Hasse diagrams in the design and analysis of experiments

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

In memory of Paul Darius,
Leuven Statistics Days, 5 December 2014

Abstract I

Paul Darius was a great advocate of the use of Hasse diagrams as an aid to thinking about factors in a designed experiment.

These diagrams show relationships between factors. They give a straightforward algorithm for calculating degrees of freedom, with (almost) no need to remember formulae. When the fixed and random factors are shown on the same diagram, it is fairly straightforward to read off the skeleton analysis of variance table, to spot confounding and to identify false replication. I recommend doing all of these things at the design stage.

I will try to explain all of this for people who have not used Hasse diagrams before.

Abstract II

There is another use of Hasse diagrams in designed experiments.

Usually the equation giving the expected value of the response is a (very) concise way of saying that we are considering several different models: for example, main effects only or full factorial. The collection of models being considered can also be shown on a Hasse diagram. When the data are available, this diagram can be redrawn with its edges scaled to gve a clear visual display of the information in the ANOVA table.

I will show this on some real examples.

Example 1: small factorial

(From Paul Darius's slides, slightly adapted.)

A field experiment is to be conducted using all combinations of three varieties (of some cereal) and two methods of fertilization. Each combination will replicated four times, so that 24 plots of land (observational units) are needed.

Factor	Number of levels	Levels
Variety (A)	3	$A1, A2, A3$
Method of Fertilization (B)	2	$B1, B2$
Observational unit	24	$1, 2, 3, 4, 5, \ldots, 24$

What is a factor?

(From Paul Darius's slides, slightly adapted.)

A factor assigns a level to each observational unit.

The subsets corresponding to the levels of a factor form together a partition of the set of observational units. These subsets are called parts.

There are two trivial factors (or partitions).

Description	PD notation	RAB notation
Every observational unit is in a different part	ε or E	E
There is only one part	μ or M	U

A partial order: "is finer than"

(From Paul Darius's slides, slightly adapted.)

A partition A is finer than a partition B if A and B are different and each pair of observational units that belong to the same part of A also belong to the same part of B.

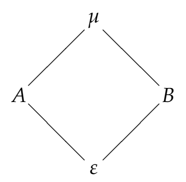
We can then say that partition B is coarser than partition A.

(I use the same words to describe the corresponding factors.)

Technically, the relation "is finer than" induces a partial order in the set of partitions of the set of observational units. A Hasse diagram is a general tool to represent partial orders graphically.

Our version of the Hasse diagram will display the different partitions associated with the experiment from "coarse" on top to "fine" at the bottom.

Example 1: PD's "Hasse diagram of factor structure"



ε is finer than everything else.
Everything else is finer than μ.
Neither A nor B is finer than the other.

Combining two factors: I

In general, if A and B are two factors
then we can make a new factor whose levels are
all combinations of a level of A with a level of B
(restricting to only those combinations that occur in the experiment).

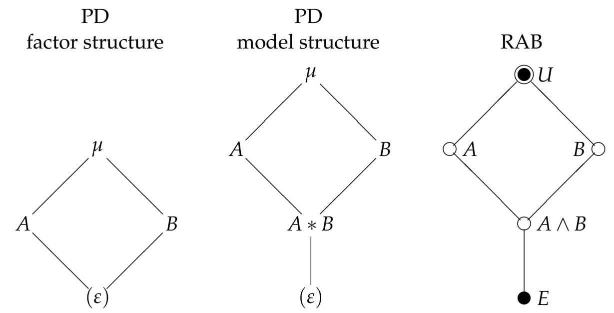
In Example 1, the new factor is Treatment, which has 6 levels:

$$(A1, B1) \quad (A1, B2) \quad (A2, B1) \quad (A2, B2) \quad (A3, B1) \quad (A3, B2)$$

	PD	RAB
Notation	$A * B$ or $A \times B$ or AB	$A \wedge B$
Name	Interaction between A and B	Infimum of A and B

I shall use my notation and name, and explain why later.

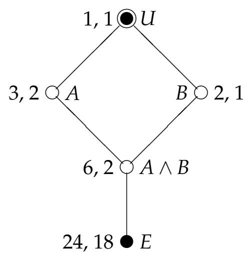
Example 1: three Hasse diagrams



( ) denotes random

● denotes random

From Hasse diagram to degrees of freedom to anova



Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
E	A	2
	B	1
	$A \wedge B$	2
	residual	18

Show the number of levels of each factor.
Calculate degrees of freedom recursively by subtraction.
Each ● gives a stratum (eigenspace of the covariance matrix).
Each ○ gives a source for a fixed effect, contained in the stratum given by the highest ● below or equal to it.
◉ shows a stratum with no residual.

