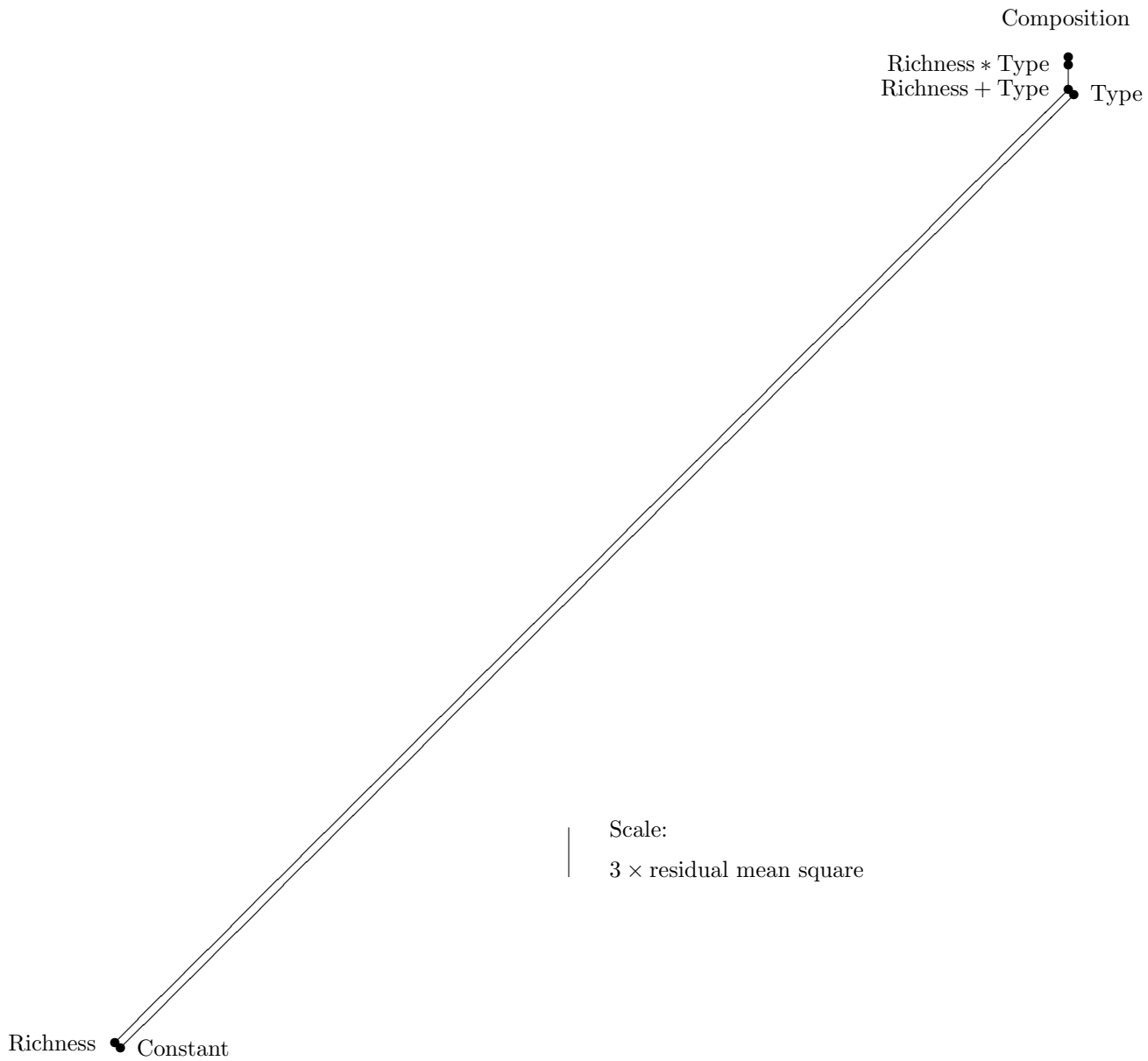Figure 17: Hasse diagram of expectation models in Example 13

Figure 18: Scaled Hasse diagram in Example 13

Example 14 (Habitat complexity) Flores et al. (2016) describe some experiments to investigate how habitat complexity affects biodiversity. In one experiment, the treatments were all combinations of levels of two factors. This experiment was limited to monocultures, and so factor M was simply the type of monoculture, with three levels.

