Optimal design of experiments with very low average replication

R. A. Bailey

Queen Mary University of London

r.a.bailey@qmul.ac.uk

LinStat, July 2012

Motivation

Poster at LinStat 2010, Tomar, Portugal, July 2010:
'Unreplicated experiments in early stage breeding programs'
by Katarzyna Marczyńska and Stanisław Mejza.

Motivation

Poster at LinStat 2010, Tomar, Portugal, July 2010:
'Unreplicated experiments in early stage breeding programs'
by Katarzyna Marczyńska and Stanisław Mejza.

Single replication of new treatments; many check plots.
How to use the information from the check plots?

Motivation

Poster at LinStat 2010, Tomar, Portugal, July 2010:
'Unreplicated experiments in early stage breeding programs'
by Katarzyna Marczyńska and Stanisław Mejza.

Single replication of new treatments; many check plots.
How to use the information from the check plots?

Question from Brian Cullis and David Butler, Adelaide, Australia, November 2010:
If we are not interested in control treatments, we know that we can get more information by replacing the check plots by second replicates of some treatments—can you prove this?

Abstract

Trials of new crop varieties usually have very low average replication.

Abstract

Trials of new crop varieties usually have very low average replication. Thus one possiblity is to have a single plot for each new variety and several plots for a control variety, with the latter well spread out over the field.

Abstract

Trials of new crop varieties usually have very low average replication. Thus one possiblity is to have a single plot for each new variety and several plots for a control variety, with the latter well spread out over the field. A more recent proposal is to ignore the control, and instead have two plots for each of a small proportion of the new varieties.

Abstract

Trials of new crop varieties usually have very low average replication. Thus one possiblity is to have a single plot for each new variety and several plots for a control variety, with the latter well spread out over the field. A more recent proposal is to ignore the control, and instead have two plots for each of a small proportion of the new varieties.

Variation in the field may be accounted for by a polynomial trend, by spatial correlation, or by blocking.

Abstract

Trials of new crop varieties usually have very low average replication. Thus one possiblity is to have a single plot for each new variety and several plots for a control variety, with the latter well spread out over the field. A more recent proposal is to ignore the control, and instead have two plots for each of a small proportion of the new varieties.

Variation in the field may be accounted for by a polynomial trend, by spatial correlation, or by blocking. However, if the experiment has a second phase, such as making bread from flour milled from the grain produced in the first phase, then that second phase usually has blocks.

Abstract

Trials of new crop varieties usually have very low average replication. Thus one possiblity is to have a single plot for each new variety and several plots for a control variety, with the latter well spread out over the field. A more recent proposal is to ignore the control, and instead have two plots for each of a small proportion of the new varieties.

Variation in the field may be accounted for by a polynomial trend, by spatial correlation, or by blocking. However, if the experiment has a second phase, such as making bread from flour milled from the grain produced in the first phase, then that second phase usually has blocks.

The optimality criterion used is usually the A criterion: the average variance of the pairwise differences between the new varieties.

Abstract

Trials of new crop varieties usually have very low average replication. Thus one possiblity is to have a single plot for each new variety and several plots for a control variety, with the latter well spread out over the field. A more recent proposal is to ignore the control, and instead have two plots for each of a small proportion of the new varieties.

Variation in the field may be accounted for by a polynomial trend, by spatial correlation, or by blocking. However, if the experiment has a second phase, such as making bread from flour milled from the grain produced in the first phase, then that second phase usually has blocks.

The optimality criterion used is usually the A criterion: the average variance of the pairwise differences between the new varieties.

I shall compare designs under the A criterion when the average replication is much less than two.

Agricultural plant-breeding trials

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Example

There are 224 new varieties, with very little seed of each.

Agricultural plant-breeding trials

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Example

There are 224 new varieties, with very little seed of each.
There are 280 plots available, in a 14×20 rectangle.

Agricultural plant-breeding trials

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Example

There are 224 new varieties, with very little seed of each.
There are 280 plots available, in a 14×20 rectangle.
How do you design the experiment?

Agricultural plant-breeding trials

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Example

There are 224 new varieties, with very little seed of each.
There are 280 plots available, in a 14×20 rectangle.
How do you design the experiment?

Assume that

$$\text{number of varieties} < \text{number of plots}$$

and

$$\text{number of plots} << 2 \times (\text{number of varieties}).$$

Some notation

$$f(\omega) = \text{variety on plot } \omega.$$

Some notation

$$f(\omega) = \text{variety on plot } \omega.$$

$$Y_\omega = \text{response on plot } \omega.$$

Some notation

$$f(\omega) = \text{variety on plot } \omega.$$

$$Y_\omega = \text{response on plot } \omega.$$

$$\tau_i = \text{effect of variety } i.$$

Some notation

$$f(\omega) = \text{variety on plot } \omega.$$

$$Y_\omega = \text{response on plot } \omega.$$

$$\tau_i = \text{effect of variety } i.$$

Assume that

$$Y_\omega = \tau_{f(\omega)} + \text{stuff depending on plots.}$$

Some notation

$$f(\omega) = \text{variety on plot } \omega.$$

$$Y_\omega = \text{response on plot } \omega.$$

$$\tau_i = \text{effect of variety } i.$$

Assume that

$$Y_\omega = \tau_{f(\omega)} + \text{stuff depending on plots.}$$

We want to minimize

$$\sum_i \sum_{j \neq i} \text{Var}(\hat{\tau}_i - \hat{\tau}_j).$$

Simplest model

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \quad \text{if } \omega \neq \omega'.$$

Simplest model

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \quad \text{if } \omega \neq \omega'.$$

The A-optimal design has
2 plots for some varieties and 1 plot for all other varieties,
and is completely randomized.

