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

R. A. Bailey
University of St Andrews
QMUL (emerita)

LINSTAT 2018, Będlewo, Poland, August 2018

Abstract I

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?).

Abstract I

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?).
Nelder explained this in a 1977 RSS paper.

Abstract I

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?). Nelder explained this in a 1977 RSS paper.

I find it helpful to show the family of models on a Hasse diagram.

Abstract I

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?). Nelder explained this in a 1977 RSS paper.

I find it helpful to show the family of models on a Hasse diagram. I will show examples from experiments in agriculture, industry and ecology.

Abstract I

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?). Nelder explained this in a 1977 RSS paper.

I find it helpful to show the family of models on a Hasse diagram. I will show examples from experiments in agriculture, industry and ecology.

In a more recent development, I show how we can go a stage further. By changing the lengths of the edges in this Hasse diagram, we can use it as a visual display of the analysis of variance.

Abstract II

Some scientists find the original Hasse diagram
easier to understand than a single equation
(or even a collection of equations) for the models,

Abstract II

Some scientists find the original Hasse diagram
easier to understand than a single equation
(or even a collection of equations) for the models,
and they find the scaled Hasse diagram
easier to interpret than the standard analysis-of-variance table.

Linear model for two factors

Given two treatment factors A and B, the linear model for response Y_ω on unit ω is often written as follows.
If $A(\omega) = i$ and $B(\omega) = j$ then

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega,$$

where the ε_ω are random variables with zero means and a covariance matrix whose eigenspaces we know.

Linear model for two factors

Given two treatment factors A and B, the linear model for response Y_ω on unit ω is often written as follows.
If $A(\omega) = i$ and $B(\omega) = j$ then

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega,$$

where the ε_ω are random variables with zero means and a covariance matrix whose eigenspaces we know.

Some authors: “Too many parameters! Let’s impose constraints.”

(a) $\displaystyle\sum_i \alpha_i = 0$, and so on, or

(b) $\displaystyle\sum_i r_i\alpha_i = 0$, where $r_i = |\{\omega : A(\omega) = i\}|$, and so on, or

(c) $\alpha_1 = 0$, or …

Linear model with constraints: some bad consequences

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

(a) $\sum_i \alpha_i = 0$, and so on, or

(b) $\sum_i r_i \alpha_i = 0$, where $r_i = |\{\omega : A(\omega) = i\}|$, and so on, or …

Linear model with constraints: some bad consequences

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

(a) $\sum_i \alpha_i = 0$, and so on, or

(b) $\sum_i r_i \alpha_i = 0$, where $r_i = |\{\omega : A(\omega) = i\}|$, and so on, or …

- It is too easy to give all parameters the same status, and then the conclusions "$\beta_j = 0$ for all j" and "$\gamma_{ij} = 0$ for all i and j" appear to be comparable.

Linear model with constraints: some bad consequences

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

(a) $\sum_i \alpha_i = 0$, and so on, or

(b) $\sum_i r_i\alpha_i = 0$, where $r_i = |\{\omega : A(\omega) = i\}|$, and so on, or …

- It is too easy to give all parameters the same status, and then the conclusions "$\beta_j = 0$ for all j" and "$\gamma_{ij} = 0$ for all i and j" appear to be comparable.
- If some parameters are, after testing, deemed to be zero, the estimated values of the others may not give the vector of fitted values.

  For example, if both main effects and interaction are deemed to be zero, then $\hat{\mu}$ under constraint (a) is not the fitted overall mean if replications are unequal.

Linear model with constraints: some bad consequences

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

(a) $\sum_i \alpha_i = 0$, and so on, or

(b) $\sum_i r_i\alpha_i = 0$, where $r_i = |\{\omega : A(\omega) = i\}|$, and so on, or …

- It is too easy to give all parameters the same status, and then the conclusions "$\beta_j = 0$ for all j" and "$\gamma_{ij} = 0$ for all i and j" appear to be comparable.
- If some parameters are, after testing, deemed to be zero, the estimated values of the others may not give the vector of fitted values.

  For example, if both main effects and interaction are deemed to be zero, then $\hat{\mu}$ under constraint (a) is not the fitted overall mean if replications are unequal.

Linear model with constraints: some bad consequences

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

(a) $\sum_i \alpha_i = 0$, and so on, or

(b) $\sum_i r_i\alpha_i = 0$, where $r_i = |\{\omega : A(\omega) = i\}|$, and so on, or …

- It is too easy to give all parameters the same status, and then the conclusions "$\beta_j = 0$ for all j" and "$\gamma_{ij} = 0$ for all i and j" appear to be comparable.
- If some parameters are, after testing, deemed to be zero, the estimated values of the others may not give the vector of fitted values.
  For example, if both main effects and interaction are deemed to be zero, then $\hat{\mu}$ under constraint (a) is not the fitted overall mean if replications are unequal.

Popular software allows both of these.

Say goodbye to linear models with incremental parameters and constraints

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

(a) $\sum_i \alpha_i = 0$, and so on, or

(b) $\sum_i r_i\alpha_i = 0$, where $r_i = |\{\omega : A(\omega) = i\}|$, and so on, or

(c) $\alpha_1 = 0$, or …

Nelder's approach to such linear models

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

John Nelder had a rant about the constraints on parameters in his 1977 paper 'A reformulation of linear models', *Journal of the Royal Statistical Society, Series A* **140**, 48–76, and various later papers too.

Nelder's approach to such linear models

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

John Nelder had a rant about the constraints on parameters in his 1977 paper 'A reformulation of linear models', *Journal of the Royal Statistical Society, Series A* **140**, 48–76, and various later papers too.

Essentially he said:

- if $\gamma_{ij} = 0$ for all i and j then the model simplifies to

  $$Y_\omega = \mu + \alpha_i + \beta_j + \varepsilon_\omega$$

  so that the expectation of the vector $\mathbf{Y}$ lies in a subspace of dimension at most $n + m - 1$,
  where n and m are the numbers of levels of A and B;

Nelder's approach to such linear models

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

John Nelder had a rant about the constraints on parameters in his 1977 paper 'A reformulation of linear models', *Journal of the Royal Statistical Society, Series A* **140**, 48–76, and various later papers too.

Essentially he said:

- if $\gamma_{ij} = 0$ for all i and j then the model simplifies to

  $$Y_\omega = \mu + \alpha_i + \beta_j + \varepsilon_\omega$$

  so that the expectation of the vector $\mathbf{Y}$ lies in a subspace of dimension at most $n + m - 1$,
  where n and m are the numbers of levels of A and B;
- if $\beta_j = 0$ for all j, but the γ_{ij} are not all zero,
  then the model does not simplify at all.

RAB's approach to such linear models

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

This equation is a short-hand for saying that there are FIVE subspaces which we might suppose to contain the vector $\mathbb{E}(\mathbf{Y})$.

RAB's approach to such linear models

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

This equation is a short-hand for saying that there are FIVE subspaces which we might suppose to contain the vector $\mathbb{E}(\mathbf{Y})$.

Let us parametrize these subspaces separately,
and consider the relationships between them.

RAB's approach to such linear models

$$Y_\omega = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_\omega$$

This equation is a short-hand for saying that there are FIVE subspaces which we might suppose to contain the vector $\mathbb{E}(\mathbf{Y})$.

Let us parametrize these subspaces separately,
and consider the relationships between them.

This is the approach which I always use in teaching
and in consulting, and in my 2008 book.

Expectation subspaces

$$\mathbb{E}(\mathbf{Y}) \in V_A \quad \Longleftrightarrow \quad \begin{array}{l}\text{there are constants } \alpha_i \text{ such that} \\ \mathbb{E}(Y_\omega) = \alpha_i \text{ whenever } A(\omega) = i.\end{array}$$

Expectation subspaces

$$\begin{aligned} \mathbb{E}(\mathbf{Y}) \in V_A &\iff \text{there are constants } \alpha_i \text{ such that} \\ & \qquad \mathbb{E}(Y_\omega) = \alpha_i \text{ whenever } A(\omega) = i. \\ \dim(V_A) &= \text{number of levels of } A = n. \end{aligned}$$

Expectation subspaces

$$\begin{aligned}
\mathbb{E}(\mathbf{Y}) \in V_A \quad &\iff \quad \text{there are constants } \alpha_i \text{ such that} \\
& \quad \mathbb{E}(Y_\omega) = \alpha_i \text{ whenever } A(\omega) = i. \\
\dim(V_A) \quad &= \quad \text{number of levels of } A = n. \\
\mathbb{E}(\mathbf{Y}) \in V_B \quad &\iff \quad \text{there are constants } \beta_j \text{ such that} \\
& \quad \mathbb{E}(Y_\omega) = \beta_j \text{ whenever } B(\omega) = j.
\end{aligned}$$

Expectation subspaces

$$\begin{aligned}
\mathbb{E}(\mathbf{Y}) \in V_A \quad &\iff \quad \text{there are constants } \alpha_i \text{ such that} \\
& \qquad \mathbb{E}(Y_\omega) = \alpha_i \text{ whenever } A(\omega) = i. \\
\dim(V_A) \quad &= \quad \text{number of levels of } A = n.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\mathbf{Y}) \in V_B \quad &\iff \quad \text{there are constants } \beta_j \text{ such that} \\
& \qquad \mathbb{E}(Y_\omega) = \beta_j \text{ whenever } B(\omega) = j.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\mathbf{Y}) \in V_U \quad &\iff \quad \text{there is a constant } \mu \text{ such that} \\
& \qquad \mathbb{E}(Y_\omega) = \mu \text{ for all } \omega.
\end{aligned}$$

Expectation subspaces

$$
\begin{aligned}
\mathbb{E}(\mathbf{Y}) \in V_A &\iff \text{there are constants } \alpha_i \text{ such that} \\
&\qquad \mathbb{E}(Y_\omega) = \alpha_i \text{ whenever } A(\omega) = i. \\
\dim(V_A) &= \text{number of levels of } A = n. \\
\mathbb{E}(\mathbf{Y}) \in V_B &\iff \text{there are constants } \beta_j \text{ such that} \\
&\qquad \mathbb{E}(Y_\omega) = \beta_j \text{ whenever } B(\omega) = j. \\
\mathbb{E}(\mathbf{Y}) \in V_U &\iff \text{there is a constant } \mu \text{ such that} \\
&\qquad \mathbb{E}(Y_\omega) = \mu \text{ for all } \omega. \\
\mathbb{E}(\mathbf{Y}) \in V_A + V_B &\iff \text{there are constants } \theta_i \text{ and } \phi_j \text{ such that} \\
&\qquad \mathbb{E}(Y_\omega) = \theta_i + \phi_j \text{ if } A(\omega) = i \text{ and } B(\omega) = j.
\end{aligned}
$$

