Blocking in multi-stage experiments

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

Designed Experiments:
Recent Advances in Methods and Applications,
Sydney, December 2015

Abstract

In a multi-stage experiment,
the same experimental units are used in each stage
but different treatment factors are applied at different stages.

Constraints on processing imply that
these units must be partitioned into blocks at each stage.

However, unlike in the classical situation,
the blocks are not inherent, and the designer of the experiment
can choose the partition into blocks at each stage.

Is it better to align the Stage 2 blocks with the Stage 1 blocks as far as possible
or to make them as orthogonal to each other as possible?

In either case, how should treatments be assigned?

Example 1: the problem

There are 36 experimental units.

In Stage 1, these must be processed in 6 batches of size 6.
Treatment factor F has 2 levels, which are applied in Stage 1,
and these can be changed within each batch.

In Stage 2, the units must be processed in 6 lots of size 6.
Treatment factor G has 3 levels, which are applied in Stage 2,
and these can be changed within each lot.

How should we design the experiment?

Example 1: design 1a

Design 1a Align batches with lots,
and allocate the 6 combinations of levels of F and G
in a randomized complete-block design.

units		treatments		
source	df	source	df	EMS
Mean	1	Mean	1	$6\sigma_B^2 + 6\sigma_L^2 + \sigma^2 + q_0$
Blocks	5			$6\sigma_B^2 + 6\sigma_L^2 + \sigma^2$
Units[B]	30	F	1	$\sigma^2 + q(F)$
		G	2	$\sigma^2 + q(G)$
		F#G	2	$\sigma^2 + q(FG)$
		residual	25	σ^2

Example 1: design 1b

Design 1b Cross batches with lots to form a square array,
and allocate the 6 combinations of levels of F and G
in a Latin square.

units		treatments		
source	df	source	df	EMS
Mean	1	Mean	1	$6\sigma_B^2 + 6\sigma_L^2 + \sigma^2 + q_0$
Batches	5			$6\sigma_B^2 + \sigma^2$
Lots	5			$6\sigma_L^2 + \sigma^2$
B#L	25	F	1	$\sigma^2 + q(F)$
		G	2	$\sigma^2 + q(G)$
		F#G	2	$\sigma^2 + q(FG)$
		residual	20	σ^2

We have lost 5 residual degrees of freedom, and gained nothing.

An easy lesson

Lesson
If treatments can be orthogonal to blocks in both stages,
and the blocks from the two stages can be aligned,
this gives more residual degrees of freedom
without increasing the variance of treatment contrasts.

This extends to three or more stages.

Brien, Harch, Correll and Bailey (2011) call this
"confounding big with big".

Example 2: the problem

There are 36 experimental units.

In Stage 1, these must be processed in 6 batches of size 6.
Treatment factor F has 2 levels, which are applied in Stage 1,
and these must be applied to whole batches.

In Stage 2, the units must be processed in 6 lots of size 6.
Treatment factor G has 3 levels, which are applied in Stage 2,
and these can be changed within each lot.

How should we design the experiment?

Example 2: design 2a

Design 2a Align batches with lots;
allocate the 2 levels of F to whole batches
in a completely randomized design,
and allocate the 3 levels of G to two random units per block.

units		treatments		
source	df	source	df	EMS
Mean	1	Mean	1	$6\sigma_B^2 + 6\sigma_L^2 + \sigma^2 + q_0$
Blocks	5	F	1	$6\sigma_B^2 + 6\sigma_L^2 + \sigma^2 + q(F)$
		residual	4	$6\sigma_B^2 + 6\sigma_L^2 + \sigma^2$
Units[B]	30	G	2	$\sigma^2 + q(G)$
		F#G	2	$\sigma^2 + q(FG)$
		residual	26	σ^2

The variance for the contrast between levels of F involves
σ_L^2 as well as σ_B^s, so it is larger than it needs to be.

There are 4 residual degrees of freedom for testing
the main effect of F, and this cannot be increased.

