Square array designs for unreplicated test treatments with replicated controls

R. A. Bailey
University of St Andrews

Thirteenth Working Seminar on
Statistical Methods in Variety Testing,
COBORU, Słupia Wielka, Poland, September 2025

Joint work with Linda Haines (University of Cape Town)

Outline

1. Background
2. Our construction
3. Valid randomization
4. Optimality
5. References

Chapter 1

Background

The setting

In plant breeding experiments, the quantity of seed available for each test line is usually sufficient for only a single plot. Therefore it is common to use augmented designs, in which several control treatments are replicated.

The setting

In plant breeding experiments, the quantity of seed available for each test line is usually sufficient for only a single plot. Therefore it is common to use augmented designs, in which several control treatments are replicated.

Here we consider square array designs.
There are t^2 plots, arranged in a $t \times t$ square.
There are k control treatments,
each occurring once in each row and once in each column.
The remaining $t(t-k)$ plots are each allocated a different test line.

The setting

In plant breeding experiments, the quantity of seed available for each test line is usually sufficient for only a single plot. Therefore it is common to use augmented designs, in which several control treatments are replicated.

Here we consider square array designs.
There are t^2 plots, arranged in a $t \times t$ square.
There are k control treatments,
each occurring once in each row and once in each column.
The remaining $t(t-k)$ plots are each allocated a different test line.

Rob Kempton (1984) suggested that the ratio k/t should be in the range of 20 to 25%.

The setting

In plant breeding experiments, the quantity of seed available for each test line is usually sufficient for only a single plot.
Therefore it is common to use augmented designs,
in which several control treatments are replicated.

Here we consider square array designs.
There are t^2 plots, arranged in a $t \times t$ square.
There are k control treatments,
each occurring once in each row and once in each column.
The remaining $t(t-k)$ plots are each allocated a different test line.

Rob Kempton (1984) suggested that the ratio k/t should be in the range of 20 to 25%.
However, the number of error degrees of freedom is a multiple of $k-2$, so we always assume that $k \geq 3$.

Previous work

In the 1970s, Federer, Raghavarao and colleagues worked on what they called augmented row–column designs.

Previous work

In the 1970s, Federer, Raghavarao and colleagues worked on what they called augmented row–column designs.

Here is an example with $t = 7$ and $k = 3$.

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

Previous work

In the 1970s, Federer, Raghavarao and colleagues worked on what they called augmented row–column designs.

Here is an example with $t = 7$ and $k = 3$.

A	1	2	3	*C*	4	*B*
B	*A*	5	6	7	*C*	8
9	*B*	*A*	10	11	12	*C*
C	13	*B*	*A*	14	15	16
17	*C*	18	*B*	*A*	19	20
21	22	*C*	23	*B*	*A*	24
25	26	27	*C*	28	*B*	*A*

There are three controls treatments: A, B and C.

Previous work

In the 1970s, Federer, Raghavarao and colleagues worked on what they called augmented row–column designs.

Here is an example with $t = 7$ and $k = 3$.

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

There are three controls treatments: A, B and C.
Each occurs once in each row and once in each column.

Previous work

In the 1970s, Federer, Raghavarao and colleagues worked on what they called augmented row–column designs.

Here is an example with $t = 7$ and $k = 3$.

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

There are three controls treatments: A, B and C.
Each occurs once in each row and once in each column.

Previous work

In the 1970s, Federer, Raghavarao and colleagues worked on what they called augmented row–column designs.

Here is an example with $t = 7$ and $k = 3$.

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

There are three controls treatments: A, B and C.
Each occurs once in each row and once in each column.
The remaining plots are occupied by test treatments 1–28, each occurring only once.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Similarly, for each pair of columns, there is a single row where they both have control treatments.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Similarly, for each pair of columns, there is a single row where they both have control treatments.

The relationship between rows and columns is like that between blocks and treatments in a symmetric balanced incomplete-block design for 7 treatments in 7 blocks of size 3;

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Similarly, for each pair of columns, there is a single row where they both have control treatments.

The relationship between rows and columns is like that between blocks and treatments in a symmetric balanced incomplete-block design for 7 treatments in 7 blocks of size 3; this is called the auxiliary block design.

