Adaptive Randomization for Field Trials

Design Optimization in Field Experiments

Peter Claussen
Gylling Data Management
https://gdmdata.com/Resources/Meeting%20Handouts

What do we mean by optimal design?

Consider a simple t-test

$$\frac{\delta}{\sigma/\sqrt{\frac{2}{n}}} \geq t_{\alpha/2}$$

- δ difference between two means $\bar{y}_i - \bar{y}_j$
- σ standard deviation
- n number of replicates
- $t_{\alpha/2}$ critical value to declare statistical significance

What do we mean by optimal design?

Statistical power

$$1 - F\left(t_{\alpha/2} - \frac{\delta}{\sigma/\sqrt{\frac{2}{n}}}\right) \geq 0.80$$

- F is probability distribution function
  - F(…) becomes smaller as $t_{\alpha/2} - \frac{\delta}{\sigma/\sqrt{\frac{2}{n}}}$ becomes more negative
  - larger observed t values implies a smaller probability of rejecting a true difference $\delta > 0$

What do we mean by optimal design?

Statistical power

$$1 - F\left(t_{\alpha/2} - \frac{\delta}{\sigma/\sqrt{\frac{2}{n}}}\right)$$

- increase power by
  - decreasing critical value $t_{\alpha/2}$
  - increasing the effect size (increase δ, decrease σ)
  - increasing replicates n
- What's the easiest to change?

Eleven strategies for increasing statistical power

- Adding subjects
- Assigning more subjects to groups which are cheaper to run
- Choosing a less stringent significance or alpha level
- Increasing the size of the hypothesized ES
- Employing as few groups as possible
- Employing covariates and/or blocking variables
- Employing a cross-over or repeated measures/within subject design
- Hypothesizing mean effects rather than interactions
- Employing measures with are sensitive to change
- Employing reliable measures
- Using direct rather than indirect dependent measures

Power Analysis for Experimental Research (2002) Bausell and Li

Eleven strategies for increasing statistical power

- Adding subjects
- Assigning more subjects to groups which are cheaper to run
- Choosing a less stringent significance or alpha level
- Increasing the size of the hypothesized ES
- Employing as few groups as possible
- **Employing covariates and/or blocking variables**
- Employing a cross-over or repeated measures/within subject design
- Hypothesizing mean effects rather than interactions
- Employing measures with are sensitive to change
- Employing reliable measures
- Using direct rather than indirect dependent measures

Power Analysis for Experimental Research (2002) Bausell and Li

Employing covariates **or** blocking variables

- Employing covariates
    - Analysis of Covariance
- Employing blocking variables
    - “~~Post-blocking~~” “Scattered blocks”

European and Mediterranean Plant Protection Organization (2012), Design and analysis of efficacy evaluation trials. EPPO Bull, 42: 367-381. doi:10.1111/epp.2610

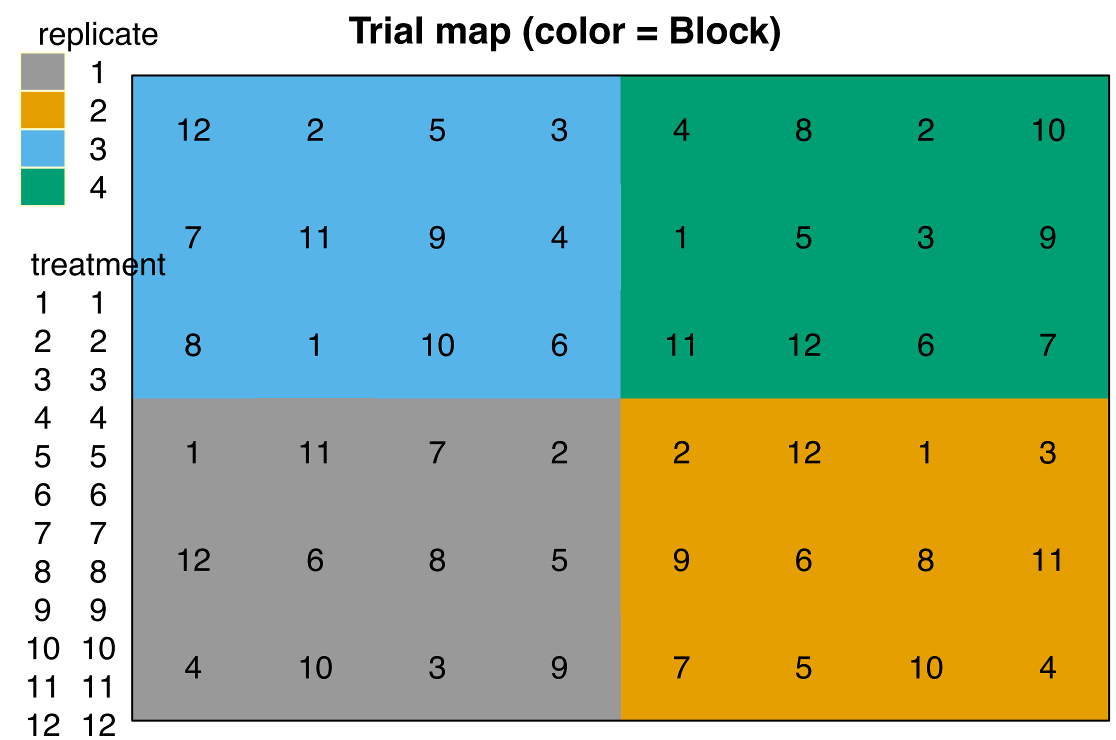
Pause a moment

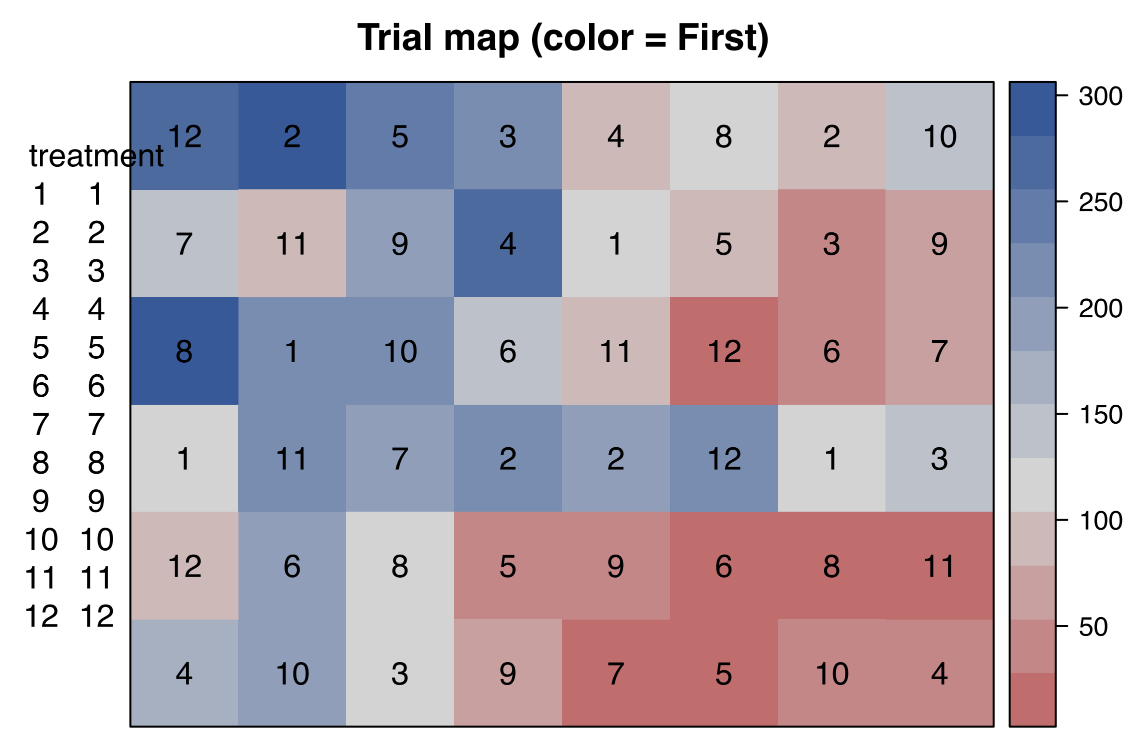
Here's a picture of a butterfly on a flower

Analysis of Covariance

Example

- Test of fumigants to control soil nematodes in a randomized complete block design.
- Response (Second) is nematode count after fumigation.
- Covariate (First) is nematode count before fumigation.





Cochran, W. G., & Cox, G. M. (1957). Experimental Design (Vol. 2). John Wiley and Sons, Inc.

Analysis of Variance/Covariance

- Analysis of Variance

$$y_{ij} = \mu + \tau_i + b_j + e_{ij}$$

Source	D F	SS	MS	F	P(F)
Treatmen	11	313234	28476	2.41628	0.025
Block	3	289426	96476	8.18632	<0.001
Residual	33	388904	11785		

- Analysis of Covariance

$$y_{ij} = \mu + \tau_i + b_j + \beta x_{ij} + e_{ij}$$

Source	D F	SS	MS	F	P(F)
Treatmen	11	313234	28476	5.86725	<0.0001
Block	3	289426	96476	19.87811	<0.0001
First	1	233597	233597	48.13098	<0.0001
Residual	32	155307	4853		

Analysis of Variance/Covariance

- Analysis of Variance (Second)

Source	D F	SS	MS	F	P(F)
Treatment	11	313234	28476	2.41628	0.025
Block	3	289426	96476	8.18632	<0.001
Residual	**33**	**388904**	**11785**		

- Analysis of Covariance (Second)

Source	D F	SS	MS	F	P(F)
Treatment	11	313234	28476	5.86725	<0.0001
Block	3	289426	96476	19.87811	<0.0001
First	**1**	**233597**	**233597**	48.13098	<0.0001
Residual	**32**	**155307**	**4853**		

- Analysis of covariance reduces $\boldsymbol{\sigma}$ by decomposing RCB residual error.
- $\downarrow \sigma \Longrightarrow \uparrow$ power

Comparing like-to-like

- To compare treatments in an RCB, we can use the difference between arithmetic means, since treatments are uniformly sampled over replicates.

$$\delta_{RCB} = \bar{y}_i - \bar{y}_j$$

- Treatments may not be uniformly sampled over covariates, so we adjust treatment difference according to differences in covariate and the strength of association between covariate and response.

$$\delta_{COV} = \bar{y}_i - \bar{y}_j - \widehat{\beta}(\bar{x}_i - \bar{x}_j)$$

Comparing like-to-like

- Similarly, in computing standard error, we need to adjust for differences in covariate.

$$s.e.(\delta_{RCB}) = \frac{2\hat{\sigma}^2_{RCB}}{r}$$

$$s.e.(\delta_{COV}) = \frac{2\sigma^2_{COV}}{r} + (\bar{x}_i - \bar{x}_j)^2\sigma^2_{COV}$$

Mean comparisons

$$\frac{2\sigma^2}{r} + (\bar{x}_i - \bar{x}_j)^2\sigma^2$$

Residual Error

Regression Error

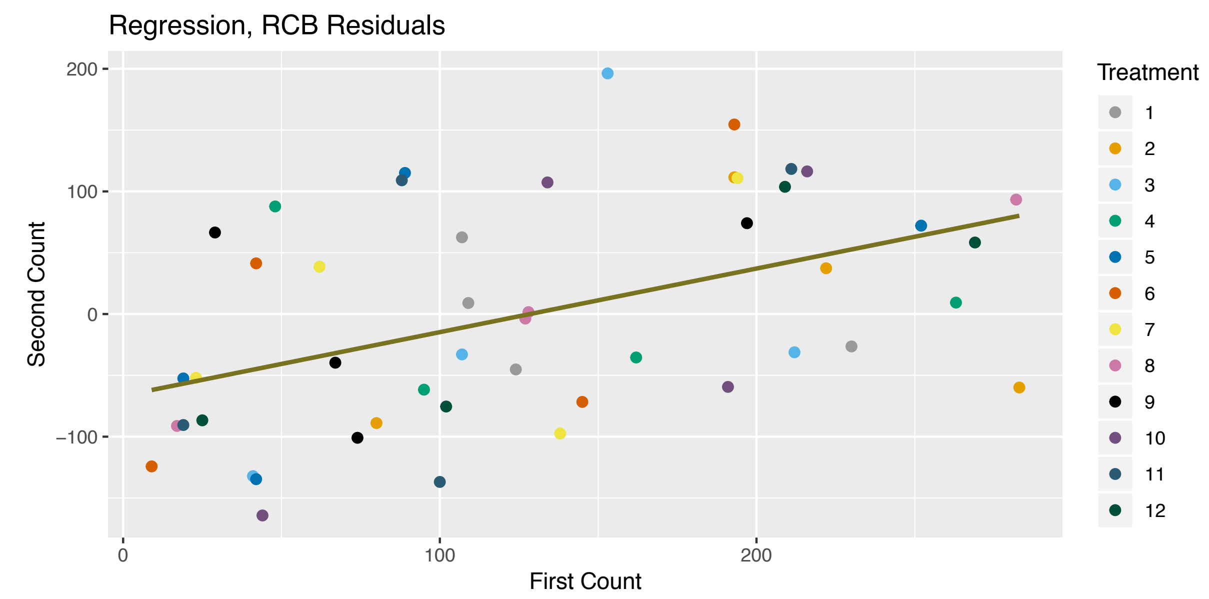
- When the covariates are balanced across treatments, the regression term is relatively small and precision is increased.
- If covariates are assigned to treatments haphazardly, we compare treatments with different levels of accuracy.
  - Increased precision in error is offset by decreased precision in covariate sampling

