Semi-Latin Squares and their Extensions

R. A. Bailey
University of St Andrews

Model-Oriented Data Analysis and Optimum Design,
Drongen Abbey, Belgium, 18 June 2026

What is a semi-Latin Square?

The plot structure of a semi-Latin square has the form

$$((n \text{ rows }) \times (n \text{ columns }))/(k \text{ plots }),$$

using the notation of Nelder (1965).

What is a semi-Latin Square?

The plot structure of a semi-Latin square has the form

$$((n \text{ rows }) \times (n \text{ columns }))/(k \text{ plots }),$$

using the notation of Nelder (1965).
Call the intersection of a row and column a cell.
Thus there are n^2 cells, arranged in a $n \times n$ square array.

What is a semi-Latin Square?

The plot structure of a semi-Latin square has the form

$$((n \text{ rows }) \times (n \text{ columns }))/(k \text{ plots }),$$

using the notation of Nelder (1965).
Call the intersection of a row and column a cell.
Thus there are n^2 cells, arranged in a $n \times n$ square array.
Each cell contains k experimental units, which I always call plots.

What is a semi-Latin Square?

The plot structure of a semi-Latin square has the form

$$((n \text{ rows}) \times (n \text{ columns}))/(k \text{ plots}),$$

using the notation of Nelder (1965).
Call the intersection of a row and column a cell.
Thus there are n^2 cells, arranged in a $n \times n$ square array.
Each cell contains k experimental units, which I always call plots.

There are nk treatments, each applied to one plot in each row and one plot in each column.

What is a semi-Latin Square?

The plot structure of a semi-Latin square has the form

$$((n \text{ rows }) \times (n \text{ columns }))/(k \text{ plots }),$$

using the notation of Nelder (1965).
Call the intersection of a row and column a cell.
Thus there are n^2 cells, arranged in a $n \times n$ square array.
Each cell contains k experimental units, which I always call plots.

There are nk treatments, each applied to one plot in each row and one plot in each column.

Each treatment has replication n,
so we always assume that $n > 1$.

What is a semi-Latin Square?

The plot structure of a semi-Latin square has the form

$$((n \text{ rows }) \times (n \text{ columns }))/(k \text{ plots }),$$

using the notation of Nelder (1965).
Call the intersection of a row and column a cell.
Thus there are n^2 cells, arranged in a $n \times n$ square array.
Each cell contains k experimental units, which I always call plots.

There are nk treatments, each applied to one plot in each row and one plot in each column.

Each treatment has replication n,
so we always assume that $n > 1$.

In the case that $k = 1$, we just get a Latin square of order n,
so we always assume that $k > 1$.

An example of a semi-Latin square

$A\ \alpha\ a$	$B\ \beta\ b$	$C\ \gamma\ c$	$D\ \delta\ d$	$E\ \varepsilon\ e$
$E\ \delta\ c$	$A\ \varepsilon\ d$	$B\ \alpha\ e$	$C\ \beta\ a$	$D\ \gamma\ b$
$D\ \beta\ e$	$E\ \gamma\ a$	$A\ \delta\ b$	$B\ \varepsilon\ c$	$C\ \alpha\ d$
$C\ \varepsilon\ b$	$D\ \alpha\ c$	$E\ \beta\ d$	$A\ \gamma\ e$	$B\ \delta\ a$
$B\ \gamma\ d$	$C\ \delta\ e$	$D\ \varepsilon\ a$	$E\ \alpha\ b$	$A\ \beta\ c$

An example of a semi-Latin square

$A \quad \alpha \quad a$	$B \quad \beta \quad b$	$C \quad \gamma \quad c$	$D \quad \delta \quad d$	$E \quad \varepsilon \quad e$
$E \quad \delta \quad c$	$A \quad \varepsilon \quad d$	$B \quad \alpha \quad e$	$C \quad \beta \quad a$	$D \quad \gamma \quad b$
$D \quad \beta \quad e$	$E \quad \gamma \quad a$	$A \quad \delta \quad b$	$B \quad \varepsilon \quad c$	$C \quad \alpha \quad d$
$C \quad \varepsilon \quad b$	$D \quad \alpha \quad c$	$E \quad \beta \quad d$	$A \quad \gamma \quad e$	$B \quad \delta \quad a$
$B \quad \gamma \quad d$	$C \quad \delta \quad e$	$D \quad \varepsilon \quad a$	$E \quad \alpha \quad b$	$A \quad \beta \quad c$

Here there are 5 rows, 5 columns, and 3 plots per cell.