Different artificial habitats were added to the jars. Their complexity was measured by their fractal dimension, which had five levels. The habitats were structures constructed from plastic rings with various shapes and quantities of plastic fronds. One habitat had no structure: this gives a factor S with two levels which distinguishes structure from no structure. Of those with genuine structure, two were made from one ring and two from two rings. This gives a three-level factor for the number of rings. Using a convention different from that in Examples 6–7, this factor is denoted by $A(S)$ and the factor for fractal dimension by $F(A(S))$. The treatments were all combinations of levels of M and $F(A(S))$.

The Hasse diagram for the family of expectation models is in Figure 19.

The anova for one of the responses given by Flores et al. (2016) gives the scaled Hasse diagram in Figure 20. To avoid confusion, two models have not been named. The reader should be able to work out that the lower one is $M + A(S)$ as it is the lowest one joined to both Monculture and Amount by downwards edges. Similarly, the upper one is $S*M+A(S)$. Alternatively, the reader can compare Figure 20 to Figure 19.

Figure 20 shows very clearly that the model for $\mathbb{E}(\mathbf{Y})$ can be simplified to the additive model Monoculture + Structure but no further.

Example 15 (Composition and Temperature) Perkins et al. (2015) conducted an experiment similar to the one in Example 13, using the 15 compositions in Table 12. Only four types of individual were used. Richness level 4 was also included: this has three individuals of each type. The model Richness $*$ Type needs only a single parameter for this composition, and so the dimension of this model subpace is $4 + 4 + 4 + 1 = 13$, while $\dim(V_C) = 15$.

After the experiment had been performed, every response measured gave a value for SSF(Composition) hardly any bigger than SSF(Richness$*$Type). Given that their dimensions are also so close, it was decided to omit the model Richness $*$ Type.

This experiment had an extra complication. Each composition was combined with three levels of Temperature, so there were 45 treatments altogether, each replicated twice. Perkins et al. (2015) listed only the irreducible models for $\mathbb{E}(Y_\omega)$, which are shown in Table 13. Here the factor Temperature is shown as P, to avoid confusion with Type. Also, the model names use $\times$ instead of $*$.

