The design key in single- and multi-phase experiments

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

Biometrics by the Harbour:
Meeting of the Australasian Region of the International Biometric Society,
30 November 2015

Abstract

Desmond Patterson introduced the design key in 1965 in the context of experiments on crop rotations.

It can be used whenever the treatments have factorial structure, the experimental units have a poset block structure, and an orthogonal design is required.

The design key gives an algorithm for allocating treatments to experimental units, and another algorithm for identifying which stratum contains which treatment effect.

These two properties make it a very useful tool when extended to multi-phase experiments.

H. Desmond Patterson

Introduction of the design key

H. D. Patterson:
The factorial combination of treatments in rotation experiments.
Journal of Agricultural Science, **65** (1965), 171–182.

This paper introduced the design key.

The number of levels of each factor must be a power of a single prime number p. All examples have $p = 2$, but it is mentioned that the method can also be used with $p = 3$.

I had intended to include an example from this paper, but they are all far too complicated.
I am amazed that any referee understood it at the time.
Parts of it are almost as if Desmond is thinking aloud.

Example 1 (Graeco-Latin square): the key

Factors

	Factors	Factors with five levels
Experimental units	Rows	R
	Columns	C
Treatments	Variety of wheat	W
	Quantity of nitrogen	N

Every factor is represented by a single letter.
Levels are integers modulo 5.

Constraints The treatment factors W and N should both be orthogonal to rows and to columns.

Design key The design key expresses each treatment factor as a linear combination of factors on the experimental units.

$$W = R + C \qquad N = R + 2C$$

Example 1 (Graeco-Latin square): construction

$$W = R + C \qquad N = R + 2C$$

	C				
R	0	1	2	3	4
0	0,0	1,2	2,4	3,1	4,3
1	1,1	2,3	3,0	4,2	0,4
2	2,2	3,4	4,1	0,3	1,0
3	3,3	4,0	0,2	1,4	2,1
4	4,4	0,1	1,3	2,0	3,2

The experimental units are defined by all combinations of levels of R and C.

The level of W is shown first in each cell.

The level of N is shown second in each cell.

Example 1 (Graeco-Latin square): confounding

$$W = R + C \qquad N = R + 2C$$

Stratum	unit effect	df	tmt factor	tmt effect
Rows	R	4	$W + 2N$	interaction
Columns	C	4	$W + 4N$	interaction
Rows-by-Columns	$R + C$	4	W	variety main
	$R + 2C$	4	N	nitrogen main
	$R + 3C$	4	$W + 3N$	interaction
	$R + 4C$	4	$W + N$	interaction

$$W + N = 2R + 3C \equiv 6R + 9C = R + 4C$$

Treatment factors

There is a set $\mathcal{F}$ of treatment factors.

There is one potential treatment for each combination of levels of all the factors in $\mathcal{F}$.

At first, we assume that every factor in $\mathcal{F}$ has p levels, where p is prime. The levels are the integers modulo p. All addition is done modulo p.

Each non-zero linear combination of factors in $\mathcal{F}$ gives a treatment pseudofactor with p levels.
This gives $p - 1$ degrees of freedom for contrasts between treatments, all belonging to the interaction of those genuine treatment factors whose coefficient is non-zero.

If one such linear combination is a non-zero multiple of another, then they correspond to the same df; otherwise the corresponding sets of contrasts are orthogonal to each other.

Factors on the experimental units

There is a set $\mathcal{G}$ of unit factors.

We assume that the real factors on the experimental units form a poset block structure.

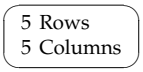
This means that they can be defined by a panel diagram, showing
- the list of factors $G_1, \ldots, G_m$ in $\mathcal{G}$
- for each G_i, its number n_i of levels;
- for each G_i, what it is nested in.

There are $n_1 \times \cdots \times n_m$ experimental units, one for each combination of levels of $G_1, \ldots, G_m$.

"G_i is nested in G_j" means
"if two objects have the same level of G_i then this has no significance unless they have the same level of G_j".

The real factors are combinations of levels of none or more of $G_1, \ldots, G_m$ subject to the rule that
if G_i is included and G_i is nested in G_j then G_j must be included.

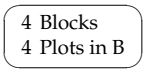
Poset block structure in Example 1



This panel diagram tells us that
- there are factors R and C, each with 5 levels;
- there are 25 experimental units, one for each combination of levels of R and C;
- there is no nesting;
- the real factors on the experimental units are

$\emptyset$	with 1 level;
R	with 5 levels;
C	with 5 levels;
RC	with 25 levels.