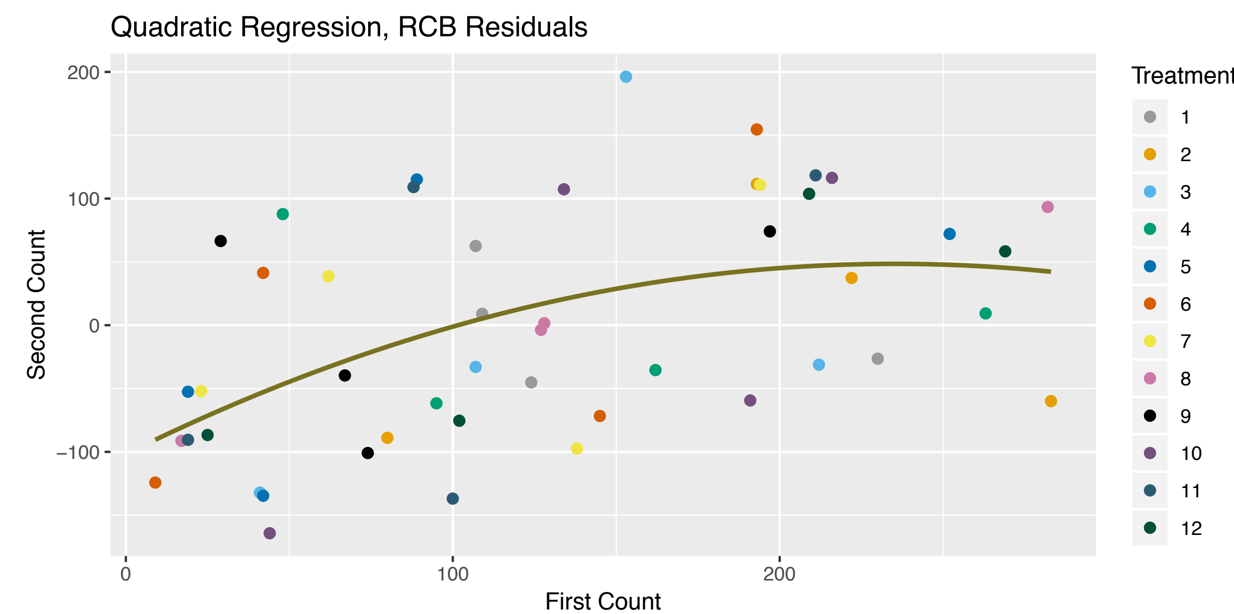
Analysis of Covariance

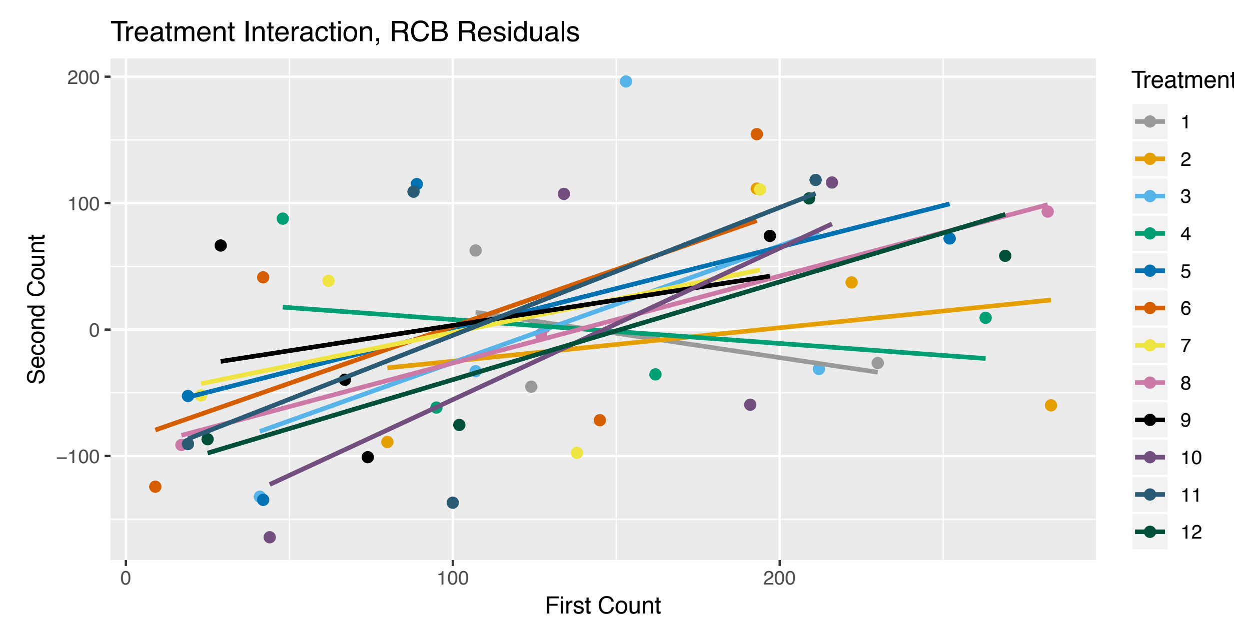
- In some cases, nuisance variables are not captured in blocks.
- Covariates that are measured after treatments are applied can help explain what happened.
- However, we must be aware of the hazards of regression on happenstance data …

Hazards of Regression on Happenstance Data

- Is the response simply linear, or is it polynomial or non-linear?
- Are the treatment responses uniform?







Box, Hunter and Hunter (1978) Statistics for experimenters: an introduction to design, data analysis, and model building. Wiley

Part 2

Here’s another bug on a flower

Treatment dispersion over covariates

- If we can measure x_{ij} before randomization, we can control the differences $\bar{x}_i - \bar{x}_j$.
- One method for controlling treatment dispersion is to use the covariate as a blocking variable.

Blocking Variables

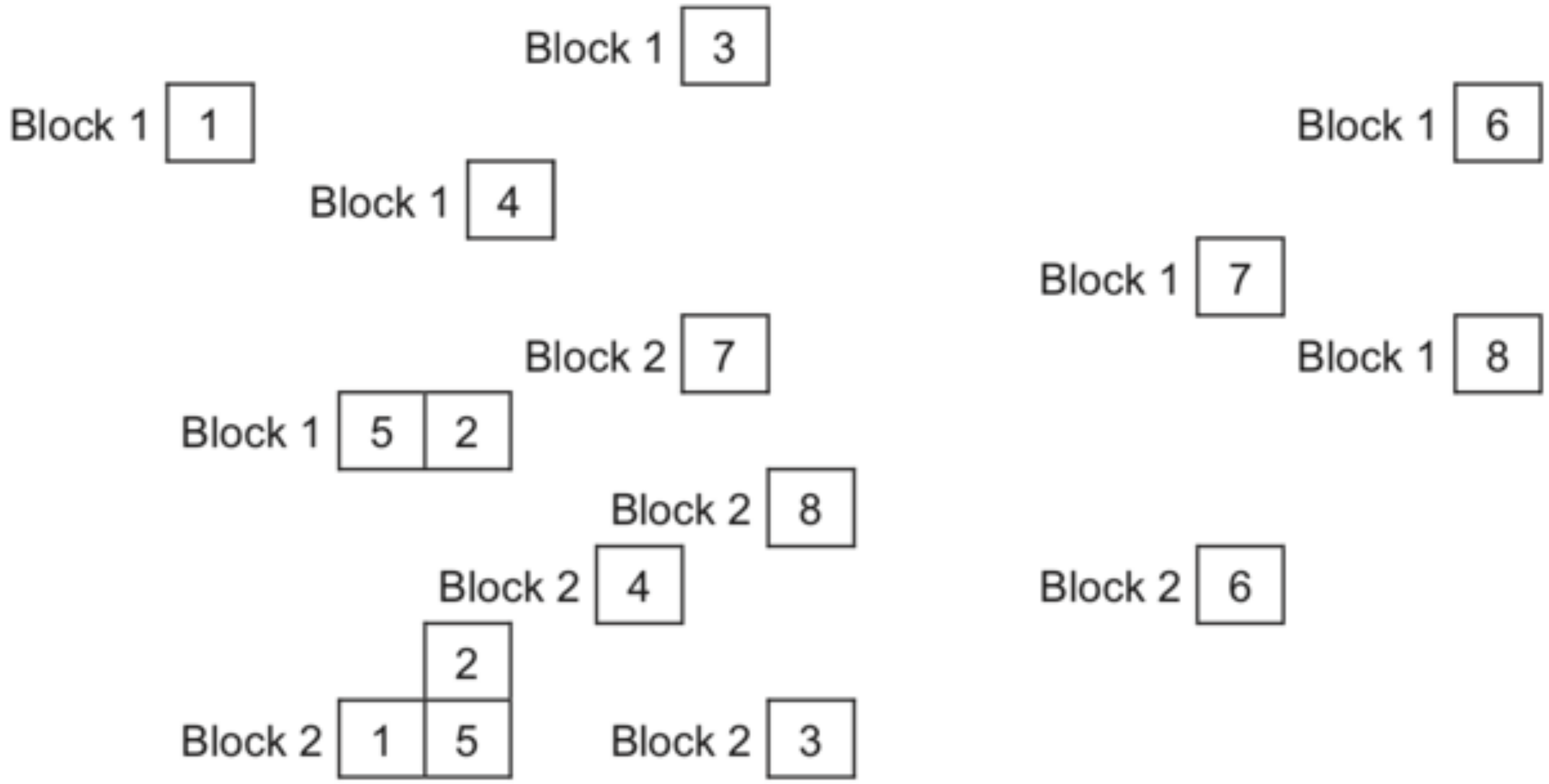


Fig. 6 Possible arrangement of blocks and plots in randomized blocks in field trials. Blocks scattered across the field, according to complex, previously observed heterogeneity.

- In extreme cases, plots may be haphazardly scattered in space.

European and Mediterranean Plant Protection Organization (2012), Design and analysis of efficacy evaluation trials. EPPO Bull, 42: 367-381. doi:10.1111/epp.2610

Simple Variable Blocking

1. Rank each plot by covariate.

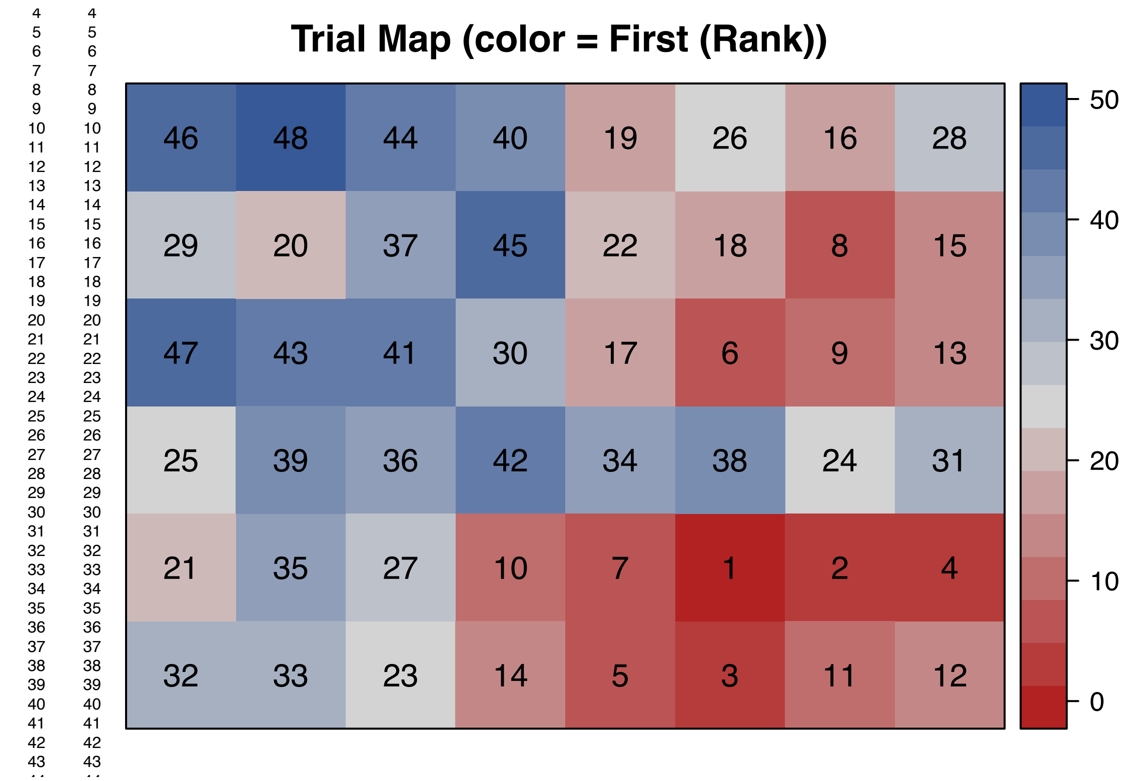
2. Divide rank by the number of treatments, rounding up.

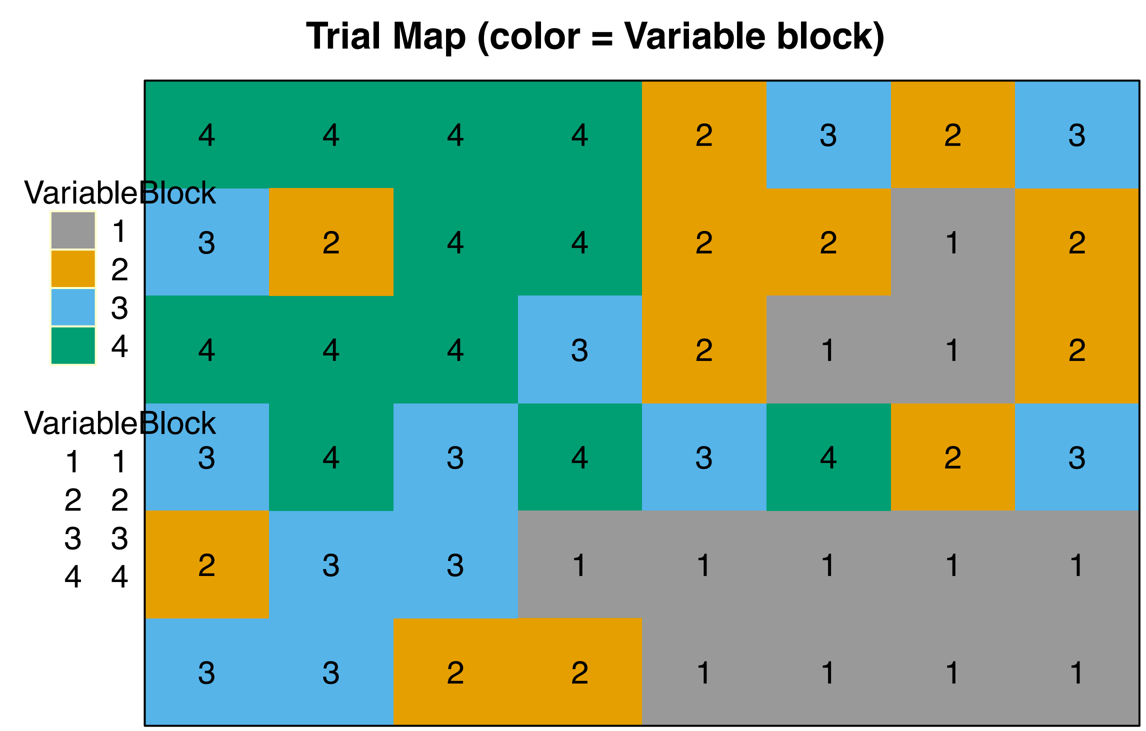
3. This number is now the block.

