IMS Medallion Lecture, JSM, Washington DC, August 2009

From R. A. Fisher to Microarrays

Why 70 Year Old Theory is Relevant Today

George Casella
Department of Statistics
University of Florida
casella@ufl.edu

Overview

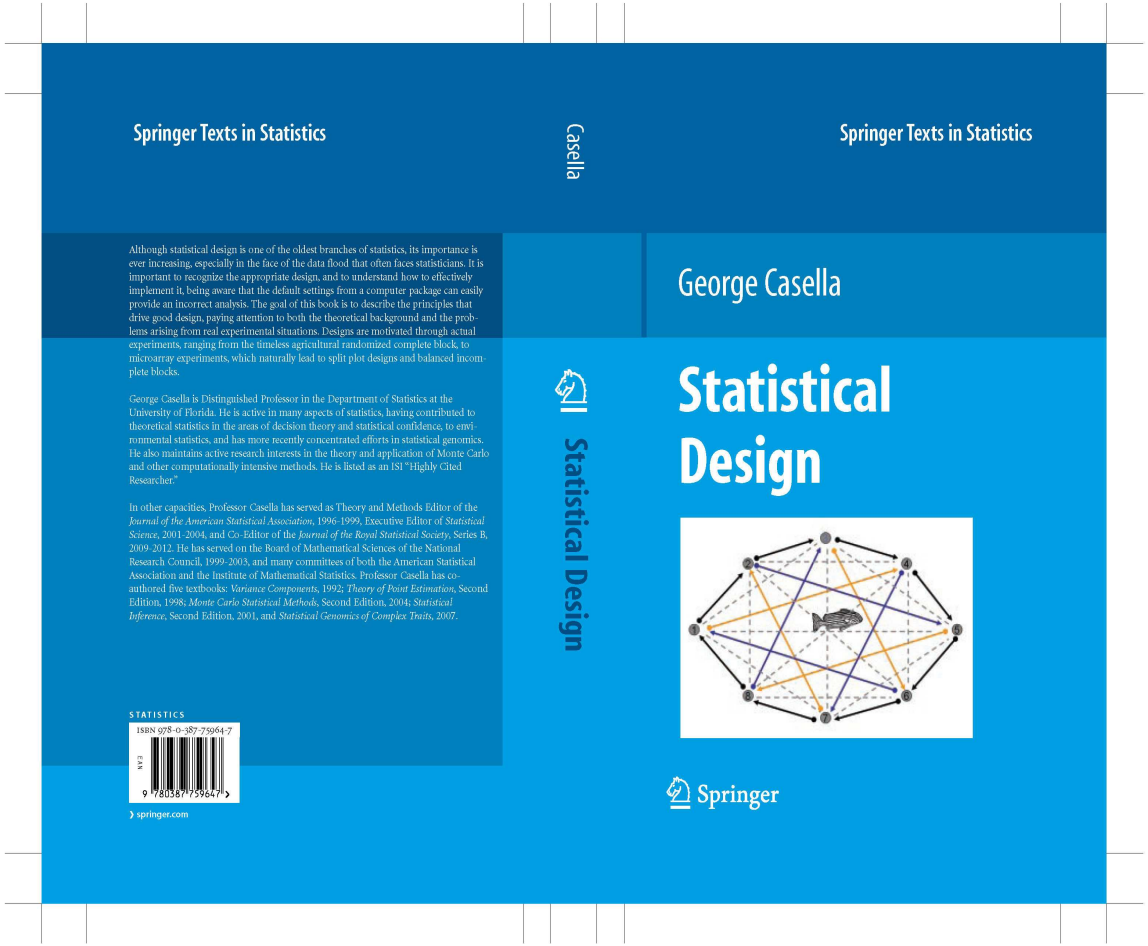
- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

- There is nothing here
  that your don't already know
- (or knew)
- but maybe not in this context

Overview

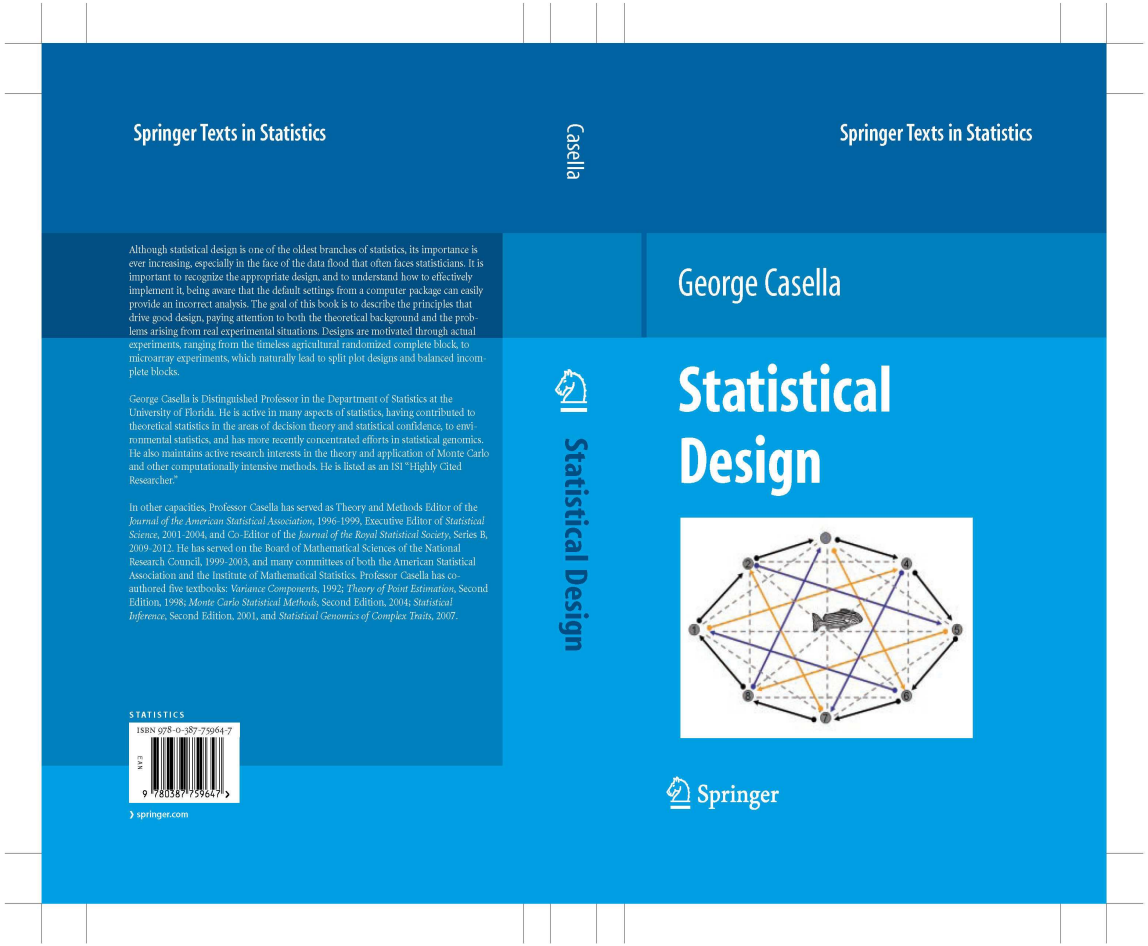
- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