Example 2: design 2b

Design 2b Cross batches with lots to form a square array;
allocate the 2 levels of F to whole batches,
allocate the 3 levels of G to two units in each row and column,
and then randomize rows and columns.

units		treatments		
source	df	source	df	EMS
Mean	1	Mean	1	$6\sigma_B^2 + 6\sigma_L^2 + \sigma^2 + q_0$
Batches	5	F	1	$6\sigma_B^2 + \sigma^2 + q(F)$
		residual	4	$6\sigma_B^2 + \sigma^2$
Lots	5			$6\sigma_L^2 + \sigma^2$
B#L	25	G	2	$\sigma^2 + q(G)$
		F#G	2	$\sigma^2 + q(FG)$
		residual	21	σ^2

Main effect of F has smaller variance than before, and same df.
Other treatment effects have same (small) variance,
and df reduced from 26 to 21.

Some principles

Principle

If a treatment factor has to be applied to large units such as blocks in one stage,
then try to make it orthogonal to blocks in other stages.

Principle

If a treatment factor has to be applied to large units such as blocks in one stage,
then it will have relatively few residual degrees of freedom.
In order not to reduce these further, try to confound
the whole of this block term with the same term in other stages.

Example 2: design 2c

Design 2c Make 3 squares by crossing pairs of batches
(shown as rows) with pairs of lots (shown as columns).

$F1$	$G1, G2, G3$	$G1, G2, G3$
$F2$	$G1, G2, G3$	$G1, G2, G3$

$F1$	$G1, G2, G3$	$G1, G2, G3$
$F2$	$G1, G2, G3$	$G1, G2, G3$

$F1$	$G1, G2, G3$	$G1, G2, G3$
$F2$	$G1, G2, G3$	$G1, G2, G3$

Randomize levels of F to rows within each square;
randomize levels of G within each corner of each square.

Remark Squares is the supremum of Batches and Lots:
Squares $=$ Batches $\vee$ Lots.

Example 2: skeleton anova for design 2c

units		treatments		
source	df	source	df	EMS
Mean	1	Mean	1	$6\sigma_B^2 + 6\sigma_L^2 + \sigma_1^2 + q_0$
Squares	2			$6\sigma_B^2 + 6\sigma_L^2 + \sigma_1^2$
Batches[S]	3	F	1	$6\sigma_B^2 + \sigma_1^2 + q(F)$
		residual	2	$6\sigma_B^2 + \sigma_1^2$
Lots[S]	3			$6\sigma_L^2 + \sigma_1^2$
B#L[S]	3			σ_1^2
Units[B,L,S]	24	G	2	$\sigma^2 + q(G)$
		F#G	2	$\sigma^2 + q(FG)$
		residual	20	σ^2

The randomization argument suggests that $\sigma_1^2 \neq \sigma^2$.
Both residual df have decreased, and nothing has been gained.

Example 3: the problem

There are 36 experimental units.

In Stage 1, these must be processed in 6 batches of size 6.
Treatment factor F has 2 levels, which are applied in Stage 1,
and these must be applied to whole batches.

In Stage 2, the units must be processed in 6 lots of size 6.
Treatment factor G has 3 levels, which are applied in Stage 2,
and these must be applied to whole lots.

How should we design the experiment?

Example 3: design 3

Design 3 Cross batches with lots to form a square array;
randomize the 2 levels of F to whole batches,
randomize the 3 levels of G to whole lots.

units		treatments		
source	df	source	df	EMS
Mean	1	Mean	1	$6\sigma_B^2 + 6\sigma_L^2 + \sigma^2 + q_0$
Batches	5	F	1	$6\sigma_B^2 + \sigma^2 + q(F)$
		residual	4	$6\sigma_B^2 + \sigma^2$
Lots	5	G	2	$\sigma_L^2 + \sigma^2 + q(G)$
		residual	3	$6\sigma_L^2 + \sigma^2$
B#L	25	$F\#G$	2	$\sigma^2 + q(FG)$
		residual	23	σ^2

This is quite a standard design, sometimes called a strip-plot design or criss-cross design.
Any futher blocking of either batches or lots
reduces the already-small residual df for main effects.

More general numbers

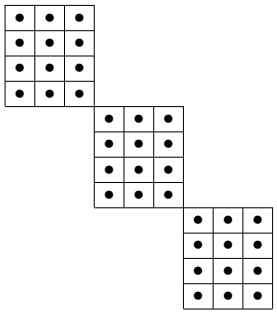
Suppose that there are N experimental units.
Stage 1 groups them into b batches of size s;
Stage 2 groups them into c lots of size k.