Explanation of the design

A	1	2	3	C	4	B
B	A	5	6	7	C	8
9	B	A	10	11	12	C
C	13	B	A	14	15	16
17	C	18	B	A	19	20
21	22	C	23	B	A	24
25	26	27	C	28	B	A

For each pair of rows, there is a single column where they both have control treatments.

Similarly, for each pair of columns, there is a single row where they both have control treatments.

The relationship between rows and columns is like that between blocks and treatments in a symmetric balanced incomplete-block design for 7 treatments in 7 blocks of size 3; this is called the auxiliary block design. If we remove the test treatments, this is called a Youden square.

Our construction

An example with $t = 16$ and $k = 4$

We start with an equireplicate auxiliary block design for t treatments in t blocks of size k.

An example with $t = 16$ and $k = 4$

We start with an equireplicate auxiliary block design for t treatments in t blocks of size k.
Here is an example with $t = 16$ and $k = 4$.

$$
\begin{array}{cccc}
\{1,2,3,4\} & \{5,6,7,8\} & \{9,10,11,12\} & \{13,14,15,16\} \\
\{1,5,9,13\} & \{2,6,10,14\} & \{3,7,11,15\} & \{4,8,12,16\} \\
\{1,6,11,16\} & \{2,5,12,15\} & \{3,8,9,14\} & \{4,7,10,13\} \\
\{1,7,12,14\} & \{2,8,11,13\} & \{3,5,10,16\} & \{4,6,9,15\}
\end{array}
$$

An example with $t = 16$ and $k = 4$

We start with an equireplicate auxiliary block design for t treatments in t blocks of size k.
Here is an example with $t = 16$ and $k = 4$.

$$\begin{array}{cccc}
\{1,2,3,4\} & \{5,6,7,8\} & \{9,10,11,12\} & \{13,14,15,16\} \\
\{1,5,9,13\} & \{2,6,10,14\} & \{3,7,11,15\} & \{4,8,12,16\} \\
\{1,6,11,16\} & \{2,5,12,15\} & \{3,8,9,14\} & \{4,7,10,13\} \\
\{1,7,12,14\} & \{2,8,11,13\} & \{3,5,10,16\} & \{4,6,9,15\}
\end{array}$$

A technique known as *Hall's Marriage Theorem*, using a result of Hall (1935), shows that we can present this block design as a $k \times t$ rectangle where the columns are labelled by the blocks, the entries in each column are the treatments in the corresponding block, and each treatment occurs exactly once in each row.

Using Hall's Marriage Theorem

$$\begin{array}{cccc}
\{1,2,3,4\} & \{5,6,7,8\} & \{9,10,11,12\} & \{13,14,15,16\} \\
\{1,5,9,13\} & \{2,6,10,14\} & \{3,7,11,15\} & \{4,8,12,16\} \\
\{1,6,11,16\} & \{2,5,12,15\} & \{3,8,9,14\} & \{4,7,10,13\} \\
\{1,7,12,14\} & \{2,8,11,13\} & \{3,5,10,16\} & \{4,6,9,15\}
\end{array}$$

Using Hall's Marriage Theorem

$\{1,2,3,4\}$	$\{5,6,7,8\}$	$\{9,10,11,12\}$	$\{13,14,15,16\}$
$\{1,5,9,13\}$	$\{2,6,10,14\}$	$\{3,7,11,15\}$	$\{4,8,12,16\}$
$\{1,6,11,16\}$	$\{2,5,12,15\}$	$\{3,8,9,14\}$	$\{4,7,10,13\}$
$\{1,7,12,14\}$	$\{2,8,11,13\}$	$\{3,5,10,16\}$	$\{4,6,9,15\}$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	8	10	15	9	2	7	16	6	12	3	13	14	11	5	4
4	5	11	14	1	10	15	8	16	2	9	7	12	13	3	6
3	6	12	13	5	14	11	4	1	15	8	10	7	2	16	9
2	7	9	16	13	6	3	12	11	5	14	4	1	8	10	15