4. Randomize treatments within blocks.

Variable Blocks

- Cochran and Cox, blocked by First
- This trial would be analyzed as an RCB, but Variable Block instead of Complete Block

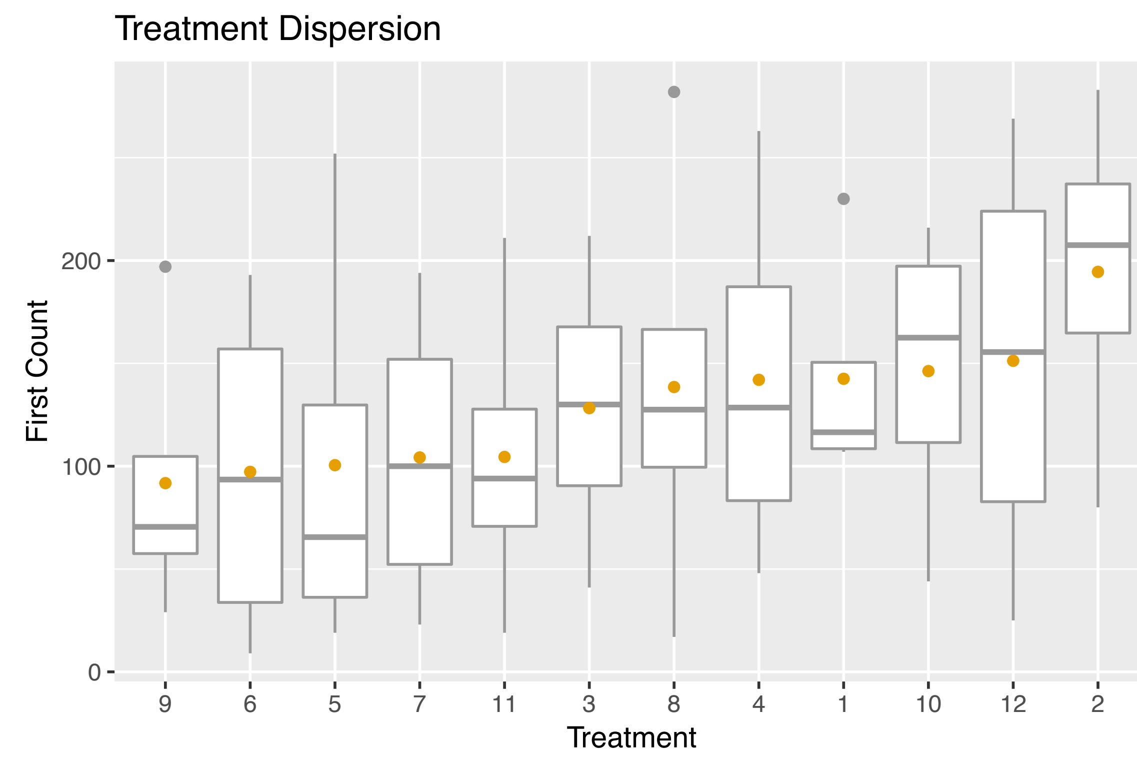




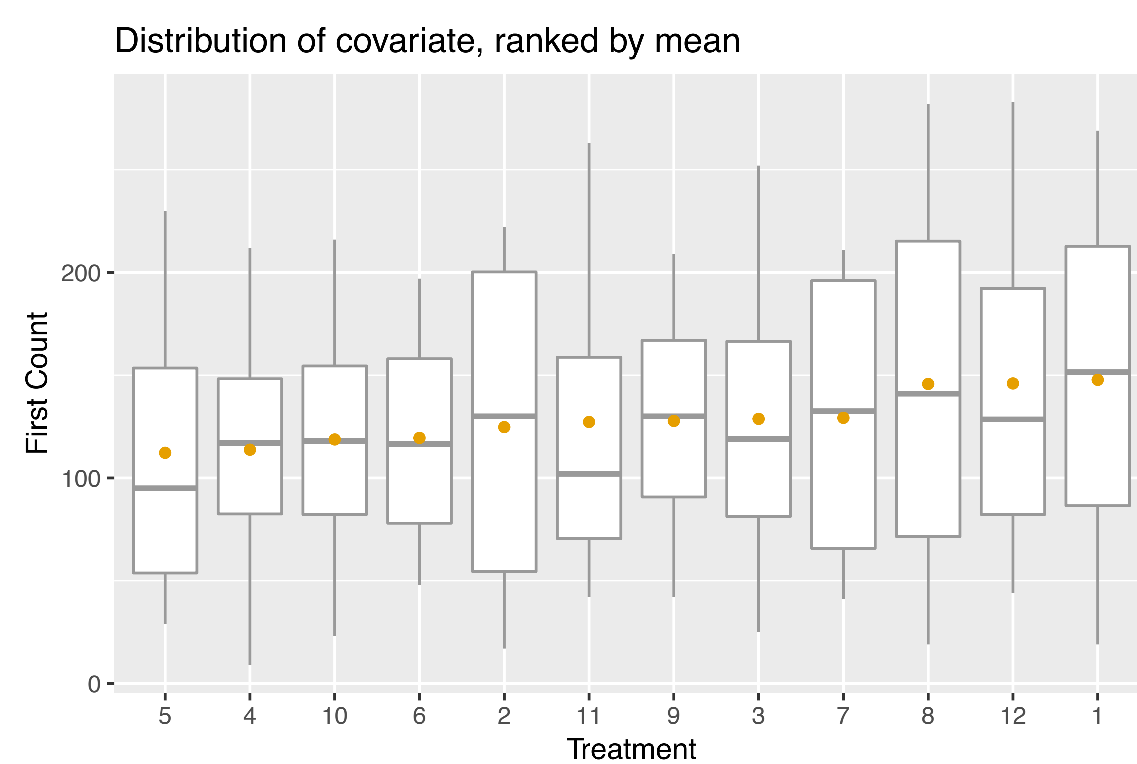
Treatment dispersion

Randomizing by variable improves sampling over treatments:

Original RCB



Variable Blocking

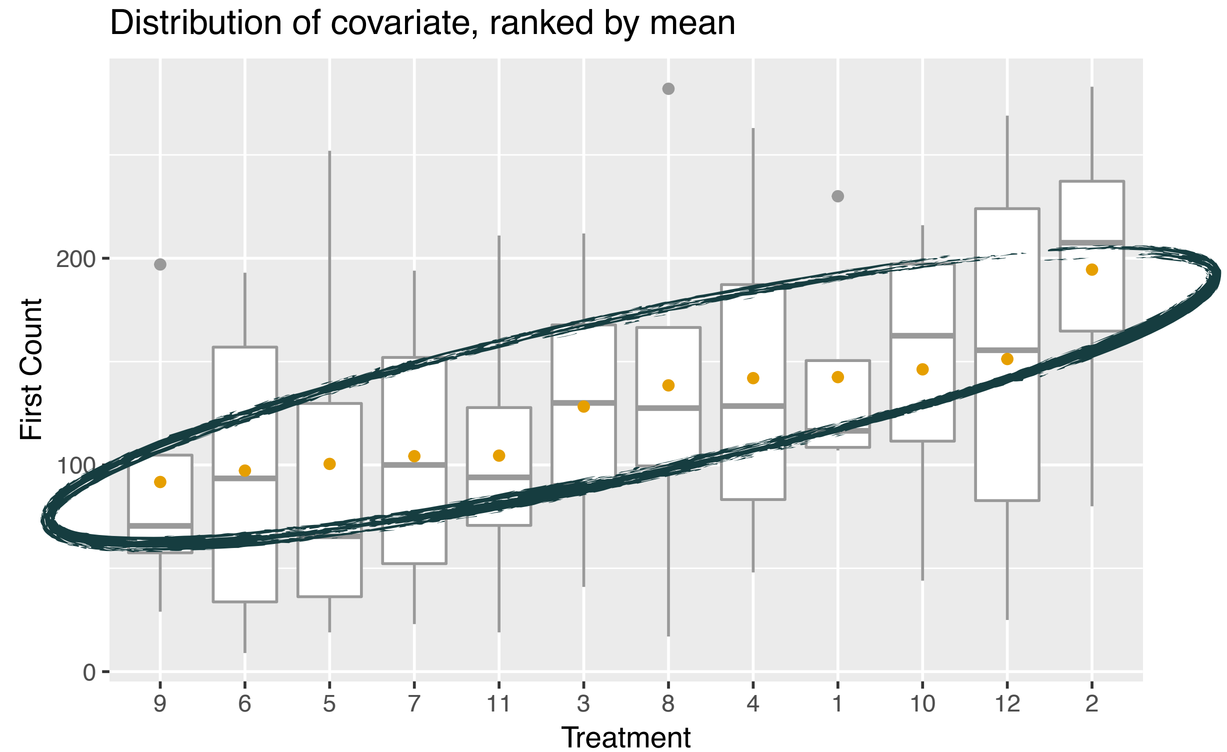


Uniform sampling

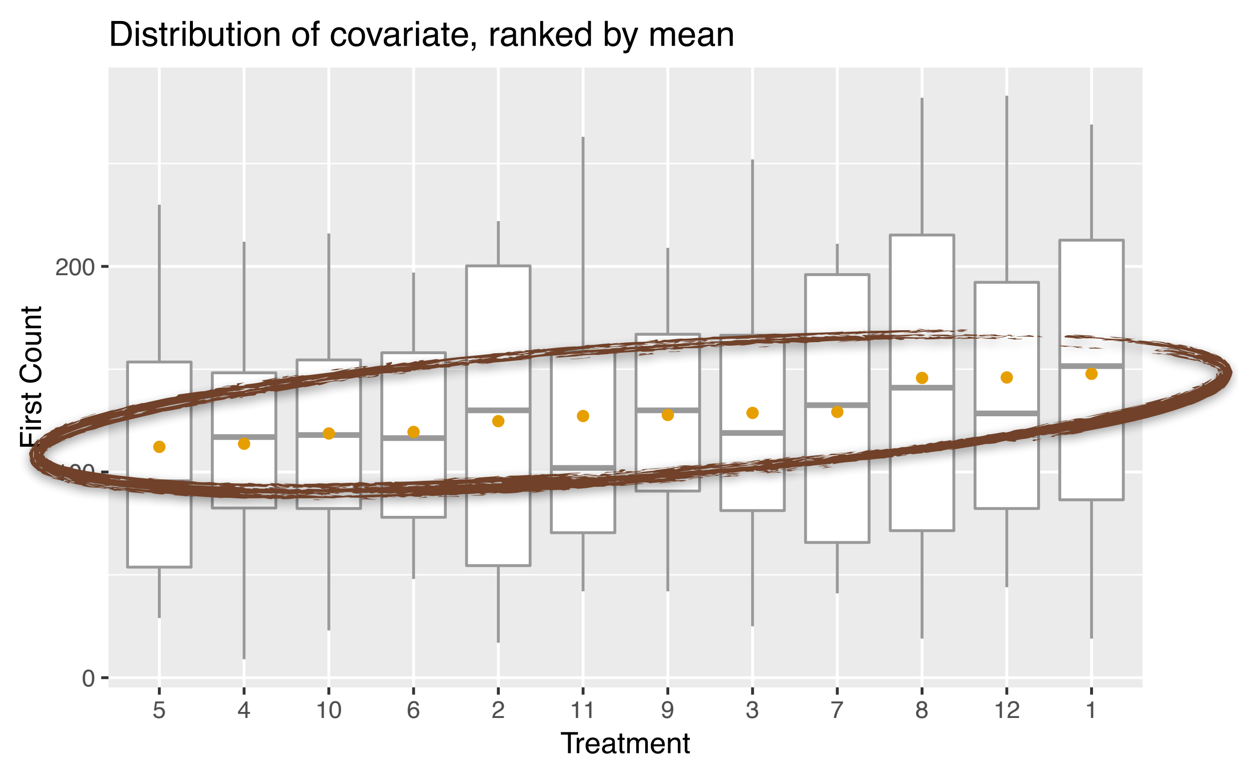
Randomizing by variable improves sampling over treatments:

- minimize covariate means over treatments (Largest Mean Difference)