Simplest model: example

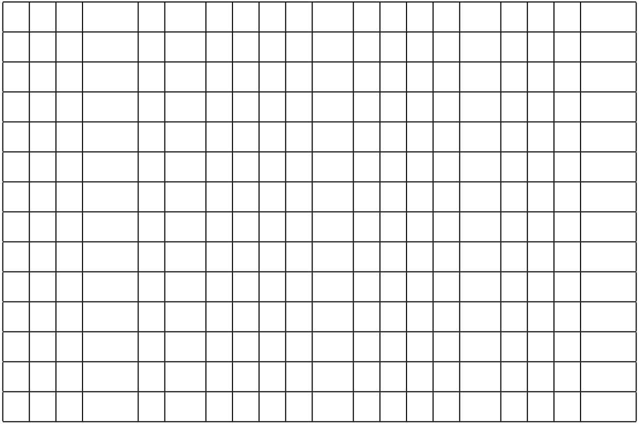
56 varieties have replication 2;

168 varieties have replication 1.

Simplest model: example

56 varieties have replication 2;

168 varieties have replication 1.



Simplest model: example

56 varieties have replication 2;

168 varieties have replication 1.

															1				
			2																
					1														
										2									

Simplest model: example

56 varieties have replication 2;

168 varieties have replication 1.

															1				
			2																
																			57
					1														
			58																
									2										

A breeder says . . .

Unfair!

The single plot with my variety was in an infertile part of the field.

Fixed spatial trend

$$Y_\omega = \tau_{f(\omega)} + g(\omega) + \epsilon_\omega$$

where

$$g(\omega) \text{ is a two-dimensional low-degree polynomial in } \omega,$$

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \text{ if } \omega \neq \omega'.$$

Fixed spatial trend

$$Y_\omega = \tau_{f(\omega)} + g(\omega) + \epsilon_\omega$$

where

$g(\omega)$ is a two-dimensional low-degree polynomial in ω,

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \text{ if } \omega \neq \omega'.$$

Caliński, Mejza, Marczyńska . . . :

Fixed spatial trend

$$Y_\omega = \tau_{f(\omega)} + g(\omega) + \epsilon_\omega$$

where

$g(\omega)$ is a two-dimensional low-degree polynomial in ω,

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \text{ if } \omega \neq \omega'.$$

Caliński, Mejza, Marczyńska . . . :
use one plot for each new variety

Fixed spatial trend

$$Y_\omega = \tau_{f(\omega)} + g(\omega) + \epsilon_\omega$$

where

$g(\omega)$ is a two-dimensional low-degree polynomial in ω,

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \text{ if } \omega \neq \omega'.$$

Caliński, Mejza, Marczyńska . . . :
use one plot for each new variety
and several plots for a well-established but uninteresting "control";

Fixed spatial trend

$$Y_\omega = \tau_{f(\omega)} + g(\omega) + \epsilon_\omega$$

where

$g(\omega)$ is a two-dimensional low-degree polynomial in ω,

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \text{ if } \omega \neq \omega'.$$

Caliński, Mejza, Marczyńska . . . :
use one plot for each new variety
and several plots for a well-established but uninteresting "control";
place the "control" plots in a grid;

Fixed spatial trend

$$Y_\omega = \tau_{f(\omega)} + g(\omega) + \epsilon_\omega$$

where

$g(\omega)$ is a two-dimensional low-degree polynomial in ω,

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \text{ if } \omega \neq \omega'.$$

Caliński, Mejza, Marczyńska . . . :
use one plot for each new variety
and several plots for a well-established but uninteresting "control";
place the "control" plots in a grid;
use the "control" responses to estimate the polynomial trend;

Fixed spatial trend

$$Y_\omega = \tau_{f(\omega)} + g(\omega) + \epsilon_\omega$$

where

$g(\omega)$ is a two-dimensional low-degree polynomial in ω,

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \text{ if } \omega \neq \omega'.$$

Caliński, Mejza, Marczyńska . . . :
use one plot for each new variety
and several plots for a well-established but uninteresting "control";
place the "control" plots in a grid;
use the "control" responses to estimate the polynomial trend;
estimate each variety effect by subtracting the trend value from its response.

Spatial trend: example

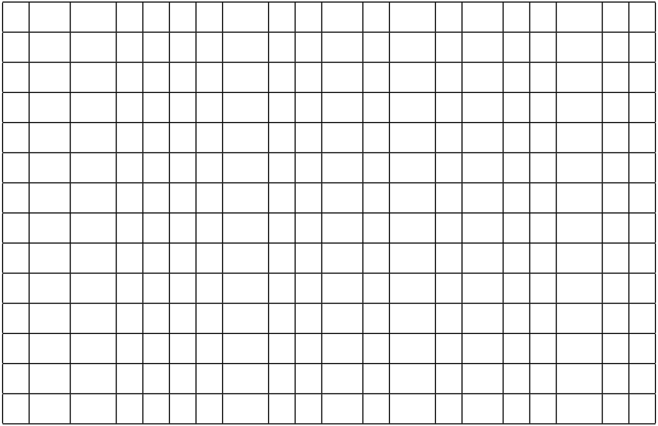
56 plots for “control”

224 new varieties have replication 1.