Using Hall's Marriage Theorem

$\{1,2,3,4\}$	$\{5,6,7,8\}$	$\{9,10,11,12\}$	$\{13,14,15,16\}$
$\{1,5,9,13\}$	$\{2,6,10,14\}$	$\{3,7,11,15\}$	$\{4,8,12,16\}$
$\{1,6,11,16\}$	$\{2,5,12,15\}$	$\{3,8,9,14\}$	$\{4,7,10,13\}$
$\{1,7,12,14\}$	$\{2,8,11,13\}$	$\{3,5,10,16\}$	$\{4,6,9,15\}$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	1	8	10	15	9	2	7	16	6	12	3	13	14	11	5	4
B	4	5	11	14	1	10	15	8	16	2	9	7	12	13	3	6
C	3	6	12	13	5	14	11	4	1	15	8	10	7	2	16	9
D	2	7	9	16	13	6	3	12	11	5	14	4	1	8	10	15

Label the rows of this rectangle with k letters which represent the control treatments.

Using Hall's Marriage Theorem

$\{1,2,3,4\}$	$\{5,6,7,8\}$	$\{9,10,11,12\}$	$\{13,14,15,16\}$
$\{1,5,9,13\}$	$\{2,6,10,14\}$	$\{3,7,11,15\}$	$\{4,8,12,16\}$
$\{1,6,11,16\}$	$\{2,5,12,15\}$	$\{3,8,9,14\}$	$\{4,7,10,13\}$
$\{1,7,12,14\}$	$\{2,8,11,13\}$	$\{3,5,10,16\}$	$\{4,6,9,15\}$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	1	8	10	15	9	2	7	16	6	12	3	13	14	11	5	4
B	4	5	11	14	1	10	15	8	16	2	9	7	12	13	3	6
C	3	6	12	13	5	14	11	4	1	15	8	10	7	2	16	9
D	2	7	9	16	13	6	3	12	11	5	14	4	1	8	10	15

Label the rows of this rectangle with k letters which represent the control treatments.

As Federer and Raghavarao (1975) showed, interchanging the rows and letters gives a $t \times t$ square array with each of the k controls occuring once in each row and once in each column.

Using Hall's Marriage Theorem

$$\begin{array}{cccc}
\{1,2,3,4\} & \{5,6,7,8\} & \{9,10,11,12\} & \{13,14,15,16\} \\
\{1,5,9,13\} & \{2,6,10,14\} & \{3,7,11,15\} & \{4,8,12,16\} \\
\{1,6,11,16\} & \{2,5,12,15\} & \{3,8,9,14\} & \{4,7,10,13\} \\
\{1,7,12,14\} & \{2,8,11,13\} & \{3,5,10,16\} & \{4,6,9,15\}
\end{array}$$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	1	8	10	15	9	2	7	16	6	12	3	13	14	11	5	4
B	4	5	11	14	1	10	15	8	16	2	9	7	12	13	3	6
C	3	6	12	13	5	14	11	4	1	15	8	10	7	2	16	9
D	2	7	9	16	13	6	3	12	11	5	14	4	1	8	10	15

Label the rows of this rectangle with k letters which represent the control treatments.

As Federer and Raghavarao (1975) showed, interchanging the rows and letters gives a $t \times t$ square array with each of the k controls occuring once in each row and once in each column.

We modifed this to change
(row A, column 2, number 8) to (row 2, column 8, letter A).

The ensuing square array

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	A	D	C	B												
2					B	C	D	A								
3									D	A	B	C				
4													C	B	A	D
5	B				C				A				D			
6		A				D				B				C		
7			D				A				C				B	
8				C				B				D				A
9	C					A					D					B
10		B			D							A			C	
11			A					C	B					D		
12				D			B			C			A			
13	D						C					B		A		
14		C						D			A		B			
15			B		A					D						C
16				A		B			C						D	

Finishing the design

What we have done so far leaves $t(t-k)$ empty plots.

Now we fill them with test treatments, with a different one on each plot.