Original RCB



Variable Blocking

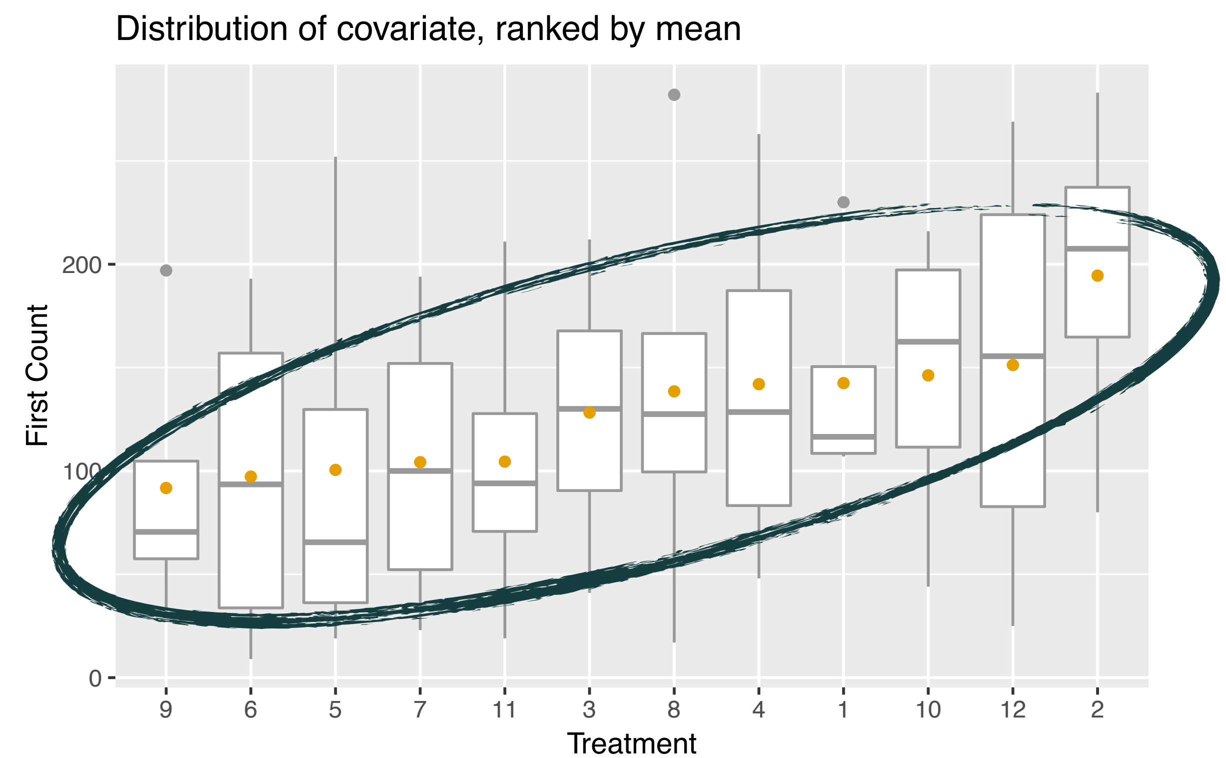


Uniform sampling

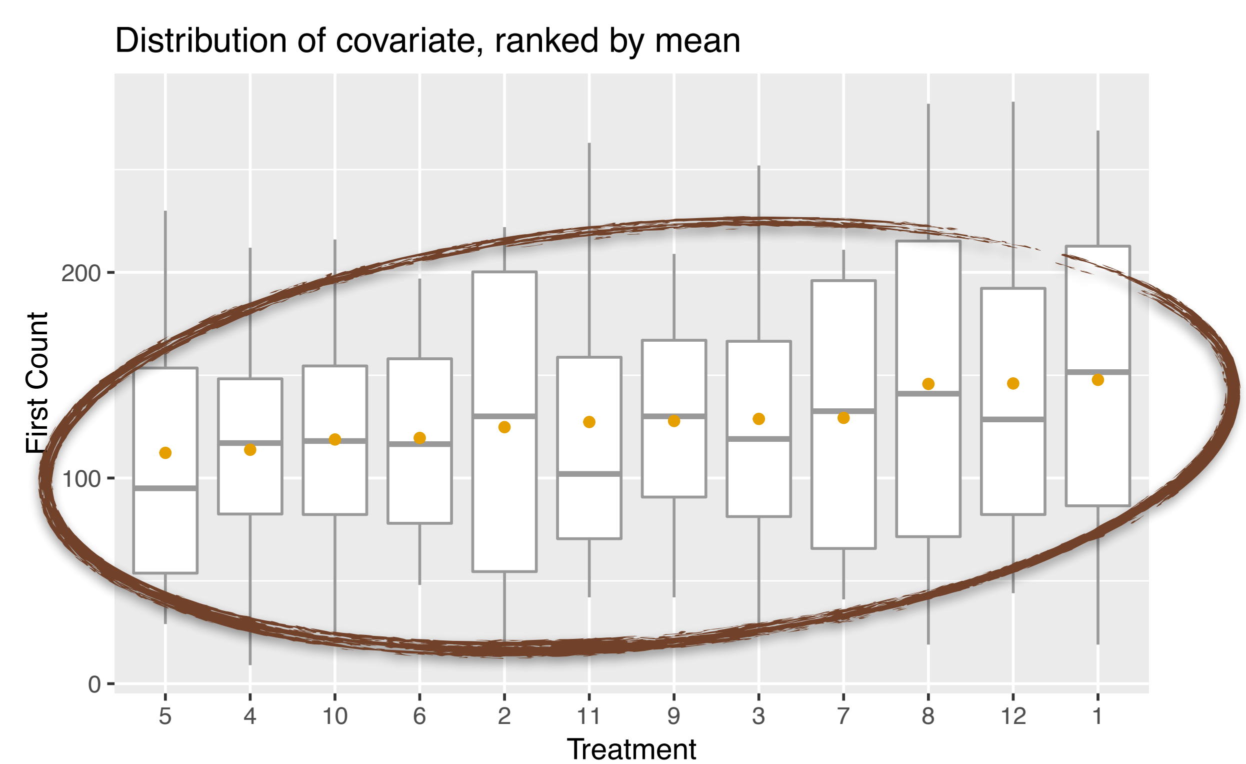
Randomizing by variable improves sampling over treatments:

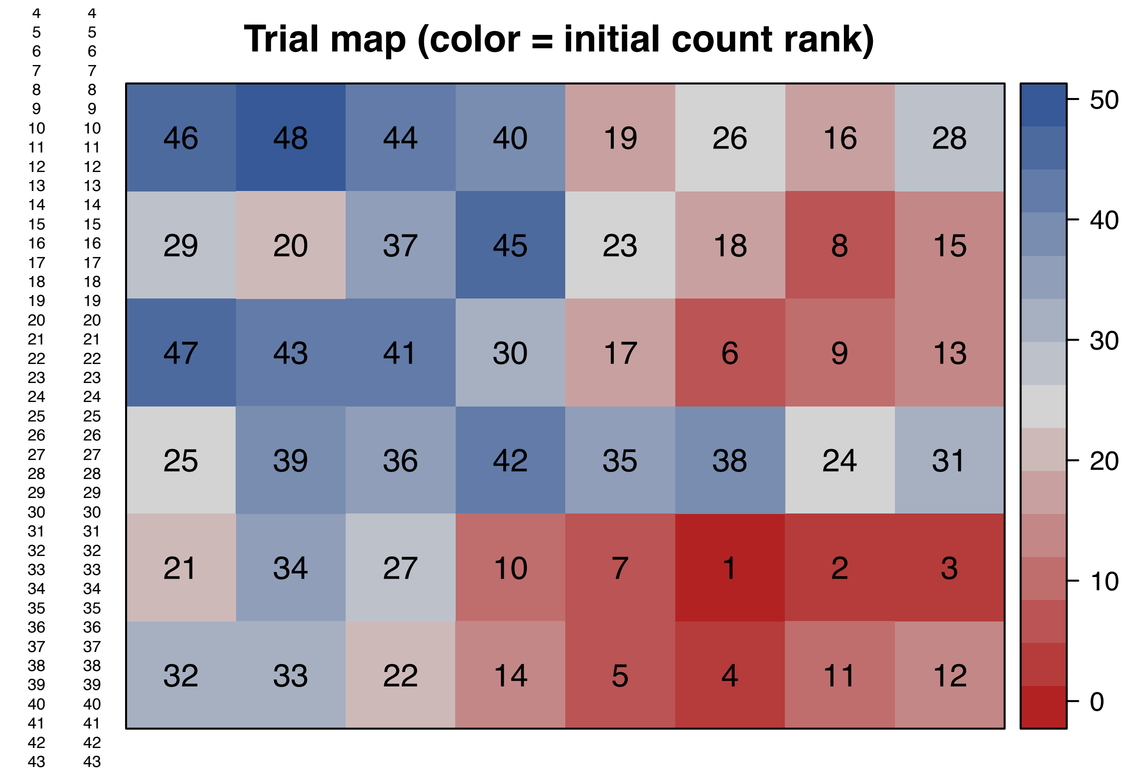
- minimize covariate means over treatments (Largest Mean Difference)
- maximize covariate dispersion over treatments (Average SD)

Original RCB



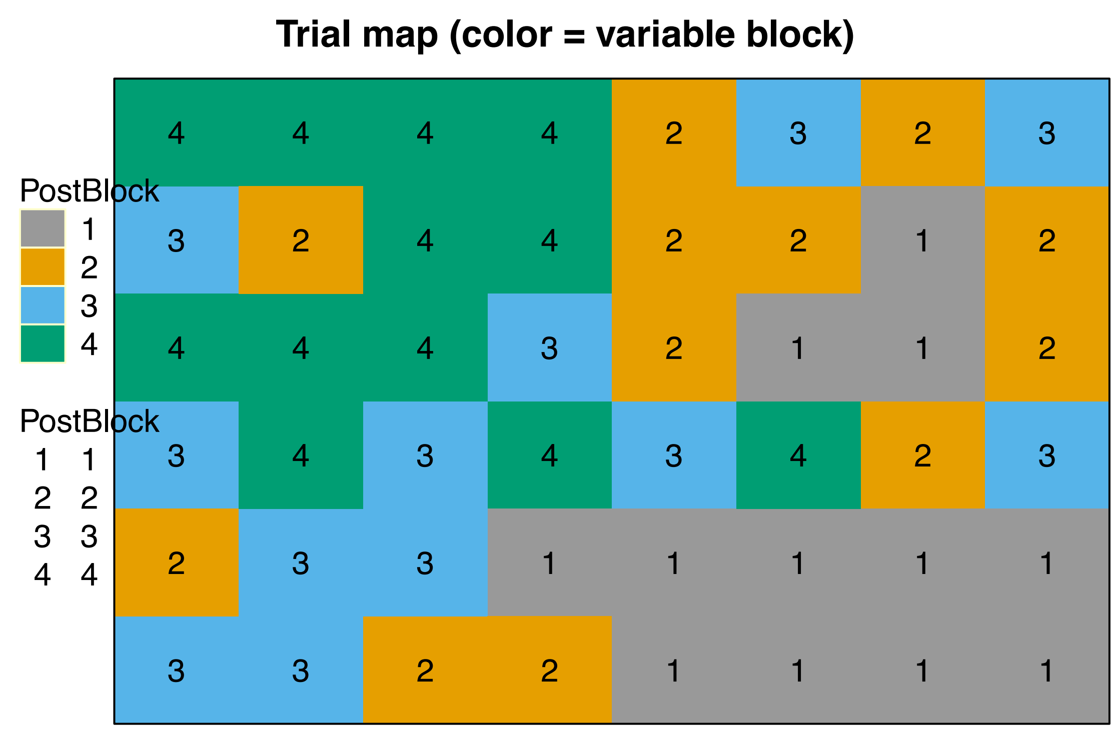
Variable Blocking





Variable Blocks

- However, blocks are no longer spatially contiguous, so blocks may not capture other spatially varying nuisance variables.



Act 3

Why not both?

Employing covariates **and** blocking variables

- Can we combine the traditional randomized complete blocks with variable blocking?
- Can we maintain optimal treatment sampling for covariate regression, in the constraint of complete blocks?
  - Rerandomization

Rerandomization

- *…Rubin recounts the following conversation with his advisor Bill Cochran:*
    - *Rubin: What if, in a randomized experiment, the chosen randomized allocation exhibited substantial imbalance on a prognostically important baseline covariate?*
    - *Cochran: Why didn't you block on that variable?*
    - *Rubin: Well, there were many baseline covariates, and the correct blocking wasn't obvious; and I was lazy at that time.*
    - *Cochran: This is a question that I once asked Fisher, and his reply was unequivocal:*
    - *Fisher (recreated via Cochran): Of course, if the experiment had not been started, I would rerandomize.*

Morgan, K. L., & Rubin, D. B. (2012). Rerandomization to Improve Covariate Balance in Experiments.

Optimal design/randomization

- How do we know when a chosen randomized allocation exhibits substantial imbalance?

- How do we know when a proposed rerandomization improves balance?

Wall of math

- Let's go back to the simple covariate model

$$y_{ij} = \mu + \alpha_i + \beta x_{ij} + e_{ij}$$

- We represent this in matrix form

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ \vdots \\ y_{ij} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & x_{11} \\ 1 & 1 & 0 & \dots & 0 & x_{12} \\ 1 & 1 & 0 & \dots & 0 & x_{13} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 1 & x_{ij} \end{bmatrix} [\mu \quad \tau_1 \quad \tau_2 \quad \dots \quad \tau_i \quad \beta]^t + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ \vdots \\ e_{ij} \end{bmatrix}$$

- which simplifies to

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$$

- We sometimes use the normal form

$$\mathbf{X}^t\mathbf{X}\beta = \mathbf{X}\mathbf{y}$$

- to find a solution for

$$\hat{\beta} = \left(\mathbf{X}^t\mathbf{X}\right)^- \mathbf{X}\mathbf{y}$$

- We can go on to compute 'least-square' means using a linear contrast

$$y_{1.} = [1 \quad 1 \quad 0 \quad \dots \quad 0 \quad \bar{x}] \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \dots \\ \hat{\tau}_i \\ \hat{\beta} \end{bmatrix} = L_1\hat{\beta}$$

- and the difference between means with a set of linear contrasts

$$\bar{y}_{1.} - \bar{y}_{2.} = ([1 \quad 1 \quad 0 \quad \dots \quad 0 \quad \bar{x}] - [1 \quad 0 \quad 1 \quad \dots \quad 0 \quad \bar{x}]) \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \dots \\ \hat{\tau}_i \\ \hat{\beta} \end{bmatrix} = (L_1 - L_2)\hat{\beta}$$

- We can show that the variance of our estimates is given by

$$Var(\hat{\beta}) = \left(\mathbf{X}^t\mathbf{X}\right)^- \sigma^2$$

- Further, we can write the variance for means and the difference between means by

$$Var(\bar{y}_1) = L_1\left(\mathbf{X}^t\mathbf{X}\right)^- L_1^t\sigma^2$$

$$Var(\bar{y}_1 - \bar{y}_2) = (L_1 - L_2)\left(\mathbf{X}^t\mathbf{X}\right)^-(L_1 - L_2)^t\sigma^2$$

- since

$$\mathbf{X}^t\mathbf{X} = \begin{bmatrix} n & 0 & 0 & \dots & 0 \\ 0 & n & 0 & \dots & 0 \\ 0 & 0 & n & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & n \end{bmatrix}$$

$$\left(\mathbf{X}^t\mathbf{X}\right)^- = \begin{bmatrix} 1/n & 0 & 0 & \dots & 0 \\ 0 & 1/n & 0 & \dots & 0 \\ 0 & 0 & 1/n & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1/n \end{bmatrix}$$

- it follows

$$s.e.(\bar{y}_1) = \sqrt{L_1\left(\mathbf{X}^t\mathbf{X}\right)^- L_1^t\sigma^2} = \sqrt{\frac{\sigma^2}{n}}$$

$$s.e.(\bar{y}_1 - \bar{y}_2) = (L_1 - L_2)\left(\mathbf{X}^t\mathbf{X}\right)^-(L_1 - L_2)^t\sigma^2 = \sqrt{\frac{2\sigma^2}{n}}$$

- If we increase the 'size' of

$$\mathbf{X}^t\mathbf{X}$$

- we decrease the magnitude of standard error

$$s.e.(\bar{y}_1) = \sqrt{\frac{\sigma^2}{n}}$$

$$s.e.(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{2\sigma^2}{n}}$$

- When X contains a covariate

$$X = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & x_{11} \\ 1 & 1 & 0 & \dots & 0 & x_{12} \\ 1 & 1 & 0 & \dots & 0 & x_{13} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 1 & x_{ij} \end{bmatrix}$$

- increasing

$$\mathbf{X}^t\mathbf{X}$$

- reduces error by reducing

$$(\bar{x}_i - \bar{x}_j)^2$$

- leading to a smaller standard error

$$s.e.(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{2\sigma^2}{r} + (\bar{x}_i - \bar{x}_j)^2\sigma^2}$$

- D-optimality minimizing variance by finding

$$\mathbf{X}$$

- that maximizes the determinant

$$\left|\mathbf{X}^t\mathbf{X}\right|$$

- Other criteria exist. For example, a measure of average variance (and A-optimality) is given by

$$trace\left(\mathbf{X}^t\mathbf{X}\right)$$

-

TL;DR

- **D-optimal** designs improve balance and minimize :

$$s.e.(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{2\sigma^2}{r} + (\bar{x}_i - \bar{x}_j)^2\sigma^2}$$

- by maximizing the determinant of the information matrix :

$$D = \left|\mathbf{X}^t\mathbf{X}\right|$$

- where $\mathbf{X}$ is the design matrix for a linear model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$$

What do D-optimal randomizations look like?