Spatial trend: example

56 plots for "control"

224 new varieties have replication 1.



Spatial trend: example

56 plots for “control”

224 new varieties have replication 1.

		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		

Spatial trend: example

56 plots for "control"

224 new varieties have replication 1.

		X					X					X					X		
		X					X			3		X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
	2	X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X		1			X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		
		X					X					X					X		

Spatial trend: example

56 plots for “control”

224 new varieties have replication 1.

		X					**X**					**X**					**X**		
		X					**X**			**3**		**X**					**X**		
		X					**X**					**X**					**X**		
		X					**X**					**X**					**X**		
		X					**X**					**X**					**X**		
	2	**X**					**X**					**X**					**X**		
		X					**X**					**X**					**X**		
		X					**X**					**X**					**X**		
		X					**X**					**X**					**X**		
		X					**X**					**X**		**1**			**X**		
		X					**X**					**X**					**X**		
		X					**X**					**X**					**X**		
		X					**X**					**X**					**X**		
		X					**X**					**X**					**X**		

Controls are on every fifth plot, working along rows.

Spatial trend: example, another layout

56 plots for "control"

224 new varieties have replication 1.

Spatial trend: example, another layout

56 plots for "control"

224 new varieties have replication 1.

	X							X			X							X	
				X	X									X	X				
X									X	X									X
			X			X							X			X			
		X					X					X					X		
	X							X			X							X	
				X	X									X	X				
X									X	X									X
			X			X							X			X			
		X					X					X					X		
	X							X			X							X	
				X	X									X	X				
X									X	X									X
			X			X							X			X			

Controls are on every 5th plot, working boustrophedon along columns.

Spatial trend: example, a third layout

56 plots for "control"

224 new varieties have replication 1.

Spatial trend: example, a third layout

56 plots for "control"

224 new varieties have replication 1.

	X		X			X		X			X		X			X		X	
	X		X			X		X			X		X			X		X	
	X		X			X		X			X		X			X		X	
	X		X			X		X			X		X			X		X	
	X		X			X		X			X		X			X		X	
	X		X			X		X			X		X			X		X	
	X		X			X		X			X		X			X		X	

Controls are on a complete sub-rectangle.

Spatial trend: example, what should we optimize?

56 plots for "control"

224 new varieties have replication 1.

Spatial trend: example, what should we optimize?

56 plots for "control"

224 new varieties have replication 1.

X	X				X	X							X	X				X	X
X	X				X	X							X	X				X	X
X	X																	X	X
X	X				X									X				X	X
									X	X									
									X	X									
X	X				X									X				X	X
X	X																	X	X
X	X				X	X							X	X				X	X
X	X				X	X							X	X				X	X

Controls are positioned to make the average variance of prediction small if the trend is a polynomial of degree three.

Spatial trend: example, what should we optimize/assume?

56 plots for “control”

224 new varieties have replication 1.

Spatial trend: example, what should we optimize/assume?

56 plots for "control"

224 new varieties have replication 1.

X	X	X						X	X	X	X						X	X	X
X	X	X															X	X	X
X																			X
X																			X
X									X	X									X
X									X	X									X
X									X	X									X
X									X	X									X
X																			X
X																			X
X	X	X															X	X	X
X	X	X						X	X	X	X						X	X	X

Controls are positioned to make the maximum variance of prediction small if the trend is a polynomial of degree two.

Spatial trend: comments on optimality criteria

Minimizing the maximum variance of prediction is G-optimalty.

Spatial trend: comments on optimality criteria

Minimizing the maximum variance of prediction is G-optimalty.

Minimizing the average variance of prediction is I-optimalty. Should we include the check plots in the average?

Spatial trend: comments on optimality criteria

Minimizing the maximum variance of prediction is G-optimalty.

Minimizing the average variance of prediction is I-optimalty. Should we include the check plots in the average?

These critiera give different designs.

Spatial trend: comments on optimality criteria

Minimizing the maximum variance of prediction is G-optimalty.

Minimizing the average variance of prediction is I-optimalty. Should we include the check plots in the average?

These critiera give different designs.

Each criterion gives different designs for different assumptions about the trend.

Spatial trend: comments on optimality criteria

Minimizing the maximum variance of prediction is G-optimalty.

Minimizing the average variance of prediction is I-optimalty. Should we include the check plots in the average?

These critiera give different designs.

Each criterion gives different designs for different assumptions about the trend.

Thanks to Bradley Jones, who found these optimal designs.

Spatial trend: comments on optimality criteria

Minimizing the maximum variance of prediction is G-optimalty.

Minimizing the average variance of prediction is I-optimalty. Should we include the check plots in the average?

These critiera give different designs.

Each criterion gives different designs for different assumptions about the trend.

Thanks to Bradley Jones, who found these optimal designs.

If the plots are not square, do the optimal designs change?