Expectation subspaces

$$\mathbb{E}(\mathbf{Y}) \in V_A \iff \text{there are constants } \alpha_i \text{ such that } \mathbb{E}(Y_\omega) = \alpha_i \text{ whenever } A(\omega) = i.$$

$$\dim(V_A) = \text{number of levels of } A = n.$$

$$\mathbb{E}(\mathbf{Y}) \in V_B \iff \text{there are constants } \beta_j \text{ such that } \mathbb{E}(Y_\omega) = \beta_j \text{ whenever } B(\omega) = j.$$

$$\mathbb{E}(\mathbf{Y}) \in V_U \iff \text{there is a constant } \mu \text{ such that } \mathbb{E}(Y_\omega) = \mu \text{ for all } \omega.$$

$$\mathbb{E}(\mathbf{Y}) \in V_A + V_B \iff \text{there are constants } \theta_i \text{ and } \phi_j \text{ such that } \mathbb{E}(Y_\omega) = \theta_i + \phi_j \text{ if } A(\omega) = i \text{ and } B(\omega) = j.$$

$$\mathbb{E}(\mathbf{Y}) \in V_{A \wedge B} \iff \text{there are constants } \gamma_{ij} \text{ such that } \mathbb{E}(Y_\omega) = \gamma_{ij} \text{ if } A(\omega) = i \text{ and } B(\omega) = j.$$

Dimensions when A has n levels and B has m levels

For general factors A and B:

$$\dim(V_A + V_B) = \dim(V_A) + \dim(V_B) - \dim(V_A \cap V_B).$$

Dimensions when A has n levels and B has m levels

For general factors A and B:

$$\dim(V_A + V_B) = \dim(V_A) + \dim(V_B) - \dim(V_A \cap V_B).$$

If all combinations of levels of A and B occur, then

$$V_A \cap V_B = V_U,$$

which has dimension 1,

Dimensions when A has n levels and B has m levels

For general factors A and B:

$$\dim(V_A + V_B) = \dim(V_A) + \dim(V_B) - \dim(V_A \cap V_B).$$

If all combinations of levels of A and B occur, then

$$V_A \cap V_B = V_U,$$

which has dimension 1, so

$$\dim(V_A + V_B) = \dim(V_A) + \dim(V_B) - 1 = n + m - 1,$$

Dimensions when A has n levels and B has m levels

For general factors A and B:

$$\dim(V_A + V_B) = \dim(V_A) + \dim(V_B) - \dim(V_A \cap V_B).$$

If all combinations of levels of A and B occur, then

$$V_A \cap V_B = V_U,$$

which has dimension 1, so

$$\dim(V_A + V_B) = \dim(V_A) + \dim(V_B) - 1 = n + m - 1,$$

and $A \wedge B$ has nm levels so

$$\dim(V_{A \wedge B}) = nm.$$

The partial order on subspaces

If V_1 and V_2 are two subspaces,
write $V_1 < V_2$ to indicate that V_1 is a subspace of V_2
but $V_1 \neq V_2$.

The partial order on subspaces

If V_1 and V_2 are two subspaces,
write $V_1 < V_2$ to indicate that V_1 is a subspace of V_2
but $V_1 \neq V_2$.
Write $V_1 \leq V_2$ to mean that V_1 is a subspace of V_2
(including the possibility that $V_1 = V_2$).

The partial order on subspaces

If V_1 and V_2 are two subspaces,
write $V_1 < V_2$ to indicate that V_1 is a subspace of V_2
but $V_1 \neq V_2$.
Write $V_1 \leq V_2$ to mean that V_1 is a subspace of V_2
(including the possibility that $V_1 = V_2$).

The relation “is a subspace of” is a partial order,
which means that

- $V \leq V$ for all subspaces V;

The partial order on subspaces

If V_1 and V_2 are two subspaces,
write $V_1 < V_2$ to indicate that V_1 is a subspace of V_2
but $V_1 \neq V_2$.
Write $V_1 \leq V_2$ to mean that V_1 is a subspace of V_2
(including the possibility that $V_1 = V_2$).

The relation "is a subspace of" is a partial order,
which means that

- $V \leq V$ for all subspaces V;
- if $V_1 \leq V_2$ and $V_2 \leq V_1$ then $V_1 = V_2$;

The partial order on subspaces

If V_1 and V_2 are two subspaces,
write $V_1 < V_2$ to indicate that V_1 is a subspace of V_2
but $V_1 \neq V_2$.
Write $V_1 \leq V_2$ to mean that V_1 is a subspace of V_2
(including the possibility that $V_1 = V_2$).

The relation "is a subspace of" is a partial order,
which means that

- $V \leq V$ for all subspaces V;
- if $V_1 \leq V_2$ and $V_2 \leq V_1$ then $V_1 = V_2$;
- if $V_1 \leq V_2$ and $V_2 \leq V_3$ then $V_1 \leq V_3$.

Hasse diagram

Every partially ordered set (poset)
can be shown on a Hasse diagram.

Hasse diagram

Every partially ordered set (poset)
can be shown on a Hasse diagram.

Put a symbol for each object (here, a subspace).

Hasse diagram

Every partially ordered set (poset)
can be shown on a Hasse diagram.

Put a symbol for each object (here, a subspace).

If $V_1 < V_2$ then
the symbol for V_1 is lower in the diagram
than the symbol for V_2,
and is joined to it by lines that are traversed upwards.

Hasse diagram

Every partially ordered set (poset)
can be shown on a Hasse diagram.

Put a symbol for each object (here, a subspace).

If $V_1 < V_2$ then
the symbol for V_1 is lower in the diagram
than the symbol for V_2,
and is joined to it by lines that are traversed upwards.

So we can use a Hasse diagram to show the subspaces
which are being considered to model the expectation of $\mathbf{Y}$.

Hasse diagram

Every partially ordered set (poset)
can be shown on a Hasse diagram.

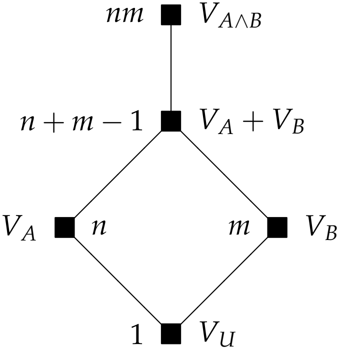
Put a symbol for each object (here, a subspace).

If $V_1 < V_2$ then
the symbol for V_1 is lower in the diagram
than the symbol for V_2,
and is joined to it by lines that are traversed upwards.

So we can use a Hasse diagram to show the subspaces
which are being considered to model the expectation of $\mathbf{Y}$.

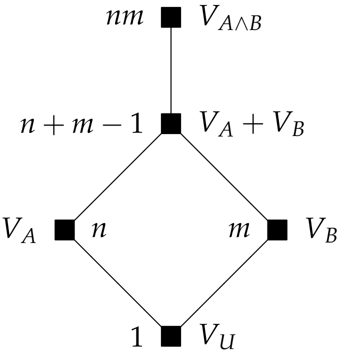
Now it is helpful to show the dimension of each subspace
at the appropriate place in the diagram.

Hasse diagram for model subspaces



If one subspace is contained in another then it is joined to it by an upwards line, or a sequence of such lines.

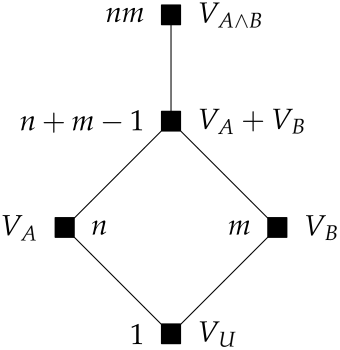
Hasse diagram for model subspaces



null model

If one subspace is contained in another then it is joined to it by an upwards line, or a sequence of such lines.

Hasse diagram for model subspaces

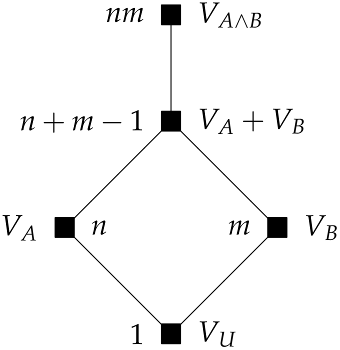


only factor B makes any difference

null model

If one subspace is contained in another then it is joined to it by an upwards line, or a sequence of such lines.

Hasse diagram for model subspaces



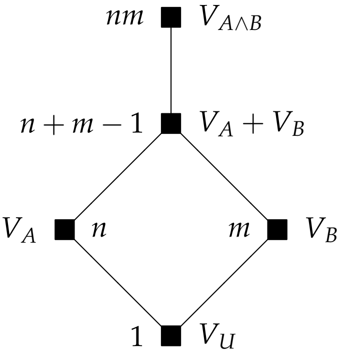
additive model

only factor *B* makes any difference

null model

If one subspace is contained in another then it is joined to it by an upwards line, or a sequence of such lines.

Hasse diagram for model subspaces



full model

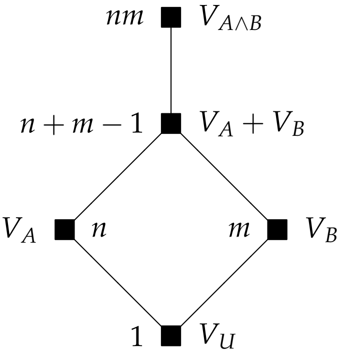
additive model

only factor B makes any difference

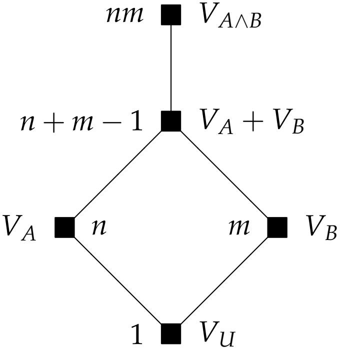
null model

If one subspace is contained in another then it is joined to it by an upwards line, or a sequence of such lines.

Main effects and interaction

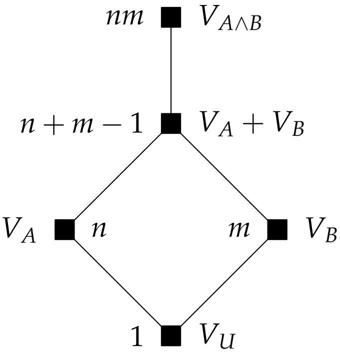


Main effects and interaction



The vector of fitted values in V_U has the grand mean in every coordinate.

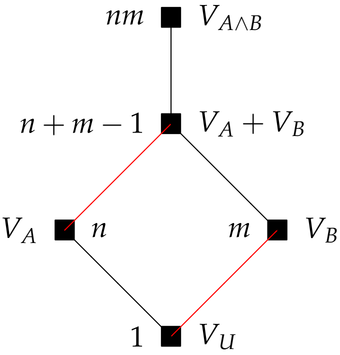
Main effects and interaction



Each coordinate in the vector of fitted values in V_B is the mean on that level of B.

The vector of fitted values in V_U has the grand mean in every coordinate.

Main effects and interaction

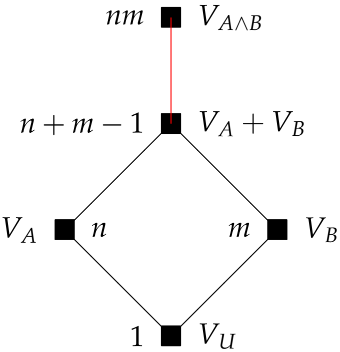


The main effect of factor B is the difference between the vector of fitted values in V_B and the vector of fitted values in V_U.