- There is nothing here
  that your don't already know
- (or knew)
- but maybe not in this context



This is stuff I learned when writing this

▷ Shameless Advertising



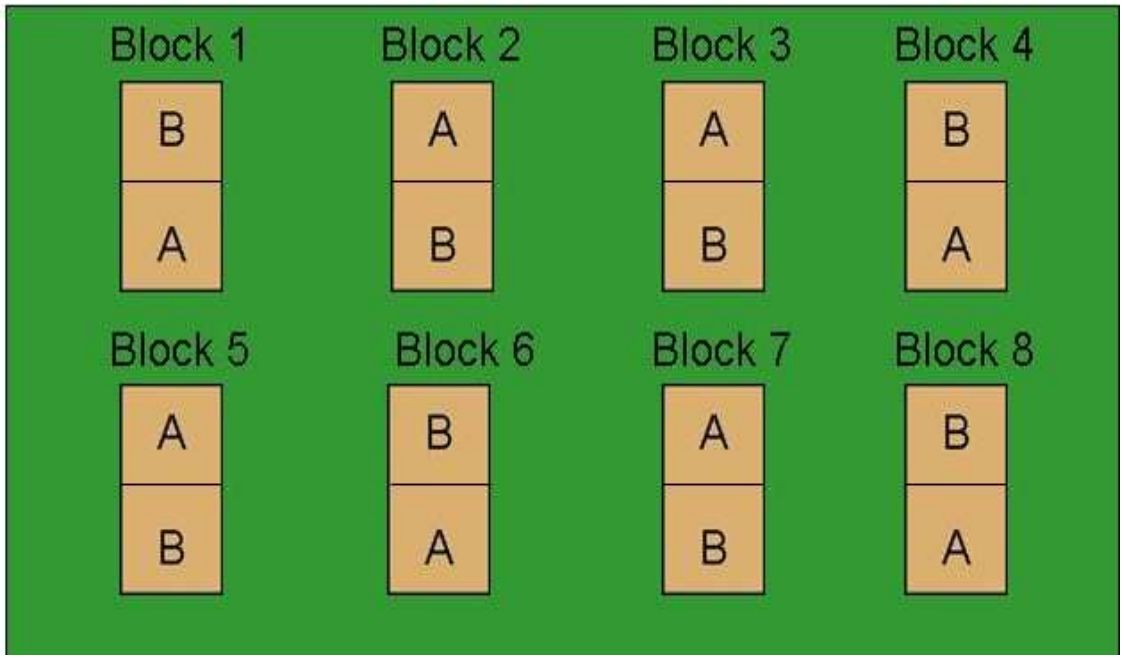
This is stuff I learned when writing this

▷ Shameless Advertising

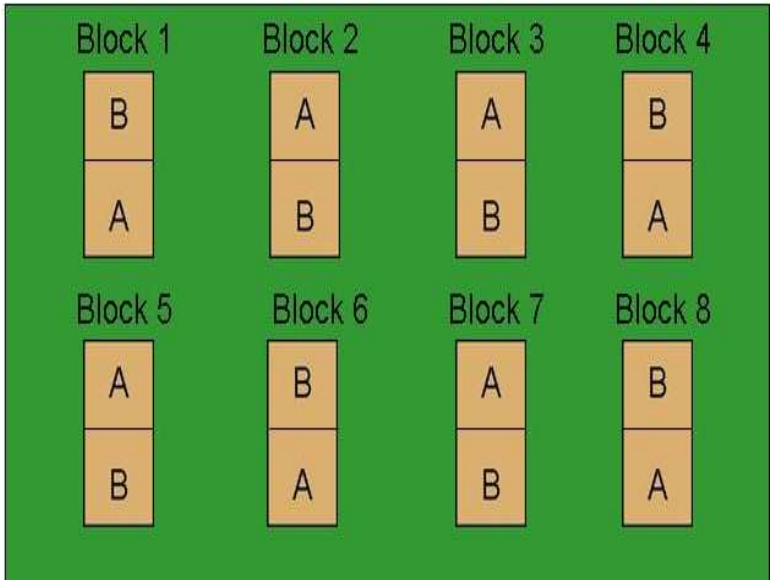
- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