Spatial trend: controls as covariates

Yates (1936), Atiqullah and Cox (1962) consider controls spread throughout the field. In their analysis, a weighted mean of the response on the nearest controls is used as a covariate, rather than being simply subtracted.

Spatial trend: controls as covariates

Yates (1936), Atiqullah and Cox (1962) consider controls spread throughout the field. In their analysis, a weighted mean of the response on the nearest controls is used as a covariate, rather than being simply subtracted.

This is a rather different model.

Spatial correlation

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

and

$\mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'})$ depends on the spatial relationship between ω and ω'.

Spatial correlation

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

and

$\mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'})$ depends on the spatial relationship between ω and ω'.

Kempton, Talbot, Besag, Martin, Eccleston . . . :

Spatial correlation

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

and

$\mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'})$ depends on the spatial relationship between ω and ω'.

Kempton, Talbot, Besag, Martin, Eccleston . . . :
use one plot for each new variety and several plots for "control";

Spatial correlation

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

and

$\mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'})$ depends on the spatial relationship between ω and ω'.

Kempton, Talbot, Besag, Martin, Eccleston ...:
use one plot for each new variety and several plots for "control";
place the "control" plots in some kind of grid;

Spatial correlation

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

and

$\mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'})$ depends on the spatial relationship between ω and ω'.

Kempton, Talbot, Besag, Martin, Eccleston . . . :
use one plot for each new variety and several plots for "control";
place the "control" plots in some kind of grid;
analyse all the data with GLS or REML.

Spatial correlation

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

and

$\mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'})$ depends on the spatial relationship between ω and ω'.

Kempton, Talbot, Besag, Martin, Eccleston . . . :
use one plot for each new variety and several plots for "control";
place the "control" plots in some kind of grid;
analyse all the data with GLS or REML.

Cullis, Smith, Lim, Gilmour, Butler . . . :

Spatial correlation

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

and

$\mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'})$ depends on the spatial relationship between ω and ω'.

Kempton, Talbot, Besag, Martin, Eccleston . . . :
use one plot for each new variety and several plots for "control";
place the "control" plots in some kind of grid;
analyse all the data with GLS or REML.

Cullis, Smith, Lim, Gilmour, Butler . . . :
use 2 plots for some varieties and 1 plot for all other varieties,

Spatial correlation

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

and

$\mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'})$ depends on the spatial relationship between ω and ω'.

Kempton, Talbot, Besag, Martin, Eccleston . . . :
use one plot for each new variety and several plots for "control";
place the "control" plots in some kind of grid;
analyse all the data with GLS or REML.

Cullis, Smith, Lim, Gilmour, Butler . . . :
use 2 plots for some varieties and 1 plot for all other varieties,
optimize the design by computer search,

Spatial correlation

$$Y_\omega = \tau_{f(\omega)} + \epsilon_\omega$$

where

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

and

$\mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'})$ depends on the spatial relationship between ω and ω'.

Kempton, Talbot, Besag, Martin, Eccleston . . . :
use one plot for each new variety and several plots for "control";
place the "control" plots in some kind of grid;
analyse all the data with GLS or REML.

Cullis, Smith, Lim, Gilmour, Butler . . . :
use 2 plots for some varieties and 1 plot for all other varieties,
optimize the design by computer search,
analyse all the data with GLS or REML.

Blocks

The field is partitioned into homogeneous blocks.
(One block has all the stony plots;
one block has all the plots near the trees;
one block has all the plots near the rabbit warren, ...)

Blocks

The field is partitioned into homogeneous blocks.
(One block has all the stony plots;
one block has all the plots near the trees;
one block has all the plots near the rabbit warren, ...)

$$Y_\omega = \tau_{f(\omega)} + \beta_{h(\omega)} + \epsilon_\omega$$

where

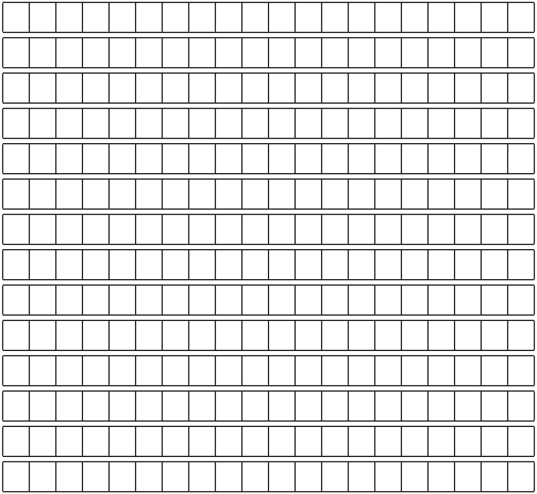
$$h(\omega) = \text{block containing } \omega,$$

$$\mathbb{E}(\epsilon_\omega) = 0, \qquad \mathrm{Var}(\epsilon_\omega) = \sigma^2,$$

$$\text{and} \quad \mathrm{Cov}(\epsilon_\omega, \epsilon_{\omega'}) = 0 \text{ if } \omega \neq \omega'.$$

Blocks: example

Rows are blocks, so there are 14 blocks, each with 20 plots.