The ensuing square array design

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	A	D	C	B	1	2	3	4	5	6	7	8	9	10	11	12
2	13	14	15	16	B	C	D	A	17	18	19	20	21	22	23	24
3	25	26	27	28	29	30	31	32	D	A	B	C	33	34	35	36
4	37	38	39	40	41	42	43	44	45	46	47	48	C	B	A	D
5	B	49	50	51	C	52	53	54	A	55	56	57	D	58	59	60
6	61	A	62	63	64	D	65	66	67	B	68	69	70	C	71	72
7	73	74	D	75	76	77	A	78	79	80	C	81	82	83	B	84
8	85	86	87	C	88	89	90	B	91	92	93	D	94	95	96	A
9	C	97	98	99	100	A	101	102	103	104	D	105	106	107	108	B
10	109	B	110	111	D	112	113	114	115	116	117	A	118	119	C	120
11	121	122	A	123	124	125	126	C	B	127	128	129	130	D	131	132
12	133	134	135	D	136	137	B	138	139	C	140	141	A	142	143	144
13	D	145	146	147	148	149	C	150	151	152	153	B	154	A	155	156
14	157	C	158	159	160	161	162	D	163	164	A	165	B	166	167	168
15	169	170	B	171	A	172	173	174	175	D	176	177	178	179	180	C
16	181	182	183	A	184	B	185	186	C	187	188	189	190	191	D	192

Remark

Federer, Raghavarao and colleagues developed this approach in the 1970s using auxiliary block designs that are balanced (and often which are cyclic).

Remark

Federer, Raghavarao and colleagues developed this approach in the 1970s using auxiliary block designs that are balanced (and often which are cyclic).

Linda Haines and I are astonished that no one had extended the idea to general incomplete block designs with the same number of blocks and treatments.

Valid randomization

Randomization

The "ensuing square array design" on a previous slide has some rows where all the control treatments occur in a line of contiguous plots. This is probably not a good idea.

Randomization

The “ensuing square array design” on a previous slide has some rows where all the control treatments occur in a line of contiguous plots. This is probably not a good idea.

However, in order to draw valid conclusions from the experiment, we need to randomize in a valid way.

Randomization

The "ensuing square array design" on a previous slide has some rows where all the control treatments occur in a line of contiguous plots. This is probably not a good idea.

However, in order to draw valid conclusions from the experiment, we need to randomize in a valid way.

Grundy and Healy (1950) showed that one way to do this is to start with a carefully chosen array and then permute rows, and permute columns, in each case choosing a random element of a given doubly transitive permutation group.

Randomization

The “ensuing square array design” on a previous slide has some rows where all the control treatments occur in a line of contiguous plots. This is probably not a good idea.

However, in order to draw valid conclusions from the experiment, we need to randomize in a valid way.

Grundy and Healy (1950) showed that one way to do this is to start with a carefully chosen array and then permute rows, and permute columns, in each case choosing a random element of a given doubly transitive permutation group.

This technique is called restricted randomization by statisticians in the UK,
and constrained randomization by statisticians in the USA.

Randomization

The “ensuing square array design” on a previous slide has some rows where all the control treatments occur in a line of contiguous plots. This is probably not a good idea.

However, in order to draw valid conclusions from the experiment, we need to randomize in a valid way.

Grundy and Healy (1950) showed that one way to do this is to start with a carefully chosen array and then permute rows, and permute columns, in each case choosing a random element of a given doubly transitive permutation group.

This technique is called restricted randomization by statisticians in the UK,
and constrained randomization by statisticians in the USA.

For many practical values of t and k, we were able to find a suitable doubly transitive permutation group and a good starting array such that the positioning of the control treatments never gave a bad value of the space-filling criterion.

Randomization: Some useful doubly transitive groups

A permutation group is doubly transitive if whenever (α, β) and (γ, δ) are pairs of distinct elements then the group contains a permutation taking α to γ and taking β to δ.

Randomization: Some useful doubly transitive groups

A permutation group is doubly transitive if whenever (α, β) and (γ, δ) are pairs of distinct elements then the group contains a permutation taking α to γ and taking β to δ.

First suppose that $t = 11$.
Renumber the rows and the columns as $0, 1, \ldots, 10$ modulo 11.

