Simple orthogonal block structures, nesting and marginality

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

John Nelder Workshop in Methodological Statistics,
Imperial College London, 28 March 2015

Abstract I

John Nelder introduced simple orthogonal block structures in one of his famous 1965 papers. They provide a compact description of many of the structures in common use in experiments, so much so that some people find it hard to understand a structure that cannot be expressed in this way.

Terry Speed and I later generalized them to poset block structures.

But there are still misunderstandings.

- If there are 5 blocks of 4 plots each, should the plot factor have 4 levels or 20?
- What is the difference between nesting and marginality?
- What is the difference between a factor, the effect of that factor (this effect may be called an interaction in some cases), and the smallest model which includes that factor whilst respecting marginality?

Abstract II

John himself expressed strong views about people who ignored marginality in the model-fitting process.

My take on this is that there are three different partial orders involved: I will try to explain the difference.

Labelling plots in blocks

Suppose that there are five blocks of four plots each.
How should we label them?

B	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5
P	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

B	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5
Q	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Advantage of using P: fewer levels, so less computer storage space.

Advantage of using Q: if the data are analysed by someone who did not design the experiment, they cannot make the mistake of thinking that all plots ω with $P(\omega) = 1$ have something in common.

Terminology

B	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5
P	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

B	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5
Q	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

We say that P is nested in B because the information that $P(\omega_1) = P(\omega_2)$ is irrelevant unless $B(\omega_1) = B(\omega_2)$.

We say that Q is finer than P because we know that if $Q(\omega_1) = Q(\omega_2)$ then $B(\omega_1) = B(\omega_2)$.

These relationships are different, and need different words, but many people confuse them.

P and Q are different types of thing, and play different roles, so I shall call P a pre-factor and Q a factor, but many people confuse them, or use different terminology.

Pre-factors and nesting

Write $P \sqsubset B$ to indicate that P is nested in B.
Write $P \sqsubseteq B$ to mean that either $P \sqsubset B$ or $P = B$.

Nesting is a partial order, which means that

- $F \sqsubseteq F$ for all pre-factors F;
- if $F \sqsubseteq G$ and $G \sqsubseteq F$ then $F = G$;
- if $F \sqsubseteq G$ and $G \sqsubseteq H$ then $F \sqsubseteq H$.

Hasse diagram

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

Put a symbol for each object (here, a pre-factor).

If $F \sqsubset G$ then the symbol for F is lower in the diagram than the symbol for G, and is joined to it by lines that are traversed upwards.

5 □ B

4 □ P

Show the numbers of levels.

If we have three rows (R) and eight columns (C) with no nesting then we get

3 □ R 8 □ C

Combining two factors or pre-factors

If A and B are two factors then their infimum $A \wedge B$ is the factor whose levels are all combinations of levels of A and B that occur.

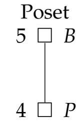
$$(A \wedge B)(\omega) = (A(\omega), B(\omega))$$

Other notations: $A.B$ or $A : B$.

Crossing and nesting

Operation	Formula	Poset
crossing	$(3\,R) * (8\,C)$	3 □ R 8 □ C

Experimental units	$\{1,2,3\} \times \{1,2,3,4,5,6,7,8\}$
Factors:	U with one level
	R with 3 levels (1st coordinate)
	C with 8 levels (2nd coordinate)
	$R \wedge C$ with 24 levels

Operation	Formula	Poset
nesting	$(5\,B)/(4\,P)$	

Experimental units	$\{1,2,3,4,5\} \times \{1,2,3,4\}$
Factors:	U with one level
	B with 5 levels (1st coordinate)
	$B \wedge P$ with 20 levels

From crossing and nesting to simple orthogonal block structures

The key ingredient of John Nelder's 1965 paper on 'Block structure and the null analysis of variance' was to realise that crossing and nesting could be iterated (maybe with some steps of each sort).

He developed an almost-complete theory, notation and algorithms based on this.

He called the resulting sets of experimental units with their factor lists simple orthogonal block structures.

Factors and refinement

If B and Q are factors on the same set,
write $Q \prec B$ to indicate that Q is finer than B.
Write $Q \preceq B$ to mean that either $Q \prec B$ or $Q = B$.

Refinement is another partial order, because

- $F \preceq F$ for all factors F;
- if $F \preceq G$ and $G \preceq F$ then $F = G$;
- if $F \preceq G$ and $G \preceq H$ then $F \preceq H$.

(For simplicity here, I am ignoring the possibility of aliasing.)

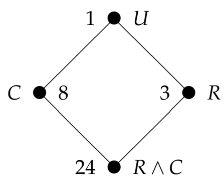
So we can show factors on a Hasse diagram too!