Blocks: example, continued

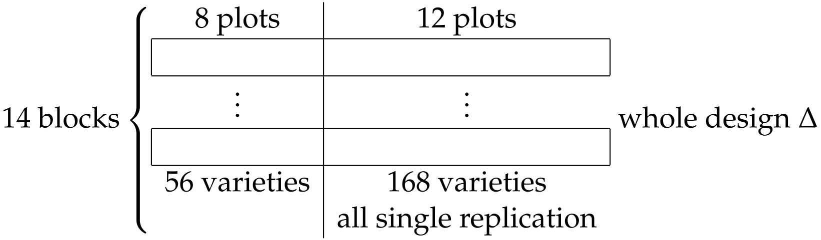
224 varieties in 14 blocks of size 20.

Blocks: example, continued

224 varieties in 14 blocks of size 20.
$(280 - 224 = 56$ and $224 - 56 = 168$,
so at least 168 varieties must have single replication.)

Blocks: example, continued

224 varieties in 14 blocks of size 20.
$(280 - 224 = 56$ and $224 - 56 = 168$,
so at least 168 varieties must have single replication.)



Subdesign Γ has 56 varieties
in 14 blocks of size 8.

Blocks: remember that replication is very low

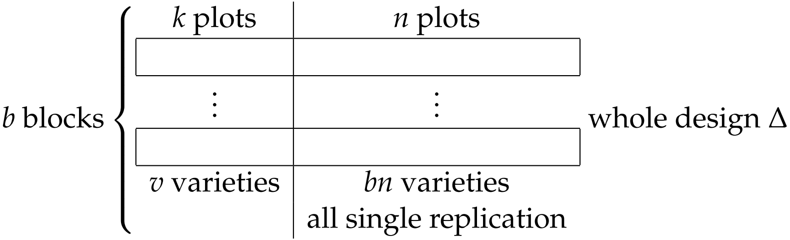
Yates (1936) concluded that square-lattice incomplete-block designs are superior to the inclusion of controls, but all of his examples had average replication equal to three or more.

Blocks: remember that replication is very low

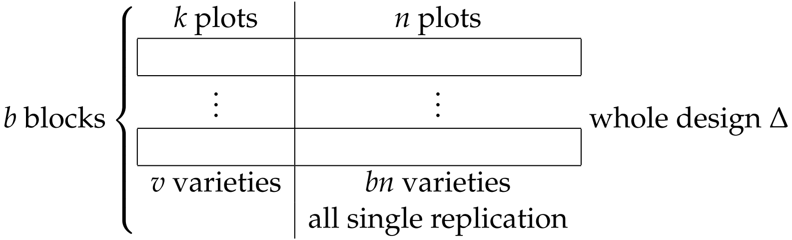
Yates (1936) concluded that square-lattice incomplete-block designs are superior to the inclusion of controls, but all of his examples had average replication equal to three or more.

Here we assume that average replication is (much) less than two.

A general block design with average replication less than 2

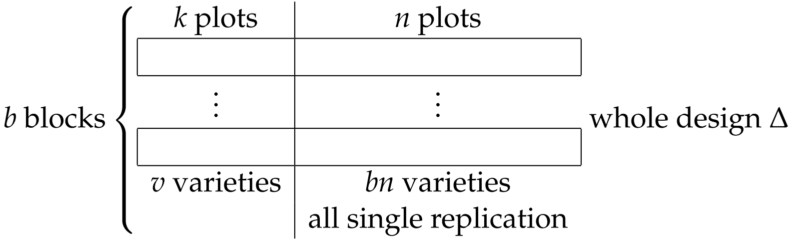


A general block design with average replication less than 2



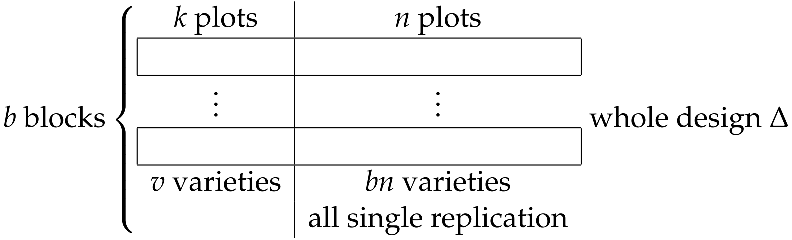
Whole design Δ has $v + bn$ varieties in b blocks of size $k + n$;

A general block design with average replication less than 2



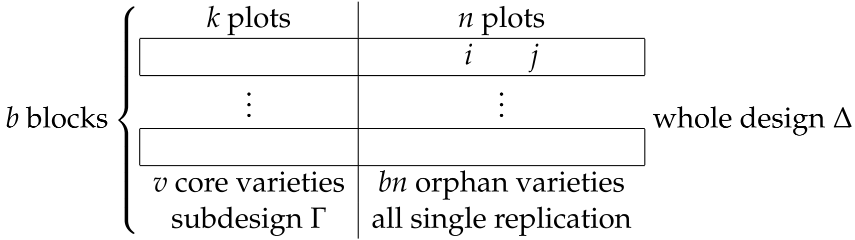
Whole design Δ has $v + bn$ varieties in b blocks of size $k + n$;
the subdesign Γ has v core varieties in b blocks of size k;

A general block design with average replication less than 2

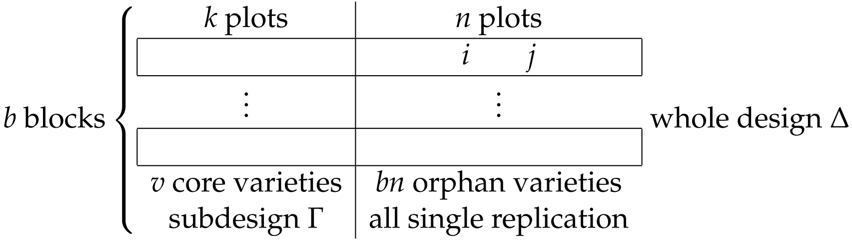


Whole design Δ has $v + bn$ varieties in b blocks of size $k + n$;
the subdesign Γ has v core varieties in b blocks of size k;
call the remaining varieties orphans.