Randomization: Some useful doubly transitive groups

A permutation group is doubly transitive if whenever (α, β) and (γ, δ) are pairs of distinct elements then the group contains a permutation taking α to γ and taking β to δ.

First suppose that $t = 11$.
Renumber the rows and the columns as $0, 1, \ldots, 10$ modulo 11.
One such group consists of the permutations

$$x \mapsto ax + b \bmod 11 \qquad \text{where } a \neq 0 \bmod 11.$$

Randomization: Some useful doubly transitive groups

A permutation group is doubly transitive if whenever (α, β) and (γ, δ) are pairs of distinct elements then the group contains a permutation taking α to γ and taking β to δ.

First suppose that $t = 11$.
Renumber the rows and the columns as $0, 1, \ldots, 10$ modulo 11.
One such group consists of the permutations

$$x \mapsto ax + b \bmod 11 \qquad \text{where } a \neq 0 \bmod 11.$$

For example, when $a = 2$ and $b = 5$ this gives the following permutation.

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 7 & 9 & 0 & 2 & 4 & 6 & 8 & 10 & 1 & 3 \end{pmatrix}$$

Randomization: Some useful doubly transitive groups

A permutation group is doubly transitive if whenever (α, β) and (γ, δ) are pairs of distinct elements then the group contains a permutation taking α to γ and taking β to δ.

First suppose that $t = 11$.
Renumber the rows and the columns as $0, 1, \ldots, 10$ modulo 11.
One such group consists of the permutations

$$x \mapsto ax + b \bmod 11 \qquad \text{where } a \neq 0 \bmod 11.$$

For example, when $a = 2$ and $b = 5$ this gives the following permutation.

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 7 & 9 & 0 & 2 & 4 & 6 & 8 & 10 & 1 & 3 \end{pmatrix}$$

This method works whenever t is a prime number.

Randomization: Some useful doubly transitive groups

A permutation group is doubly transitive if whenever (α, β) and (γ, δ) are pairs of distinct elements then the group contains a permutation taking α to γ and taking β to δ.

First suppose that $t = 11$.
Renumber the rows and the columns as 0, 1, …, 10 modulo 11.
One such group consists of the permutations

$$x \mapsto ax + b \bmod 11 \qquad \text{where } a \neq 0 \bmod 11.$$

For example, when $a = 2$ and $b = 5$ this gives the following permutation.

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 7 & 9 & 0 & 2 & 4 & 6 & 8 & 10 & 1 & 3 \end{pmatrix}$$

This method works whenever t is a prime number.
A slightly more complicated version works when t is a power of a prime number.

Randomization: More useful doubly transitive groups

If $t-1$ is a prime p (or a power of a prime number), then this slightly more complicated method works.
Relabel the rows and the columns as $0, 1, \ldots, p-1$ modulo p, followed by ∞.

Randomization: More useful doubly transitive groups

If $t-1$ is a prime p (or a power of a prime number), then this slightly more complicated method works.
Relabel the rows and the columns as $0, 1, \ldots, p-1$ modulo p, followed by ∞.
Now we use permutations of the form

$$x \mapsto \frac{ax+b}{cx+d} \bmod p \qquad \text{where } ad-bc = 1 \bmod p.$$

Randomization: More useful doubly transitive groups

If $t-1$ is a prime p (or a power of a prime number), then this slightly more complicated method works.
Relabel the rows and the columns as $0, 1, \ldots, p-1$ modulo p, followed by ∞.
Now we use permutations of the form

$$x \mapsto \frac{ax+b}{cx+d} \bmod p \qquad \text{where } ad-bc = 1 \bmod p.$$

(Here, $\infty + 1 = \infty$ and $1/\infty = 0$.)

Randomization: More useful doubly transitive groups

If $t-1$ is a prime p (or a power of a prime number), then this slightly more complicated method works.
Relabel the rows and the columns as $0, 1, \ldots, p-1$ modulo p, followed by ∞.
Now we use permutations of the form

$$x \mapsto \frac{ax+b}{cx+d} \bmod p \qquad \text{where } ad - bc = 1 \bmod p.$$

(Here, $\infty + 1 = \infty$ and $1/\infty = 0$.)

How do we find such groups?