Crossing

Hasse diagram for pre-factors: 3 □ R 8 □ C

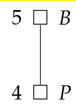
Experimental units	$\{1,2,3\} \times \{1,2,3,4,5,6,7,8\}$
Factors:	U with one level
	R with 3 levels (1st coordinate)
	C with 8 levels (2nd coordinate)
	$R \wedge C$ with 24 levels

Hasse diagram for factors:



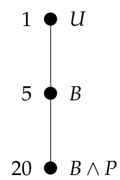
Nesting

Hasse diagram for pre-factors:

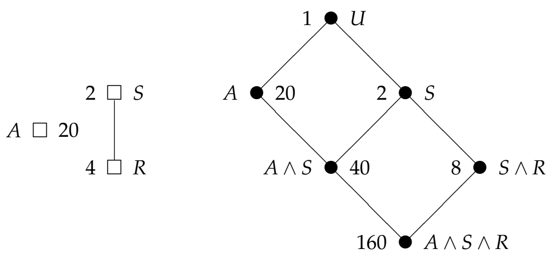


Experimental units	$\{1,2,3,4,5\} \times \{1,2,3,4\}$
Factors:	U with one level
	B with 5 levels (1st coordinate)
	$B \wedge P$ with 20 levels

Hasse diagram for factors:



Iteration: (20 Athletes $*$ ((2 Sessions) / (4 Runs)))



Start with the first poset

Terry Speed and I found that
you can start with the nesting poset and use it to directly construct the set Ω of experimental units and its factors.

Given pre-factors $P_1, \ldots, P_m$ with $n_1, \ldots, n_m$ levels,
and a nesting relation $\sqsubset$:

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_m \quad \text{where } \Omega_i = \{1, 2, \ldots, n_i\}.$$

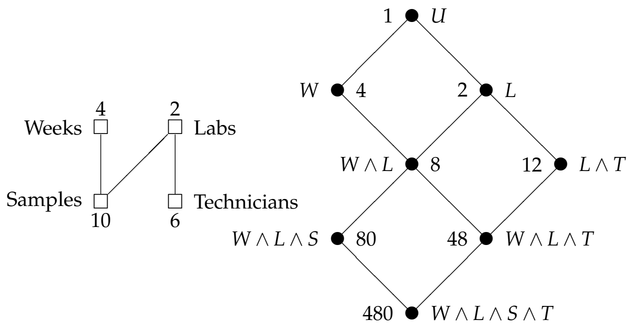
If $\mathcal{A}$ is any subset of $\{1, 2, \ldots, m\}$ satisfying

$$\text{if } i \in \mathcal{A} \text{ and } P_i \sqsubset P_j \text{ then } j \in \mathcal{A}$$

then include the factor $\bigwedge_{i \in \mathcal{A}} P_i$.

Poset block structures

These poset block structures have all John Nelder's properties, even when the first poset cannot be made by iterated crossing and nesting.



Too successful

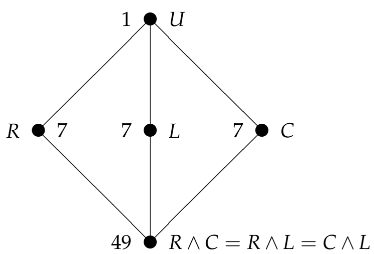
John Nelder's theory of simple orthogonal block structures, and the ensuing algorithms developed with Graham Wilkinson,
have been enormously successful, but perhaps too much so.

B	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5
P	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Q	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

As factors, $B \wedge P = Q = B \wedge Q$,
but does your software think so?
Some software cannot detect that $Q \prec B$,
because B is not in the name of Q.
Some software thinks that $B \wedge Q$ has 100 levels,
and tries to make 100×100 matrices to deal with this.

Other orthogonal block structures

There are still other collections of mutually orthogonal factors which obey most of the theory but do not come from pre-factors.
For example, the Rows (R), Columns (C) and Letters (L) of a 7×7 Latin square give the following.



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 level of factor $A \vee B$ combines levels of A
and also combines levels of B,
and has replication 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.

Factorial treatments plus control

Chemical

Dose		Z	N	S	K	M
	0	✓				
	1		✓	✓	✓	✓
	2		✓	✓	✓	✓

$$\text{Dose} \vee \text{Chemical} = \text{Fumigant},$$

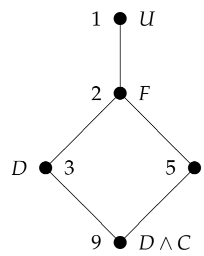
which is the two-level factor distinguishing zero treatment from the rest.

If you do not fit Fumigant, its effect will be included in whichever of Dose and Chemical you fit first.

Hasse diagram including supremum

Chemical

Dose		Z	N	S	K	M
	0	✓				
	1		✓	✓	✓	✓
	2		✓	✓	✓	✓



With F included, all the usual nice results apply.

Heiko Großmann's software includes suprema
(as well as checking which factors are finer than which others).

Does yours?

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) $\sum_i \alpha_i = 0$, and so on

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

Linear model with constraints: bad consequences

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

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

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

- 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" are 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 constraints

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

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

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

JAN'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' 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 $\mathbf{Y}$ 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 then the model does not simplify at all.

(I read this in one of his papers, but could not find it again when preparing these slides.)

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

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

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

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

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

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

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

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

$$V_A \cap V_B = V_U,$$

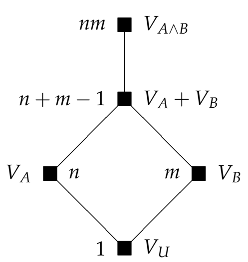
which has dimension 1, so

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

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.