- Randomize the Cochran data set based on First count.
- 2000 randomizations:
  - Variable Blocking (“scattered”)
  - Randomized Complete Block
- Compute D-optimality for each randomization.

What do D-optimal randomizations look like?

- Focus on two models:
  - Regression only

$$y_{ij} = \mu + \tau_i + \beta x_{ij} + e_{ij}$$

  - RCB + Covariate (ANCOVA)

$$y_{ij} = \mu + \tau_i + b_j + \beta x_{ij} + e_{ij}$$

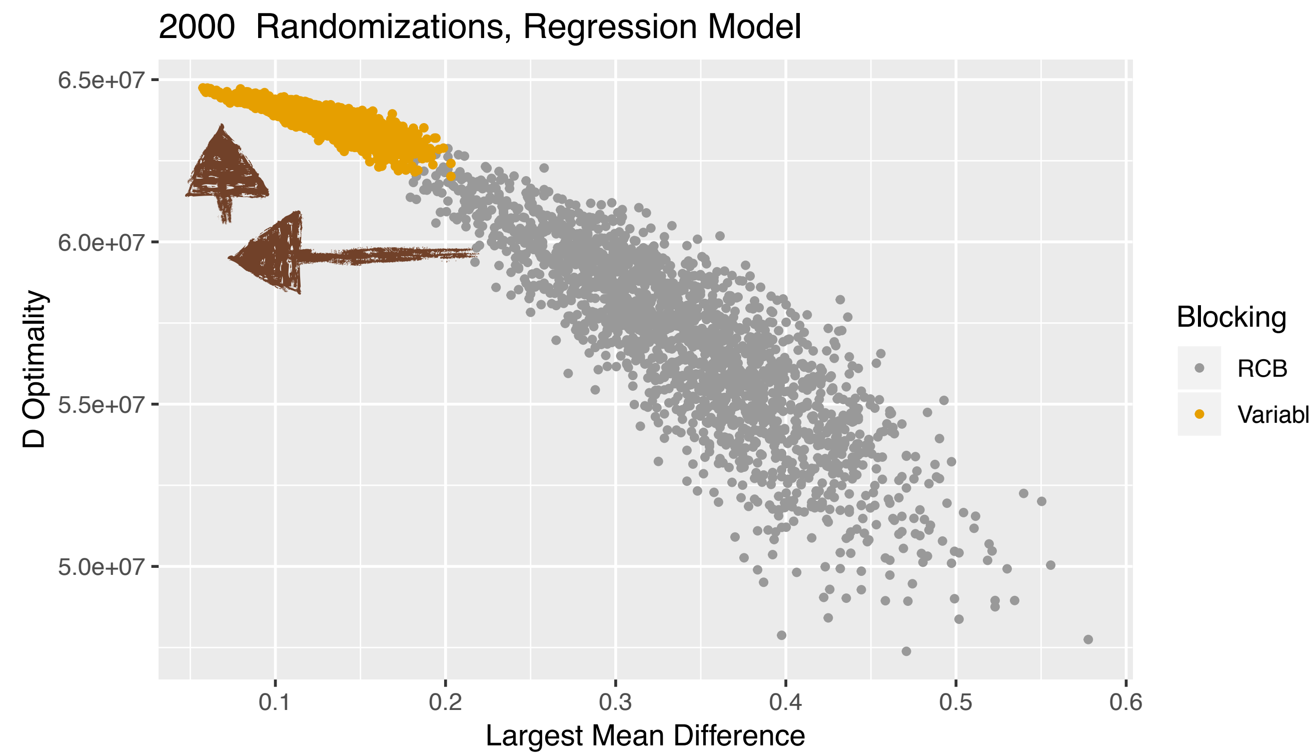
  - We can ignore RCB or Variable Block only models; they will all have the same optimality.

$$y_{ij} = \mu + \tau_i + b_j + e_{ij}$$

Optimality for Regression

D-optimal designs tend to minimize the differences of covariate means over treatments

These differences are smaller when treatments are blocked by covariate

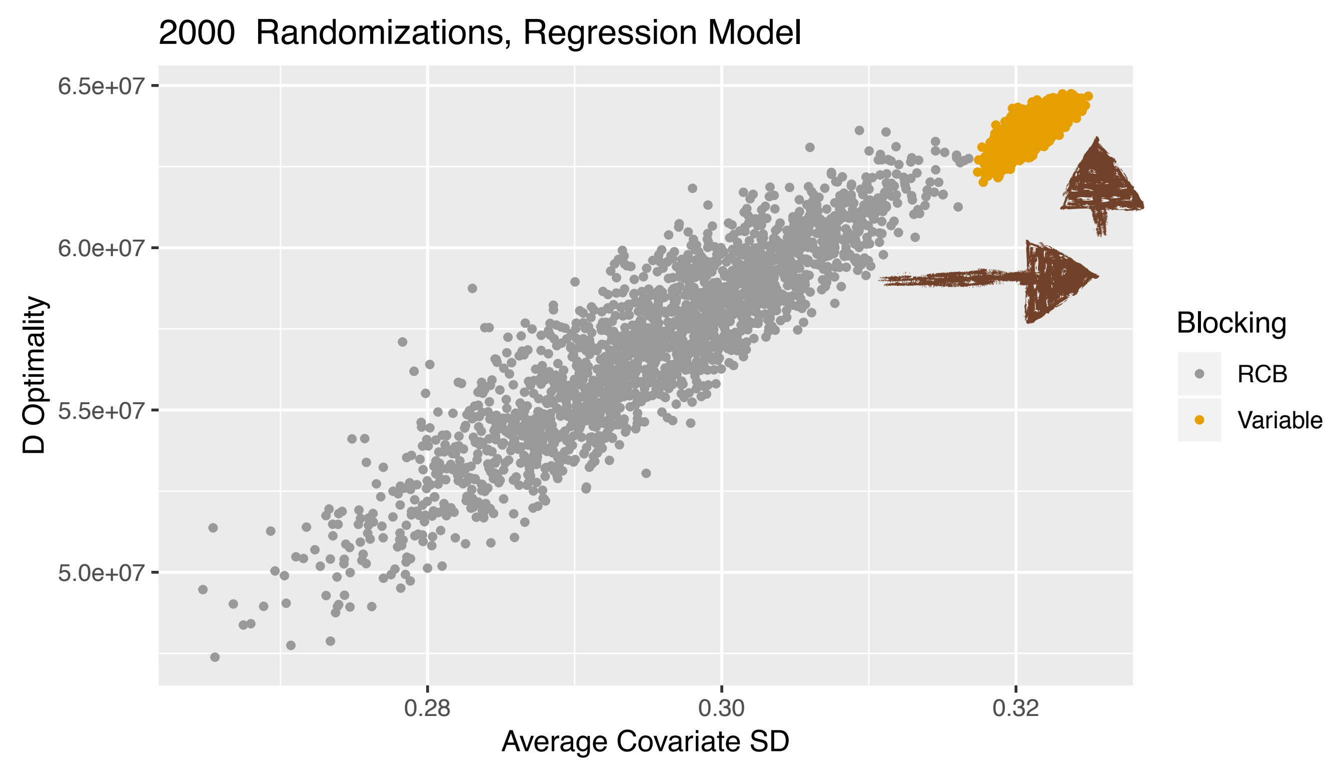


$$max(\bar{x}_i - \bar{x}_j), i, j = 1...t$$

Optimality for Regression

D-optimal designs also tend to maximize the dispersion of covariates over treatments.

These differences are greater when treatments are blocked by covariate

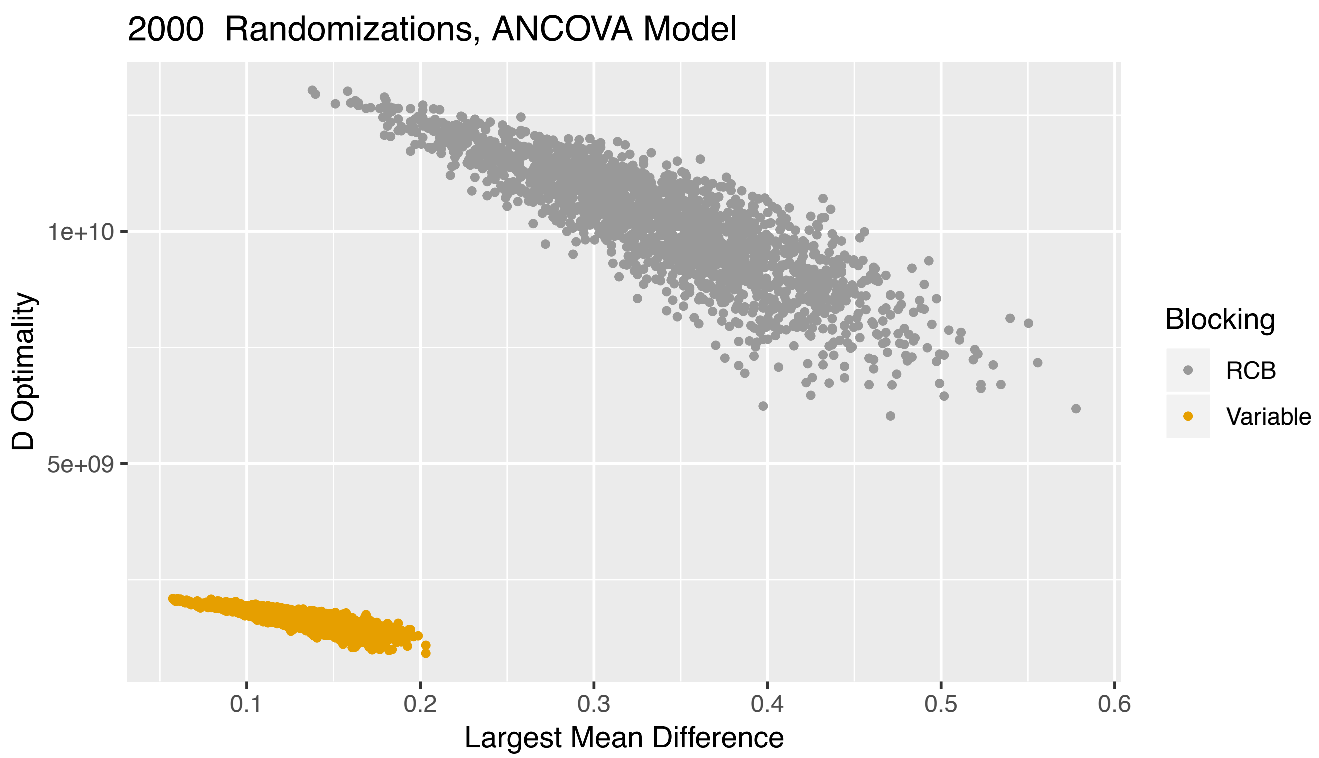


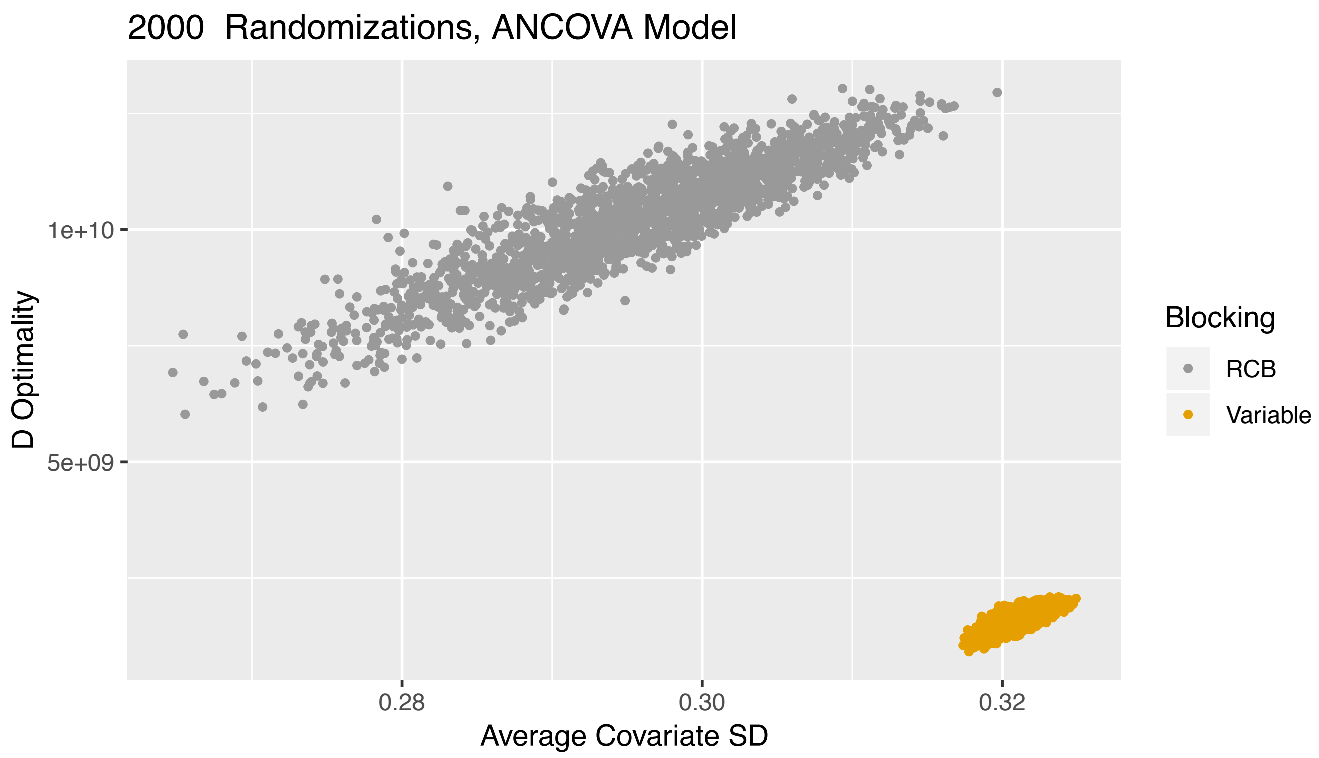
$$mean(\sigma_{x_1}, \ldots, \sigma_{x_t})$$