Each coordinate in the vector of fitted values in V_B is the mean on that level of B.

The vector of fitted values in V_U has the grand mean in every coordinate.

Main effects and interaction



The interaction between factors A and B is the difference between the vector of fitted values in $V_{A\wedge B}$ and the vector of fitted values in $V_A + V_B$.

The main effect of factor B is the difference between the vector of fitted values in V_B and the vector of fitted values in V_U.

Each coordinate in the vector of fitted values in V_B is the mean on that level of B.

The vector of fitted values in V_U has the grand mean in every coordinate.

Different approaches in different software

The command `anova` in GenStat gives, as part of its output,
the list of estimated parameters (with their standard errors)
for each expectation model
that is not a sum of two smaller ones.

Different approaches in different software

The command `anova` in GenStat gives, as part of its output,
the list of estimated parameters (with their standard errors)
for each expectation model
that is not a sum of two smaller ones.

Compare this with the command `aov` in R,
which gives a list of estimated incremental parameters,
subject to some side conditions
over which the user has limited control.

That comparison again

I fit all the models separately,
then look at the differences in SS(fitted values) and in df
between each pair of models joined by an edge.

That comparison again

I fit all the models separately,
then look at the differences in SS(fitted values) and in df between each pair of models joined by an edge.

Most other software chooses a route from the bottom to the top.
At each stage,
it fits incremental parameters for the bigger model,
allowing for what has already been fitted for the smaller model,
so it needs some side constraints on the parameters for the larger model.

Nothing new here

Using a Hasse diagram to show the family of expectation models is not new. For example, see

- C. J. Brien: A model comparison approach to linear models, *Utilitas Mathematica* **36** (1989), 225–254.
- R. A. Bailey: Principles of designed experiments in J. A. Nelder's papers, in *Methods and Models in Statistics* (eds. Niall Adams, Martin Crowder, David J. Hand and David Stephens), Imperial College Press (2004), 171–194.
- R. A. Bailey: *Design of Comparative Experiments*, Cambridge University Press, 2008.
- Brenton R. Clarke: *Linear Models: the Theory and Application of Analysis of Variance*, John Wiley & Sons, 2008.

Nothing new here

Using a Hasse diagram to show the family of expectation models is not new. For example, see

- C. J. Brien: A model comparison approach to linear models, *Utilitas Mathematica* **36** (1989), 225–254.
- R. A. Bailey: Principles of designed experiments in J. A. Nelder's papers, in *Methods and Models in Statistics* (eds. Niall Adams, Martin Crowder, David J. Hand and David Stephens), Imperial College Press (2004), 171–194.
- R. A. Bailey: *Design of Comparative Experiments*, Cambridge University Press, 2008.
- Brenton R. Clarke: *Linear Models: the Theory and Application of Analysis of Variance*, John Wiley & Sons, 2008.

If the family of expectation models contains $V_1 + V_2$ and $V_1 \cap V_2$ whenever it contains both V_1 and V_2 then the partially ordered set is called a lattice.
Brien and Clarke use the word lattice for the Hasse diagram.

Polynomial models

Suppose that we apply quantity x_i of something to unit i, and then measure quantity Y_i of something else on unit i, for $i = 1, \ldots, n$.

Polynomial models

Suppose that we apply quantity x_i of something to unit i,
and then measure quantity Y_i of something else on unit i,
for $i = 1, \ldots, n$.
We want a model that predicts Y as a function of x.

Polynomial models

Suppose that we apply quantity x_i of something to unit i,
and then measure quantity Y_i of something else on unit i,
for $i = 1, \ldots, n$.
We want a model that predicts Y as a function of x.
We might try a cubic polynomial:

$$Y_i = a + bx_i + cx_i^2 + dx_i^3 + \varepsilon_i.$$

Polynomial models

Suppose that we apply quantity x_i of something to unit i, and then measure quantity Y_i of something else on unit i, for $i = 1, \ldots, n$.
We want a model that predicts Y as a function of x.
We might try a cubic polynomial:

$$Y_i = a + bx_i + cx_i^2 + dx_i^3 + \varepsilon_i.$$

A special case of this is the quadratic polynomial:

$$Y_i = e + fx_i + gx_i^2 + \varepsilon_i,$$

Polynomial models

Suppose that we apply quantity x_i of something to unit i, and then measure quantity Y_i of something else on unit i, for $i = 1, \ldots, n$.
We want a model that predicts Y as a function of x.
We might try a cubic polynomial:

$$Y_i = a + bx_i + cx_i^2 + dx_i^3 + \varepsilon_i.$$

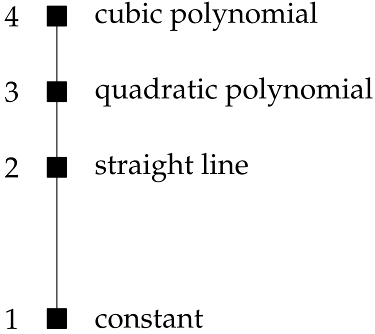
A special case of this is the quadratic polynomial:

$$Y_i = e + fx_i + gx_i^2 + \varepsilon_i,$$

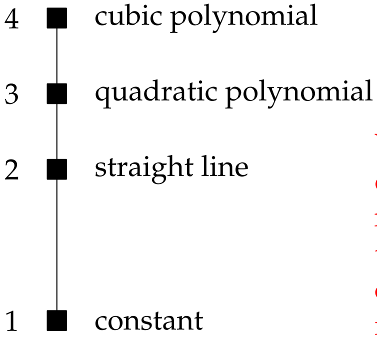
and a special case of that is the straight line:

$$Y_i = h + kx_i + \varepsilon_i.$$

Hasse diagram of polynomial models



Hasse diagram of polynomial models


4 cubic polynomial

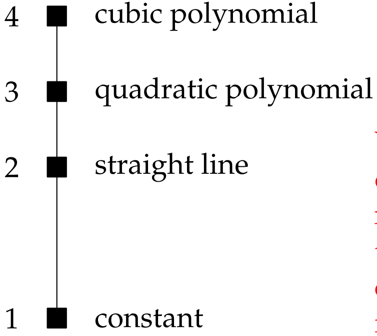
3 quadratic polynomial

2 straight line

1 constant

Warning: the best-fitting quadratic polynomial is not usually obtained by taking the best-fitting cubic polynomial and removing the term in x^3.

Hasse diagram of polynomial models



Warning: the best-fitting quadratic polynomial is not usually obtained by taking the best-fitting cubic polynomial and removing the term in x^3.

Another warning: although there is only one route from the bottom to the top of the diagram, some software allows you to calculate the sum of squares for x after allowing for x^2.

Algorithm for model fitting

1. Start at the top of the Hasse diagram.

Algorithm for model fitting

1. Start at the top of the Hasse diagram.
2. At point V, assume that $\mathbb{E}(\mathbf{Y})$ is in, or close to, V. Then

Algorithm for model fitting

1. Start at the top 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.

Algorithm for model fitting

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

Algorithm for model fitting

1. Start at the top 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.
    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 the residual mean square in the appropriate stratum.

Algorithm for model fitting

1. Start at the top of the Hasse diagram.
2. At point V, assume that $\mathbb{E}(\mathbf{Y})$ is in, or close to, V. Then
    1. Choose a unused downwards edge.
    2. Suppose that the point at the bottom of the edge is W.
    3. Perform a test of the hypothesis that $P_V(\mathbb{E}(\mathbf{Y})) - P_W(\mathbb{E}(\mathbf{Y})) = 0$, using the residual mean square in the appropriate stratum.
    4. If the hypothesis is not rejected then

Algorithm for model fitting

1. Start at the top 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.
   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 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;

Algorithm for model fitting

1. Start at the top 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.
   - 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 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;

Algorithm for model fitting

1. Start at the top 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.
    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 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.

Algorithm for model fitting

1. Start at the top 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.
    - 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 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.

Algorithm for model fitting

1. Start at the top 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.
    - 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 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

Algorithm for model fitting

1. Start at the top 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.
    - 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 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;

Algorithm for model fitting

1. Start at the top 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.
    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 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;

Algorithm for model fitting

1. Start at the top 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.
   - 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 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;

Algorithm for model fitting

1. Start at the top 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.
   - 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 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;

Algorithm for model fitting

1. Start at the top 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.
    - 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 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.

What about orthogonality?

Vector subspaces V_1 and V_2 are geometrically orthogonal to each other if

$$V_1 \cap (V_1 \cap V_2)^\perp \quad \text{is orthogonal to} \quad V_2 \cap (V_1 \cap V_2)^\perp .$$

What about orthogonality?

Vector subspaces V_1 and V_2 are geometrically orthogonal to each other if

$$V_1 \cap (V_1 \cap V_2)^\perp \quad \text{is orthogonal to} \quad V_2 \cap (V_1 \cap V_2)^\perp .$$

If every pair of subspaces in our model collection is geometrically orthogonal,
and the model collection is a lattice (it contains $V_1 \cap V_2$ and $V_1 + V_2$ whenever it contains V_1 and V_2),
then all routes from the top of the Hasse diagram to the bottom will give the same result.

What about orthogonality?

Vector subspaces V_1 and V_2 are geometrically orthogonal to each other if

$$V_1 \cap (V_1 \cap V_2)^\perp \quad \text{is orthogonal to} \quad V_2 \cap (V_1 \cap V_2)^\perp .$$

If every pair of subspaces in our model collection is geometrically orthogonal,
and the model collection is a lattice (it contains $V_1 \cap V_2$ and $V_1 + V_2$ whenever it contains V_1 and V_2),
then all routes from the top of the Hasse diagram to the bottom will give the same result.

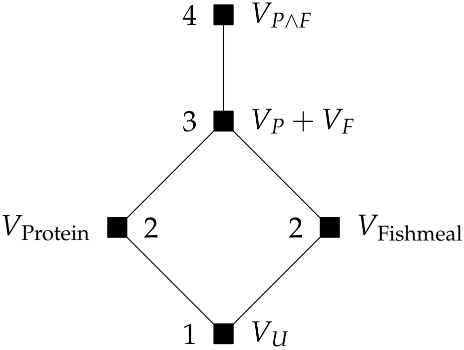
Otherwise, different routes can give different conclusions for some data vectors.

Example with two treatment factors: feeding chickens

Four diets for feeding newly-hatched chickens were compared. The diets consisted of all levels of Protein (groundnuts or soya bean) with two levels of Fishmeal (added or not). Each diet was fed to two chickens, and they were weighed at the end of six weeks.

Example with two treatment factors: feeding chickens

Four diets for feeding newly-hatched chickens were compared. The diets consisted of all levels of Protein (groundnuts or soya bean) with two levels of Fishmeal (added or not). Each diet was fed to two chickens, and they were weighed at the end of six weeks.