To align batches with lots, we need $b = c$.
To nest batches within lots, we need $c|b$.
To nest lots within batches, we need $b|c$.
To cross batches with lots, we need $bc|N$.

For more general orthogonality between batches and lots,
we need m arrays of b/m batches crossed with c/m lots,
with each intersection containing Nm/bc experimental units,
where m divides b, m divides c and bc divides Nm,
so m divides $\gcd\{b, c\}$, and $\gcd\{b, c\}$ divides $Nm/\mathrm{lcm}\{b, c\}$.

Example 4: arrays

There are 36 experimental units.
In Stage 1, these must be processed in 12 batches of size 3.
In Stage 2, the units must be processed in 9 lots of size 4.
For orthogonality, the only possibililty is three 4×3 arrays.



m Arrays

The between-Arrays stratum variance is

$$s\sigma_B^2 + k\sigma_L^2 + \sigma^2,$$

so these $m - 1$ degrees of freedom are typically not used for inference or estimation.

It is desirable to keep m small.

Powers of the same small prime

Suppose that the number of batches and the number of lots are both powers of p, where $p = 2$ or $p = 3$,
that several p-level treatment factors $F_1, F_2, \ldots$ must be applied to whole batches in Stage 1, and several p-level treatment factors $G_1, G_2, \ldots$ must be applied to whole lots in Stage 2.

If each stage is a single replicate of the relevant treatments, we may be able to take $m = p$ and sacrifice information only on high-order interactions.

For example, if $p = 2$ and F_1, F_2, F_3 and F_4 are applied to whole batches in Stage 1 while G_1, G_2 and G_3 are applied to whole lots in Stage 2, we can use two 8×4 arrays, and confound $F_1F_2F_3F_4$ and $G_1G_2G_3$ with each other and with arrays.

This technique is called post-fractionation by Bisgaard (1997) and Vivacqua and Bisgaard (2009).

Three stages

Suppose that there are three stages, and that treatment factors F, G and H are applied in Stages 1, 2, 3 respectively.

Suppose that each treatment factor must be applied to whole blocks in its stage.

We already know that we should try to make the blocks from each stage as orthogonal as possible to blocks from every other stage.

An example was investigated by Mee and Bates (1998).

Example 5: the problem

There are 16 experimental units.

In Stage 1, these must be processed in 4 batches of size 4.
Treatment factor F has 2 levels, which are applied in Stage 1, and these must be applied to whole batches.

In Stage 2, the units must be processed in 4 lots of size 4.
Treatment factor G has 2 levels, which are applied in Stage 2, and these must be applied to whole lots.

In Stage 3, the units must be processed in 4 pods of size 4.
Treatment factor H has 2 levels, which are applied in Stage 3, and these must be applied to whole pods.

How should we design the experiment?

Example 5: design 5a

Design 5a Form the experimental units into 2 arrays of size $2 \times 2 \times 2$. The first coordinate indicates the batch, the second coordinate indicates the lot, and the third coordinate indicates the pod. Within each array, randomize levels of F to batches, levels of G to lots, and levels of H to pods.

The single df between Arrays is wasted.

All 7 treatment df have different variances, each with just one residual df.

Example 5: design 5b

Design 5b Let Λ be a 4×4 Latin square.
Stage 1: identify the batches with the rows of Λ.
Stage 2: identify the lots with the columns of Λ.
Stage 3: identify the pods with the letters Λ.

Can we apply levels of F, G, H to rows, columns, letters respectively in such a way that all treatment interactions are orthogonal to rows, columns and letters?

Mee and Bates found a cunning way of doing this, using the non-cyclic Latin square of order 4.