Optimality for ANCOVA

The additional information from blocks makes RCB designs more optimal than variable block designs

- Variable blocks allow us to recover information about the covariate only, and don't allow for other spatial variation

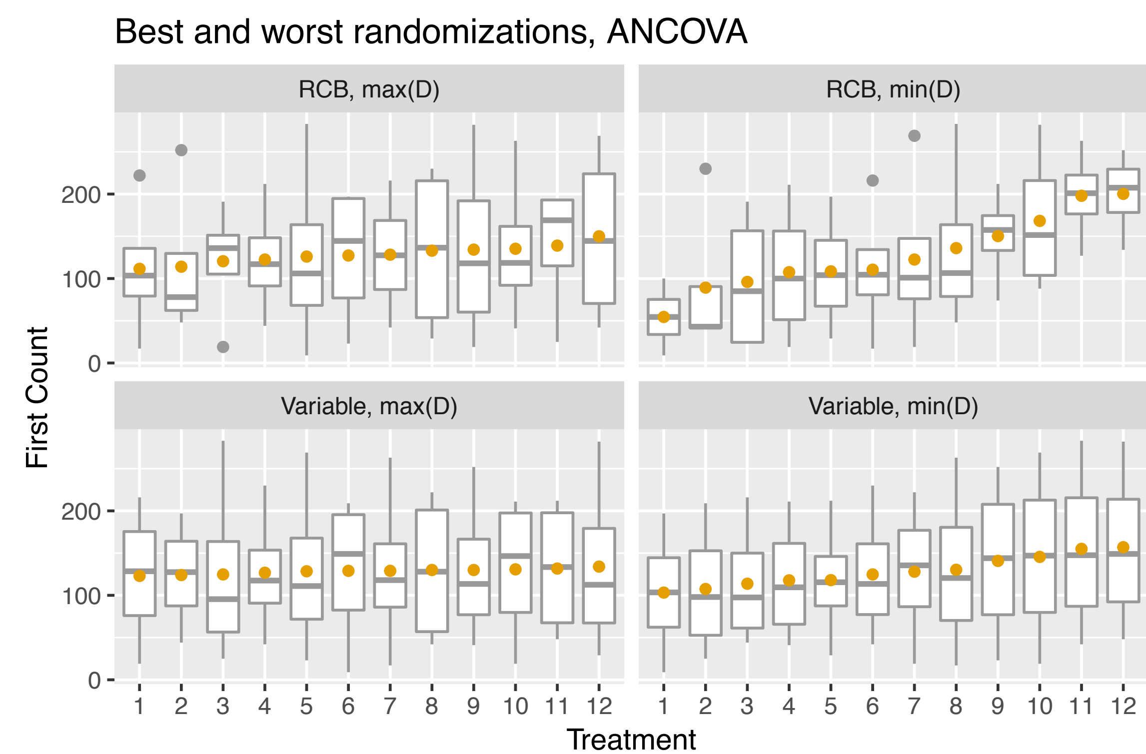




Treatment means and dispersions

Selecting between RCB or variable blocking requires a decision:

- Will measurable spatially-varying covariates have a greater impact on experimental outcome than unmeasurable variables?



Employing covariates **and** blocking variables

- Variable blocking generally produces more optimal designs, with respect to the analysis treatment response to covariate.

- Some RCB variable blocked designs can have treatment dispersions comparable to variable blocking.

- How can we find an optimal variable blocked design without selecting from an unknown number of possible randomizations?

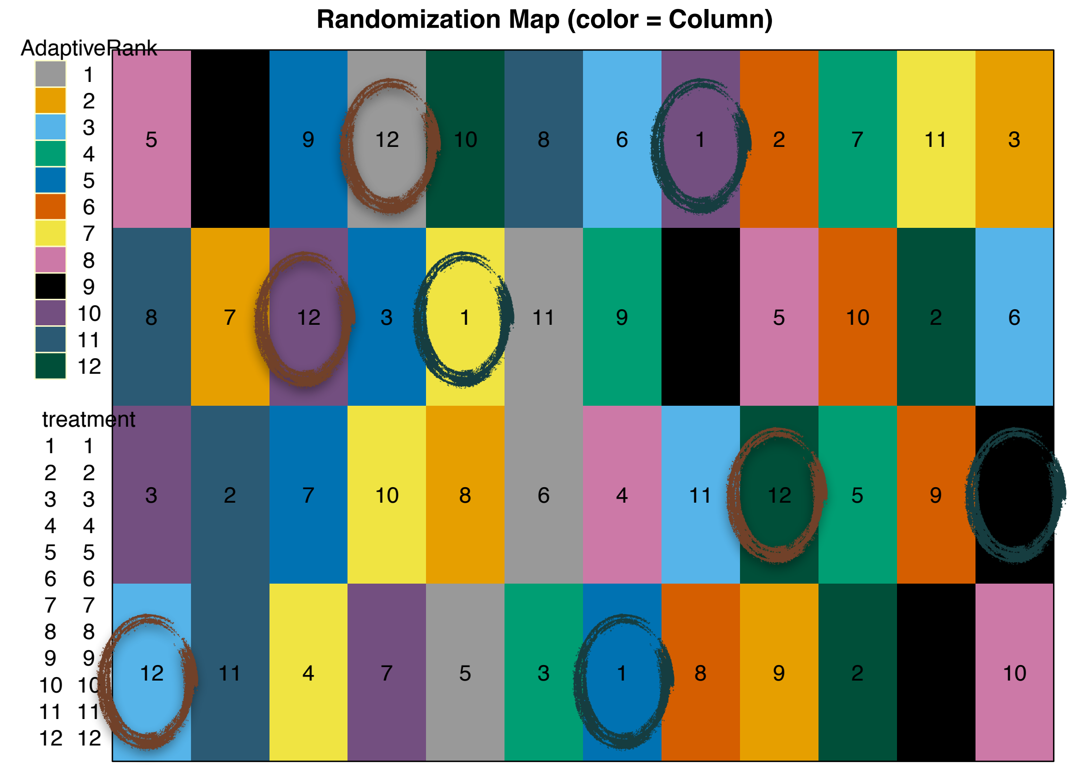
Two-factor blocking

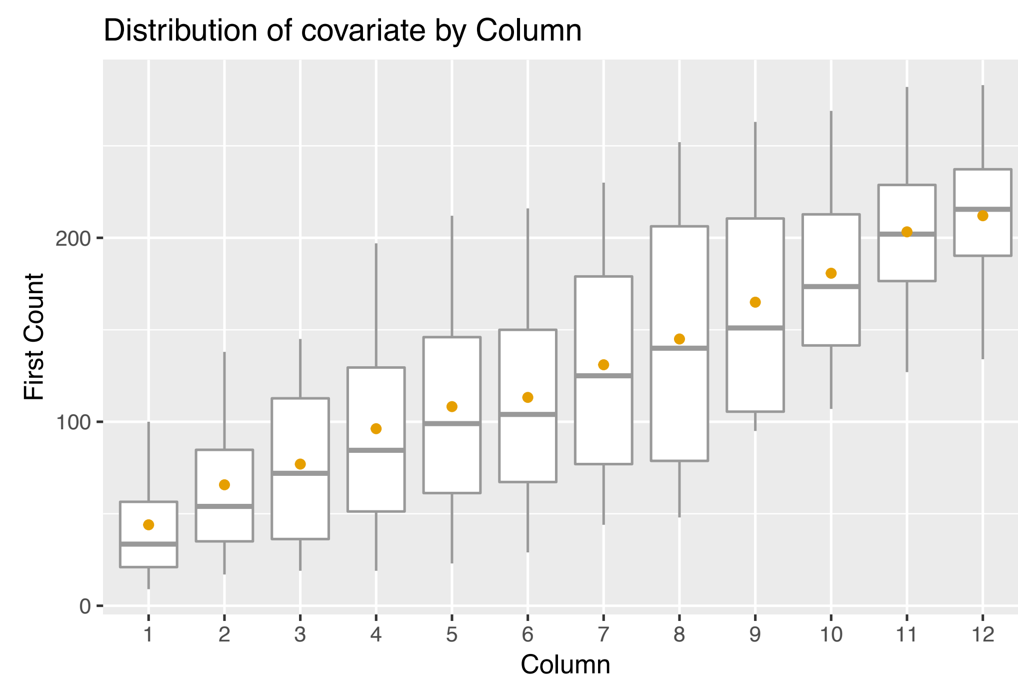
- Latin square blocks in two dimensions to eliminate two sources of nuisance variability
- Youden squares or partial Latin squares are complete blocks in one dimension and incomplete blocks in another dimension
    - If a treatment is applied to the plot with the lowest ranked covariate in one replicate, it will not be applied to the lowest ranked plot in any other replicate
- Will blocking with complete blocks and with variable blocks improve optimality, w.r.t covariate regression?

Variable Block Youden Square

No treatment is assigned to plots of the same covariate rank.

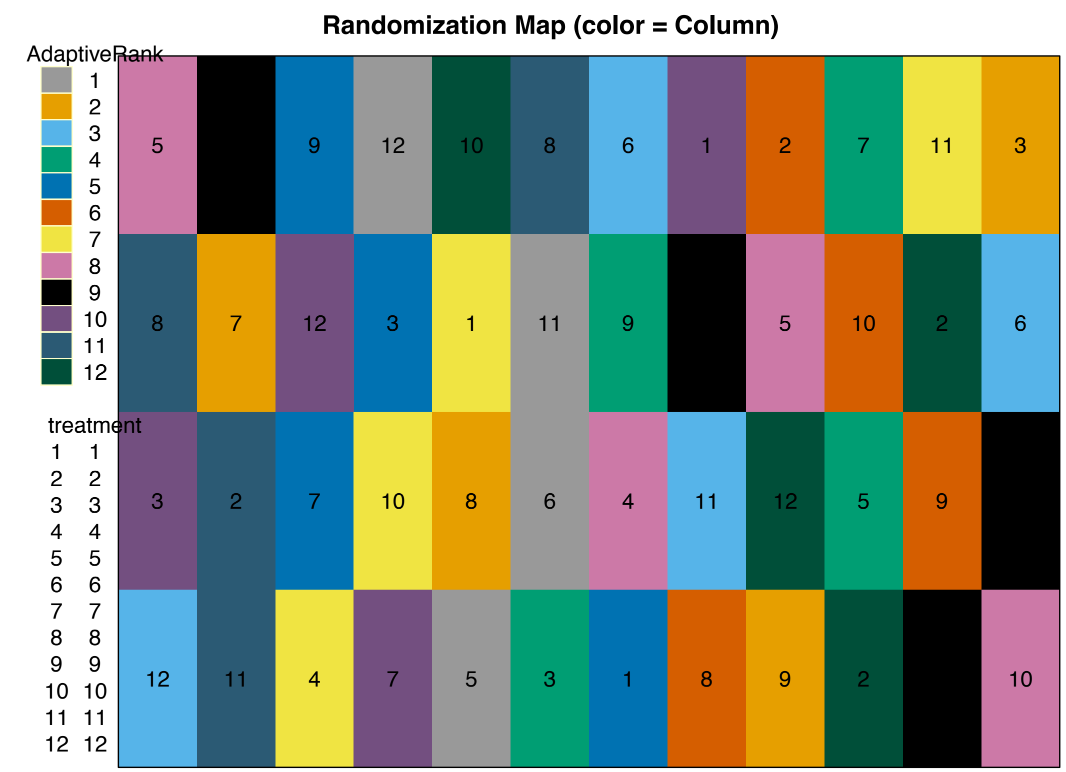
- Treatment 1 assigned to plots with ranks 5,7,9,10
- Treatment 12 assigned to plots with ranks 1,3,9,12





Variable Block Youden Square

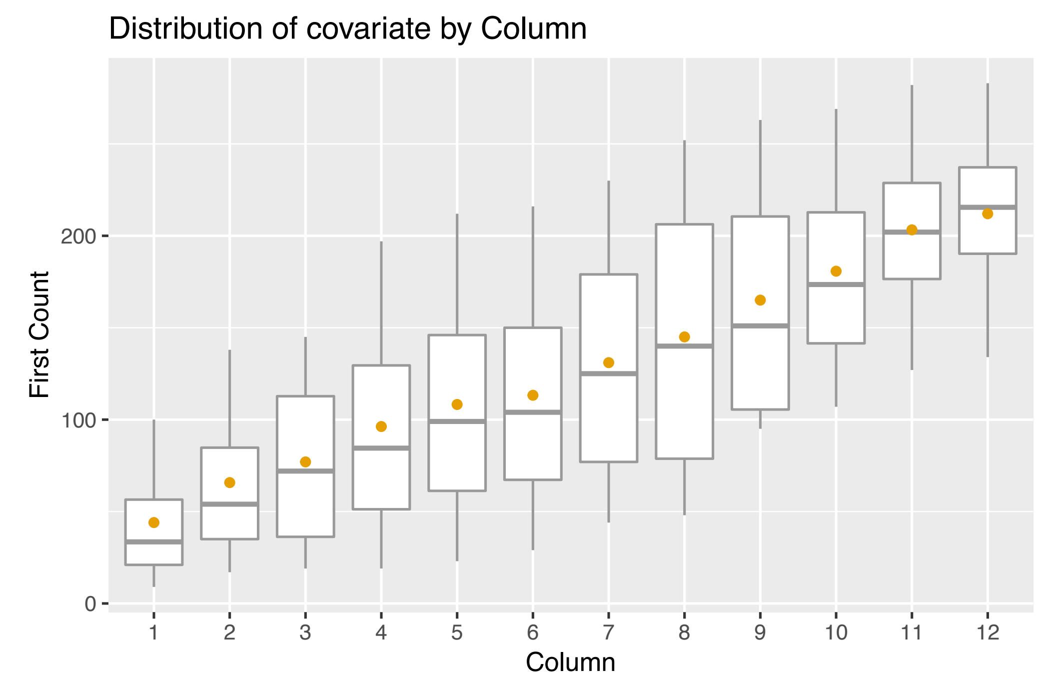
Using variables as a second blocking criteria improves upon RCB, w.r.t covariate regression, but does not approach the optimality of unconstrained variable blocking.