Randomization: More useful doubly transitive groups

If $t-1$ is a prime p (or a power of a prime number), then this slightly more complicated method works.
Relabel the rows and the columns as $0, 1, \ldots, p-1$ modulo p, followed by ∞.
Now we use permutations of the form

$$x \mapsto \frac{ax+b}{cx+d} \bmod p \qquad \text{where } ad-bc = 1 \bmod p.$$

(Here, $\infty + 1 = \infty$ and $1/\infty = 0$.)

How do we find such groups?

- Ask a group theorist in a nearby Pure Maths department.
- Use the software GAP.

Chapter 4

Optimality

What do we seek to optimize?

Denote by A_{cc} the average variance of the estimate of the difference between the effects of control treatments.

What do we seek to optimize?

Denote by A_{cc} the average variance of the estimate of the difference between the effects of control treatments.
This is equal to $2/t$, because contrasts between control treatments are orthogonal to rows and to columns,
and each control treatment has replication t.

What do we seek to optimize?

Denote by A_{cc} the average variance of the estimate of the difference between the effects of control treatments.
This is equal to $2/t$, because contrasts between control treatments are orthogonal to rows and to columns, and each control treatment has replication t.

Denote by A_{tt} the average variance of the estimate of the difference between the effects of test treatments,

What do we seek to optimize?

Denote by A_{cc} the average variance of the estimate of the difference between the effects of control treatments.
This is equal to $2/t$, because contrasts between control treatments are orthogonal to rows and to columns, and each control treatment has replication t.

Denote by A_{tt} the average variance of the estimate of the difference between the effects of test treatments,
by A_{ct} the average variance of the estimate of the difference between the effect of a test treatment and the effect of a control treatment,

What do we seek to optimize?

Denote by A_{cc} the average variance of the estimate of the difference between the effects of control treatments.
This is equal to $2/t$, because contrasts between control treatments are orthogonal to rows and to columns,
and each control treatment has replication t.

Denote by A_{tt} the average variance of the estimate of the difference between the effects of test treatments,
by A_{ct} the average variance of the estimate of the difference between the effect of a test treatment and the effect of a control treatment,
and by A_{abd} the average variance of the estimate of the difference between treatments in the auxiliary block design when it is used in its usual setting.

What do we seek to optimize?

Denote by A_{cc} the average variance of the estimate of the difference between the effects of control treatments.
This is equal to $2/t$, because contrasts between control treatments are orthogonal to rows and to columns, and each control treatment has replication t.

Denote by A_{tt} the average variance of the estimate of the difference between the effects of test treatments,
by A_{ct} the average variance of the estimate of the difference between the effect of a test treatment and the effect of a control treatment,
and by A_{abd} the average variance of the estimate of the difference between treatments in the auxiliary block design when it is used in its usual setting.

We seek designs which minimize both of A_{tt} and A_{ct}.

Results on optimality

We were able to show that

$$A_{\mathrm{ct}} = \frac{k-1}{kt} + \frac{1}{k(t-k)} + \frac{t(t-k)-1}{2t(t-k)} A_{\mathrm{tt}}.$$

Results on optimality

We were able to show that

$$A_{\mathrm{ct}} = \frac{k-1}{kt} + \frac{1}{k(t-k)} + \frac{t(t-k)-1}{2t(t-k)} A_{\mathrm{tt}}.$$

Thus A_{tt} and A_{ct} are positive linear functions of each other, so there is no conflict between optimizing the design on either criterion.

Results on optimality

We were able to show that

$$A_{\text{ct}} = \frac{k-1}{kt} + \frac{1}{k(t-k)} + \frac{t(t-k)-1}{2t(t-k)} A_{\text{tt}}.$$

Thus A_{tt} and A_{ct} are positive linear functions of each other, so there is no conflict between optimizing the design on either criterion.

Then Williams and Piepho (2025) told us that they had proved that A_{tt} is a positive linear function of A_{abd}.

Results on optimality

We were able to show that

$$A_{\mathrm{ct}} = \frac{k-1}{kt} + \frac{1}{k(t-k)} + \frac{t(t-k)-1}{2t(t-k)} A_{\mathrm{tt}}.$$

Thus A_{tt} and A_{ct} are positive linear functions of each other, so there is no conflict between optimizing the design on either criterion.