An example of a semi-Latin square

A α a	B β b	C γ c	D δ d	E ε e
E δ c	A ε d	B α e	C β a	D γ b
D β e	E γ a	A δ b	B ε c	C α d
C ε b	D α c	E β d	A γ e	B δ a
B γ d	C δ e	D ε a	E α b	A β c

Here there are 5 rows, 5 columns, and 3 plots per cell.

There are 15 treatments,
each occurring once in each row and once in each column.

Data analysis

How should we analyse data from such an experiment?

Data analysis

How should we analyse data from such an experiment?
Obviously, we need to include a term for rows and another one for columns.

Data analysis

How should we analyse data from such an experiment?
Obviously, we need to include a term for rows and another one for columns.

In a paper read to the Royal Statistical Society in 1935, Frank Yates was scathing about the idea that cells should be ignored.

Data analysis

How should we analyse data from such an experiment? Obviously, we need to include a term for rows and another one for columns.

In a paper read to the Royal Statistical Society in 1935, Frank Yates was scathing about the idea that cells should be ignored.

Of course, because cells are included in the design process, they should also be included in the data analysis.

Data analysis

How should we analyse data from such an experiment?
Obviously, we need to include a term for rows and another one for columns.

In a paper read to the Royal Statistical Society in 1935, Frank Yates was scathing about the idea that cells should be ignored.

Of course, because cells are included in the design process, they should also be included in the data analysis.

They effectively form a system of incomplete blocks,
so the design needs to take account of this.

Data analysis

How should we analyse data from such an experiment? Obviously, we need to include a term for rows and another one for columns.

In a paper read to the Royal Statistical Society in 1935, Frank Yates was scathing about the idea that cells should be ignored.

Of course, because cells are included in the design process, they should also be included in the data analysis.

They effectively form a system of incomplete blocks,
so the design needs to take account of this.

Once people started to include the cells in the data analysis, Frank Yates became enthusiastic about using semi-Latin squares.

Randomization

How should such a design be randomized?

Randomization

How should such a design be randomized?

1. Randomize the order of the whole rows.
2. Randomize the order of the whole columns.
3. Within each cell independently, randomize the order of the plots.

Randomization

How should such a design be randomized?

1. Randomize the order of the whole rows.
2. Randomize the order of the whole columns.
3. Within each cell independently, randomize the order of the plots.

This randomization justifies the linear model whose variance-covariance matrix has the following eigenspaces (also called strata).

Subspace	dimension
Mean	1
Rows	$n-1$
Columns	$n-1$
Cells within Rows and Columns	$(n-1)^2$
Plots within Cells	$n^2(k-1)$

General balance

Treatment contrasts are orthogonal to the first three strata,
so any treatment contrast that is an eigenvector of
the matrix of projection onto one of the last two strata
is also an eigenvector of the matrix of projection onto the other.

General balance

Treatment contrasts are orthogonal to the first three strata,
so any treatment contrast that is an eigenvector of
the matrix of projection onto one of the last two strata
is also an eigenvector of the matrix of projection onto the other.

This property was called general balance by Nelder (1965).

General balance

Treatment contrasts are orthogonal to the first three strata, so any treatment contrast that is an eigenvector of the matrix of projection onto one of the last two strata is also an eigenvector of the matrix of projection onto the other.

This property was called general balance by Nelder (1965).

General balance is very useful if we want to combine treatment estimates using information from different strata.

Trojan squares

The design is called a Trojan square if it can be made by superposing the letters of k mutually orthogonal Latin squares, each with n different treatments.

Trojan squares

The design is called a Trojan square if it can be made by superposing the letters of k mutually orthogonal Latin squares, each with n different treatments.

Our previous example with $n = 5$ and $k = 3$ is Trojan.

$A\ \alpha\ a$	$B\ \beta\ b$	$C\ \gamma\ c$	$D\ \delta\ d$	$E\ \varepsilon\ e$
$E\ \delta\ c$	$A\ \varepsilon\ d$	$B\ \alpha\ e$	$C\ \beta\ a$	$D\ \gamma\ b$
$D\ \beta\ e$	$E\ \gamma\ a$	$A\ \delta\ b$	$B\ \varepsilon\ c$	$C\ \alpha\ d$
$C\ \varepsilon\ b$	$D\ \alpha\ c$	$E\ \beta\ d$	$A\ \gamma\ e$	$B\ \delta\ a$
$B\ \gamma\ d$	$C\ \delta\ e$	$D\ \varepsilon\ a$	$E\ \alpha\ b$	$A\ \beta\ c$

A-optimality

The design is A-optimal if it minimizes the average variance of pairwise treatment comparisons in the Plots-in-Cells stratum.