A Classic Field Experiment

- The Field Plot Design might be



Analysis



Source	df
Blocks	7
Treatments	1
T $\times$ B	7

H_0: No Treatment Effect

$$F = \frac{\text{MS(Trts)}}{\text{MS(T} \times \text{B)}}$$

The Denominator

- Although we talk of tests here
- It is all about the denominator
  - If

$$F = \frac{\text{MS(Trts)}}{\text{MS(T} \times \text{B)}}$$

  - A confidence interval on treatment differences is

$$\text{Trt}_i - \text{Trt}_j \in \bar{Y}_i - \bar{Y}_j \pm t\sqrt{\text{MS(T} \times \text{B)}}$$

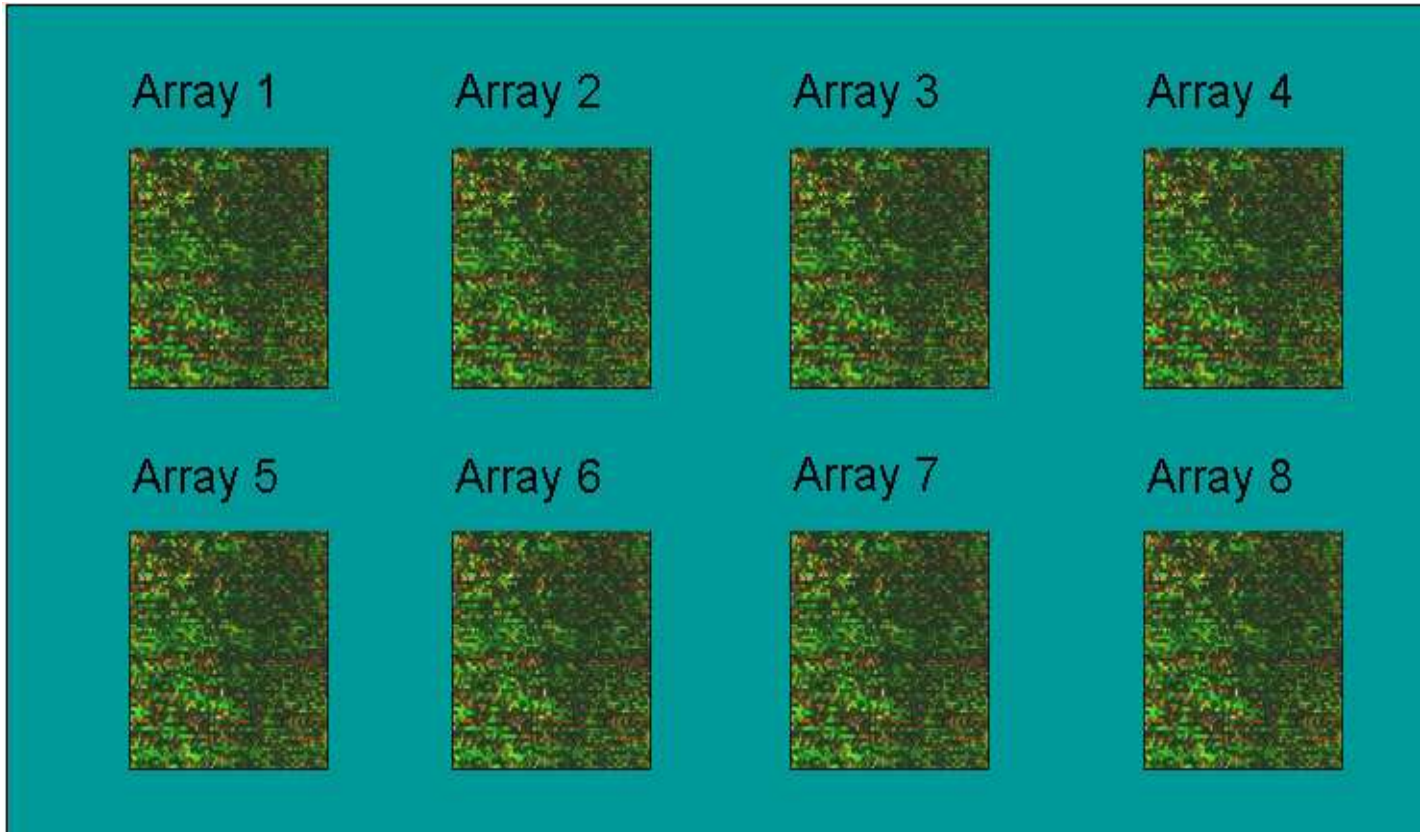
Where t is a t cutoff with MS(T $\times$ B) degrees of freedom