The collection of expection models which is obtained from Table 13 by conforming to Condition (**L2**) gives their Hasse diagram, which is reproduced in Figure 21. Because the model Richness $*$ Type has been excluded, this diagram has the same geometry as the Hasse diagram for three qualitative treatment factors given by Bailey (2008, Chapter 5). This is because the underlying partial orders are isomorphic, with Richness, Type and Temperature playing the role of the three factors, even though the model Type is not defined by a factor. There is a small error in this diagram: the model second from the top should be labelled Composition + Richness $\times$ Temp + Type $\times$ Temp.
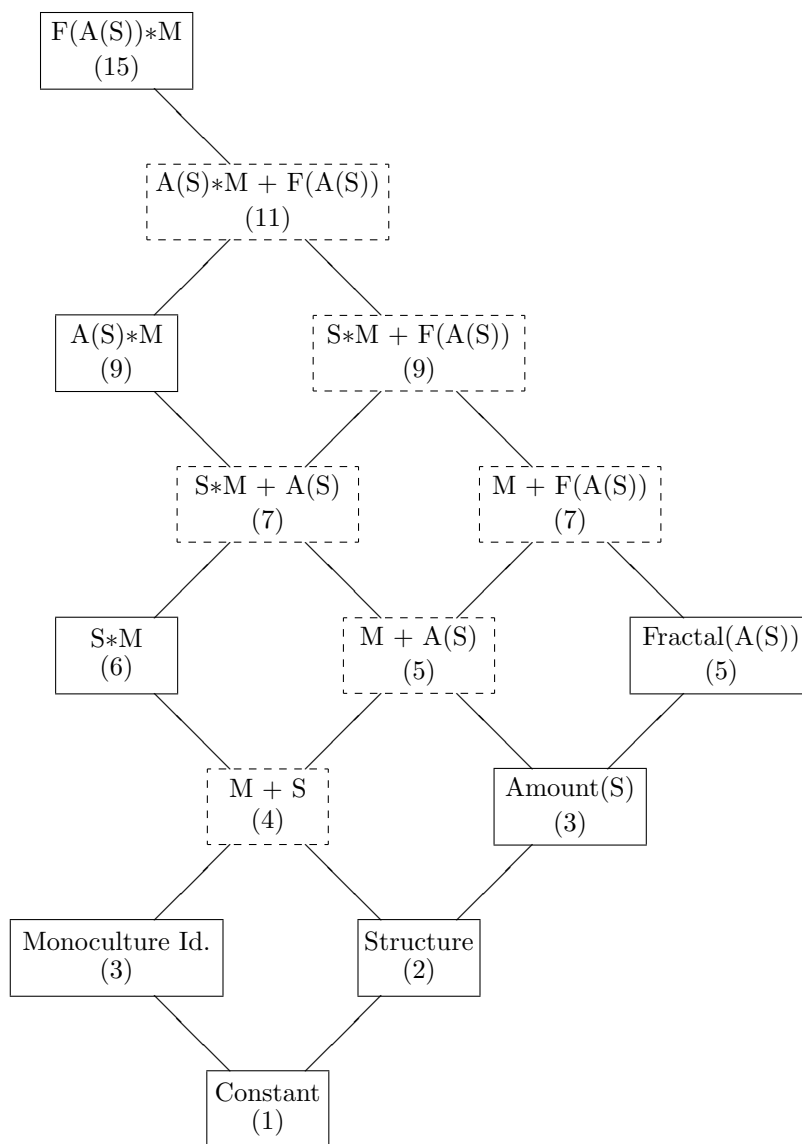
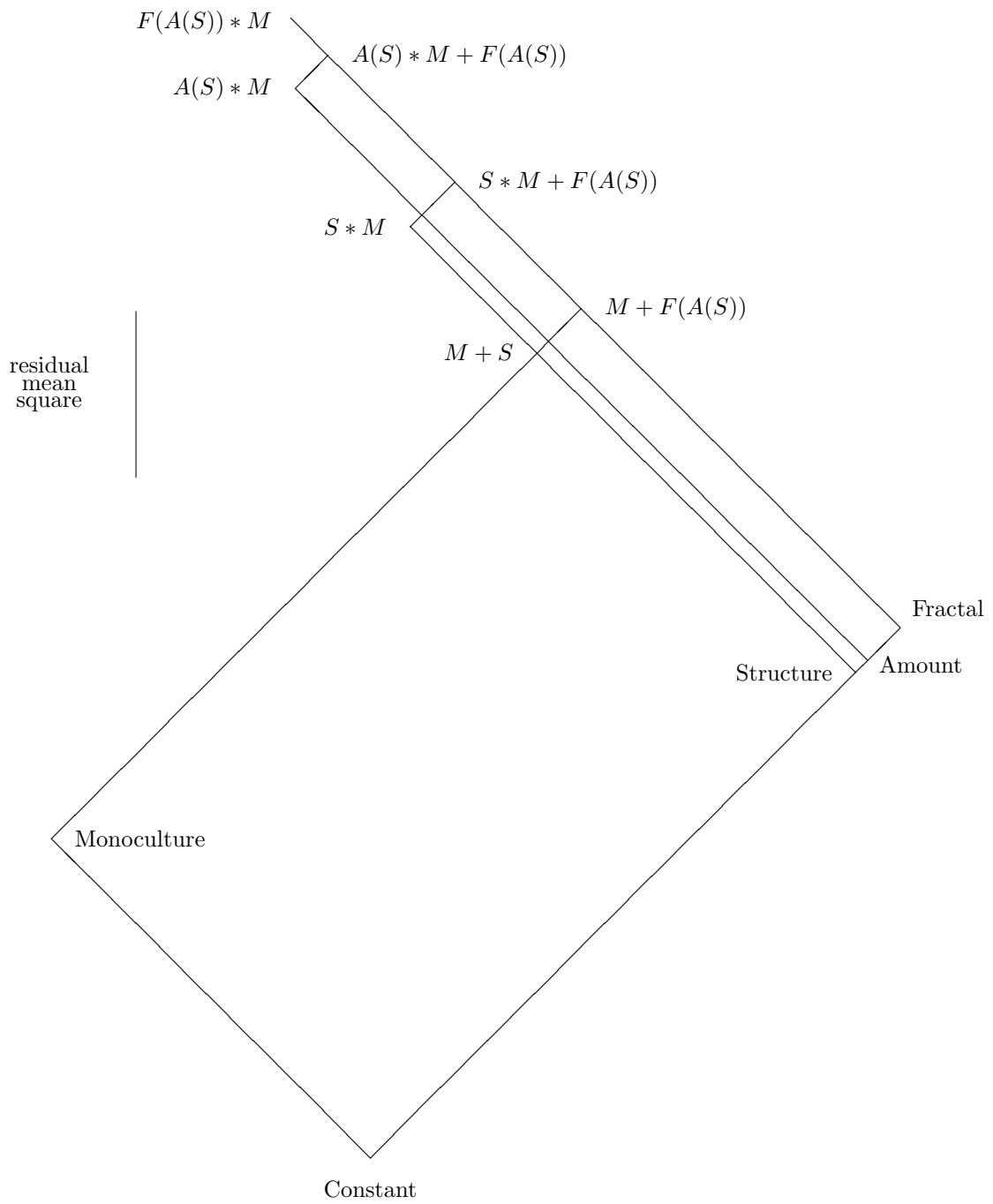Figure 19: Hasse diagram of expectation models in Example 14