A-optimality

The design is A-optimal if it minimizes the average variance of pairwise treatment comparisons in the Plots-in-Cells stratum.

Theorem

If there is a Trojan square for the given values of n and k, then it is A-optimal.

A-optimality

The design is A-optimal if it minimizes the average variance of pairwise treatment comparisons in the Plots-in-Cells stratum.

Theorem

If there is a Trojan square for the given values of n and k, then it is A-optimal.

An s-fold inflation of a semi-Latin $(n \times n)/k$ square is a semi-Latin $(n \times n)/ks$ square obtained by replacing each treatment in the first semi-Latin square by s new treatments.

A-optimality

The design is A-optimal if it minimizes the average variance of pairwise treatment comparisons in the Plots-in-Cells stratum.

Theorem

If there is a Trojan square for the given values of n and k, then it is A-optimal.

An s-fold inflation of a semi-Latin $(n \times n)/k$ square is a semi-Latin $(n \times n)/ks$ square obtained by replacing each treatment in the first semi-Latin square by s new treatments.

Theorem

If there is a Trojan square of size $(n \times n)/(n-1)$, then all of its inflations are A-optimal.

What to do when $n = 6$?

Of course, when $n = 6$ there is no pair of mutually orthogonal Latin squares.

What to do when $n = 6$?

Of course, when $n = 6$ there is no pair of mutually orthogonal Latin squares.

RAB and Gordon Royle (1997) investigated $(6 \times 6)/2$ semi-Latin squares by computer search, and recorded the A-optimal one, as well as those optimal on some other criteria.

What to do when $n = 6$?

Of course, when $n = 6$ there is no pair of mutually orthogonal Latin squares.

RAB and Gordon Royle (1997) investigated $(6 \times 6)/2$ semi-Latin squares by computer search, and recorded the A-optimal one, as well as those optimal on some other criteria.

Leonard Soicher (2012) defined a semi-Latin square to be uniform if every pair of cells which are not in the same row or same column have the same number of treatments in common.

What to do when $n = 6$?

Of course, when $n = 6$ there is no pair of mutually orthogonal Latin squares.

RAB and Gordon Royle (1997) investigated $(6 \times 6)/2$ semi-Latin squares by computer search, and recorded the A-optimal one, as well as those optimal on some other criteria.

Leonard Soicher (2012) defined a semi-Latin square to be uniform if every pair of cells which are not in the same row or same column have the same number of treatments in common. For this to be possible, k must be a multiple of $n - 1$.

What to do when $n = 6$?

Of course, when $n = 6$ there is no pair of mutually orthogonal Latin squares.

RAB and Gordon Royle (1997) investigated $(6 \times 6)/2$ semi-Latin squares by computer search, and recorded the A-optimal one, as well as those optimal on some other criteria.

Leonard Soicher (2012) defined a semi-Latin square to be uniform if every pair of cells which are not in the same row or same column have the same number of treatments in common. For this to be possible, k must be a multiple of $n - 1$.

Soicher proved that there exists a $(6 \times 6)/5s$ uniform semi-Latin square whenever $s \geq 2$.

Extended semi-Latin squares: the background

In trials on glasshouse crops, there are usually pronounced row effects and column effects.

Extended semi-Latin squares: the background

In trials on glasshouse crops, there are usually pronounced row effects and column effects.
The glasshouse is typically rectangular, with its long axis East–West.

Extended semi-Latin squares: the background

In trials on glasshouse crops, there are usually pronounced row effects and column effects.
The glasshouse is typically rectangular, with its long axis East–West.
Because North–South variation is usually greater per distance than East–West variation, statisticians in agricultural research in the UK decided in the 1980s that it is reasonable to divide the glasshouse into n rows and n columns,
where each row is an East–West line of nk plots and
each column consists of k contiguous North–South lines of plots.

Extended semi-Latin squares: the background

In trials on glasshouse crops, there are usually pronounced row effects and column effects.
The glasshouse is typically rectangular, with its long axis East–West.
Because North–South variation is usually greater per distance than East–West variation, statisticians in agricultural research in the UK decided in the 1980s that it is reasonable to divide the glasshouse into n rows and n columns,
where each row is an East–West line of nk plots and
each column consists of k contiguous North–South lines of plots.