Chicken example: anova

(Subset of data from Carpenter and Duckworth, 1941)

Source	SS	df	MS	VR
Protein	4704.5	1	4704.50	35.57
Fishmeal	3120.5	1	3120.50	23.60
Protein $\wedge$ Fishmeal	128.0	1	128.00	0.97
residual	529.0	4	132.25	

Chicken example: anova

(Subset of data from Carpenter and Duckworth, 1941)

Source	SS	df	MS	VR
Protein	4704.5	1	4704.50	35.57
Fishmeal	3120.5	1	3120.50	23.60
Protein $\wedge$ Fishmeal	128.0	1	128.00	0.97
residual	529.0	4	132.25	

You know how to interpret the anova table:
do the scientists who did the experiment know how to?

Scaling the Hasse diagram of expectation subspaces

Suppose that V_1 and V_2 are expectation subspaces,
with $V_1 < V_2$,
and an edge joining V_1 to V_2.

Scaling the Hasse diagram of expectation subspaces

Suppose that V_1 and V_2 are expectation subspaces,
with $V_1 < V_2$,
and an edge joining V_1 to V_2.

The mean square for
the extra fit in V_2 compared to the fit in V_1 is

$$\frac{\text{SS(fitted values in } V_2\text{)} \;-\; \text{SS(fitted values in } V_1\text{)}}{\dim(V_2) - \dim(V_1)}.$$

Scaling the Hasse diagram of expectation subspaces

Suppose that V_1 and V_2 are expectation subspaces,
with $V_1 < V_2$,
and an edge joining V_1 to V_2.

The mean square for
the extra fit in V_2 compared to the fit in V_1 is

$$\frac{\text{SS(fitted values in } V_2\text{)} \;-\; \text{SS(fitted values in } V_1\text{)}}{\dim(V_2) - \dim(V_1)}.$$

Scale the Hasse diagram so that each edge has length proportional to the relevant mean square,
and show the residual mean square to give a scale.

Scaling the Hasse diagram of expectation subspaces

Suppose that V_1 and V_2 are expectation subspaces,
with $V_1 < V_2$,
and an edge joining V_1 to V_2.

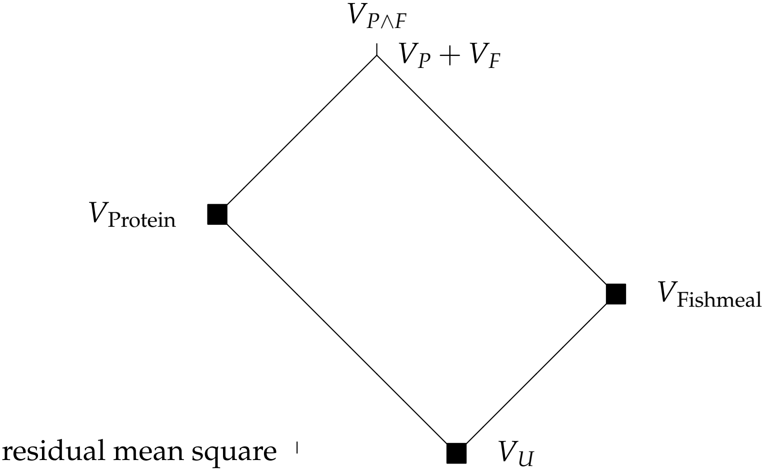
The mean square for
the extra fit in V_2 compared to the fit in V_1 is

$$\frac{\text{SS(fitted values in } V_2\text{)} \; - \; \text{SS(fitted values in } V_1\text{)}}{\dim(V_2) - \dim(V_1)}.$$

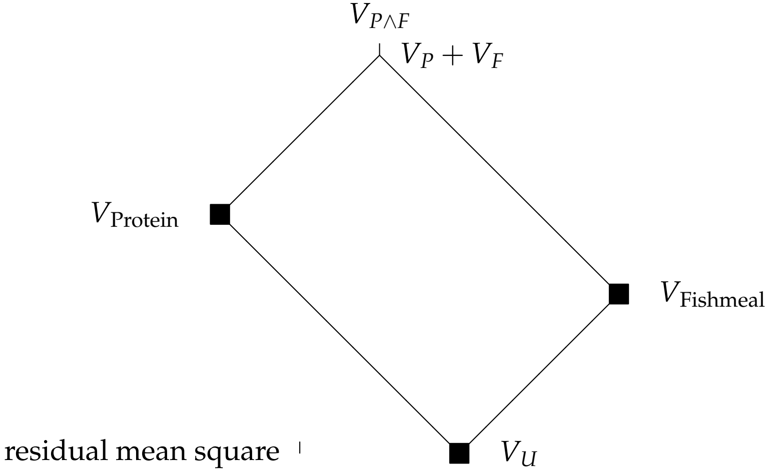
Scale the Hasse diagram so that each edge has length
proportional to the relevant mean square,
and show the residual mean square to give a scale.

This really is a new idea, which I thought of about six years ago.

Chickens: scaled Hasse diagram of expectation subspaces

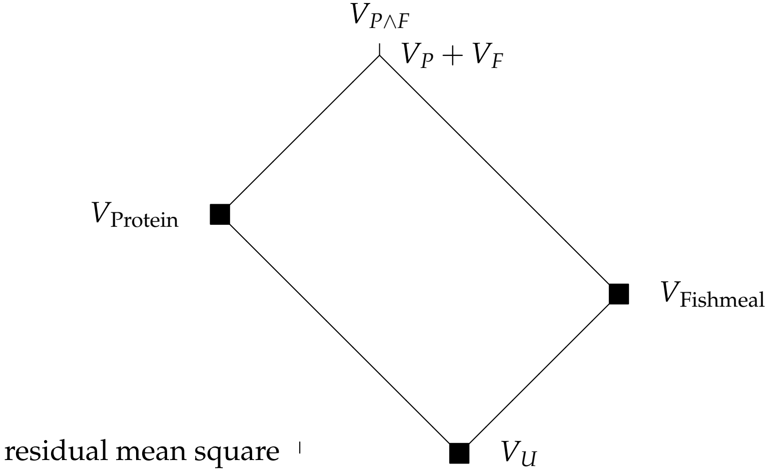


Chickens: scaled Hasse diagram of expectation subspaces



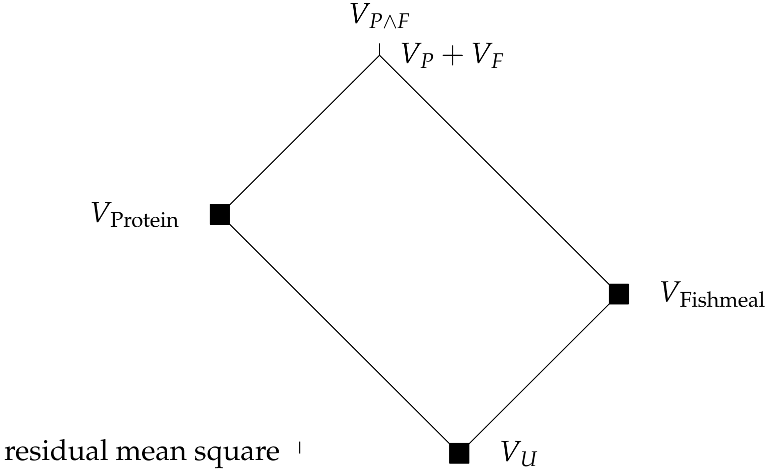
There is no evidence of any interaction,
so we can simplify to the additive model

Chickens: scaled Hasse diagram of expectation subspaces



There is no evidence of any interaction,
so we can simplify to the additive model
(but we don't change the residual mean square).

Chickens: scaled Hasse diagram of expectation subspaces



There is no evidence of any interaction,
so we can simplify to the additive model
(but we don't change the residual mean square).
Neither main effect is zero, so we cannot simplify further.

Example: an experiment about protecting metal

(Data from Crowder and Kimber, 1997)

An experiment was conducted to compare two protective dyes for metal, both with each other and with no dye. Ten braided metal cords were broken into three pieces. The three pieces of each cord were randomly allocated to the three treatments. After the dyes had been applied, the cords were left to weather for a fixed time, then their strengths were measured, and recorded as a percentage of the nominal strength specification.

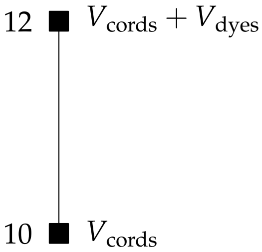
Example: an experiment about protecting metal

(Data from Crowder and Kimber, 1997)

An experiment was conducted to compare two protective dyes for metal, both with each other and with no dye. Ten braided metal cords were broken into three pieces. The three pieces of each cord were randomly allocated to the three treatments. After the dyes had been applied, the cords were left to weather for a fixed time, then their strengths were measured, and recorded as a percentage of the nominal strength specification.

Factors: Dye, with three levels (no dye, dye A, Dye B);
Cords, with ten levels;
U, with one level; E, with 30 levels.

Cords: Hasse diagram of expectation subspaces

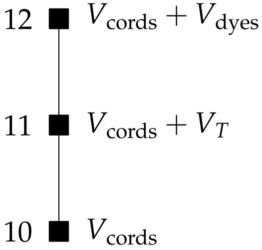


12 $V_{\text{cords}} + V_{\text{dyes}}$

10 V_{cords}

We assume that there are differences between cords, so all the models that we consider include V_{cords}.

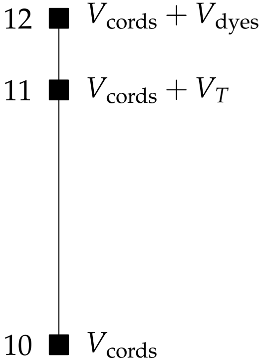
Cords: Hasse diagram of expectation subspaces



We assume that there are differences between cords,
so all the models that we consider include V_{cords}.

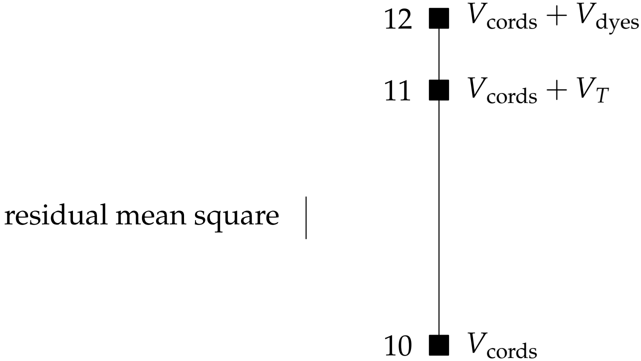
There is another factor T (To-dye-or-not-to-dye).
It has one level on ‘no dye’ and another level on both real dyes.

Cords: Scaled Hasse diagram of expectation subspaces



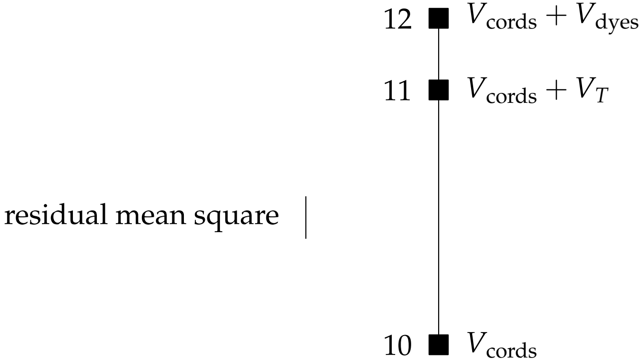
residual mean square

Cords: Scaled Hasse diagram of expectation subspaces



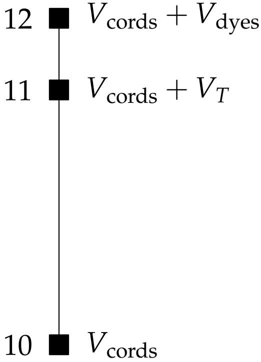
There is no evidence of a difference between dye A and dye B;

Cords: Scaled Hasse diagram of expectation subspaces



There is no evidence of a difference between dye A and dye B; so we can simplify to the model $V_{\text{cords}} + V_T$.