Example 1: Model

Denote the response on observational unit k by Y_k.
If this observational unit has level i of A and level j of B,
then we assume that

$$Y_k = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_k,$$

where the ε_k are independent normal random variables with zero mean and the same variance, and all the other symbols denote constants.

Is this equation " the model"?
or is it a concise way of saying that we are considering several different models?

What is a model?

A model for the response vector $\mathbf{Y}$ usually specifies
$E(\mathbf{Y})$ and $\mathrm{Cov}(\mathbf{Y})$ up to some unknown parameters.

The model is called linear if the possible values for $E(\mathbf{Y})$ form a vector subspace of the space of all possible response vectors.

For factor A, let V_A be the set of vectors whose coordinates are equal on all observational units with the same level of A.

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

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

Example 1: Model subspaces

$E(\mathbf{Y}) \in V_A \iff$ there are constants α_i such that $E(Y_k) = \alpha_i$ whenever $A(k) = i$.

$E(\mathbf{Y}) \in V_B \iff$ there are constants β_j such that $E(Y_k) = \beta_j$ whenever $B(k) = j$.

$E(\mathbf{Y}) \in V_U \iff$ there is a constant μ such that $E(Y_k) = \mu$ for all k.

$E(\mathbf{Y}) \in V_A + V_B \iff$ there are constants α_i and β_j such that $E(Y_k) = \alpha_i + \beta_j$ if $A(k) = i$ and $B(k) = j$.

$E(\mathbf{Y}) \in V_{A \wedge B} \iff$ there are constants λ_{ij} such that $E(Y_k) = \lambda_{ij}$ if $A(k) = i$ and $B(k) = j$.

Dimensions

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

$$V_A \cap V_B = V_U,$$

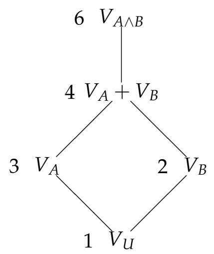
which has dimension 1, so

$$\dim(V_A + V_B) = \dim(V_A) + \dim(V_B) - 1.$$

Another partial order; another Hasse diagram

The relation "is contained in" gives a partial order on subspaces of a vector space.
So we can use a Hasse diagram to show the subspaces being considered to model the expectation of $\mathbf{Y}$.
Now it is helpful to show the dimension of each subspace on the diagram.

Example 1: Hasse diagram for model subspaces



full model

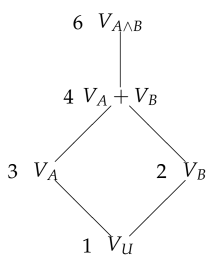
additive model

only factor B makes any difference

null model

For complicated families of models, non-mathematicians may find the Hasse diagram easier to understand than the equations.

Example 1: 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.

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

Labelling in the anova table

In the analysis-of–variance table,
the row labelled by factor B gives the calculations for testing the hypothesis that the main effect of B is zero.

The row labelled by factor $A \wedge B$ gives the calculations for testing the hypothesis that the interaction between A and B is zero.

If the interaction between A and B is zero (up to random noise), we accept the hypothesis that $E(\mathbf{Y})$ belongs to the additive model, and then see if we can further simplify. For example, is the main effect of B zero?

Can you say that a factor is zero?

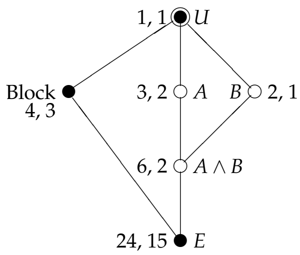
Variant of Example 1: blocks

The field is divided into 4 blocks of 6 observational units each, to take account of known or suspected differences in the soil.

To be able extrapolate our results to other plots in other fields, we need to assume that there is no interaction between the factor Block and the factor Treatment (where Treatment $= A \wedge B$).

So we do not include factors Block $\wedge$ Treatment or Block $\wedge A$ or Block $\wedge B$.

From Hasse diagram to degrees of freedom to anova



Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Block		3
E	A	2
	B	1
	$A \wedge B$	2
	residual	15

Combining two factors: II

If A and B are factors then their infimum $A \wedge B$ satisfies:

- $A \wedge B$ is finer than A, and $A \wedge B$ is finer than B;
- if any other factor is finer than A and finer than B then it is finer than $A \wedge B$.

The supremum $A \vee B$ of factors A and B is defined to satisfy:

- A is finer than $A \vee B$, and B is finer than $A \vee B$;
- if there is any other factor C with A finer than C and B finer than C, then $A \vee B$ is finer than C.

Each part of factor $A \vee B$ is a union of parts of A and is also a union of parts of B, and is as small as possible subject to this.