The Denominator

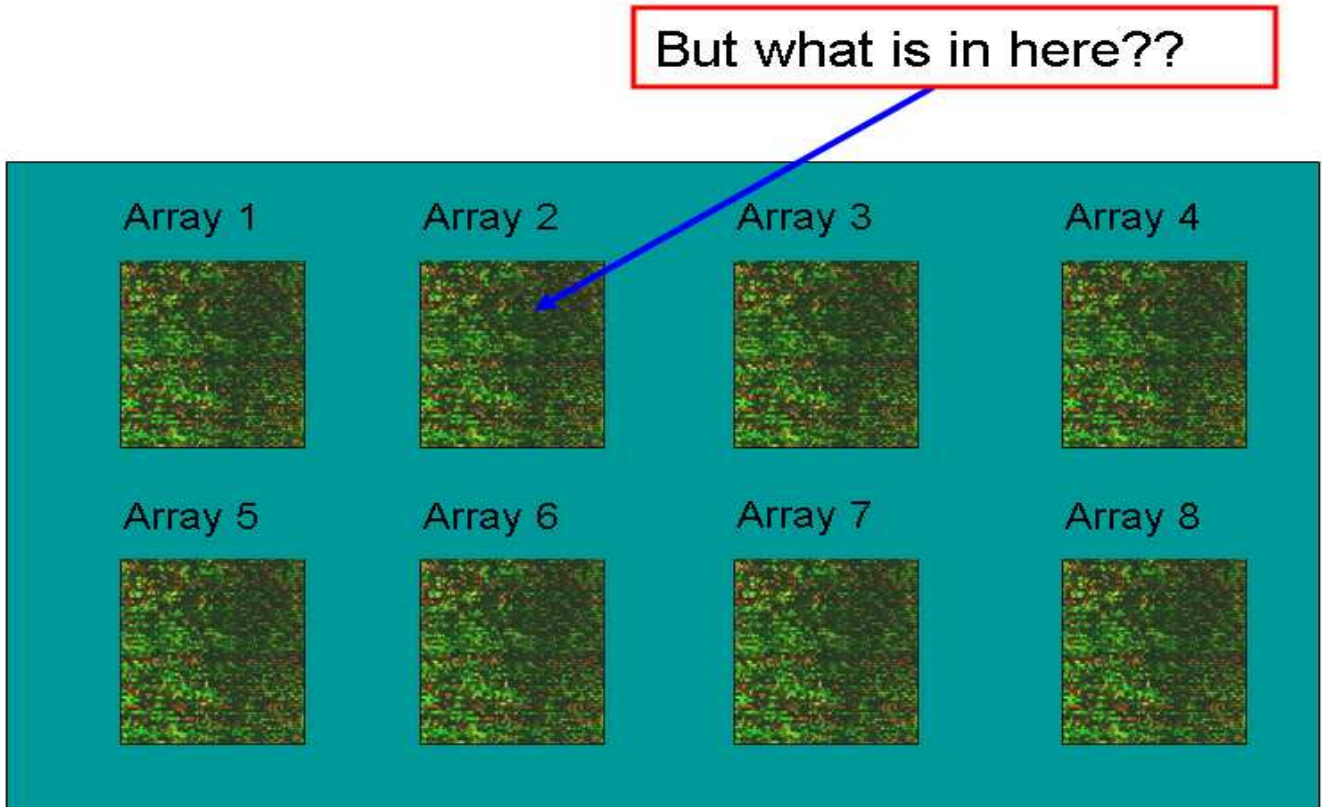
- Although we talk of tests here
- It is all about the denominator
  - If
$$F = \frac{\text{MS(Trts)}}{\text{MS(T} \times \text{B)}}$$
  - A confidence interval on treatment differences is
$$\text{Trt}_i - \text{Trt}_j \in \bar{Y}_i - \bar{Y}_j \pm t\sqrt{\text{MS(T} \times \text{B)}}$$
- Where t is a t cutoff with MS(T $\times$ B) degrees of freedom