Cords: Scaled Hasse diagram of expectation subspaces



residual mean square

There is no evidence of a difference between dye A and dye B; so we can simplify to the model $V_{\text{cords}} + V_T$.
Now there is definitely a difference between no dye and real dyes.

An experiment with a quantitative factor

(Data from Yates, 1937)

An experiment on forage crops compared five seed mixtures in the presence and absence of nitrogen fertilizer. All ten combinations were grown in plots in five different fields.

For each crop mixture in each field, the recorded response is improvement in yield, in tons per acre, if fertilizer is added.

An experiment with a quantitative factor

(Data from Yates, 1937)
An experiment on forage crops compared five seed mixtures in the presence and absence of nitrogen fertilizer. All ten combinations were grown in plots in five different fields. For each crop mixture in each field, the recorded response is improvement in yield, in tons per acre, if fertilizer is added.

Fields are like cords: we assume that there are differences between them but we do not care about their differences.

An experiment with a quantitative factor

(Data from Yates, 1937)
An experiment on forage crops compared five seed mixtures in the presence and absence of nitrogen fertilizer. All ten combinations were grown in plots in five different fields.
For each crop mixture in each field, the recorded response is improvement in yield, in tons per acre, if fertilizer is added.

Fields are like cords: we assume that there are differences between them but we do not care about their differences.

Crop mixtures are like diets or dyes: we are interested in their differences.

An experiment with a quantitative factor

(Data from Yates, 1937)
An experiment on forage crops compared five seed mixtures in the presence and absence of nitrogen fertilizer. All ten combinations were grown in plots in five different fields.
For each crop mixture in each field, the recorded response is improvement in yield, in tons per acre, if fertilizer is added.

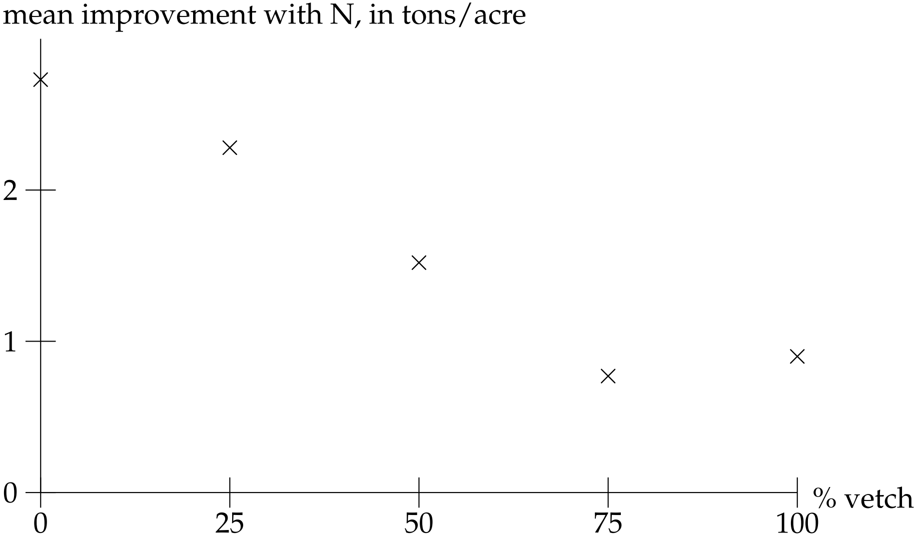
Fields are like cords: we assume that there are differences between them but we do not care about their differences.

Crop mixtures are like diets or dyes: we are interested in their differences.

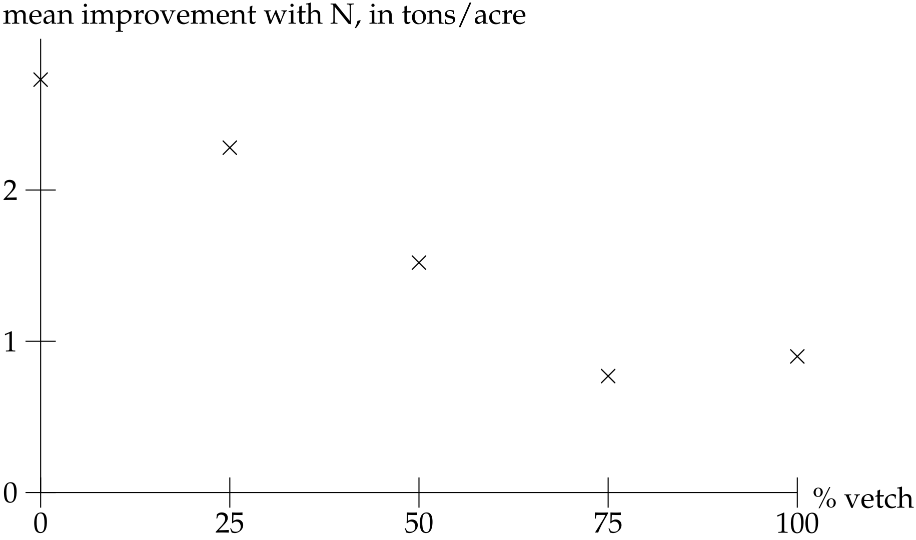
Crop mixtures are not like diets and dyes, because the levels are quantitative:

100% oats	75% oats	50% oats	25% oats	0% oats
0% vetch	25% vetch	50% vetch	75% vetch	100% vetch

An intermediate model: linear in vetch

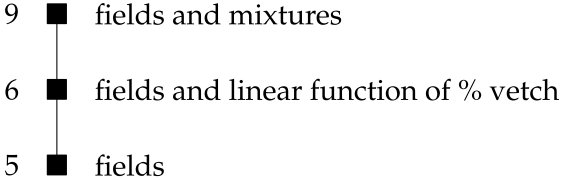


An intermediate model: linear in vetch



Intermediate model:
improvement = field parameter + linear(% vetch)

Vetch: Hasse diagram of models

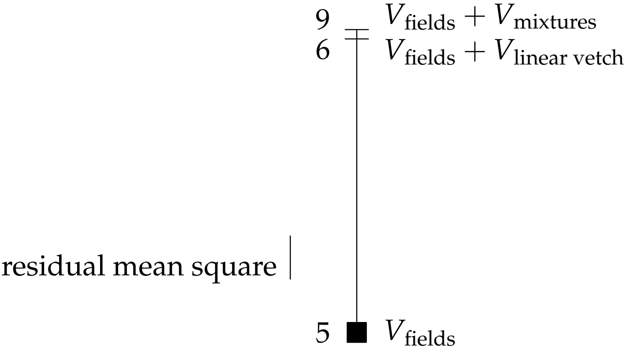


9 fields and mixtures

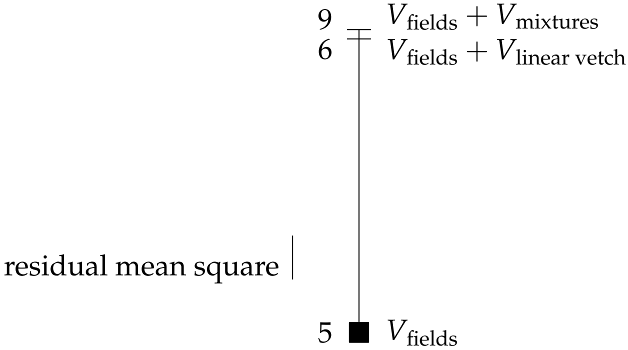
6 fields and linear function of % vetch

5 fields

Vetch: scaled Hasse diagram of models

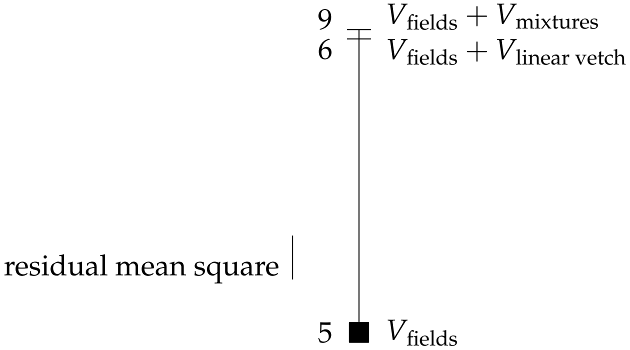


Vetch: scaled Hasse diagram of models



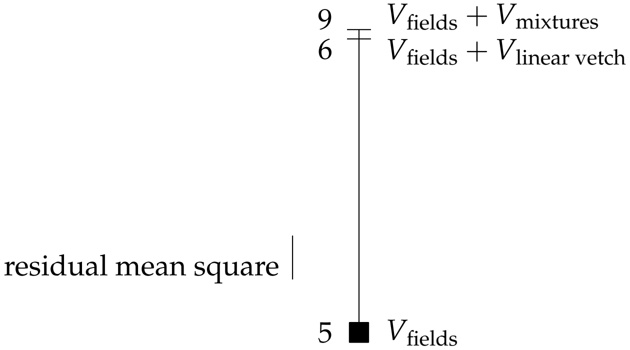
There is no evidence of any difference between mixtures other than that due to a linear trend in the proportion of vetch;

Vetch: scaled Hasse diagram of models



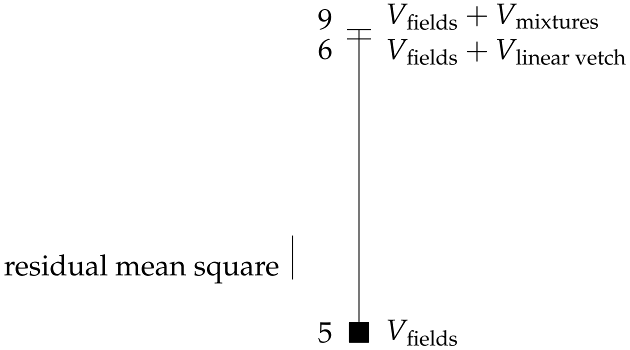
There is no evidence of any difference between mixtures other than that due to a linear trend in the proportion of vetch; so we can simplify to the model $V_{\text{fields}} + V_{\text{linear vetch}}$.

Vetch: scaled Hasse diagram of models



There is no evidence of any difference between mixtures other than that due to a linear trend in the proportion of vetch; so we can simplify to the model $V_{\text{fields}} + V_{\text{linear vetch}}$.
Now there is definitely a linear trend in the proportion of vetch.