Pairwise variance: two orphans in the same block

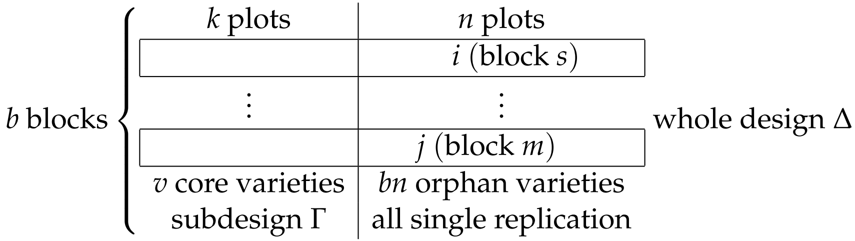


Pairwise variance: two orphans in the same block



$$\mathrm{Var}_\Delta(\hat{\tau}_i - \hat{\tau}_j) = 2\sigma^2.$$

Pairwise variance: two orphans in different blocks

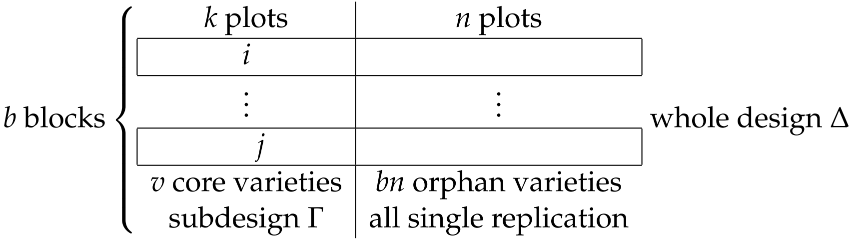


Pairwise variance: two orphans in different blocks

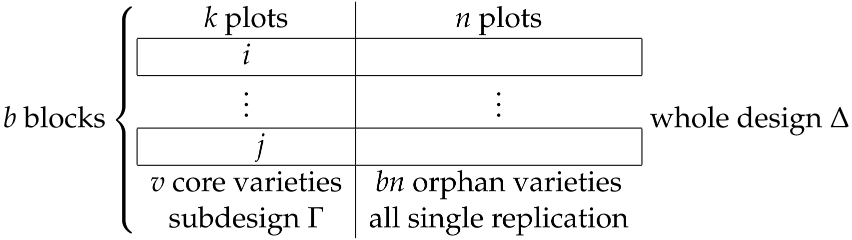
	k plots	n plots	
b blocks		i (block s)	
	$\vdots$	$\vdots$	whole design Δ
		j (block m)	
	v core varieties subdesign Γ	bn orphan varieties all single replication	

$$\mathrm{Var}_\Delta(\hat{\tau}_i - \hat{\tau}_j) = 2\sigma^2 + \mathrm{Var}_\Gamma(\hat{\beta}_s - \hat{\beta}_m).$$

Pairwise variance: two core varieties

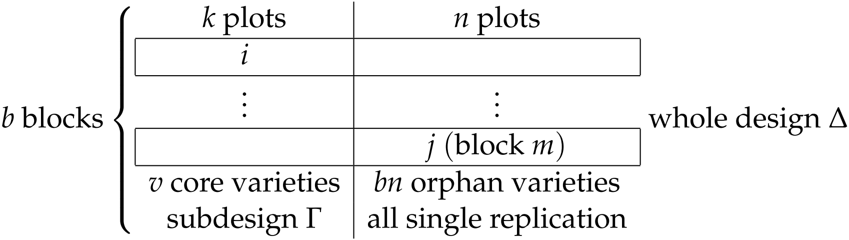


Pairwise variance: two core varieties

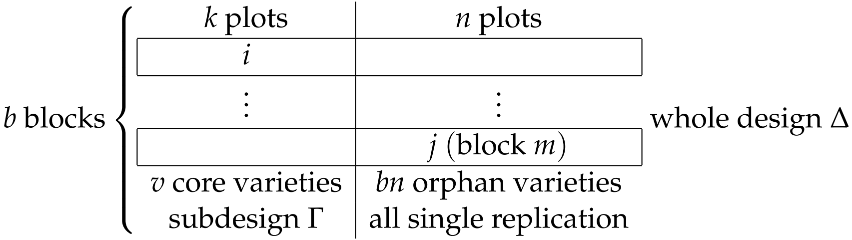


$$\mathrm{Var}_{\Delta}(\hat{\tau}_i - \hat{\tau}_j) = \mathrm{Var}_{\Gamma}(\hat{\tau}_i - \hat{\tau}_j).$$