Poset block structure in Example 2



This panel diagram tells us that
- there are factors B and P, each with 4 levels;
- there are 16 experimental units, one for each combination of levels of B and P;
- P is nested in B, so there is no real factor involving P but not B;
- the real factors on the experimental units are

$\emptyset$	with 1 level;
B	with 4 levels;
BP	with 16 levels.

Powers of a prime

If a factor has p^r levels, where $r \geq 2$, then it is represented by r pseudofactors, each with p levels.
The convention is that these pseudofactors are written with the same single letter and subscripts $1, \ldots, r$.

Identification of factorial effects

For a linear combination of factors (and pseudofactors) in $\mathcal{F}$ or a linear combination of factors (and pseudofactors) in $\mathcal{G}$, we need to identify the factorial effect containing the corresponding $p-1$ degrees of freedom.

1. Write down all the letters which occur, ignoring subscripts;
2. if factor C is nested in factor D and letter C occurs then include letter D;
3. remove any duplicate letters.

The set of letters remaining gives the factorial effect.

Example 2 (Factorial design in blocks): the key

Factors

	Factors	(Pseudo-)factors with 2 levels
Experimental units	Blocks (4)	B_1, B_2
	Plots in Blocks (4)	P_1, P_2
Treatments	S (2)	S
	T (2)	T
	U (2)	U
	V (2)	V

The pseudofactors for each factor all have the same letter. Levels are integers modulo 2.

Constraints All treatment main effects should be orthogonal to blocks. So should as many two-factor interactions as possible.

Design key

$S = P_1 \qquad T = P_2 \qquad U = B_1 + P_1 + P_2 \qquad V = B_2 + P_1 + P_2$

Example 2 (Factorial design in blocks): construction

$S = P_1 \qquad T = P_2 \qquad U = B_1 + P_1 + P_2 \qquad V = B_2 + P_1 + P_2$

The experimental units are defined by all combinations of levels of B_1, B_2, P_1 and P_2.

	Block 1				Block 2				Block 3				Block 4			
B_1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
B_2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
P_1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
P_2	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
S	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
T	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
U	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1
V	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1

Example 2 (Factorial design in blocks): confounding

$S = P_1 \qquad T = P_2 \qquad U = B_1 + P_1 + P_2 \qquad V = B_2 + P_1 + P_2$

Stratum	unit effect	df	tmt factor	tmt effect
Blocks (B)	B_1	1	$S+T+U$	3 f.i.
	B_2	1	$S+T+V$	3 f.i.
	B_1+B_2	1	$U+V$	U-by-V intn
Plots in Blocks (BP)	P_1	1	S	main S
	P_2	1	T	main T
	P_1+P_2	1	$S+T$	S-by-T intn
	B_1+P_1	1	$T+U$	T-by-U intn
	B_1+P_2	1	$S+U$	S-by-U intn
	$B_1+P_1+P_2$	1	U	main U
	B_2+P_1	1	$T+V$	T-by-V intn
	B_2+P_2	1	$S+V$	S-by-V intn
	$B_2+P_1+P_2$	1	V	main V
	$B_1+B_2+P_1$	1	$S+U+V$	3 f.i.
	$B_1+B_2+P_2$	1	$T+U+V$	3 f.i.
	$B_1+B_2+P_1+P_2$	1	$S+T+U+V$	4 f.i.

What does the design key do?

A design key is a list
giving an alias for each treatment (pseudo-)factor as
a linear combination of (pseudo-)factors for the experimental units.

This gives

- an algorithm for constructing the design;
- a design that is orthogonal;
- (if it is a fractional replicate) a fraction which is regular;
- an algorithm for identifying the confounding between treatment effects and strata defined by a poset block structure on the experimental units.

Generalization to more than one prime

Suppose that more than one prime is involved.

If any factor has a composite number of levels, express it as a product of pseudofactors, each with a prime number of levels.

For each prime p_i separately, consider only the treatment (pseudo-)factors and unit (pseudo-)factors which have p_i levels, and make a design key for them, using arithmetic modulo p_i.

Suppose that $p_1, \ldots, p_k$ are among the primes involved, and that, for $i = 1, \ldots, k$, T_i is a linear combination of treatment factors or pseudofactors with p_i levels.
Then T_i belongs to an effect defined by a subset $\mathcal{S}_i$ of the initial letters of the genuine treatment factors.
It can be shown that the $\prod_{i=1}^{k}(p_i - 1)$ df for the interaction between $T_1, \ldots, T_k$ all belong to the effect defined by the subset $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_k$.