Vetch: scaled Hasse diagram of models



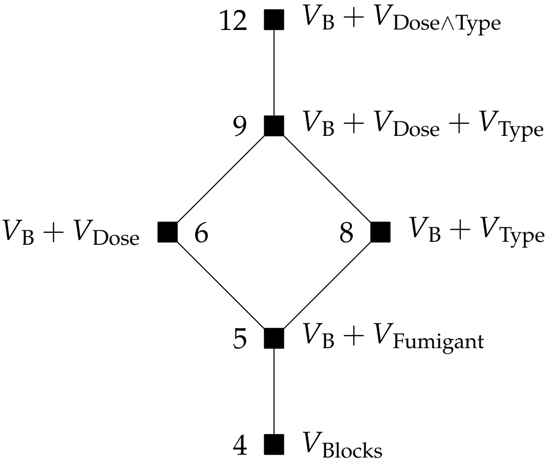
There is no evidence of any difference between mixtures other than that due to a linear trend in the proportion of vetch; so we can simplify to the model $V_{\text{fields}} + V_{\text{linear vetch}}$.
Now there is definitely a linear trend in the proportion of vetch.
(Yates does not seem to have noticed this linear component of the interaction between N and the proportion of vetch.)

Factorial treatments plus control, in 4 blocks

Type	Dose	Fumigant
0	0	None
N	1	Some
N	2	Some
M	1	Some
M	2	Some
S	1	Some
S	2	Some
K	1	Some
K	2	Some

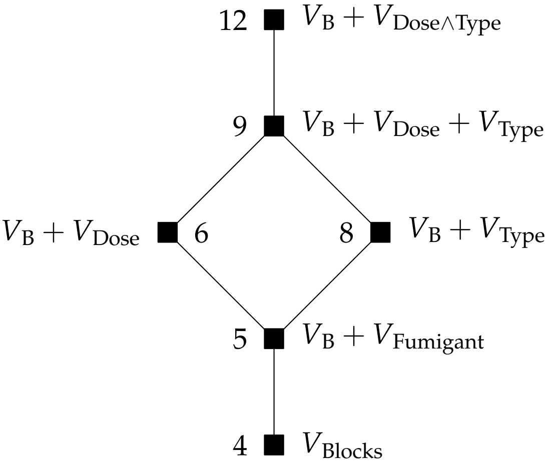
Factorial treatments plus control, in 4 blocks

Type	Dose	Fumigant
0	0	None
N	1	Some
N	2	Some
M	1	Some
M	2	Some
S	1	Some
S	2	Some
K	1	Some
K	2	Some



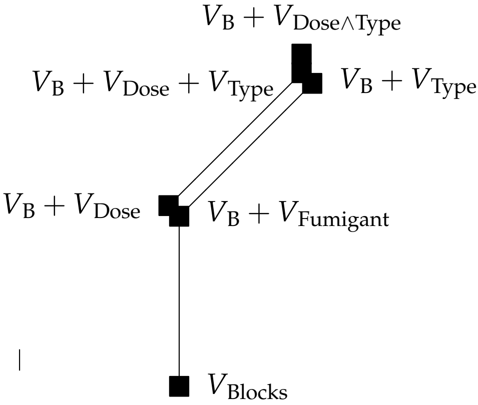
Factorial treatments plus control, in 4 blocks

Type	Dose	Fumigant
0	0	None
N	1	Some
N	2	Some
M	1	Some
M	2	Some
S	1	Some
S	2	Some
K	1	Some
K	2	Some



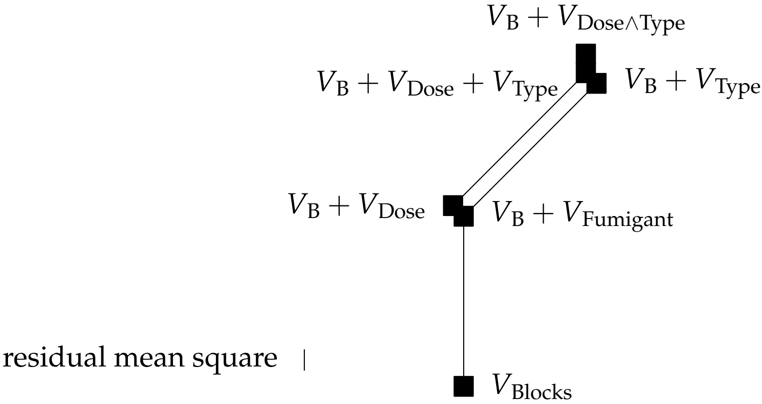
What happens if I forget to include the factor Fumigant?

Fumigation: scaled Hasse diagram



residual mean square |

Fumigation: scaled Hasse diagram



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

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

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

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

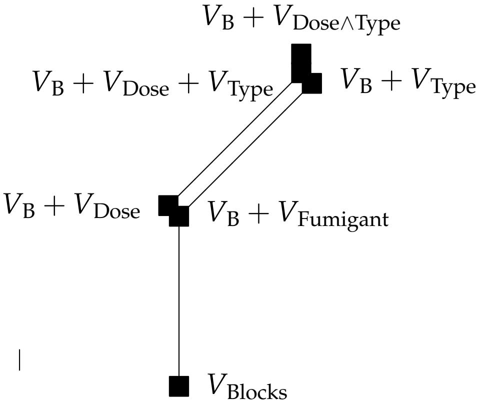
$V_{\text{B}} + V_{\text{Fumigant}}$

residual mean square |

V_{Blocks}

The differences between the five types explain all the differences between treatments.

Fumigation: scaled Hasse diagram



residual mean square |

The differences between the five types explain all the differences between treatments.

How can we automate the production of such diagrams so that (i) names do not overlap, (ii) boxes are omitted if necessary, and (iii) a visible multiple of the residual mean square is shown?

More than one residual mean square

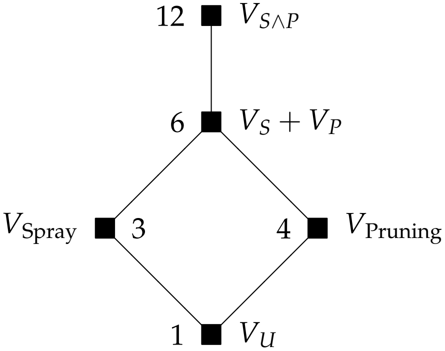
This approach can be extended to situations like split-plot designs with more than one relevant residual mean square by using different colours (or types of line) for the corresponding edges.

Split-plot example: cider apples

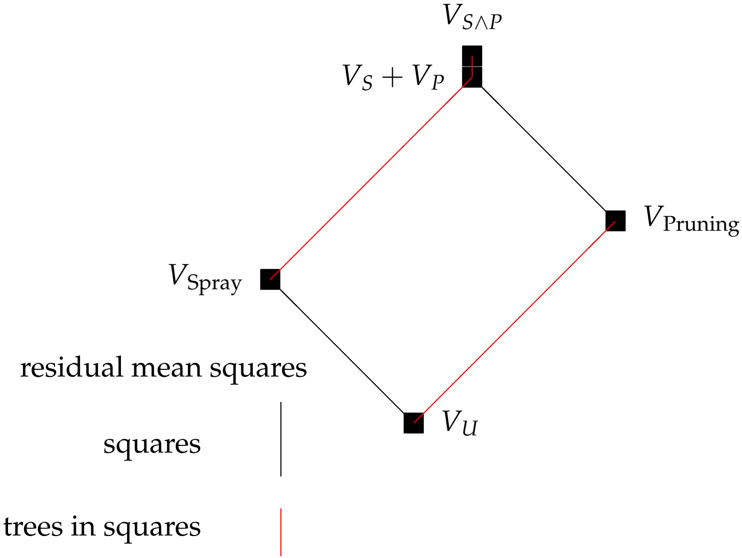
24 cider-apple trees were grouped into six squares of four trees each. Three spray treatments were applied, each to two whole squares. Four pruning methods were applied, one on each tree per square. The experimenters recorded the percentage of apples from each tree that fell after a heavy gale in October.

Split-plot example: cider apples

24 cider-apple trees were grouped into six squares of four trees each. Three spray treatments were applied, each to two whole squares. Four pruning methods were applied, one on each tree per square. The experimenters recorded the percentage of apples from each tree that fell after a heavy gale in October.



Cider apples: scaled Hasse diagram



residual mean squares

squares

trees in squares

Non-orthogonality

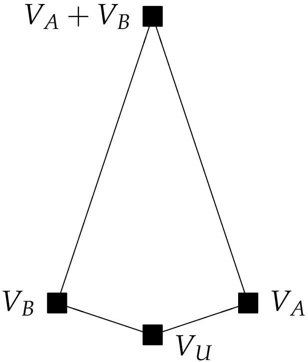
These diagrams can be extended
to deal with non-orthogonal models.

Now edges that used to have the same length
can have different lengths.

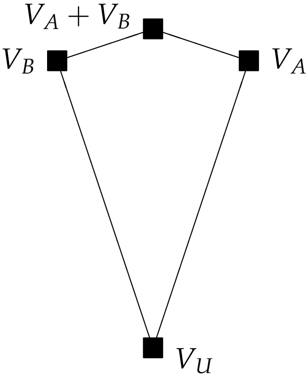
The diagram clearly shows when
fitting models in a different order can give different results.

Small non-orthogonal example

Factors A and B both have two levels. For practical or ethical reasons, one combination of levels cannot be used.
Both of the following are possible outcomes.



Neither A nor B explains much of the data, but they do together.



Either factor explains the data, with no need for the other.

A (comparatively simple) biodiversity experiment

A, B, C, D, E, F—size types of freshwater "shrimp".
Put 12 shrimps in a jar with stream water and alder leaf litter.
Measure how much leaf litter is eaten after 28 days.

Experimental unit = jar.

A (comparatively simple) biodiversity experiment

A, B, C, D, E, F—size types of freshwater "shrimp".
Put 12 shrimps in a jar with stream water and alder leaf litter.
Measure how much leaf litter is eaten after 28 days.