This view led to widespread use of semi-Latin squares in agricultural research.

Extended semi-Latin squares: definition

The plot structure of an extended semi-Latin square has the form

$$(n \text{ rows }) \times ((n \text{ columns }) / (k \text{ lines })),$$

using the notation of Nelder (1965).

Extended semi-Latin squares: definition

The plot structure of an extended semi-Latin square has the form

$$(n \text{ rows }) \times ((n \text{ columns }) / (k \text{ lines })),$$

using the notation of Nelder (1965).
Call the intersection of a row and column a cell.

Extended semi-Latin squares: definition

The plot structure of an extended semi-Latin square has the form

$$(n \text{ rows }) \times ((n \text{ columns }) / (k \text{ lines })),$$

using the notation of Nelder (1965).
Call the intersection of a row and column a cell.
Now each of the n columns is split into k lines, each of which intersects each row in a single plot.

Extended semi-Latin squares: definition

The plot structure of an extended semi-Latin square has the form

$$(n \text{ rows }) \times ((n \text{ columns })/(k \text{ lines })),$$

using the notation of Nelder (1965).
Call the intersection of a row and column a cell.
Now each of the n columns is split into k lines, each of which intersects each row in a single plot.

There are nk treatments, each applied to one plot in each row and one plot in each column.

Extended semi-Latin squares: definition

The plot structure of an extended semi-Latin square has the form

$$(n \text{ rows }) \times ((n \text{ columns })/(k \text{ lines })),$$

using the notation of Nelder (1965).

Call the intersection of a row and column a cell.

Now each of the n columns is split into k lines, each of which intersects each row in a single plot.

There are nk treatments, each applied to one plot in each row and one plot in each column.

Now the n^2 cells of size k and the nk lines of size n both form systems of incomplete-blocks, so we have to allow for both of them in our design, randomization and data analysis.

Structure

Here is an example with $n = 4$ and $k = 2$.

Structure

Here is an example with $n = 4$ and $k = 2$.

$*$	$*$	$+$	$+$	†	†	£	£

There are 4 rows, each containing 4 cells of 2 plots.

Structure

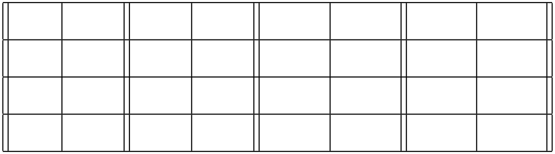
Here is an example with $n = 4$ and $k = 2$.

$*$	$+$						
$*$	$+$						
$*$	$+$						
$*$	$+$						

There are 4 rows, each containing 4 cells of 2 plots.
There are 4 columns, each containing 2 lines of 4 plots.

Structure

Here is an example with $n = 4$ and $k = 2$.



There are 4 rows, each containing 4 cells of 2 plots.
There are 4 columns, each containing 2 lines of 4 plots.
Within each column, cells are crossed with lines.

Randomization

How should such a design be randomized?

Randomization

How should such a design be randomized?

1. Randomize the order of the whole rows.
2. Randomize the order of the whole columns.
3. Within each whole column independently, randomize the order of the lines.

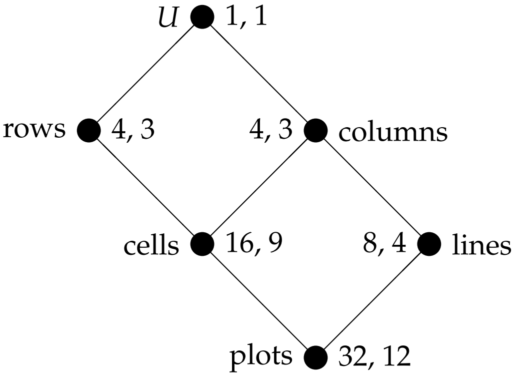
How should such a design be randomized?

1. Randomize the order of the whole rows.
2. Randomize the order of the whole columns.
3. Within each whole column independently, randomize the order of the lines.

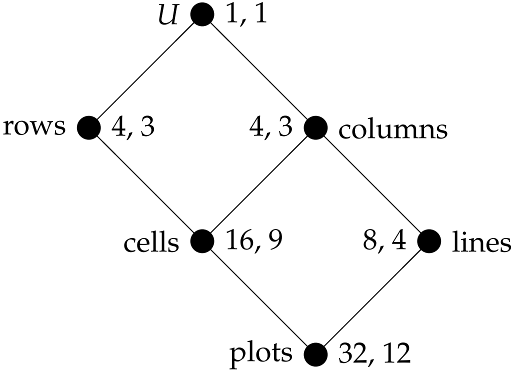
This randomization justifies the linear model whose variance-covariances matrix has the following eigenspaces (also called strata).