Pairwise variance: one core variety and one orphan



Pairwise variance: one core variety and one orphan



$$\mathrm{Var}_{\Delta}(\hat{\tau}_i - \hat{\tau}_j) = \sigma^2 + \mathrm{Var}_{\Gamma}(\hat{\tau}_i + \hat{\beta}_m).$$

Sum of the pairwise variances

Theorem (cf Herzberg and Jarrett, 2007)

The sum of the variances of treatment differences in Δ

$$= constant + V_1 + nV_3 + n^2 V_2,$$

where

$$\begin{aligned} V_1 &= \textit{the sum of the variances of treatment differences in } \Gamma \\ V_2 &= \textit{the sum of the variances of block differences in } \Gamma \\ V_3 &= \textit{the sum of the variances of sums of} \\ &\quad\ \textit{one treatment and one block in } \Gamma. \end{aligned}$$

(If Γ is equi-replicate then V_2 and V_3 are both increasing functions of V_1.)

Sum of the pairwise variances

Theorem (cf Herzberg and Jarrett, 2007)

The sum of the variances of treatment differences in Δ

$$= constant + V_1 + nV_3 + n^2 V_2,$$

where

$$\begin{aligned} V_1 &= \text{the sum of the variances of treatment differences in } \Gamma \\ V_2 &= \text{the sum of the variances of block differences in } \Gamma \\ V_3 &= \text{the sum of the variances of sums of} \\ &\quad\ \text{one treatment and one block in } \Gamma. \end{aligned}$$

(If Γ is equi-replicate then V_2 and V_3 are both increasing functions of V_1.)

Consequence

For a given choice of k, make Γ as efficient as possible.

A less obvious consequence

Consequence

If n or b is large,
it may be best to make Γ a complete block design for k' controls, even if there is no interest in comparisons between new treatments and controls, or between controls.

$5n + 10$ treatments in 5 blocks of size $4 + n$

1	2	3	4	A_1	$\cdots$	A_n
3	4	5	6	B_1	$\cdots$	B_n
5	6	7	8	C_1	$\cdots$	C_n
7	8	9	0	D_1	$\cdots$	D_n
9	0	1	2	E_1	$\cdots$	E_n

Youden and Connor (1953): "experiments in physics do not need much replication because results are not very variable" —introduced chain block designs

$5n + 10$ treatments in 5 blocks of size $4 + n$

1	2	3	4	A_1	$\cdots$	A_n
3	4	5	6	B_1	$\cdots$	B_n
5	6	7	8	C_1	$\cdots$	C_n
7	8	9	0	D_1	$\cdots$	D_n
9	0	1	2	E_1	$\cdots$	E_n

Youden and Connor (1953): "experiments in physics do not need much replication because results are not very variable" —introduced chain block designs

1	2	3	4	A_1	$\cdots$	A_n
1	5	6	7	B_1	$\cdots$	B_n
2	5	8	9	C_1	$\cdots$	C_n
3	6	8	0	D_1	$\cdots$	D_n
4	7	9	0	E_1	$\cdots$	E_n

subdesign is dual of BIBD (Herzberg and Andrews, 1978)

$5n + 10$ treatments in 5 blocks of size $4 + n$

1	2	3	4	A_1	$\cdots$	A_n
1	5	6	7	B_1	$\cdots$	B_n
2	5	8	9	C_1	$\cdots$	C_n
3	6	8	0	D_1	$\cdots$	D_n
4	7	9	0	E_1	$\cdots$	E_n

subdesign is dual of BIBD,
best subdesign for $k = 4$

$5n + 10$ treatments in 5 blocks of size $4 + n$

1	2	3	4	A_1	$\cdots$	A_n
1	5	6	7	B_1	$\cdots$	B_n
2	5	8	9	C_1	$\cdots$	C_n
3	6	8	0	D_1	$\cdots$	D_n
4	7	9	0	E_1	$\cdots$	E_n

subdesign is dual of BIBD, best subdesign for $k = 4$

1	2	3	6	A_1	$\cdots$	A_n
2	3	4	7	B_1	$\cdots$	B_n
3	4	5	8	C_1	$\cdots$	C_n
4	5	1	9	D_1	$\cdots$	D_n
5	1	2	0	E_1	$\cdots$	E_n

best subdesign for $k = 3$ is better for large n if $b \neq 5$

$5n + 10$ treatments in 5 blocks of size $4 + n$

1	2	3	6	A_1	$\cdots$	A_n
2	3	4	7	B_1	$\cdots$	B_n
3	4	5	8	C_1	$\cdots$	C_n
4	5	1	9	D_1	$\cdots$	D_n
5	1	2	0	E_1	$\cdots$	E_n

best subdesign for $k = 3$ is better for large n if $b \neq 5$

$5n + 10$ treatments in 5 blocks of size $4 + n$

1	2	3	6	A_1	$\cdots$	A_n
2	3	4	7	B_1	$\cdots$	B_n
3	4	5	8	C_1	$\cdots$	C_n
4	5	1	9	D_1	$\cdots$	D_n
5	1	2	0	E_1	$\cdots$	E_n

best subdesign for $k = 3$ is better for large n if $b \neq 5$

K_1	K_2	1	2	A_1	$\cdots$	A_n
K_1	K_2	3	4	B_1	$\cdots$	B_n
K_1	K_2	5	6	C_1	$\cdots$	C_n
K_1	K_2	7	8	D_1	$\cdots$	D_n
K_1	K_2	9	0	E_1	$\cdots$	E_n