Experimental unit = jar. R = Richness

T		Treatment	R	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$
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

Expectation subspaces for biodiversity experiment

$$\mathbb{E}(\mathbf{Y}) \in V_U \quad \Longleftrightarrow \quad \begin{array}{l}\text{there is a constant } \mu \text{ such that} \\ \mathbb{E}(Y_\omega) = \mu \text{ for all } \omega.\end{array}$$

Expectation subspaces for biodiversity experiment

$$\mathbb{E}(\mathbf{Y}) \in V_U \quad \iff \quad \text{there is a constant } \mu \text{ such that } \mathbb{E}(Y_\omega) = \mu \text{ for all } \omega.$$

$$\mathbb{E}(\mathbf{Y}) \in V_R \quad \iff \quad \text{there are constants } \alpha_j \text{ such that } \mathbb{E}(Y_\omega) = \alpha_j \text{ whenever } R(\omega) = j.$$

Expectation subspaces for biodiversity experiment

$$
\begin{aligned}
\mathbb{E}(\mathbf{Y}) \in V_U \quad &\iff \quad \text{there is a constant } \mu \text{ such that} \\
&\qquad\ \ \mathbb{E}(Y_\omega) = \mu \text{ for all } \omega. \\
\mathbb{E}(\mathbf{Y}) \in V_R \quad &\iff \quad \text{there are constants } \alpha_j \text{ such that} \\
&\qquad\ \ \mathbb{E}(Y_\omega) = \alpha_j \text{ whenever } R(\omega) = j. \\
\mathbb{E}(\mathbf{Y}) \in V_S \quad &\iff \quad \text{there are constants } \beta_i \text{ such that} \\
&\qquad\ \ \mathbb{E}(Y_\omega) = \sum_{i=1}^{6} \beta_i x_i(\omega).
\end{aligned}
$$

Expectation subspaces for biodiversity experiment

$$\mathbb{E}(\mathbf{Y}) \in V_U \iff \text{there is a constant } \mu \text{ such that } \mathbb{E}(Y_\omega) = \mu \text{ for all } \omega.$$

$$\mathbb{E}(\mathbf{Y}) \in V_R \iff \text{there are constants } \alpha_j \text{ such that } \mathbb{E}(Y_\omega) = \alpha_j \text{ whenever } R(\omega) = j.$$

$$\mathbb{E}(\mathbf{Y}) \in V_S \iff \text{there are constants } \beta_i \text{ such that } \mathbb{E}(Y_\omega) = \sum_{i=1}^{6} \beta_i x_i(\omega).$$

$$\mathbb{E}(\mathbf{Y}) \in V_R + V_S \iff \mathbb{E}(Y_\omega) = \alpha_{R(\omega)} + \sum_{i=1}^{6} \beta_i x_i(\omega).$$

Expectation subspaces for biodiversity experiment

$$\mathbb{E}(\mathbf{Y}) \in V_U \iff \text{there is a constant } \mu \text{ such that } \mathbb{E}(Y_\omega) = \mu \text{ for all } \omega.$$

$$\mathbb{E}(\mathbf{Y}) \in V_R \iff \text{there are constants } \alpha_j \text{ such that } \mathbb{E}(Y_\omega) = \alpha_j \text{ whenever } R(\omega) = j.$$

$$\mathbb{E}(\mathbf{Y}) \in V_S \iff \text{there are constants } \beta_i \text{ such that } \mathbb{E}(Y_\omega) = \sum_{i=1}^{6} \beta_i x_i(\omega).$$

$$\mathbb{E}(\mathbf{Y}) \in V_R + V_S \iff \mathbb{E}(Y_\omega) = \alpha_{R(\omega)} + \sum_{i=1}^{6} \beta_i x_i(\omega).$$

$$\mathbb{E}(\mathbf{Y}) \in V_{R*S} \iff \text{there are constants } \gamma_{ij} \text{ such that } \mathbb{E}(Y_\omega) = \sum_{i=1}^{6} \gamma_{i,R(\omega)} x_i(\omega).$$

Expectation subspaces for biodiversity experiment

$$\mathbb{E}(\mathbf{Y}) \in V_U \iff \text{there is a constant } \mu \text{ such that } \mathbb{E}(Y_\omega) = \mu \text{ for all } \omega.$$

$$\mathbb{E}(\mathbf{Y}) \in V_R \iff \text{there are constants } \alpha_j \text{ such that } \mathbb{E}(Y_\omega) = \alpha_j \text{ whenever } R(\omega) = j.$$

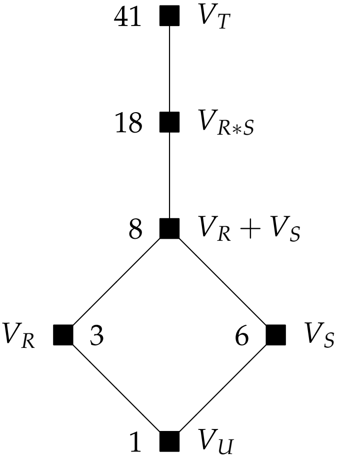
$$\mathbb{E}(\mathbf{Y}) \in V_S \iff \text{there are constants } \beta_i \text{ such that } \mathbb{E}(Y_\omega) = \sum_{i=1}^{6} \beta_i x_i(\omega).$$

$$\mathbb{E}(\mathbf{Y}) \in V_R + V_S \iff \mathbb{E}(Y_\omega) = \alpha_{R(\omega)} + \sum_{i=1}^{6} \beta_i x_i(\omega).$$

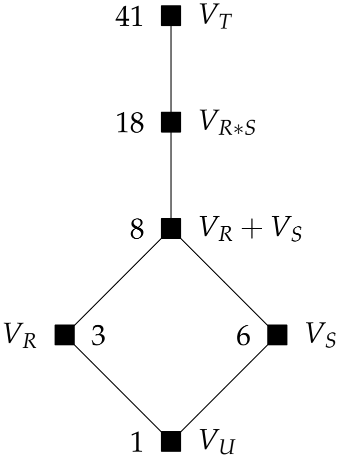
$$\mathbb{E}(\mathbf{Y}) \in V_{R*S} \iff \text{there are constants } \gamma_{ij} \text{ such that } \mathbb{E}(Y_\omega) = \sum_{i=1}^{6} \gamma_{i,R(\omega)} x_i(\omega).$$

$$\mathbb{E}(\mathbf{Y}) \in V_T \iff \mathbb{E}(Y_\omega) = \delta_{T(\omega)}.$$

Hasse diagram for model subspaces (biodiversity)

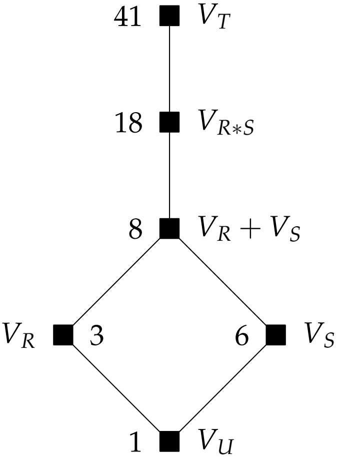


Hasse diagram for model subspaces (biodiversity)



null model

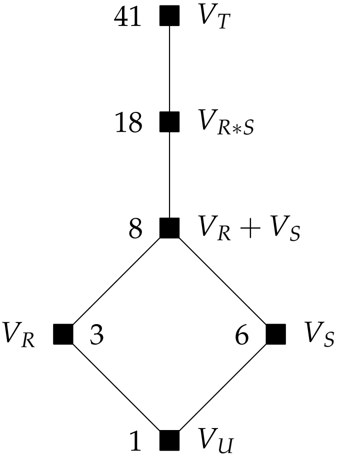
Hasse diagram for model subspaces (biodiversity)



only richness makes any difference
each shrimp has its own rate

null model

Hasse diagram for model subspaces (biodiversity)



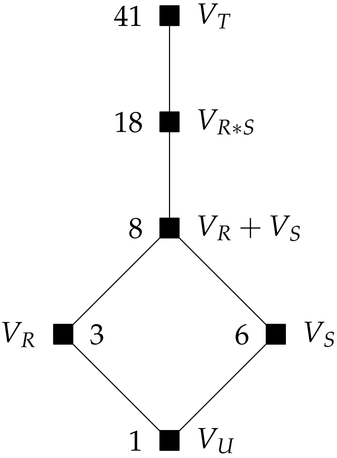
additive model

only richness makes any difference

each shrimp has its own rate

null model

Hasse diagram for model subspaces (biodiversity)



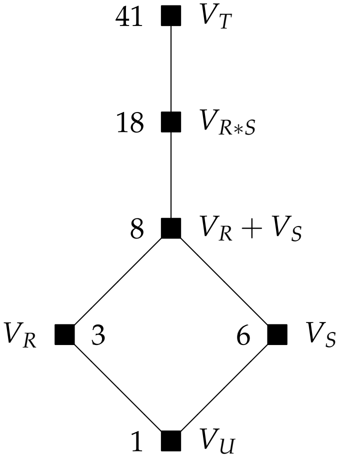
each shrimp rate depends on richness

additive model

only richness makes any difference
each shrimp has its own rate

null model

Hasse diagram for model subspaces (biodiversity)



full model

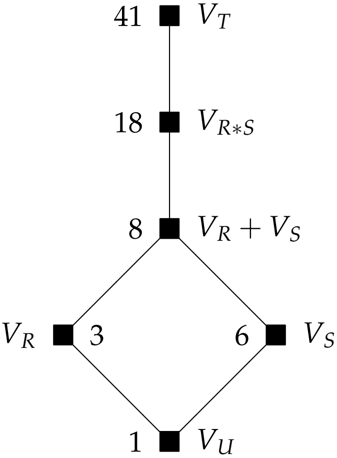
each shrimp rate depends on richness

additive model

only richness makes any difference
each shrimp has its own rate

null model

Hasse diagram for model subspaces (biodiversity)



full model

each shrimp rate depends on richness

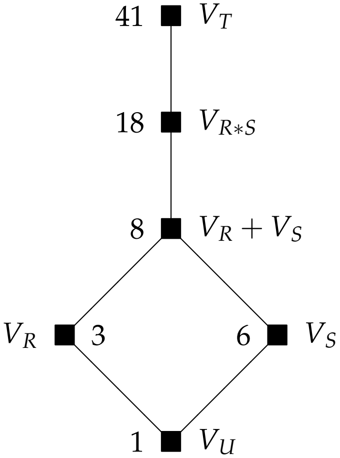
additive model

only richness makes any difference
each shrimp has its own rate

null model

$$R + (x1 + \cdots + x6) + R\!:\!(x1 + \cdots + x6) + T$$

Hasse diagram for model subspaces (biodiversity)



full model

each shrimp rate depends on richness

additive model

only richness makes any difference

each shrimp has its own rate

null model

$R + (x1 + \cdots + x6) + R\!:\!(x1 + \cdots + x6) + T$

Does your software interpret this correctly?

(Thanks to Justin Thong.)

Success!

An ecology journal published

- the Hasse diagram of the family of models
- the statement that each row of an ANOVA table is for a difference between models.

Analysis of Variance (ANOVA) table

Source	df	SS	MS	F	P
Richness	2	0.000009	0.000005	0.49	n.s.
Shrimp	5	0.003859	0.000772	81.37	< 0.0005
Richness $*$ Shrimp	10	0.000127	0.000013	1.34	n.s.
Treatment	23	0.000105	0.000005	0.48	n.s.
Block	3	0.000067	0.000022		
Error	120	0.001138	0.000009		
Total	163	0.005306			

Analysis of Variance (ANOVA) table

Source	df	SS	MS	F	P
Richness	2	0.000009	0.000005	0.49	n.s.
Shrimp	5	0.003859	0.000772	81.37	< 0.0005
Richness $*$ Shrimp	10	0.000127	0.000013	1.34	n.s.
Treatment	23	0.000105	0.000005	0.48	n.s.
Block	3	0.000067	0.000022		
Error	120	0.001138	0.000009		
Total	163	0.005306			

Each row in the ANOVA table represents not a model but the difference between a larger model and the next smaller one. See Fig. 1 for how the models are related.