A (Now) Classic Microarray Experiment

- The Microarray Design might be

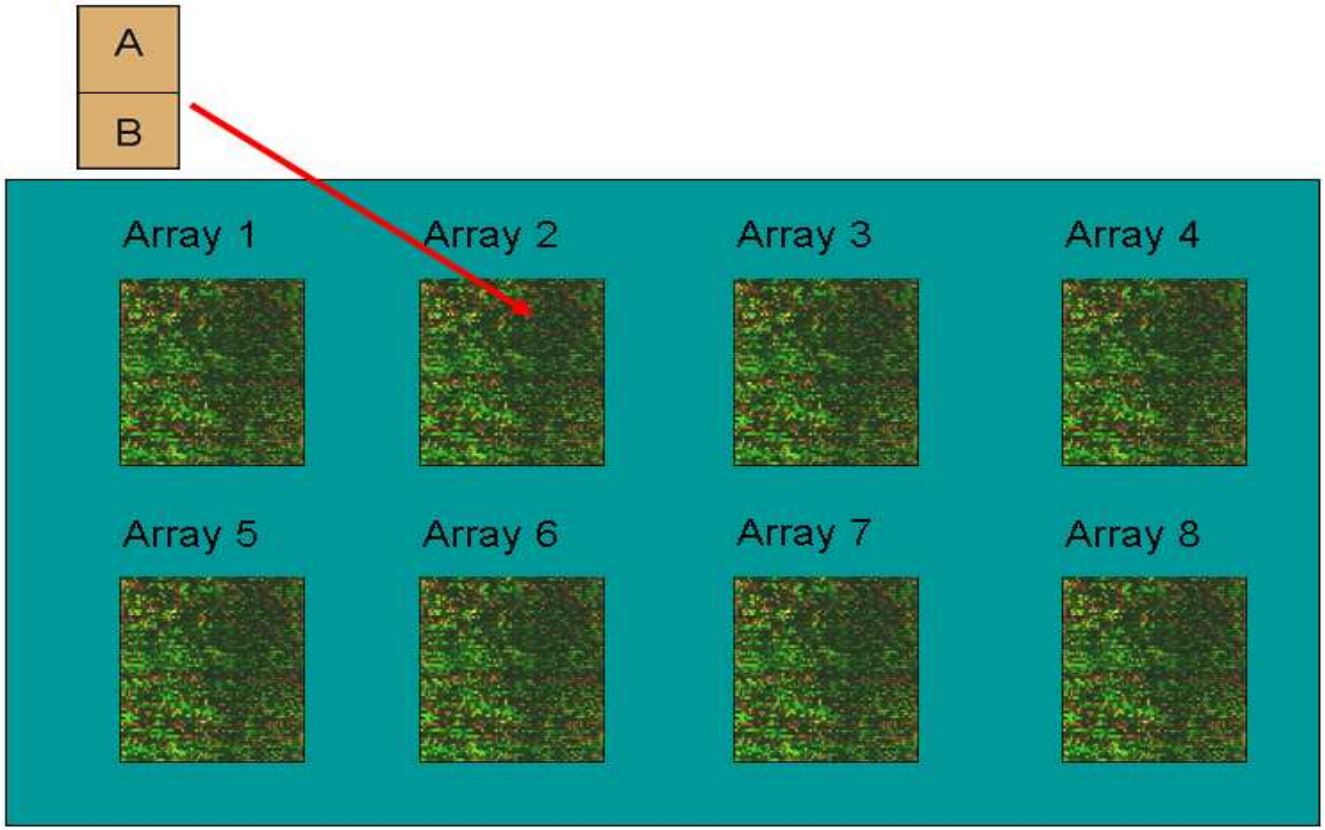


A Two-Dye Array

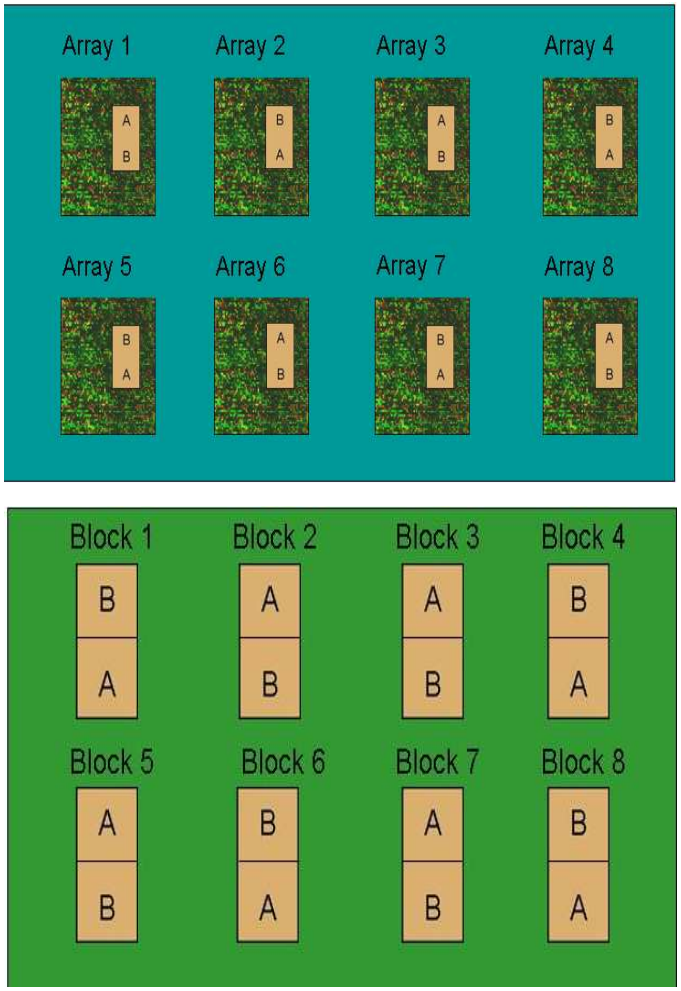


▷ The treatment comparisons!

A Randomized Block Design



▷ Look familiar?



Source	df
Blocks	7
Treatments	1
T $\times$ B	7

H_0: No Treatment Effect

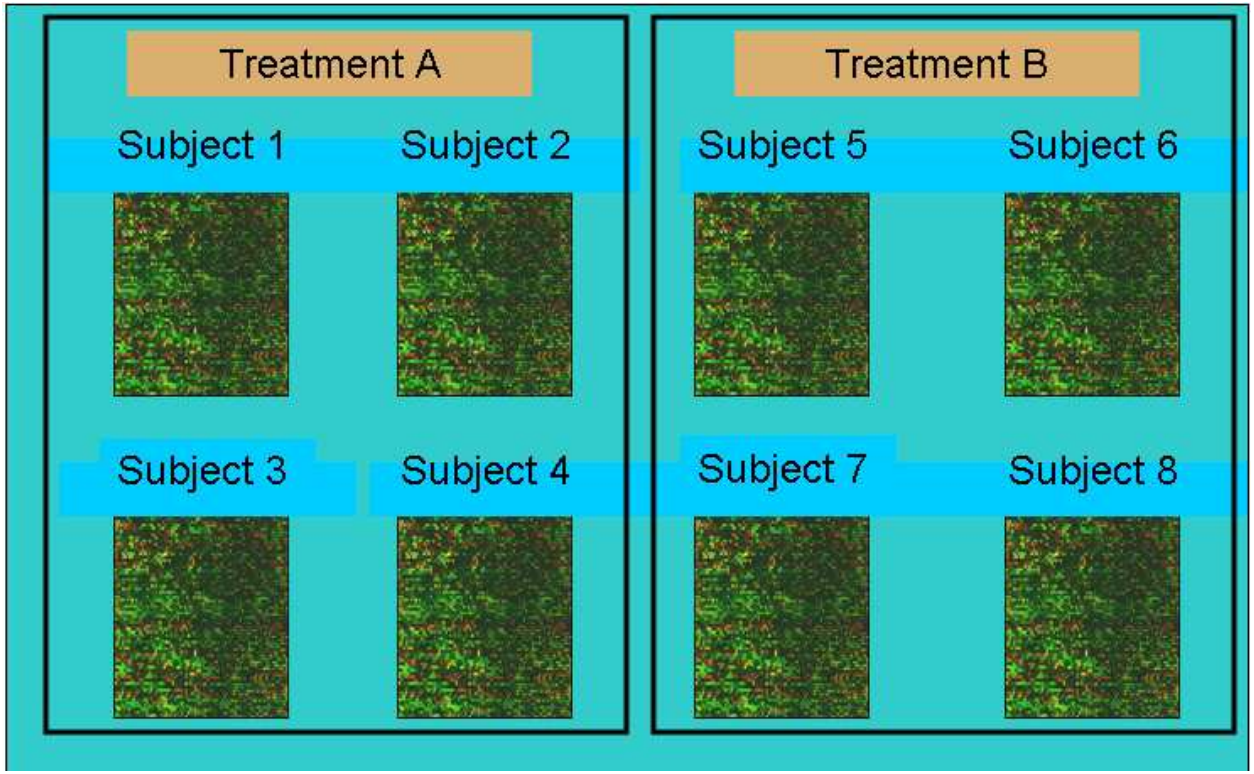
$$F = \frac{\text{MS(Trts)}}{\text{MS(T} \times \text{B)}}$$