Then Williams and Piepho (2025) told us that they had proved that A_{tt} is a positive linear function of A_{abd}.

Thus the best strategy is to use an auxiliary block design which is A-optimal in the sense that it minimizes the value of A_{abd} over all designs with the given values of t and k.

Results on optimality

We were able to show that

$$A_{\mathrm{ct}} = \frac{k-1}{kt} + \frac{1}{k(t-k)} + \frac{t(t-k)-1}{2t(t-k)} A_{\mathrm{tt}}.$$

Thus A_{tt} and A_{ct} are positive linear functions of each other, so there is no conflict between optimizing the design on either criterion.

Then Williams and Piepho (2025) told us that they had proved that A_{tt} is a positive linear function of A_{abd}.

Thus the best strategy is to use an auxiliary block design which is A-optimal in the sense that it minimizes the value of A_{abd} over all designs with the given values of t and k.

Cheng and Bailey (1991) showed that square lattice designs are A-optimal.

Results on optimality

We were able to show that

$$A_{\text{ct}} = \frac{k-1}{kt} + \frac{1}{k(t-k)} + \frac{t(t-k)-1}{2t(t-k)} A_{\text{tt}}.$$

Thus A_{tt} and A_{ct} are positive linear functions of each other, so there is no conflict between optimizing the design on either criterion.

Then Williams and Piepho (2025) told us that they had proved that A_{tt} is a positive linear function of A_{abd}.

Thus the best strategy is to use an auxiliary block design which is A-optimal in the sense that it minimizes the value of A_{abd} over all designs with the given values of t and k.

Cheng and Bailey (1991) showed that square lattice designs are A-optimal.

The design with $t = 16$ and $k = 4$ used as the auxiliary block design in previous slides is a square lattice design,

Results on optimality

We were able to show that

$$A_{\mathrm{ct}} = \frac{k-1}{kt} + \frac{1}{k(t-k)} + \frac{t(t-k)-1}{2t(t-k)} A_{\mathrm{tt}}.$$

Thus A_{tt} and A_{ct} are positive linear functions of each other, so there is no conflict between optimizing the design on either criterion.

Then Williams and Piepho (2025) told us that they had proved that A_{tt} is a positive linear function of A_{abd}.

Thus the best strategy is to use an auxiliary block design which is A-optimal in the sense that it minimizes the value of A_{abd} over all designs with the given values of t and k.

Cheng and Bailey (1991) showed that square lattice designs are A-optimal.

The design with $t = 16$ and $k = 4$ used as the auxiliary block design in previous slides is a square lattice design, so the square array design that I constructed from it is A-optimal.

References: Background

- Kempton, R. A. (1984)
  The design and analysis of unreplicated trials.
  Vorträge für Pflanzenzüchtung **7**, 219–242.
- Federer, W. T. and D. Raghavarao (1975)
  On augmented designs. *Biometrics* **31**, 29–35.
- Federer, W. T., R. C. Nair, and D. Raghavarao (1975)
  Some augmented row–column designs.
  Biometrics **31**, 361–373.

References: Construction, Randomization and Optimality

- Hall, P. (1935)
  On representations of subsets.
  Journal of the London Mathematical Society **10**, 26–30.
- Grundy, P. M. and M. J. R. Healy (1950)
  Restricted randomization and Quasi-Latin squares.
  Journal of the Royal Statistical Society Series B **12**, 286–291.
- Williams E. R. and H. P. Piepho (2025)
  A note on the construction of augmented designs in square arrays. *arXiv: 2501.08448v3*.
- Cheng, C.-S. and R. A. Bailey (1991)
  Optimality of some two-associate-class partially balanced incomplete-block designs.
  Annals of Statistics **19**, 1667–1671.

References: Our paper

- Bailey, R. A. and L. M. Haines (2025)
  Square array designs for unreplicated test treatments with replicated controls.
  Journal of Agricultural, Biological and Environmental Statistics.
  Published online on 22 June 2025,
  but not yet assigned to an issue.