Figure 20: Scaled Hasse diagram in Example 14

Table 12: Compositions in Example 15

	Composition		Richness	x_1	x_2	x_3	x_4
1	A	12 of type A	1	12	0	0	0
2	B	12 of type B	1	0	12	0	0
3	C	12 of type C	1	0	0	12	0
4	D	12 of type D	1	0	0	0	12
5	AB	6 of A, 6 of B	2	6	6	0	0
6	AC	6 of A, 6 of C	2	6	0	6	0
7	AD	6 of A, 6 of D	2	6	0	0	6
8	BC	6 of B, 6 of C	2	0	6	6	0
9	BD	6 of B, 6 of D	2	0	6	0	6
10	CD	6 of C, 6 of D	2	0	0	6	6
11	ABC	4 of A, 4 of B, 4 of C	3	4	4	4	0
12	ABD	4 of A, 4 of B, 4 of D	3	4	4	0	4
13	ACD	4 of A, 4 of C, 4 of D	3	4	0	4	4
14	BCD	4 of B, 4 of C, 4 of D	3	0	4	4	4
15	$ABCD$	3 each of A, B, C and D	4	3	3	3	3

Table 13: Family of irreducible expectation models in Example 15

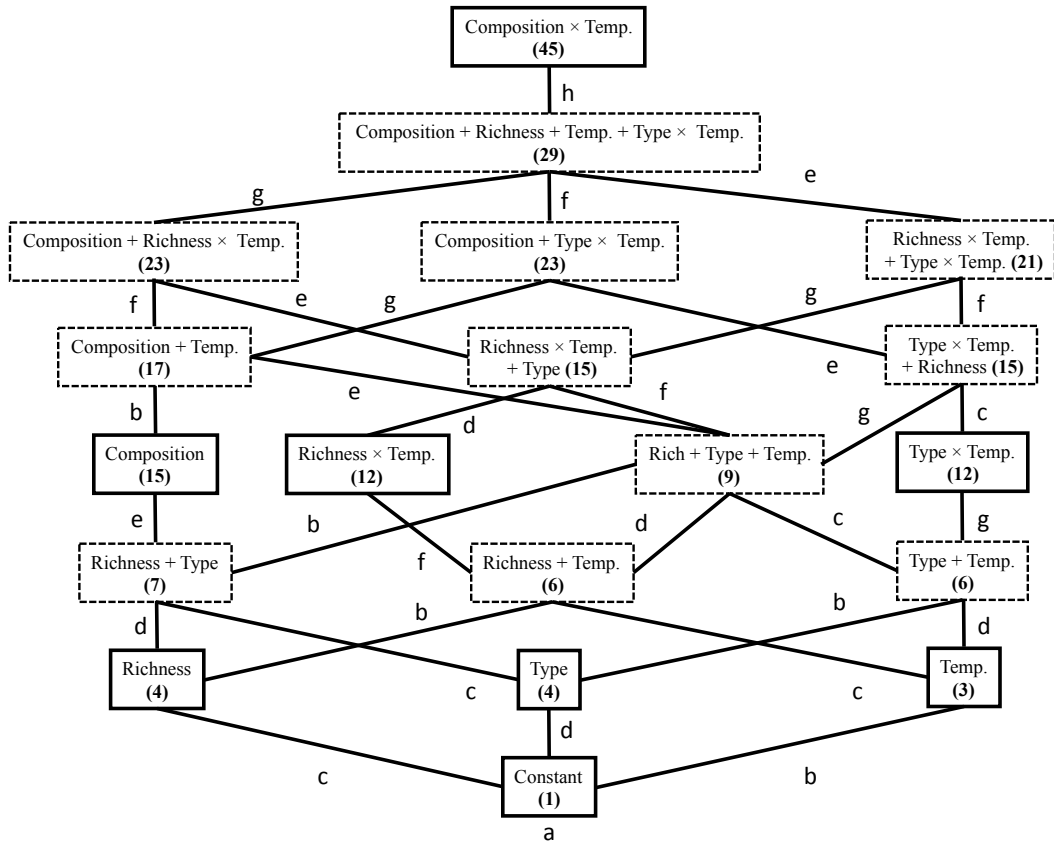| Name	Expectation model	Subspace	Dimension
Composition × Temp	$\mathbb{E}(Y_\omega) = \eta_{C(\omega),P(\omega)}$	V_{C*P}	45
Composition	$\mathbb{E}(Y_\omega) = \theta_{C(\omega)}$	V_C	15
Type × Temp	$\mathbb{E}(Y_\omega) = \sum_i \gamma_{i,P(\omega)} x_i(\omega)$	V_{T*P}	12
Richness × Temp	$\mathbb{E}(Y_\omega) = \delta_{R(\omega),P(\omega)}$	V_{R*P}	12
Richness	$\mathbb{E}(Y_\omega) = \alpha_{R(\omega)}$	V_R	4
Type	$\mathbb{E}(Y_\omega) = \sum_i \beta_i x_i(\omega)$	V_T	4
Temperature	$\mathbb{E}(Y_\omega) = \zeta_{P(\omega)}$	V_P	3
Constant	$\mathbb{E}(Y_\omega) = \mu$	V_0	1