Subspace	dimension
Mean	1
Rows	$n-1$
Columns	$n-1$
Lines within Columns	$n(k-1)$
Cells within Rows and Columns	$(n-1)^2$
Plots within Cells and Lines	$n(n-1)(k-1)$

Hasse diagram of block factors in the example



Hasse diagram of block factors in the example



Here, U is the universal factor, which has a single level.
For each factor, the first number shows its number of levels and the second number shows its number of degrees of freedom.

How can we find efficient extended semi-Latin squares?

How do we find the optimal way of allocating treatments to plots?

How can we find efficient extended semi-Latin squares?

How do we find the optimal way of allocating treatments to plots?

This is joint work with Emlyn Williams.

How can we find efficient extended semi-Latin squares?

How do we find the optimal way of allocating treatments to plots?

This is joint work with Emlyn Williams.

His approach is to use his CycDesigN software to find an optimal design.

How can we find efficient extended semi-Latin squares?

How do we find the optimal way of allocating treatments to plots?

This is joint work with Emlyn Williams.

His approach is to use his CycDesigN software to find an optimal design. This focusses on the information in the plots stratum, and usually does not produce general balance.

How can we find efficient extended semi-Latin squares?

How do we find the optimal way of allocating treatments to plots?

This is joint work with Emlyn Williams.

His approach is to use his CycDesigN software to find an optimal design. This focusses on the information in the plots stratum, and usually does not produce general balance.

RAB's approach is to use patterns and combinatorics.
Since more information is lost to the block system with the smaller block size, it is better to start by finding the best design for that block system.

An example of my approach

Suppose that $n = 4$ and $k = 2$.

An example of my approach

Suppose that $n = 4$ and $k = 2$. Considering the cells of size 2, we know that the optimal design is a Trojan square.

An example of my approach

Suppose that $n = 4$ and $k = 2$. Considering the cells of size 2, we know that the optimal design is a Trojan square.

6	2	8	4	1	7	5	3
8	1	6	3	5	2	7	4
7	3	5	1	6	4	2	8
5	4	7	2	3	8	1	6

An example of my approach

Suppose that $n = 4$ and $k = 2$. Considering the cells of size 2, we know that the optimal design is a Trojan square.

6 2	8 4	1 7	5 3
8 1	6 3	5 2	7 4
7 3	5 1	6 4	2 8
5 4	7 2	3 8	1 6

The black letters make one Latin square; the red letters make another Latin square orthogonal to the first.

An example of my approach

Suppose that $n = 4$ and $k = 2$. Considering the cells of size 2, we know that the optimal design is a Trojan square.

6 2	8 4	1 7	5 3
8 1	6 3	5 2	7 4
7 3	5 1	6 4	2 8
5 4	7 2	3 8	1 6

The black letters make one Latin square; the red letters make another Latin square orthogonal to the first.
In the first two columns, use the colours to define the lines.

An example of my approach

Suppose that $n = 4$ and $k = 2$. Considering the cells of size 2, we know that the optimal design is a Trojan square.

6	2	8	4	1	7	5	3
8	1	6	3	5	2	7	4
7	3	5	1	6	4	2	8
5	4	7	2	3	8	1	6

The black letters make one Latin square; the red letters make another Latin square orthogonal to the first.
In the first two columns, use the colours to define the lines.
In the last two columns, split the colours equally between the two lines, using a pair of orthogonal partitions of each colour into two parts of size 2.

An example of my approach

Suppose that $n = 4$ and $k = 2$. Considering the cells of size 2, we know that the optimal design is a Trojan square.

6	2	8	4	1	7	5	3
8	1	6	3	5	2	7	4
7	3	5	1	6	4	2	8
5	4	7	2	3	8	1	6

The black letters make one Latin square; the red letters make another Latin square orthogonal to the first.
In the first two columns, use the colours to define the lines.
In the last two columns, split the colours equally between the two lines, using a pair of orthogonal partitions of each colour into two parts of size 2.

In fact, Emlyn's software shows that this design is A-optimal.

Our paper

E. R. Williams and R. A. Bailey:
Extended semi-Latin squares for use in field and glasshouse trials.
Journal of Agricultural, Biological and Environmental Statistics, (2026).
doi: `10.1007/s13253-026-00729-6`