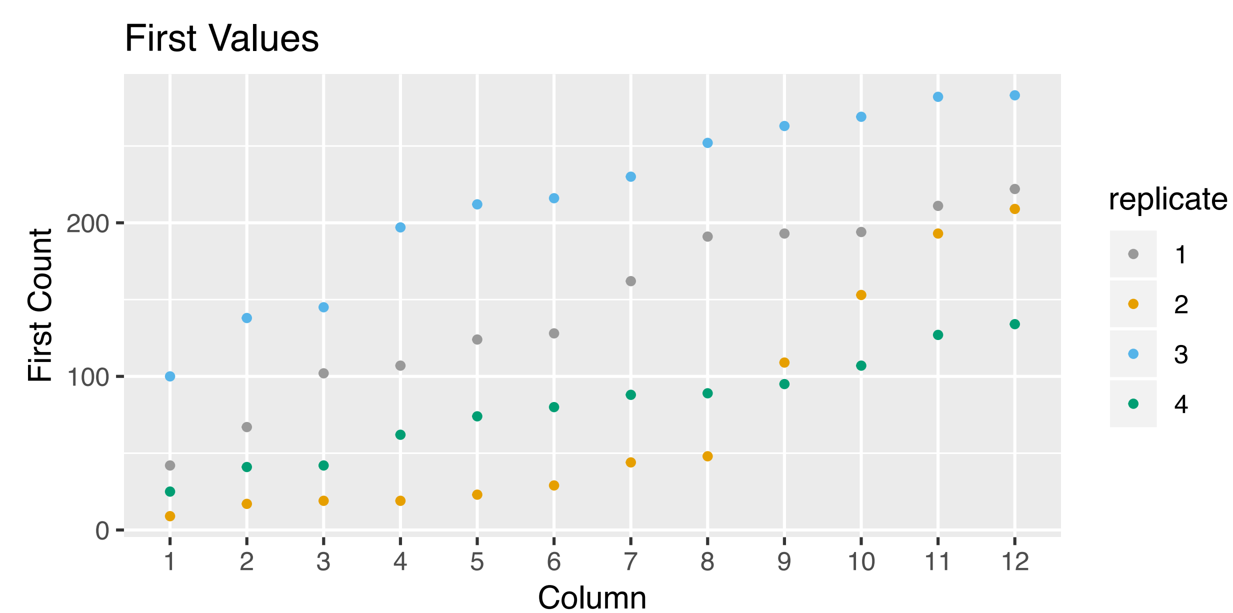


Variable Block Youden Square

Variable blocking in columns prevents treatments from being applied to plots of the same covariate rank, but does not disperse over plots of similar covariate rank.

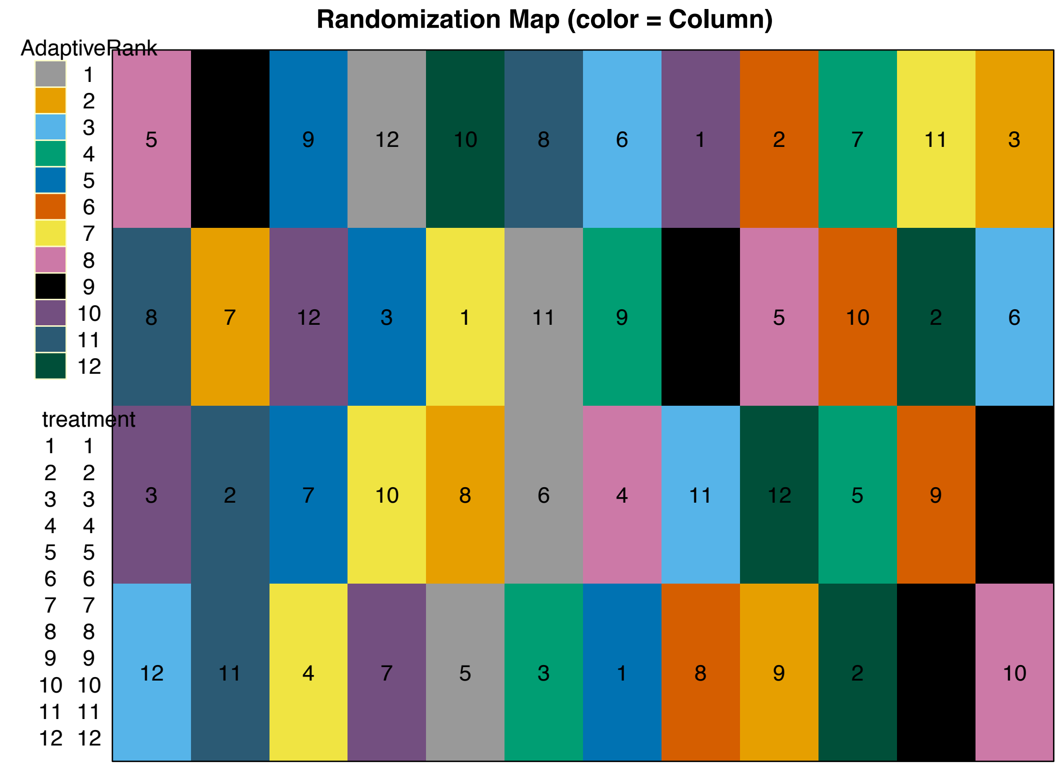
- Treatment 1 assigned to plots with ranks 5,7,9,10





Variable Block Youden Square

This design has the advantage of simplicity and relationship to other forms of row-column blocking.



Adaptive designs

- A static design applies the same randomization criteria for treatments across all phases of an experiment.
- An adaptive design alters treatment randomization based on intermediate measures.
- Can we randomize blocks in phases to maximum optimality at each phase?

An adaptive randomization

1. Randomize first replicate as usual.

2. Rank treatments in first block by covariate. Rank plots in second block by covariate. Assign treatments with the highest rank in block one to the plots with the lowest rank in block two.

3. Rank treatments in the first two blocks by covariate dispersion (standard deviation). Rank plots in the third block by deviation from the covariate mean in the first two blocks. Assign treatments with the smallest dispersion to the plots with the largest deviation.

4. Rank treatments in the first three blocks by the deviation of the treatment covariate mean from the covariate mean grand mean. Rank plots in the fourth block by deviation from the covariate mean in the first three blocks. Assign treatments with the smallest deviation to the plots with the largest deviation.

5. Alternate between 3 and 4 until all replicates have been assigned.

An adaptive randomization

1. Randomize first replicate as usual

2. Rank treatments in first block by covariate. Rank plots in second block by covariate. Assign treatments with the highest rank in block one to the plots with the lowest rank in block two.

    - This step minimizes the differences among treatment covariate means.

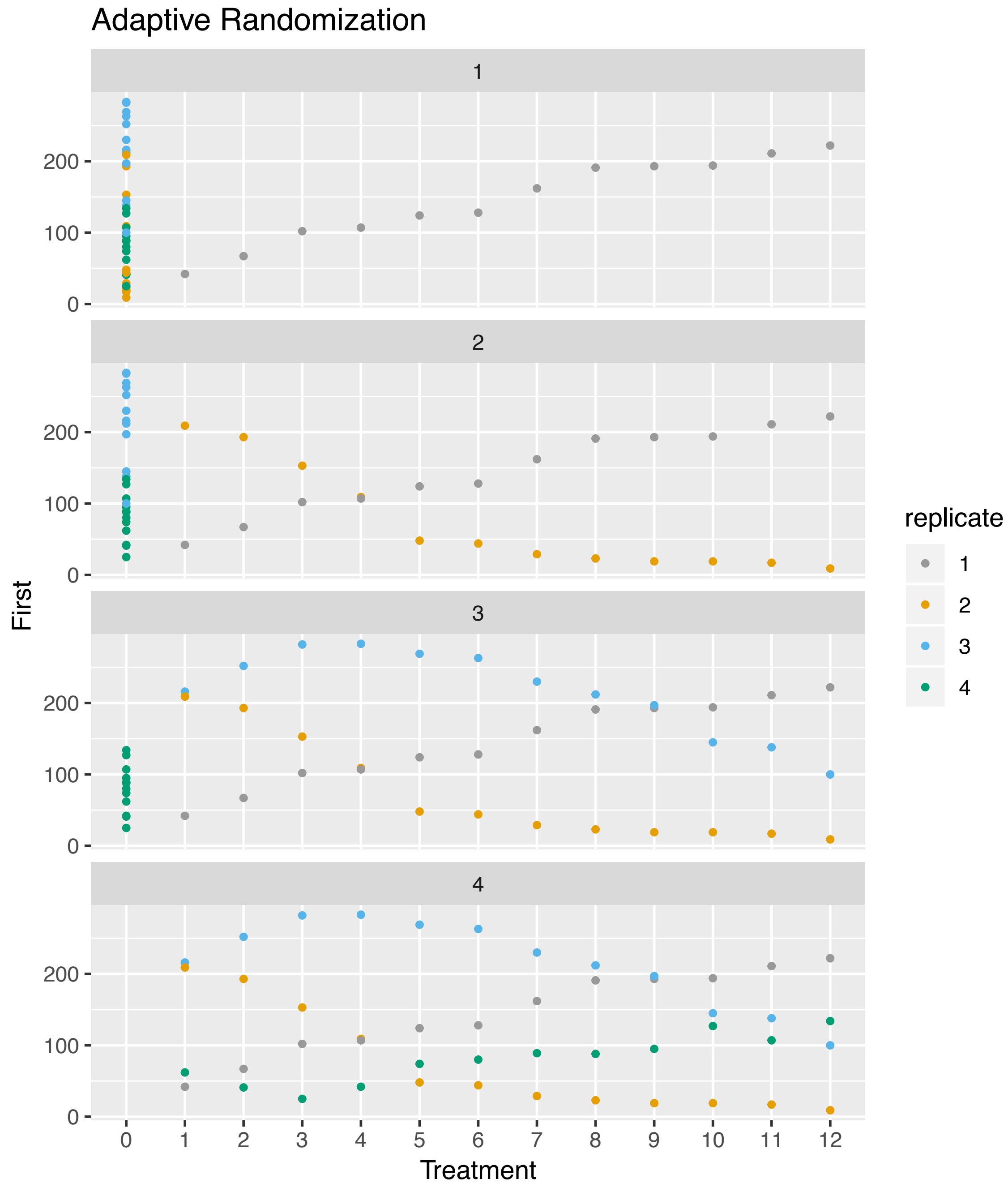
3. Rank treatments in the first two blocks by covariate dispersion (standard deviation). Rank plots in the third block by deviation from the covariate mean in the first two blocks. Assign treatments with the smallest dispersion to the plots with the largest deviation.

    - Maximize standard deviation of covariates over treatment means

4. Rank treatments in the first three blocks by the deviation of the treatment covariate mean from the covariate mean grand mean. Rank plots in the fourth block by deviation from the covariate mean in the first three blocks. Assign treatments with the smallest deviation to the plots with the largest deviation.

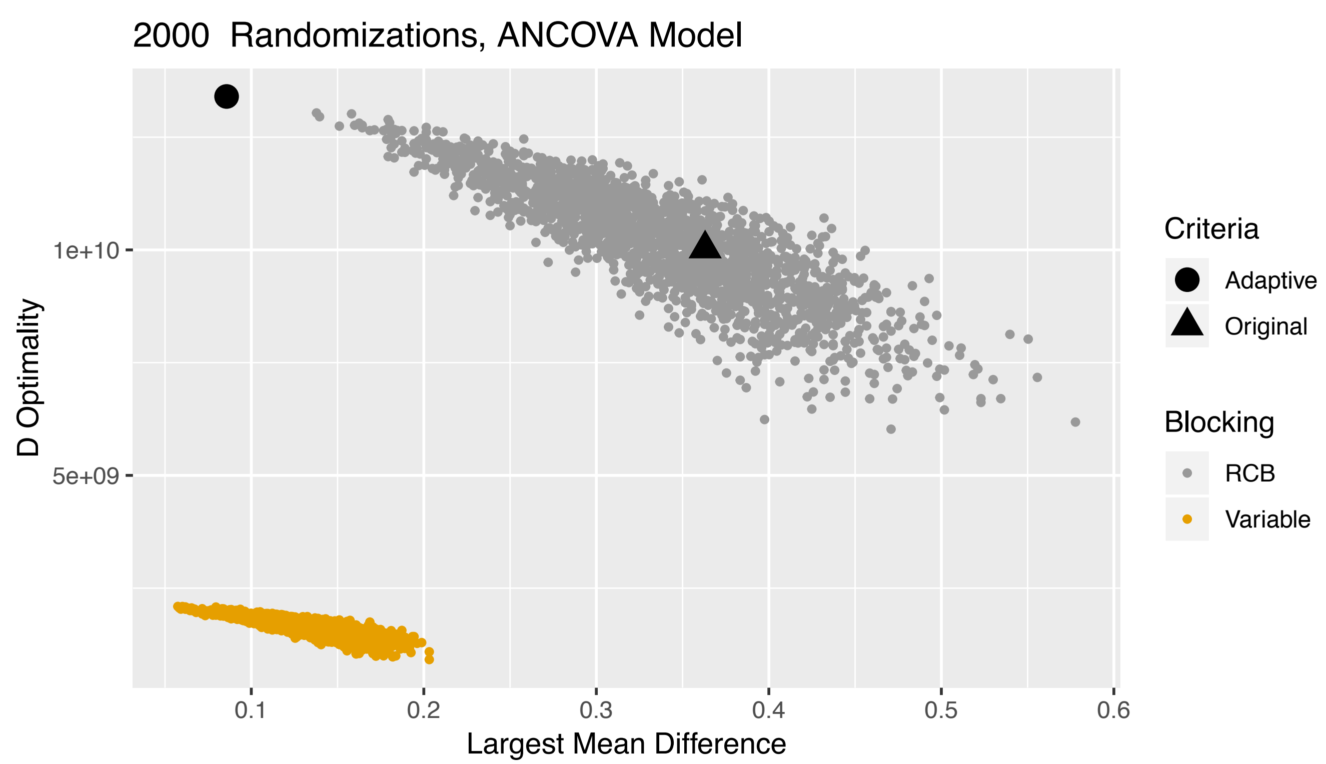
    - Minimize differences among treatment covariate means