I claim that the supremum is even more important than the infimum in designed experiments and data analysis.

$$V_A \cap V_B = V_{A \vee B}$$

Example 2: Latin squares with complications

(From Paul Darius's slides, slightly adapted.)

An experiment will be conducted to compare the effects of 4 types of petrol on CO emissions when used in cars in realistic driving circumstances. A driver will drive along a certain route with a certain car, while the emission is measured. We know that there are car-to-car differences, and we suspect that there are route-to-route differences.

There are 12 cars and 12 routes. We do not intend to use all combinations of these; instead we shall use 3 Latin squares, each with 4 routes and 4 cars.

Example 2: rows are cars; columns are routes

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

B	C	D	A
C	D	A	B
D	A	B	C
A	B	C	D

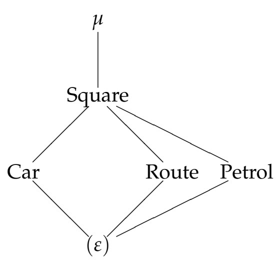
C	B	D	A
B	D	A	C
D	A	C	B
A	C	B	D

Car $\vee$ Route $=$ Square

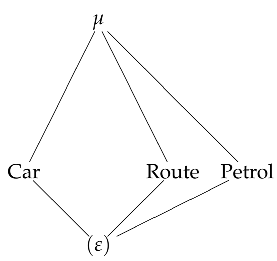
Petrol $\vee$ Car $=$ Petrol $\vee$ Route $= U$

Example 2: Hasse diagram for factor structure

PD factor structure



someone less expert



Example 2: silly data (rows are cars, columns are routes)

6	6	6	6								
6	6	6	6								
6	6	6	6								
6	6	6	6								
				1	1	1	1				
				1	1	1	1				
				1	1	1	1				
				1	1	1	1				
								8	8	8	8
								8	8	8	8
								8	8	8	8
								8	8	8	8

Statistician 1: There are differences between cars. Fit Car and subtract, then there are no differences between routes.

Statistician 2: There are differences between routes. Fit Route and subtract, then there are no differences between cars.

Statistician 3: There are differences between squares: after fitting Square there are no further differences between cars or between routes. We cannot tell whether the differences between squares are caused by cars or routes, because they are confounded.

Example 3: Split plots

(From Paul Darius's slides, slightly adapted.)

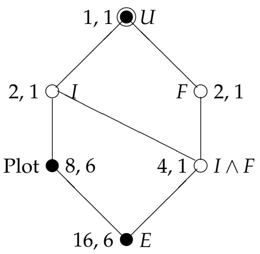
A field trial is planned to study the effect of 2 irrigation methods (factor I) and 2 fertilizers (factor F).

To irrigate a plot with a sprinkler without water spilling into the next plot, we need large plots, and there is room for only 8.

Fertilizer can be applied to much smaller areas, so we can divide each plot into two subplots, one for each level of F.

$I1, F1$	$I2, F2$	$I2, F2$	$I2, F1$	$I1, F1$	$I2, F2$	$I1, F1$	$I1, F2$
$I1, F2$	$I2, F1$	$I2, F1$	$I2, F2$	$I1, F2$	$I2, F1$	$I1, F2$	$I1, F1$

Example 2: Hasse diagram and skeleton anova



Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Plots	I	1
	residual	6
E	F	1
	$I \wedge F$	1
	residual	6

Treatment $= I \wedge F$

Plot $\vee$ Treatment $= I$

Putting in the supremum explicitly: I

Sometimes the person who analyses the data is not the person who designed the experiment, and insufficient records have been kept about the design.

Suppose that you are given data for an experiment on 4 treatments, which are applied to halves of eight plots as follows.

1	4	4	3	1	4	1	2
2	3	3	4	2	3	2	1

This is just Example 3 in disguise!

Putting in the supremum explicitly: II

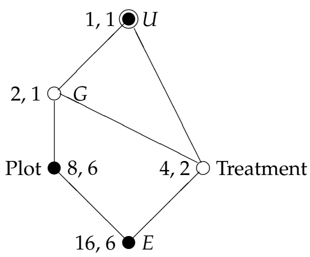
1	4	4	3	1	4	1	2
2	3	3	4	2	3	2	1

Introduce the factor G, where $G =$ Treatment $\vee$ Plot.

G partitions the treatments into two parts: $\{1, 2\}$ and $\{3, 4\}$.

(You don't know this,
but G is just the Irrigation factor that we had before.)

Putting in the supremum explicitly: III



Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Plots	G	1
	residual	6
E	Treatments	2
	residual	6

RAB's guidelines for including infima and suprema

Suppose that we include factors F and G. When should we also include $F \wedge G$ and/or $F \vee G$?