"Get it right for one gene"

Two-Dye Systems

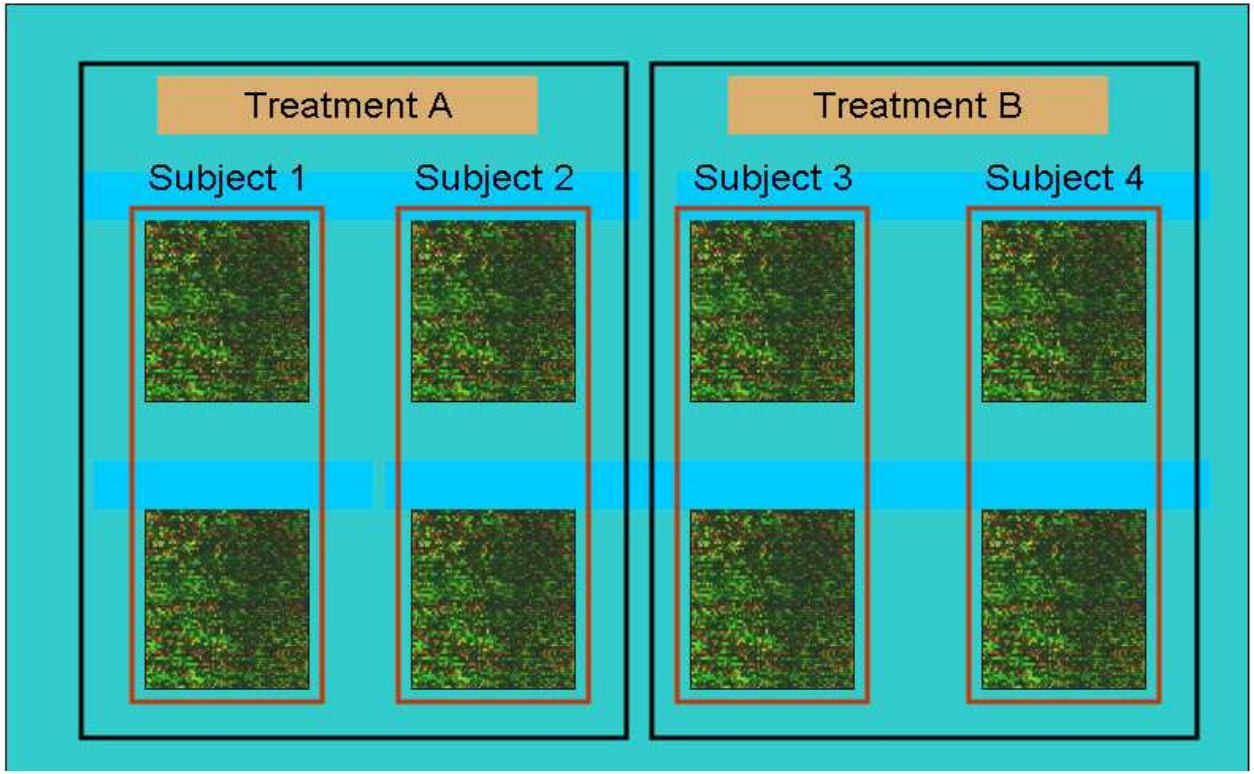
- A two-dye system
  - Leads to a blocking design, with blocks = subjects
  - Each block can have two treatments
  - Error term is from the Treatment $\times$ Subject Interaction
- Oligo arrays (Affymetrix) are a bit different.....