5. Alternate between 3 and 4 until all replicates have been assigned.



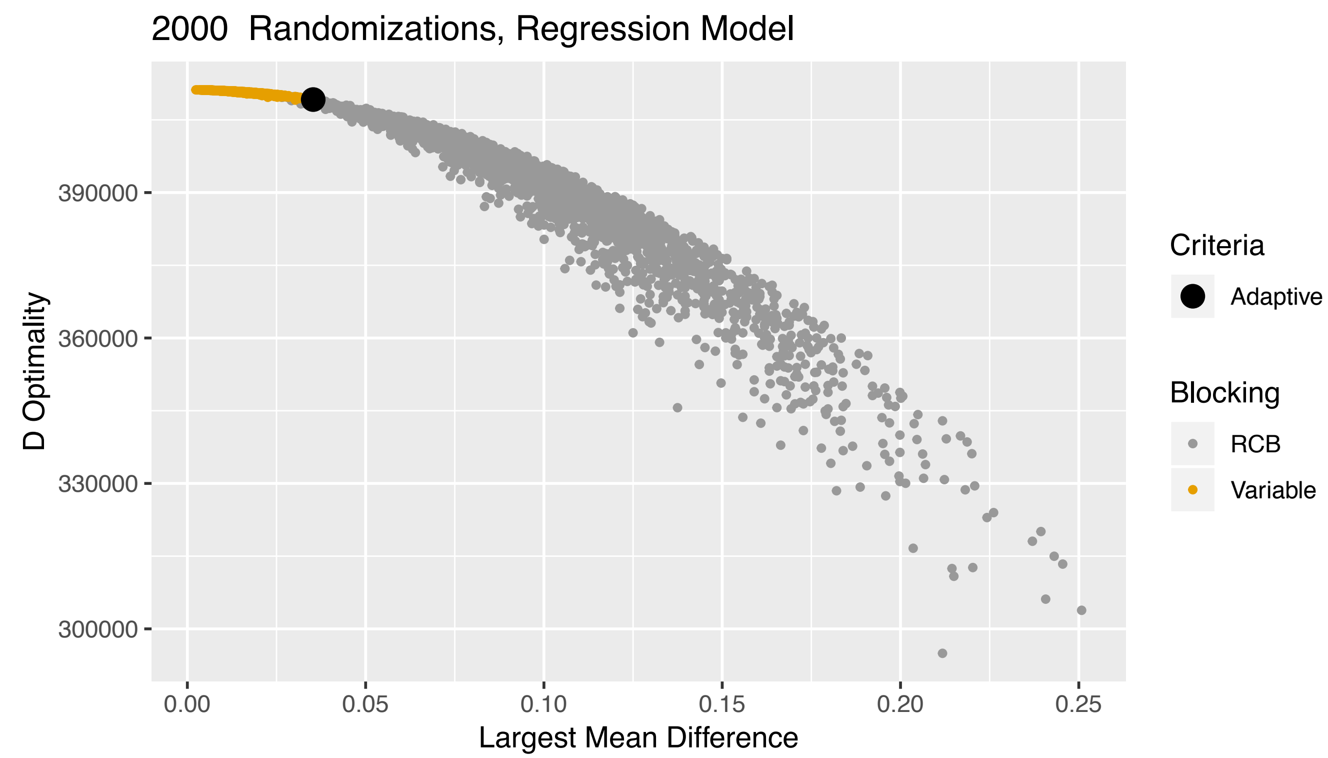
Optimality of an adaptive randomization

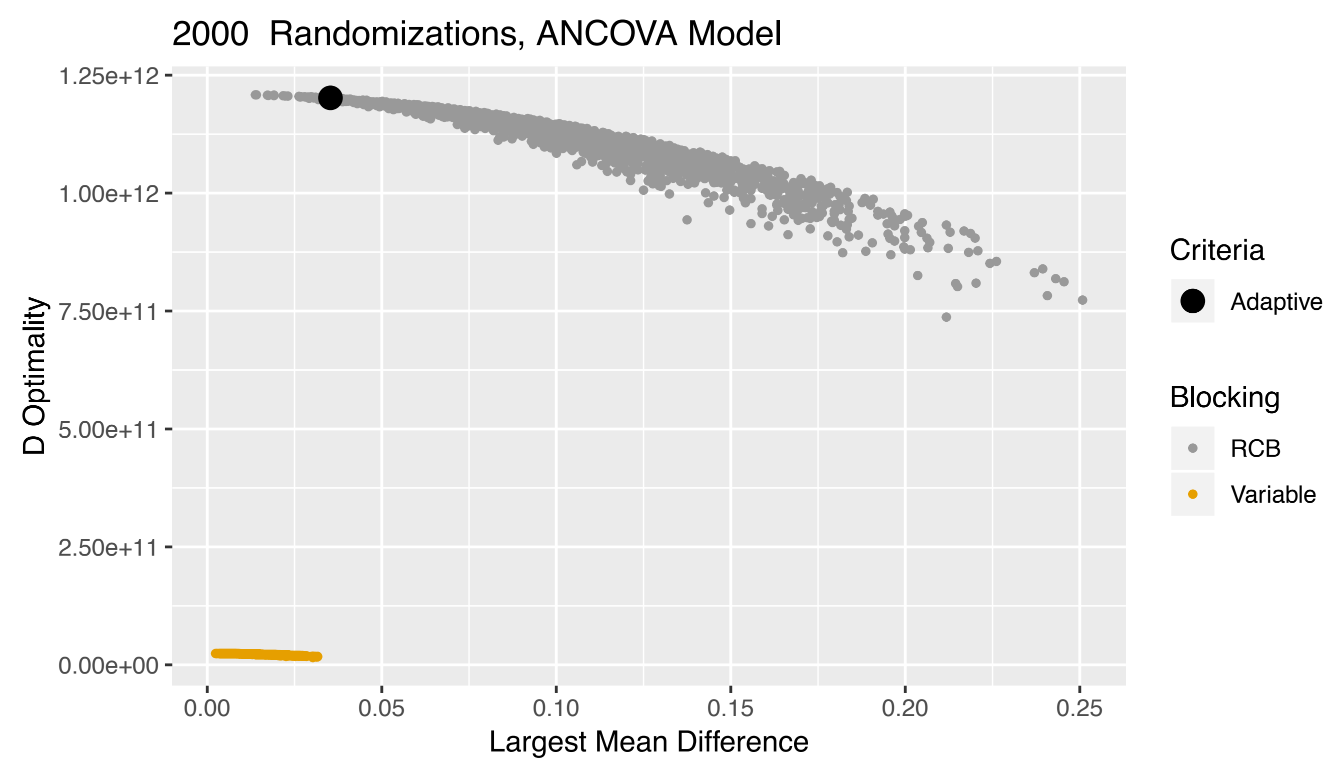
For this example, the proposed adaptive randomization finds a optimal design.



Optimality of an adaptive randomization

This is not a global optimum, so for some experimental data there can be more optimal designs.





George A. Milliken, Dallas E. Johnson (2001) Analysis of Messy Data, Volume III: Analysis of Covariance. Table 10.2

Conclusions

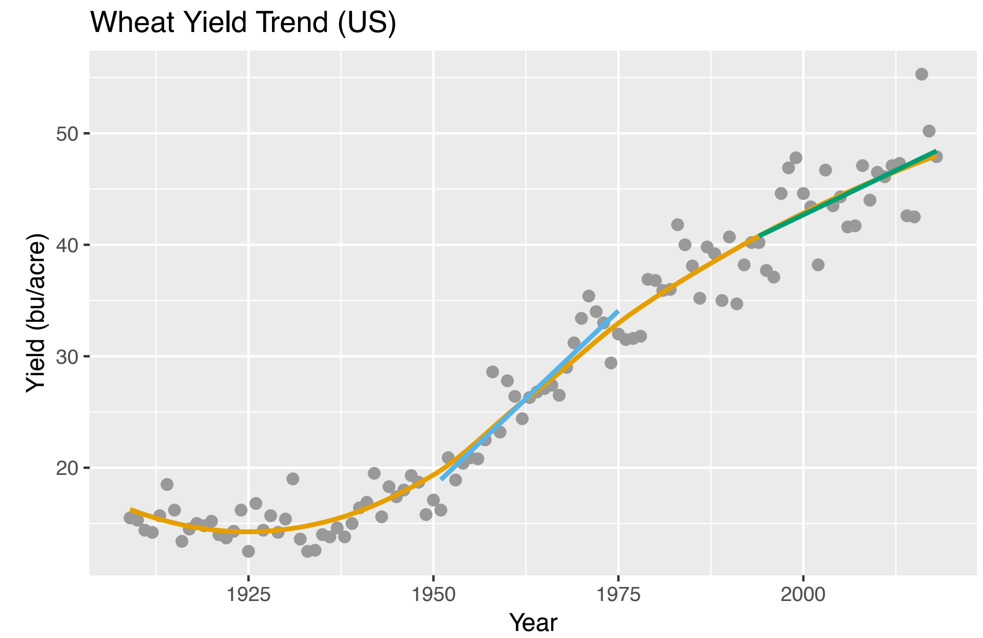
- Analysis of Covariance and Variable Blocking provide design and analysis models to control for measurable nuisance variables.

- A combination of traditional (RCB) blocking and Variable Blocking can be expected to improve upon RCB designs with respect to analysis of covariance, but not upon Variable Blocking.

- Adaptive Randomization can find RCB randomizations that optimize the Analysis of Covariance.

Appendix

- How important is decreasing σ?
- Can't we just increase δ?
- It depends…



What can we improve?

Winter Wheat, National Yield Trend

$$\delta_{1951-1975} = 0.63 \text{ bu/acre}$$
$$\delta_{1994-2018} = 0.32 \text{ bu/acre}$$

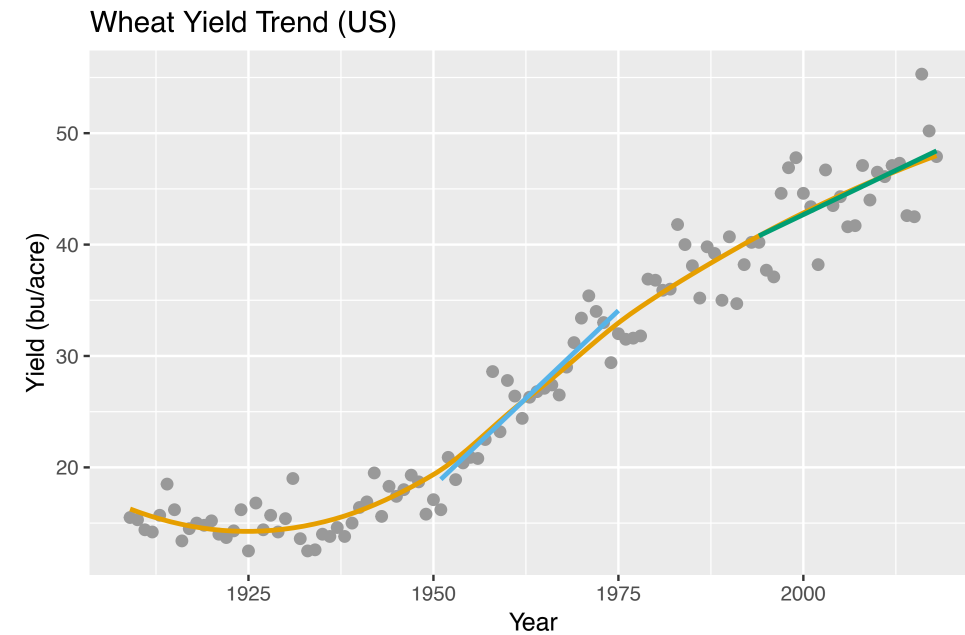
What can we improve?

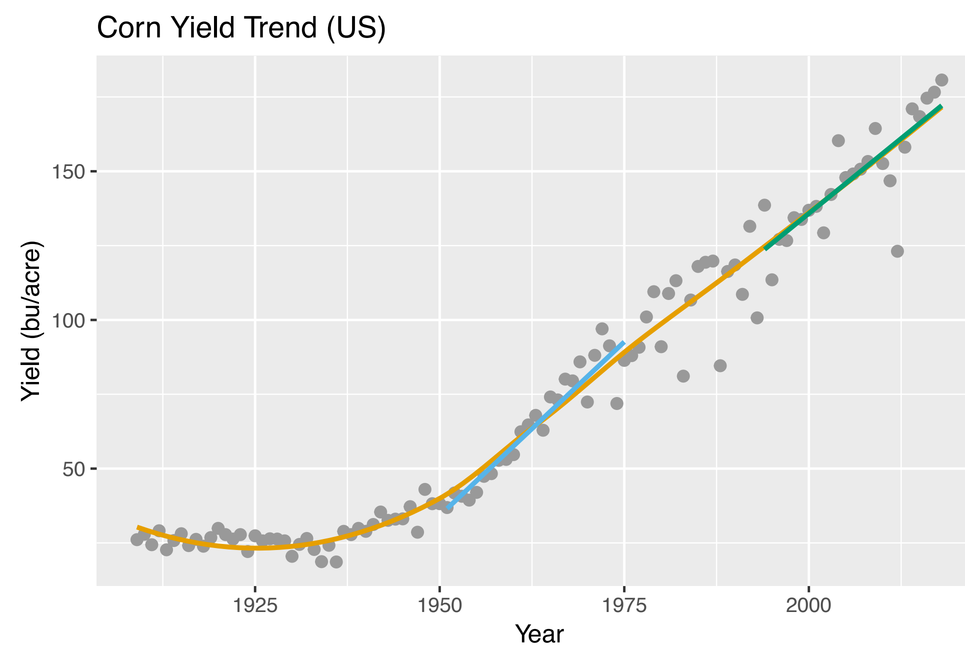
Winter Wheat, National Yield Trend

$$\delta_{1951-1975} = 0.63 \text{ bu/acre}$$
$$\delta_{1994-2018} = 0.32 \text{ bu/acre}$$

Corn, National Yield Trend

$$\delta_{1951-1975} = 2.34 \text{ bu/acre}$$
$$\delta_{1994-2018} = 2.02 \text{ bu/acre}$$





What can we improve?

Winter Wheat, National Yield Trend

$$\delta_{1951-1975} = 0.63 \text{ bu/acre}$$
$$\delta_{1994-2018} = 0.32 \text{ bu/acre}$$

Corn, National Yield Trend

$$\delta_{1951-1975} = 2.34 \text{ bu/acre}$$
$$\delta_{1994-2018} = 2.02 \text{ bu/acre}$$

Soybean, National Yield Trend

$$\delta_{1951-1975} = 0.33 \text{ bu/acre}$$
$$\delta_{1994-2018} = 0.55 \text{ bu/acre}$$