Figure 21: Hasse diagram of expectation models in Example 15

The Hasse diagram in Figure 21 has another new feature. The labelling system described at the end of Section 4.1 is shown explicitly on the edges. Letter b corresponds to the irreducible subspace V_P, using the notation in Table 13. Similarly, letter c corresponds to V_R; letter d corresponds to V_T; letter e corresponds to V_C; letter f corresponds to V_{R*P}; letter g corresponds to V_{T*P}; and letter h corresponds to V_{C*P}. Edges labelled with the same letter must have the same mean square when Condition (**L3**) is satisfied.

Three temperature-controlled rooms in a laboratory were used for this experiment. Each room had a single temperature and two of each composition. Even though Room effects were regarded as random, there was no appropriate residual mean square to compare the main effect of Temperature with. Thus the edges labelled b in Figure 21 had no appropriate residual mean square but all other effects could be assessed.

Figure 22 gives the scaled Hasse diagram for one of the responses given by Perkins et al. (2015): it is the amount of algal consumption, which they called *herbivory*.

This scaled Hasse diagram has two new features. Unlike all the previous ones, it is not possible to draw the Hasse diagram in the plane without its edges crossing. Therefore, it is
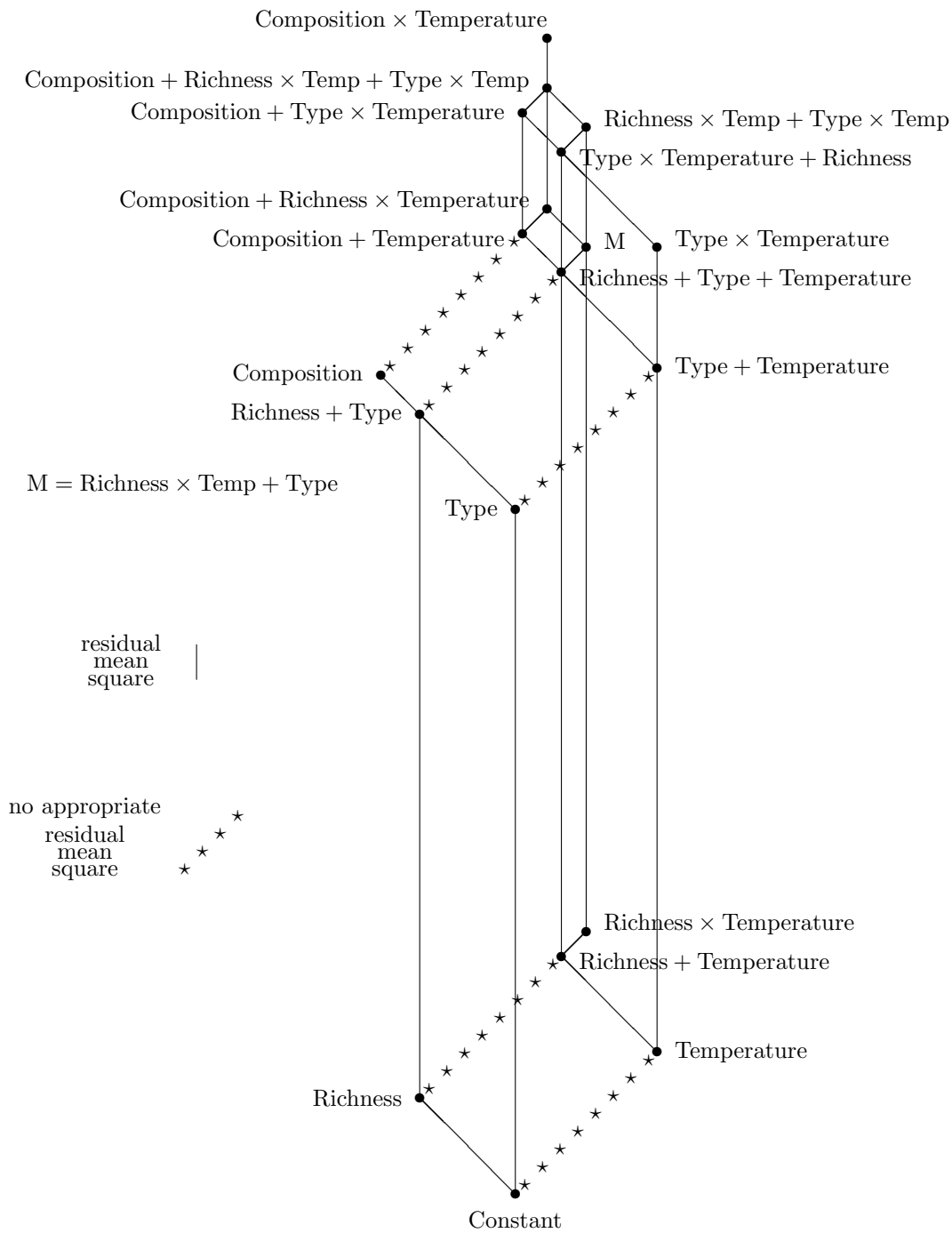
Figure 22: Scaled Hasse diagram in Example 15

essential to use a small symbol to mark each expectation model, in contrast to Figures 11, 15 and 20. Furthermore, all edges labelled b in Figure 21 are shown with a line of stars to indicate that there is no appropriate residual mean square.

The short edges at the top of the scaled Hasse diagram show that the expectation model can be simplified to Type $\times$ Temperature $+$ Richness. The vertical edge down from this point has P-value less than 0.01; the sloping edge is slightly longer but has P-value 0.017 because of smaller numerator degrees of freedom. Depending on the level of significance used, the fitted model is either Type $\times$ Temperature $+$ Richness or Type $\times$ Temperature. In neither case can we simplify to a model adjoining a starred edge, and so the pseudo-replication of the levels of Temperature does not pose a problem for model choice. However, it does mean that no standard errors of differences between levels of Temperature can be given.

7 Further work

This is a theory-and-methods paper, but bespoke diagrams will not get the ideas widely used. Nicolas Ballarini of the Medizinische Universität Wien is developing a package `HasseAOV` for R to enable users to draw both unscaled and scaled versions of the Hasse diagram. A paper about this is in preparation.

It may be possible to add further visual information to the Hasse diagram. For example, the edges could be colour-coded to indicate the P-values. The scaled version could be drawn so that the vertical height of V is proportional to $\mathrm{SSF}(V)$ to give the reader a sense of the extra sum of squares for one model compared to a smaller one. Is it possible to combine this with making the edge lengths proportional to the relevant mean squares?