The analogous result holds for factors on the experimental units.

Example 3 (Whole-plot factors): the key

2 Grass type
3 Mowing height
3 Fertilizer quantity

2 Rows
4 Columns
3 Strips in R
3 Lines in C

Constraints Grass type cannot change within any of the eight row-column combinations.
Mowing height must be the same along each strip.
Fertilizer quantity must be the same along each line.

Design key

prime	treatments	units	key
2	G	R, C_1, C_2	$G = R + C_1$
3	M, F	S, L	$M = S, F = L$

Example 3 (Whole-plot factors): confounding

prime	treatments	units	key
2	G	R, C_1, C_2	$G = R + C_1$
3	M, F	S, L	$M = S, F = L$

prime(s)	treatment lin. comb.	treatment effect	stratum
2	G	G	RC
3	M	M	RS
	F	F	CL
	$M + F$	MF	$RCSL$
	$M + 2F$	MF	$RCSL$
2 and 3	$G * M$	GM	RCS
	$G * F$	GF	RCL
	$G * (M + F)$	GMF	$RCSL$
	$G * (M + 2F)$	GMF	$RCSL$

Example 3 (Whole-plot factors): skeleton anova

units		treatments	
source	df	source	df
Mean	1	Mean	1
Rows	1		
Columns	3		
R#C	3	Grass type	1
		Residual	2
Strips[R]	4	Mowing height	2
		Residual	2
Lines[C]	8	Fertilizer quantity	2
		Residual	6
Strips[R]#C	12	G#M	2
		Residual	10
R#Lines[C]	8	G#F	2
		Residual	6
S[R]#L[C]	32	M#F	4
		G#M#F	4
		Residual	24

Two-phase experiments

Treatments Phase I units Phase II units

The first design key allocates treatments to Phase I units.
The second design key allocates Phase I units to Phase II units.

Combining these allows us to keep track of confounding all the way through,
which helps us to choose suitable design keys in the first place.

Example 4 (Proteomics): constraints

2 Interventions → 8 Cages
2 Tissues - - - → 2 Positions in A, C
?
8 Runs
8 Labels

- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

Design key

I	$I = C_1$	$C_1,\ C_2,\ C_3$	$C_i = R_i + L_i$	$R_1,\ R_2,\ R_3$
		A	$A = R_1$	
T	$T = P + C_3$	P	$P = L_2$	$L_1,\ L_2,\ L_3$

Example 4 (Proteomics): confounding

2 Interventions → 8 Cages
2 Tissues - - - → 2 Positions in A, C
key
8 Runs
8 Labels

I	$I = C_1$	$C_1,\ C_2,\ C_3$	$C_i = R_i + L_i$	$R_1,\ R_2,\ R_3$
		A	$A = R_1$	
T	$T = P + C_3$	P	$P = L_2$	$L_1,\ L_2,\ L_3$

$P + C_2 = R_2$ Positions[A,C], Runs

$T = P + C_3 = L_2 + R_3 + L_3$ T, P[A,C], R#L

$I + T = C_1 + P + C_3 = R_1 + L_1 + L_2 + R_3 + L_3$ I#T, P[A,C], R#L

Example 4 (Proteomics): skeleton anova

units		animal-bits		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 2\eta_0 + q_0$
Runs	7	Animals[C]$_1$	1			$\xi_R + 2\eta_{CA}$
		Positions[A,C]$_1$	2			$\xi_R + 2\eta_{CAP}$
		Residual	4			ξ_R
Labels	7	Animals[C]$_2$	1			$\xi_L + 2\eta_{CA}$
		Positions[A,C]$_2$	2			$\xi_L + 2\eta_{CAP}$
		Residual	4			ξ_L
R#L	49	Cages	7	Interventions	1	$\xi_{RL} + 2\eta_C + q(\mathrm{I})$
				Residual	6	$\xi_{RL} + 2\eta_C$
		Animals[C]$_3$	6			$\xi_{RL} + 2\eta_{CA}$
		Positions[A,C]$_3$	12	Tissues	1	$\xi_{RL} + 2\eta_{CAP} + q(\mathrm{T})$
				I#T	1	$\xi_{RL} + 2\eta_{CAP} + q(\mathrm{IT})$
				Residual	10	$\xi_{RL} + 2\eta_{CAP}$
		Residual	24			ξ_{RL}

Example 5 (Field then laboratory): constraints

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1,\ V_2,\ V_3$	$V_3 = R + C$	R	$B_1,\ B_2$
	$V_1 = P_1$	C	
	$V_2 = P_2$	$P_1,\ P_2$	$S_1,\ S_2$

Constraints All Variety effects should be orthogonal to Rows and orthogonal to Columns in Phase I.
Then at least 2df for Varieties must be confounded with R#C, so there is no loss of generality in taking this design key for the first phase.