better for large n if $b > 13$ even if there is no interest in controls

D-optimality

The block design is D-optimal if it minimizes
the volume of the ellipsoid of confidence for $(\tau_1, \ldots, \tau_{v+bn})$
in the hyperplane defined by

$$\sum_i \tau_i = 0.$$

Levi graph

The Levi graph of the block design has

- one vertex for each treatment

Levi graph

The Levi graph of the block design has

- one vertex for each treatment
- one vertex for each block

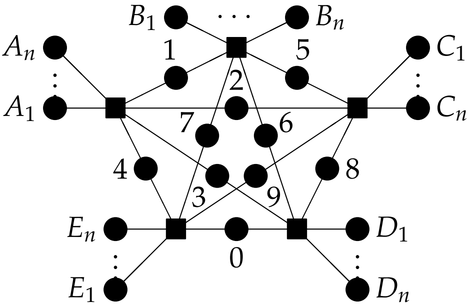
Levi graph

The Levi graph of the block design has

- one vertex for each treatment
- one vertex for each block
- one edge for each plot,
  so that the edge corresponding to plot ω joins the vertices corresponding to treatment $g(\omega)$ and block $h(\omega)$.

Levi graph: example

1	2	3	4	A_1	$\cdots$	A_n
1	5	6	7	B_1	$\cdots$	B_n
2	5	8	9	C_1	$\cdots$	C_n
3	6	8	0	D_1	$\cdots$	D_n
4	7	9	0	E_1	$\cdots$	E_n

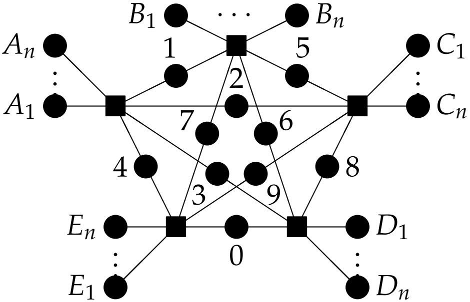


Spanning trees

A spanning tree for the Levi graph is a collection edges which provides a unique path between every pair of vertices.

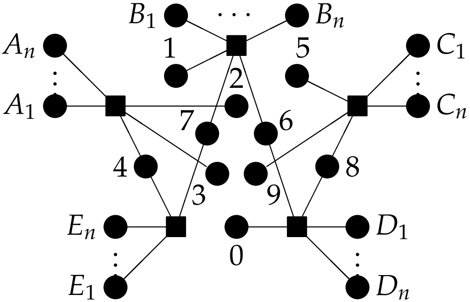
Spanning trees

A spanning tree for the Levi graph is a collection edges which provides a unique path between every pair of vertices.



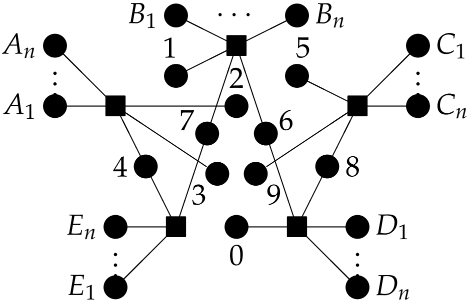
Spanning trees

A spanning tree for the Levi graph is a collection edges which provides a unique path between every pair of vertices.



Spanning trees

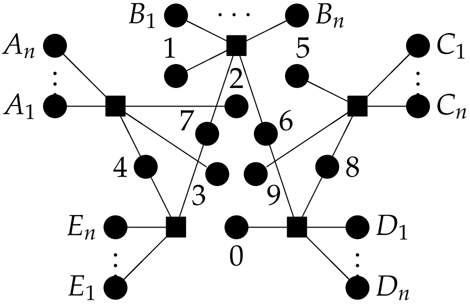
A spanning tree for the Levi graph is a collection edges which provides a unique path between every pair of vertices.



This Levi graph has 8000 spanning trees; the Levi graph for the chain block design has 5120 spanning trees.

Spanning trees

A spanning tree for the Levi graph is a collection edges which provides a unique path between every pair of vertices.



This Levi graph has 8000 spanning trees; the Levi graph for the chain block design has 5120 spanning trees.

The orphans make no difference to the number of spanning trees for the Levi graph.

Spanning trees and D-optimality

Theorem (Gaffke, 1981)

A block design is D-optimal if and only if its Levi graph maximizes the number of spanning trees.

Spanning trees and D-optimality

Theorem (Gaffke, 1981)

A block design is D-optimal if and only if
its Levi graph maximizes the number of spanning trees.

Consequence

The whole design Δ is D-optimal
for $v + bn$ treatments in b blocks of size $k + n$
if and only if the core design Γ is D-optimal
for v treatments in b blocks of size k.

Spanning trees and D-optimality

Theorem (Gaffke, 1981)

A block design is D-optimal if and only if its Levi graph maximizes the number of spanning trees.

Consequence

The whole design Δ is D-optimal for $v + bn$ treatments in b blocks of size $k + n$ if and only if the core design Γ is D-optimal for v treatments in b blocks of size k.

Consequence

Even when n or b is large, D-optimal designs do not include uninteresting controls.