The models in this paper are all linear, with normally distributed random variables. The Hasse diagrams in Section 3 do not require either of these conditions. They can be used whenever there is a partial order indicating when one model is a special case of another. It is desirable that conditions analogous to conditions (**L1**) and (**L2**) are satisfied. For example, these diagrams can be used for families of generalized linear models (GLMs) with linear predictors. It should also be possible to extend the scaled Hasse diagrams of Section 5 to show the analysis of deviance for GLMs.

When there is more than one random effect, it may be difficult to extend the methods of Section 5 to cases where conditions (**O1**) and (**O2**) are not satisfied.

References

Bailey, R. A. (2004). Principles of designed experiments in J. A. Nelder's papers. In: N. Adams, M. Crowder, D. J. Hand and D. Stephens (Eds.), *Methods and Models in Statistics* (pp. 171–194). London: Imperial College Press.

Bailey, R. A. (2008). *Design of Comparative Experiments.* Cambridge: Cambridge University Press.

Bailey, R. A. (2012). Contribution to the discussion of 'Optimum design of experiments for statistical inference' by Steven G. Gilmour and Luzia A. Trinca. *Appl. Statist.*, **61**, 374–375.

Bailey, R. A. (2015). Structures defined by factors. Chapter 10 in A. Dean, M. Morris, J. Stufken and D. Bingham (Eds.), *Handbook of Design and Analysis of Experiments* (pp. 371–414). Chapman and Hall/ CRC Handbooks of Modern Statistical Methods. Boca Raton: Chapman and Hall/ CRC.

Bailey, R. A. and Łacka, A. (2015). Nested row–column designs for near-factorial experiments with two treatment factors and one control treatment. *J. Statist. Plan. Inf.*, **165**, 63–77.

Bailey, R. A. and Reiss, J. (2014). Design and analysis of experiments testing for biodiversity effects in ecology. *J. Statist. Plan. Inf.*, **144**, 69–80.

Bowman, A. W. (2019). Graphics for uncertainty. *J. Roy. Statist. Soc.* A, **182**, 403–418.

Brien, C. J. (1989). A model comparison approach to linear models. *Utilitas Mathematica*, **36**, 225–254.

Carpenter, K. J. and Duckworth, J. (1941). Economies in the use of animal by-products in poultry rations. I. Vitamin and amino-acid provision for starting and growing chicks. *J. Agric. Sci.*, **41**, 297–308.

Chipman, H. (1996). Bayesian variable selection with related predictors. *Canad. J. Statist.*, **24**, 17–36.

Christensen, R. (1987). *Plane Answers to Complex Questions: The Theory of Linear Models.* New York: Springer-Verlag.

Clarke, B. R. (2008). *Linear Models: The Theory and Application of Analysis of Variance.* Hoboken: John Wiley & Sons.

Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*, 2nd ed. New York: Wiley.

Dagnelie, P. (1997). La planification des expériences et l'analyse de la variance: une introduction. In: J.-J. Droesbeke, J. Fine and G. Saporta (Eds.), *Plans d'Expériences: Applications à l'Enterprise* (pp. 13–67). Paris: Éditions Technip.

Dean, A. and Lewis, S. (Eds.) (2006). *Screening: Methods for Experimentation in Industry, Drug Discovery and Genetics.* New York: Springer Science.

Dean, A. and Voss, D. (1999). *Design and Analysis of Experiments.* New York: Springer-Verlag.

Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*, 3rd ed. New York: Wiley.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd.

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*, **33**, 503–513.

Fisher, R. A. (1935). *Design of Experiments.* Edinburgh: Oliver and Boyd.

Flores, L., Bailey, R. A., Elosegi, A., Larrañaga, A. and Reiss, J. (2016). Habitat complexity in aquatic microcosms affects processes driven by detrivores. *PLoS ONE*, **11**, (11):e0165065.

Gardiner, W. P. and Gettinby, G. (1998). *Experimental Design Techniques in Statistical Practice.* Chichester: Horwood.

Gerami, A. and Lewis, S. M. (1992). Comparing dual with single treatments in block designs. *Biometrika*, **79**, 603–610.

Gilmour, S. G. and Trinca, L. A. (2012). Optimum design of experiments for statistical inference (with discussion). *Appl. Statist.*, **61**, 345–401.

Grafen, A. and Hails, R. (2002). *Modern Statistics for the Life Sciences.* Oxford: Oxford University Press.