Question What should we do in the second phase, given that at least 2df for Varieties must be confounded with Batches?

Example 5 (Field then laboratory): option 1

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1,\ V_2,\ V_3$	$V_3 = R + C$	R	$R = B_1$	$B_1,\ B_2$
	$V_1 = P_1$	C	$C = B_2$	
	$V_2 = P_2$	$P_1,\ P_2$	$P_i = S_i$	$S_1,\ S_2$

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Columns	2			$\xi_B + \eta_C$
		R#C	4	V_3	2	$\xi_B + \eta_{RC} + q(V_3)$
				Residual	2	$\xi_B + \eta_{RC}$

Example 5 (Field then laboratory): option 2

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1,\ V_2,\ V_3$	$V_3 = R + C$	R	$R = B_1$	$B_1,\ B_2$
	$V_1 = P_1$	C	$P_1 = B_2$	
	$V_2 = P_2$	$P_1,\ P_2$	$C = S_1,\ P_2 = S_2$	$S_1,\ S_2$

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Plots[R,C]$_1$	6	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	4	$\xi_B + \eta_{RCP}$

Example 5 (Field then laboratory): option 3

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1,\ V_2,\ V_3$	$V_3 = R + C$	R	$R = B_1 + S_2$	$B_1,\ B_2$
	$V_1 = P_1$	C	$P_1 = B_2$	
	$V_2 = P_2$	$P_1,\ P_2$	$C = S_1,\ P_2 = S_2$	$S_1,\ S_2$

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Plots[R,C]$_1$	8	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	6	$\xi_B + \eta_{RCP}$

Possible generalizations

More phases Treatments are applied to first-phase units. For $i > 1$, in each phase i, the material from the units in phase $i-1$ is applied to units in phase i using another design key, but no further treatments are applied.

The foregoing ideas can be applied recursively, and no new concepts are involved.

Treatments in the second phase Keep two phases.
Apply one set of treatments in the first phase, and another set of treatments in the second phase. There may be interactions between the two sets of treatments.

Example coming up.

Example 6 (Potato storage): constraints

4 Cultivars 3 Fungicides	3 Blocks 12 Units in B	12 Shelves 12 Pallets in S	4 Times

prime	tmts	units	pallets	times
2	C_1, C_2	U_1, U_2	S_1, S_2, P_1, P_2	T_1, T_2
3	F	B, U_3	S_3, P_3	

Constraints The design for the first phase should be a complete-block design.
In the second phase, potatoes from each first-phase unit are harvested and put onto four different pallets on the same shelf, where they are stored for four different lengths of time.

Example 6 (Potato storage): key

4 Cultivars 3 Fungicides	3 Blocks 12 Units in B	12 Shelves 12 Pallets in S	4 Times

prime	tmts	units	pallets	times
2	C_1, C_2	U_1, U_2	S_1, S_2, P_1, P_2	T_1, T_2
3	F	B, U_3	S_3, P_3	

Design keys

prime	tmts to units	units to pallets	times to pallets
2	$C_1 = U_1$	$U_1 = S_1$	$T_1 = P_1$
	$C_2 = U_2$	$U_2 = S_2$	$T_2 = P_2$
3	$F = U_3$	$B = S_3$	
		$U_3 = P_3$	

Example 6 (Potato storage): consequences

4 Cultivars 3 Fungicides	3 Blocks 12 Units in B	12 Shelves 12 Pallets in S	4 Times

prime	tmts to units	units to pallets	times to pallets
2	$C_1 = U_1$	$U_1 = S_1$	$T_1 = P_1$
	$C_2 = U_2$	$U_2 = S_2$	$T_2 = P_2$
3	$F = U_3$	$B = S_3$	
		$U_3 = P_3$	

All combinations of levels of Cultivars, Fungicides and Times occur three times.

These 48 treatments are orthogonal to Blocks.

Cultivars are confounded with Shelves.