Example 5: skeleton anova for design 5b

units		treatments		
source	df	source	df	EMS
Mean	1	Mean	1	$4\sigma_R^2 + 4\sigma_C^2 + 4\sigma_L^2 + \sigma^2 + q_0$
Rows	3	F	1	$4\sigma_R^2 + \sigma^2 + q(F)$
		residual	2	$4\sigma_R^2 + \sigma^2$
Columns	3	G	1	$4\sigma_C^2 + \sigma^2 + q(G)$
		residual	2	$4\sigma_C^2 + \sigma^2$
Letters	3	H	1	$4\sigma_L^2 + \sigma^2 + q(H)$
		residual	2	$4\sigma_L^2 + \sigma^2$
Units[R,C,L]	6	$F\#G$	1	$\sigma^2 + q(FG)$
		$F\#H$	1	$\sigma^2 + q(FH)$
		$G\#H$	1	$\sigma^2 + q(GH)$
		$F\#G\#H$	1	$\sigma^2 + q(FGH)$
		residual	2	σ^2

For this Latin square, the decomposition into strata can be justified by randomization (Bailey, 1982).

Example 5: design 5b by design key

The design key introduced by Patterson (1965) gives a clean construction of design 5b. All factors and pseudofactors have two levels, and arithmetic is modulo 2.

Rows: R_1, R_2, $R_1 + R_2$.
Columns: C_1, C_2, $C_1 + C_2$.
Letters: $L_1 = R_1 + C_1$, $L_2 = R_2 + C_2$, $L_1 + L_2 = R_1 + R_2 + C_1 + C_2$.
$F = R_1$.
$G = C_2$.
$H = R_1 + R_2 + C_1 + C_2$.
$F + G = R_1 + C_2$.
$F + H = R_2 + C_1 + C_2$.
$G + H = R_1 + R_2 + C_1$.
$F + G + H = R_2 + C_1$.

Non-orthogonality

So far, we have assumed that treatments applied in Stage i

either must be applied to whole blocks in Stage i

or can be orthogonal to blocks in Stage i.

Suppose that neither of these conditions holds?

Example 6: the problem

There are 36 experimental units.

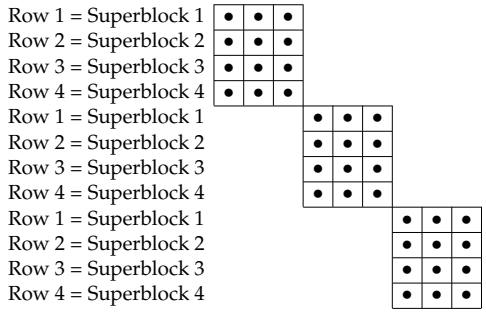
Stage 1 has 12 blocks of size 3, in 4 superblocks of 3 blocks.
Treatment factor F has 3 levels, which are applied in Stage 1, and these must be applied to whole blocks.

Stage 2 is a rectangle with 4 rows and 9 columns.
Treatment factor G has 3 levels, which are applied in Stage 2, in such a way that each combination of F and G occurs 4 times.

Forced: F applied to whole blocks.
Good idea: Apply each level of F to one block per superblock.
Good idea: Apply each level of G to one exp. unit per block.
Good idea: Align superblocks and rows.
Good idea: Form the 12 blocks and 9 columns into three arrays of size 4×3.
Good idea: Make the design in columns as efficient as possible.

Can we achieve all of this?

Example 6: align superblocks with rows, and make arrays



Example 6: design

	Array 1			Array 2			Array 3		
Row 1	$F1G1$	$F1G2$	$F1G3$	$F2G1$	$F2G2$	$F2G3$	$F3G1$	$F3G2$	$F3G3$
Row 2	$F2G3$	$F2G1$	$F2G2$	$F3G3$	$F3G1$	$F3G2$	$F1G3$	$F1G1$	$F1G2$
Row 3	$F3G3$	$F3G1$	$F3G2$	$F1G3$	$F1G1$	$F1G2$	$F2G3$	$F2G1$	$F2G2$
Row 4	$F1G2$	$F1G3$	$F1G1$	$F2G2$	$F2G3$	$F2G1$	$F3G2$	$F3G3$	$F3G1$

$F1$	$F1$	$F1$	$F2$	$F2$	$F2$	$F3$	$F3$	$F3$
$G3$	$G1$	$G2$	$G3$	$G1$	$G2$	$G3$	$G1$	$G2$

The design in columns is factorially balanced, with canonical efficiency factors 15/16 for both main effects and 3/4 for the interaction.