Hamada, M. S. and Wu, C. F. J. (1992). Analysis of designed experiments with complex aliasing. *J. Quality Technology*, **24**, 130–137.

Houtman, A. S. and Speed, T. P. (1983). Balance in designed experiments with orthogonal block structure. *Ann. Statist.*, **11**, 1069–1085.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.*, **54**, 187–211.

Janky, D. G. (2000). Sometimes pooling for analysis of variance hypothesis tests: A review and study of a split-plot model. *Am. Statistn.*, **54**, 269–279.

Kirwan, L., Connolly, J., Finn, J. A., Brophy, C., Lüscher, A., Nyfeler, D. and Sebastia, M.-T. (2009). Diversity–interaction modeling: estimating contributions of species identities and interactions to ecosystem function. *Ecology*, **90**, 2032–2038.

Montgomery, D. C. (2012). *Design and Analysis of Experiments*, 8th ed. New York: John Wiley & Sons.

Mukerjee, R. and Wu, C. F. J. (2006). *A Modern Theory of Factorial Designs*, Springer Series in Statistics. New York: Springer.

Nelder, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc. London* A, **283**, 147–162.

Nelder, J. A. (1977). A reformulation of linear models (with discussion). *J. Roy. Statist. Soc.* A, **140**, 48–77.

Nelder, J. A. (1994). The statistics of linear models: back to basics. *Statistics and Computing.* **4**, 221–234.

Nelder, J. A. (1998a). The great mixed-model muddle is alive and flourishing, alas! *Food Quality and Preference*, **9**, 157–159.

Nelder, J. A. (1998b). The selection of terms in response–surface models—How strong is the weak-heredity principle? *Am. Statistn.*, **52**, 315–318.

Nelder, J. A. (2000). Functional marginality and response–surface fitting. *J. Appl. Statist.*, **27**, 109–112.

Nelder, J. A. and Lane, P. W. (1995). The computer analysis of factorial experiments: In memorian—Frank Yates. *Am. Statistn.*, **49**, 382–385.

Peixoto, J. L. (1987). Hierarchical variable selection in polynomial regression models. *Am. Statistn.*, **41**, 311–313.

Peixoto, J. L. (1990). A property of well-formulated polynomial regression models. *Am. Statistn.*, **44**, 26–30.

Perkins, D. M., Bailey, R. A., Dossena, M., Gamfeldt, L., Reiss, J., Trimmer, M. and Woodward, G. (2015). Higher biodiversity is required to sustain multiple ecosystem processes across temperature regimes. *Global Change Biology*, **21**, 396–406.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, `http://www.R-project.org/`

Reiss, J., Bailey, R. A., Perkins, D. M., Pluchinotta, A. and Woodward, G. (2011). Testing effects of consumer richness, evenness and body size on ecosystem functioning. *J. Animal Ecology*, **80**, 1145–1154.

Saville, D. J. and Wood, G. R. (1996). *Statistical Methods: A Geometric Primer*. New York: Springer-Verlag.

Tjur, T. (1984). Analysis of variance models in orthogonal designs. *Internat. Statist. Rev.*, **52**, 33–81.

VSN International (2017). *Genstat for Windows*, 19th ed. Hemel Hempstead, UK: VSN International.

Wolde-Tsadik, G. and Afifi, A. A. (1980). A comparison of the "sometimes pool", "sometimes switch" and "never pool" procedures in the two-way ANOVA random effects model. *Technometrics*, **22**, 367–373.

Wu, C. F. J. (2015). Post-Fisherian experimentation: from physical to virtual. *J. Am. Statist. Assn.*, **110**, 612–620.

Wu, C. F. J. and Hamada, M. S. (2000). *Experiments: Planning, Analysis and Parameter Design Optimization*. New York: Wiley.

Yates, F. (1935). Complex experiments. *J. Roy. Statist. Soc. Suppl.*, **2**, 181–223.

Yates, F. (1937) *The Design and Analysis of Factorial Experiments*. Technical Communication 35. Harpenden: Imperial Bureau of Soil Science.

Yates, F. (1964). Sir Ronald Fisher and the design of experiments. *Biometrics*, **20**, 307–321.

Yates, F. (1965). A fresh look at the basic principles of the design and analysis of experiments. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4, pp. 777–790. Berkeley, CA: Univ. California Press.