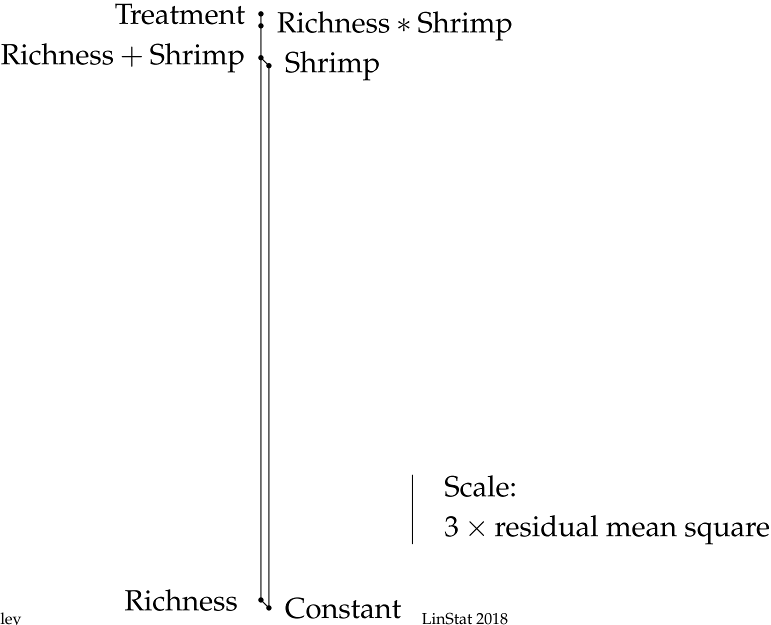
Analysis of Variance (ANOVA) table

Source	df	SS	MS	F	P
Richness	2	0.000009	0.000005	0.49	n.s.
Shrimp	5	0.003859	0.000772	81.37	< 0.0005
Richness $*$ Shrimp	10	0.000127	0.000013	1.34	n.s.
Treatment	23	0.000105	0.000005	0.48	n.s.
Block	3	0.000067	0.000022		
Error	120	0.001138	0.000009		
Total	163	0.005306			

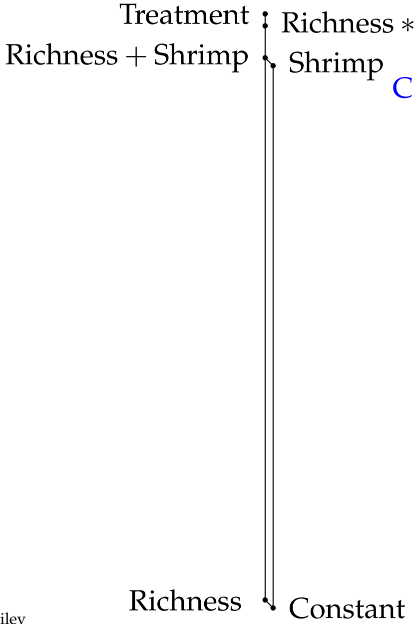
Each row in the ANOVA table represents not a model but the difference between a larger model and the next smaller one. See Fig. 1 for how the models are related.

Verbatim from *Journal of Animal Ecology*

What the data showed: mean squares



What the data showed: mean squares

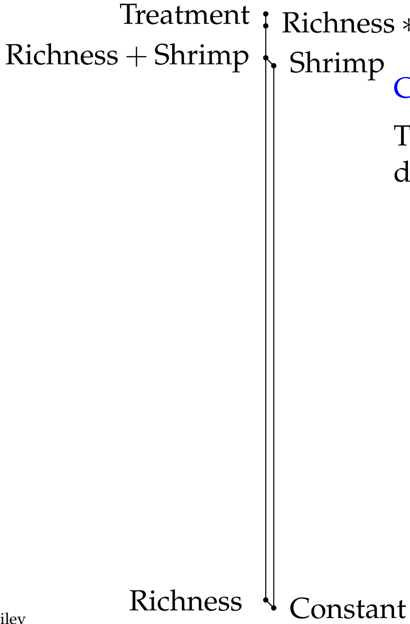


Conclusions:

Scale:

3 × residual mean square

What the data showed: mean squares



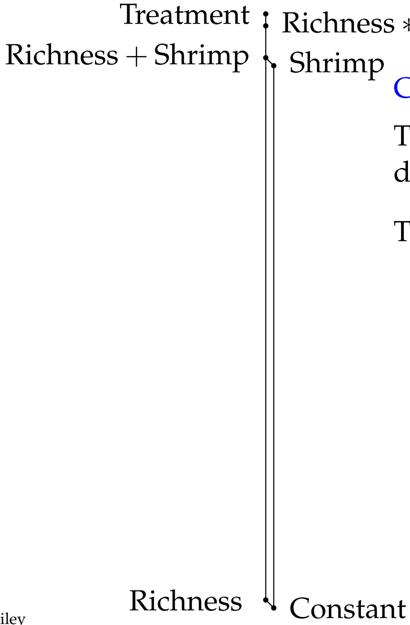
Conclusions:

The model Richness does not explain the data.

Scale:

3 × residual mean square

What the data showed: mean squares



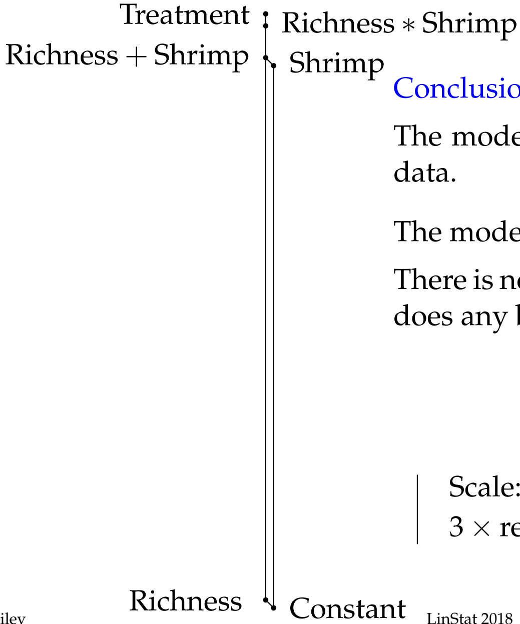
Conclusions:

The model Richness does not explain the data.

The model Shrimp explains the data well.

Scale:
3 × residual mean square

What the data showed: mean squares



Conclusions:

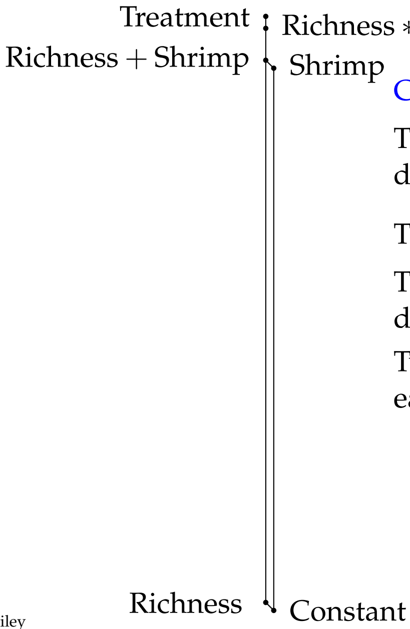
The model Richness does not explain the data.

The model Shrimp explains the data well.

There is no evidence that any larger model does any better.

Scale:
3 × residual mean square

What the data showed: mean squares



Conclusions:

The model Richness does not explain the data.

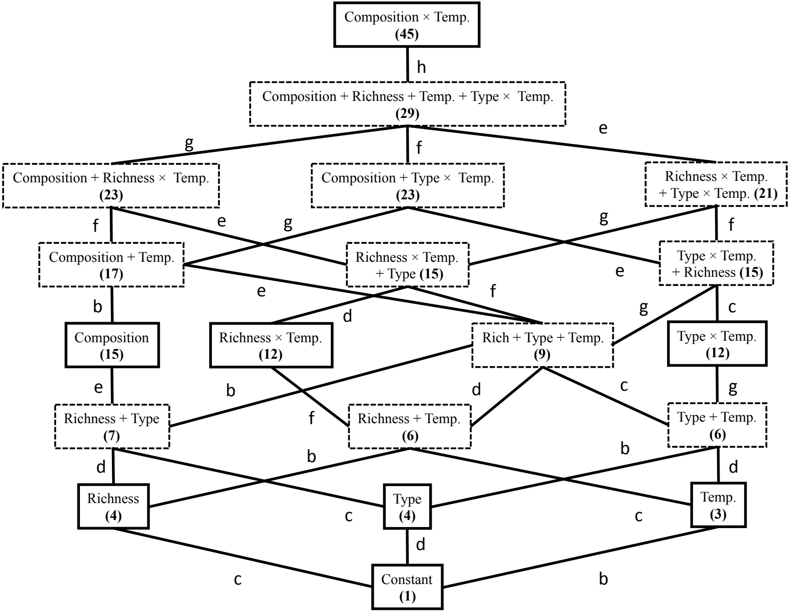
The model Shrimp explains the data well.

There is no evidence that any larger model does any better.

Two experiments, with two responses each, all led to similar conclusions.

Scale:
3 × residual mean square

Diagram from a paper in *Global Change Biology*



Composition × Temp. **(45)**

Composition + Richness + Temp. + Type × Temp. **(29)**

Composition + Richness × Temp. **(23)**

Composition + Type × Temp. **(23)**

Richness × Temp. + Type × Temp. **(21)**

Composition + Temp. **(17)**

Richness × Temp. + Type **(15)**

Type × Temp. + Richness **(15)**

Composition **(15)**

Richness × Temp. **(12)**

Rich + Type + Temp. **(9)**

Type × Temp. **(12)**

Richness + Type **(7)**

Richness + Temp. **(6)**

Type + Temp. **(6)**

Richness **(4)**

Type **(4)**

Temp. **(3)**

Constant **(1)**

Using (scaled) Hasse diagrams

I have found that non-mathematicians find

- Hasse diagrams easier to interpret than equations with side conditions,

Using (scaled) Hasse diagrams

I have found that non-mathematicians find

- Hasse diagrams easier to interpret than equations with side conditions,
- and scaled Hasse diagrams easier to interpret than anova tables,

especially for complicated families of models.

Summary

- Stop pretending that the expectation part of the linear model is a single model with side conditions on its parameters, and recognize that it is, almost always, a family of possible models to describe the expectation.

Summary

- Stop pretending that the expectation part of the linear model is a single model with side conditions on its parameters, and recognize that it is, almost always, a family of possible models to describe the expectation.
- Recognize the relationships between the different expectation models being considered, and show these on a Hasse diagram.

Summary

- Stop pretending that the expectation part of the linear model is a single model with side conditions on its parameters, and recognize that it is, almost always, a family of possible models to describe the expectation.
- Recognize the relationships between the different expectation models being considered, and show these on a Hasse diagram.
- After the data are collected, scale the lengths of the edges of the Hasse diagram to show the relevant mean squares, as a visual summary of the analysis of variance.

Summary

- Stop pretending that the expectation part of the linear model is a single model with side conditions on its parameters, and recognize that it is, almost always, a family of possible models to describe the expectation.
- Recognize the relationships between the different expectation models being considered, and show these on a Hasse diagram.
- After the data are collected, scale the lengths of the edges of the Hasse diagram to show the relevant mean squares, as a visual summary of the analysis of variance.
- Use the Hasse diagram, recursively from the top, to analyse the data and fit a model.