Example 6: skeleton anova

units		treatments			
source	df	cef	source	df	EMS
Mean	1	1	Mean	1	$9\sigma^2_R + 9\sigma^2_S + 4\sigma^2_C + 3\sigma^2_B + \sigma^2 + q_0$
Rows	3				$9\sigma^2_R + 9\sigma^2_S + 3\sigma^2_B + \sigma^2$
Arrays	2	$\frac{1}{16}$	F	2	$4\sigma^2_C + 3\sigma^2_B + \sigma^2 + \frac{1}{16}q(F)$
Blocks[R,A]	6	$\frac{15}{16}$	F	2	$3\sigma^2_B + \sigma^2 + \frac{15}{16}q(F)$
			residual	4	$3\sigma^2_B + \sigma^2$
Columns[A]	6	$\frac{1}{16}$	G	2	$4\sigma^2_C + \sigma^2 + \frac{1}{16}q(G)$
		$\frac{1}{4}$	$F\#G$	4	$4\sigma^2_C + \sigma^2 + \frac{1}{4}q(FG)$
Units[B,C,R,A]	18	$\frac{15}{16}$	G	2	$\sigma^2 + \frac{15}{16}q(G)$
		$\frac{3}{4}$	$F\#G$	4	$\sigma^2 + \frac{3}{4}q(FG)$
			residual	12	σ^2

A strategy for non-orthogonality

Suppose that, in Stage 1, treatment factor F has canonical efficiency factors

p in Batches
q in Units[Batches],

where $p + q = 1$.
Possible strategies for Stage 2:

$(F \perp L)$ make F orthogonal to Lots;

(BigwithBig) confound (the F-part of) Batches with Lots and the F-part of $U[B]$ with $U[L]$;

(Compensate) confound the F-part of Batches with $U[L]$ and the F-part of $U[B]$ with Lots.

If $p = 1$ then F is confounded with Batches, so $(F \perp L)$ is the same as (Compensate) and this is best.
If $q = 1$ then F is orthogonal to Batches, so $(F \perp L)$ is the same as (BigwithBig) and this is best.
For all values of p, $(F \perp L)$ is best if it is possible.

A strategy for non-orthogonality, continued

Theorem
If Stage 1 has b batches of size s and Stage 2 has c lots of size k, then (BigwithBig) is better than (Compensate) in the sense of giving smaller variances for the estimators of contrasts between levels of F if and only if

$$\frac{q}{p} > \frac{\sigma^2}{(s\sigma_B^2 + \sigma^2)} \frac{(k\sigma_L^2 + \sigma^2)}{(s\sigma_B^2 + k\sigma_L^2 + \sigma^2)}.$$

If $q > p$ then (BigwithBig) is better;
otherwise, prior estimates of the relatives magnitudes of σ^2, σ_B^2 and σ_L^2 are required to make the decision.

Another strategy for non-orthogonality

Stage 1 has b batches of size s and Stage 2 has c lots of size k, so $N = bs = ck$.
If s divides k then we could nest batches within lots.
If sk divides N then we could cross batches with lots in N/sk arrays.
If $k = ts$ and $N = msk$ then we could do either.

Which is better?

Another strategy for non-orthogonality, continued

Stage 1 has mts batches of size s and
Stage 2 has ms lots of size ts.
Let Δ be a design for F in mts batches of size s.
Let Γ be a design for F in ms lots of size ts.
Let Nest(Δ, Γ) be a design where each lot contains t batches, the design in batches is Δ and the design in lots is Γ.
Let Cross(Δ, Γ) be a design with m arrays of ts batches crossed with s lots,
the design in batches is Δ and the design in lots is Γ.

Theorem
The variances for the estimators of contrasts between levels of F are no bigger for Nest(Δ, Γ) than for Cross(Δ, Γ).

This is not quite the whole story,
because it may be possible to construct a Cross(Δ, Γ) for better block designs Δ and Γ than a Nest(Δ, Γ).

And finally ...

In classical experiments, the block structure is inherent and we have to do our best from that starting point.

In multi-stage experiments, the relationship between the block structures in the different stages is not fixed in advance, so the design problem includes the question of how to match the block structures as well as how to allocate sets of treatments.
Many multi-phase experiments are similar.

This all leads to interesting questions.
Thank you for listening.