A Single Dye System

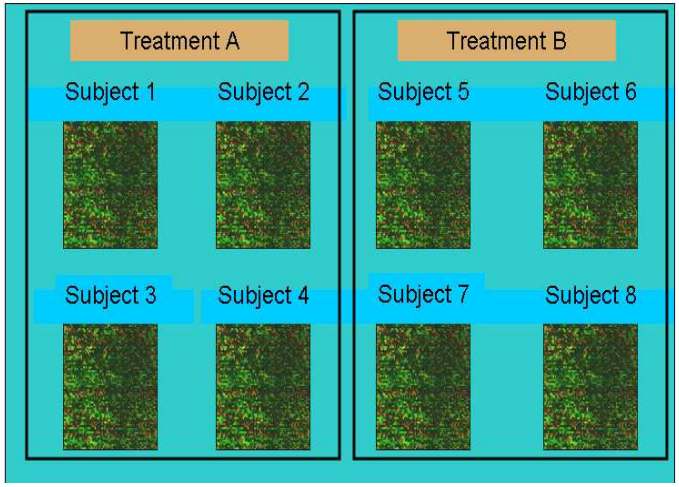


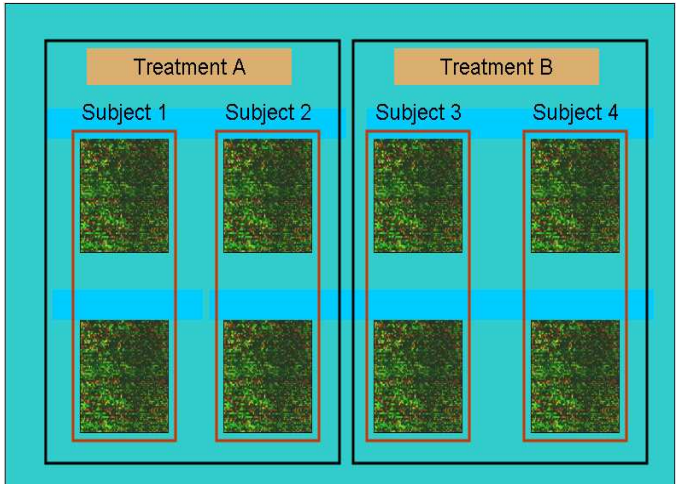
- A oneway anova

Avoid This Mistake



- A nested design - lost degrees of freedom





Oneway

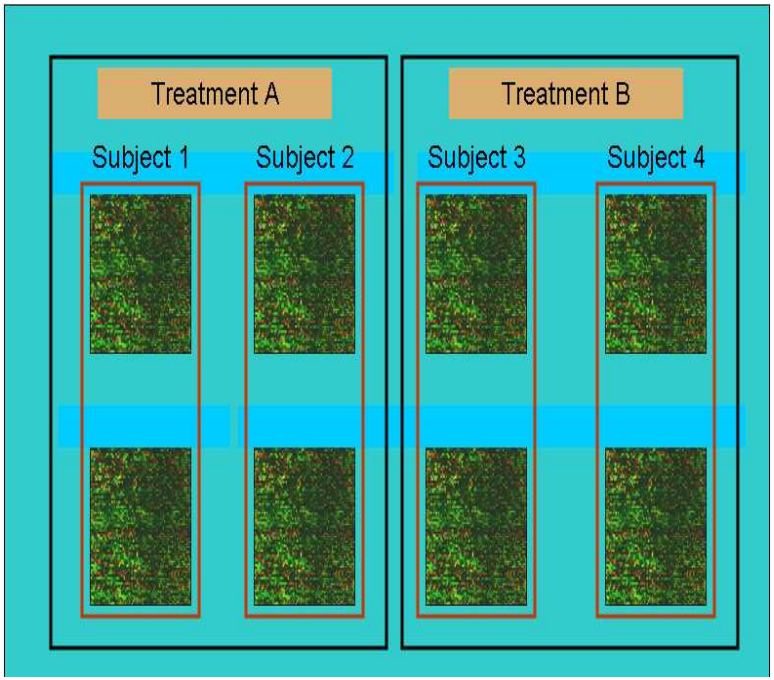
Source	df
Treatments	1
Subjects (in Treatments)	6

Nested

Source	df
Treatments	1
Subjects (in Treatments)	2
Arrays (in Subjects)	4

Error df in Red

Analysis of the Nested Design



Source	df
Treatments	1
Subjects (in Treatments)	2
Arrays (in Subjects)	4

- These are Technical Reps
- They measure Array Variability
- Useless for Treatment Variability

Design and Analysis

- Remember - We are talking of Design, not Analysis
  - We can design for one gene
  - Analysis is a bit more complicated
  - Often need to consider all (or a group of) genes
- But - a good design always helps the analysis
- Also - this talk is not just about microarrays
  - It is about Design Principles
  - Application in Life Sciences, Social Sciences, etc...

Design and Analysis

- Remember - We are talking of Design, not Analysis
  - We can design for one gene
  - Analysis is a bit more complicated
  - Often need to consider all (or a group of) genes
- But - a good design always helps the analysis
- Also - this talk is not just about microarrays
  - It is about Design Principles
  - Application in Life Sciences, Social Sciences, etc...

Design and Analysis

- Remember - We are talking of Design, not Analysis
  - We can design for one gene
  - Analysis is a bit more complicated
  - Often need to consider all (or a group of) genes
- But - a good design always helps the analysis
- Also - this talk is not just about microarrays
  - It is about Design Principles
  - Application in Life Sciences, Social Sciences, etc...

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

WWSD

- OR, why do I need statistics when I have a computer?
- Watch out for the default analysis
- Even more scary when the computer becomes the teacher!
  - This is not the computer's fault
  - We need to get the correct error term!

WWSD

- OR, why do I need statistics when I have a computer?
- Watch out for the default analysis
- Even more scary when the computer becomes the teacher!
  - This is not the computer's fault
  - We need to get the correct error term!

WWSD

- OR, why do I need statistics when I have a computer?
- Watch out for the default analysis
- Even more scary when the computer becomes the teacher!
  - This is not the computer's fault
  - We need to get the correct error term!

RCB models

- A "no interaction" RCB model

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad i = 1, \ldots, t, \quad j = 1, \ldots, b,$$

$\tau_i =$ Treatments

$\beta_j =$ Blocks

(1) $\varepsilon_{ij} \sim$ iid $\mathsf{N}(0, \sigma^2_\varepsilon)$,

(2) $\beta_1, \ldots, \beta_b$, are iid $\mathsf{N}(0, \sigma^2_\beta)$,

(3) Everything is independent.