Hasse diagram for model subspaces



full model

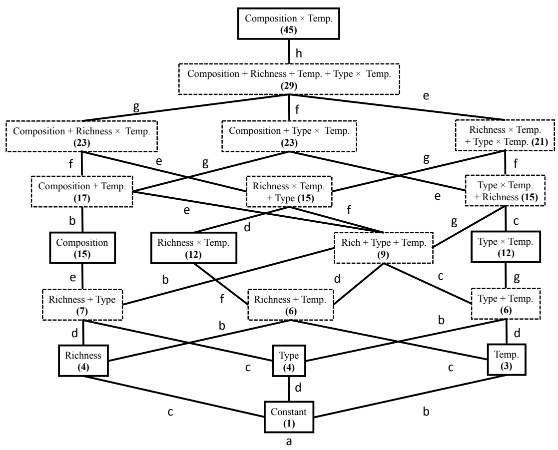
additive model

only factor B makes any difference

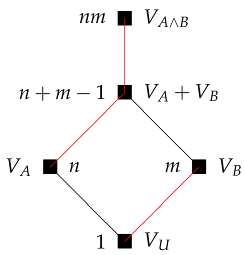
null model

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

Diagram from a paper in *Global Change Biology*



Main effects and interaction



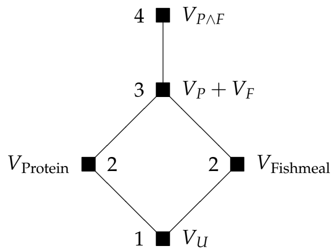
The interaction between factors A and B is the difference between the vector of fitted values in $V_{A\wedge B}$ and the vector of fitted values in $V_A + V_B$.

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.

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

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

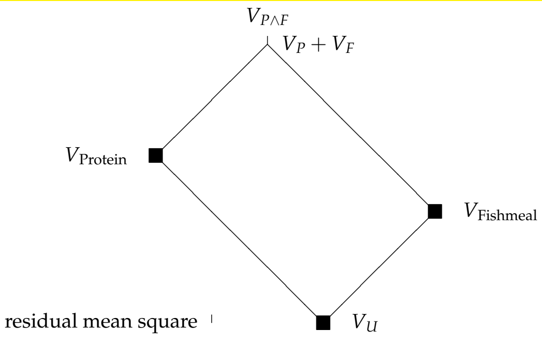
Suppose that V_1 and V_2 are expectation subspaces,
with $V_1 < V_2$,
and an edge joining V_1 to V_2.

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

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

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.

Chickens: scaled Hasse diagram of expectation 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: 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.

Cords: Hasse diagram of expectation subspaces

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

11 ■ $V_{\text{cords}} + V_T$

10 ■ V_{cords}

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

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

11 ■ $V_{\text{cords}} + V_T$

residual mean square

10 ■ V_{cords}

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).

References I

- J. A. Nelder: The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proceedings of the Royal Society of London, Series A* **282** (1965), 147–162.
- T. P. Speed & R. A. Bailey: On a class of association schemes derived from lattices of equivalence relations. In *Algebraic Structures and Applications* (eds. P. Schultz, C. E. Praeger & R. P. Sulllivan), Marcel Dekker, New York, 1982, pp. 55–74.
- H. Großmann: Automating the analysis of variance of orthogonal designs. *Computational Statistics and Data Analysis* **70** (2014), 1–18.

References II

- J. A. Nelder: A reformulation of linear models (with discussion). *Journal of the Royal Statistical Society, Series A* **140** (1977), 48–77.
- J. A. Nelder: The statistics of linear models: back to basics. *Statistics and Computing* **4** (1994), 221–234.
- J. A. Nelder & P. W. Lane: The computer analysis of factorial experiments: In memorian—Frank Yates. *The American Statistician* **49** (1995), 382–385.
- J. A. Nelder: The great mixed-model muddle is alive and flourishing, alas! *Food and Quality Preference* **9** (1998), 157–159.
- R. A. Bailey: *Design of Comparative Experiments*. Cambridge University Press, Cambridge, 2008.

References III

- D. M. Perkins, R. A. Bailey, M. Dossena, L. Gamfeldt, J. Reiss, M. Trimmer & G. Woodward: Higher biodiversity is required to sustain multiple ecosystem processes across temperature regimes. *Global Change Biology* **21** (2015), 396–406.
- R. A. Bailey & J. Reiss: Design and analysis of experiments testing for biodiversity effects in ecology. *Journal of Statistical Planning and Inference* **144** (2014), 69–80.
- K. J. Carpenter & J. Duckworth: Economies in the use of animal by-products in poultry rations. I. Vitamin and amino-acid provision for starting and growing chicks. *Journal of Agricultural Science* **41** (1941), 297–308.
- M. Crowder & A. Kimber: A score test for the multivariate Burr and other Weibull mixture distributions. *Scandinavian Journal of Statistics* **24** (1997), 419–432.