- Always include $F \vee G$.
  If either F or G is fixed then so is $F \vee G$.
- If F and G are both random, then usually include $F \wedge G$.
  (Otherwise, the covariance model cannot be justified by randomization.)
- If F is inherent and local (such as rows in a field) and G is a treatment factor with fixed effects, then assume that we can omit $F \wedge G$.
  (Otherwise, we cannot generalize our results.
  We may need to transform the responses so that the additive assumption is reasonable.)

Grand general theory

If each pair of factors is orthogonal,
and all suprema are included, then the following hold.

- We get orthogonal anova;
  in particular, changing the order of fitting makes no difference.
- The previous algorithm for calculating degrees of freedom is correct.
- The same algorithm can also be used to calculate sums of squares recursively.
- The previous rule for allocating fixed effects to strata is correct.

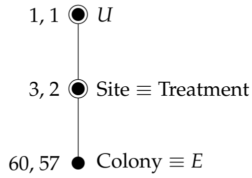
Example 4: false replication

In 2012 the UK's Food and Environment Research Agency conducted an experiment to find out "the effects of neonicotinoid seed treatments on bumble bee colonies under field conditions" (from a DEFRA report available on the web, Crown copyright 2013).
`fera.co.uk/ccss/documents/defraBumbleBeereportPS2371V4A.pdf`

Site	Treatment of oilseed rape seeds
Site A, near Lincoln	no treatment
Site B, near York	Modesto™
Site C, near Scunthorpe	Chinook™

Twenty colonies of bumble bees were placed at each site.
Various outcomes were measured on each colony.

Example 4: Hasse diagram and skeleton anova



Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

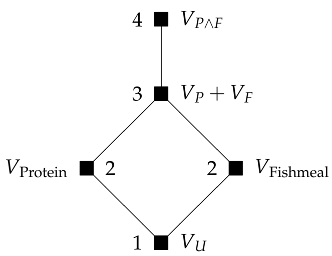
There is no residual mean square in the stratum containing Treatments.
Therefore, there is no way of giving confidence intervals for the estimates of treatment differences, or of giving P values for testing the hypothesis of no treatment difference.

The official report does claim to give confidence intervals and P values.

The Hasse diagram can clearly show such false replication before the experiment is carried out.

Example 5: Two treatment factors

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 5: anova

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 model subspaces

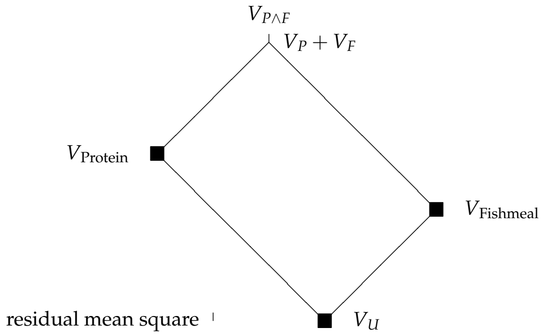
Suppose that M_1 and M_2 are model subspaces,
with $M_1 < M_2$,
and an edge joining M_1 to M_2.

The mean square for
the extra fit in M_2 compared to the fit in M_1 is

$$\frac{\text{SS(fitted values in } M_2) - \text{SS(fitted values in } M_1)}{\dim(M_2) - \dim(M_1)}.$$

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.

Example 5: scaled Hasse diagram of model subspaces



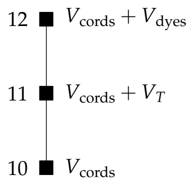
There is no evidence of any interaction, so we can simplify to the additive model. Neither main effect is zero, so we cannot simplify further.

Example 6: an experiment about protecting metal

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.

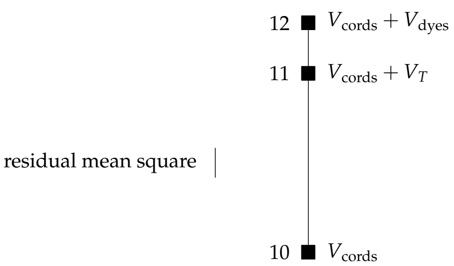
Example 6: Hasse diagram of model subspaces



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.

Example 6: Scaled Hasse diagram of model subspaces



There is no evidence of a difference between dye A and dye B; but there is definitely a difference between no dye and real dyes.

Using scaled Hasse diagrams

I have found that non-mathematicians find
scaled Hasse diagrams easier to interpret than anova tables,
especially for complicated families of models.

These diagrams can be extended to deal
with non-orthogonal models,
and with situations with more than one residual mean square
(use different colours for the corresponding edges).