RCB models

- An "interaction" RCB model

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk},$$
$$i = 1, \ldots, t, \quad j = 1, \ldots, b, \quad k = 1, \ldots, r,$$

(1) $\varepsilon_{ijk} \sim \quad \mathsf{N}(0, \sigma^2)$,

(2) $\beta_1, \ldots, \beta_b, \sim \quad \mathsf{N}(0, \sigma^2_\beta)$,

(3) $(\tau\beta)_{11}, \ldots, (\tau\beta)_{tb}$, are $\mathsf{N}(0, \sigma^2_{\tau\beta})$,

(4) Everything is independent

Interactions in RCB models

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad \text{or} \qquad Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

- Are These Really Different Models?
- Some Think So
- Some Textbooks Think So

As Seen in Some Textbooks

No interaction RCB model

Source	df
Blocks	b-1
Treatments	t-1
Residual	(b-1)(t-1)

H_0: No Treatment Effect

$$F = \frac{\text{MS(Trts)}}{\text{MS(Residual)}}$$

Interaction RCB model

Source	df
Blocks	b-1
Treatments	t-1
Interaction	(b-1)(t-1)
Error	bt(r-1)

H_0: No Treatment Effect

$$F = \frac{\text{MS(Trts)}}{\text{MS(Error)}} \quad \text{or} \quad F = \frac{\text{MS(Trts)}}{\text{MS(Pooled)}}$$

"The case of the jumping denominator"

$$F = \frac{\text{MS(Trts)}}{\text{MS(Residual)}}$$

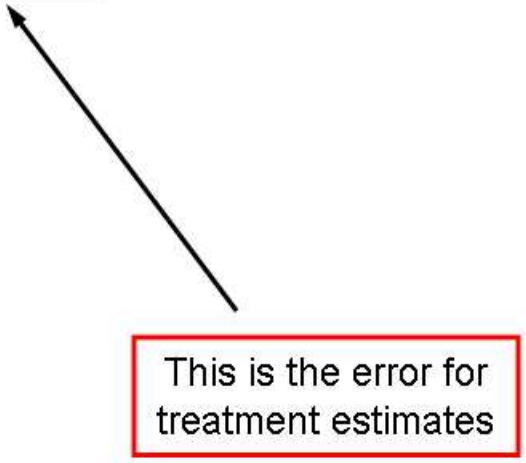
or

$$F = \frac{\text{MS(Trts)}}{\text{MS(Error)}}$$

- If the model changes does the truth change?
  - No interaction model $\not\Rightarrow$ No interaction!

Expected Mean Squares - Randomized Complete Blocks

Source	df	EMS
Blocks	$b-1$	$\sigma^2_\varepsilon + r\sigma^2_{\tau\beta} + rt\sigma^2_\beta$
Treatments	$t-1$	$\sigma^2_\varepsilon + r\sigma^2_{\tau\beta} + \frac{rt}{t-1}\sum_i(\tau_i - \bar{\tau})^2$
T × B	$(t-1)(b-1)$	$\sigma^2_\varepsilon + r\sigma^2_{\tau\beta}$
Within	$bt(r-1)$	σ^2_ε



- Works for
  - $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$
  - $Y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk}$

Fisher knew the correct errors

- Fisher's Advice:

  "We shall need to judge of the magnitude of the differences introduced by testing our treatments upon the different plots by the discrepancies between the performances of the same treatment in different blocks."

R. A. Fisher *The Design of Experiments, Section* 26

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

Revisiting the Field and the Lab

- Alfalfa Variety Trials
  - Six Blocks
  - Four Varieties
  - Three Plants/ Variety/Block
- Varieties
  - Ladak
  - Narragansett
  - DuPuits
  - Flamand
- Objective: Increase Yield

- Microarray Experiment
  - Six Subjects
  - Four Treatments
  - Three Arrays/Trt/Subject
- Treatment to Stem Cells
  - Control
  - Chemical
  - GFP(Green Flor. Protein)
  - GFP+transplant
- Convert Stem Cells to Neurons

Revisiting the Field and the Lab

- Alfalfa Variety Trials
  - Six Blocks
  - Four Varieties
  - Three Plants/ Variety/Block
- Varieties
  - Ladak
  - Narragansett
  - DuPuits
  - Flamand
- Objective: Increase Yield

- Microarray Experiment
  - Six Subjects
  - Four Treatments
  - Three Arrays/Trt/Subject
- Treatment to Stem Cells
  - Control
  - Chemical
  - GFP(Green Flor. Protein)
  - GFP+transplant
- Convert Stem Cells to Neurons

Alfalfa Variety Trial

		Blocks			
		1	2	$\cdots$	6
Varieties	1	y_{111}	y_{121}	$\cdots$	y_{161}
		y_{112}	y_{122}	$\cdots$	y_{162}
		y_{113}	y_{123}	$\cdots$	y_{163}
		$\vdots$	$\vdots$	$\vdots$	$\vdots$
	4	y_{411}	y_{421}	$\cdots$	y_{461}
		y_{412}	y_{422}	$\cdots$	y_{462}
		y_{413}	y_{423}	$\cdots$	y_{463}

Source	df	Sum Sq	Mean Sq	F	p
Block	5	3.98	1.33		
Variety	3	37.20	12.40	26.07	$< .0001$
V $\times$ B	15	4.28	0.48	1.88	0.050
Within	48	8.10	0.25		

Microarray Stem Cell

		Subjects			
		1	2	$\cdots$	6
Treatments	1	y_{111}	y_{121}	$\cdots$	y_{161}
		y_{112}	y_{122}	$\cdots$	y_{162}
		y_{113}	y_{123}	$\cdots$	y_{163}
		$\vdots$	$\vdots$	$\vdots$	$\vdots$
	4	y_{411}	y_{421}	$\cdots$	y_{461}
		y_{412}	y_{422}	$\cdots$	y_{462}
		y_{413}	y_{423}	$\cdots$	y_{463}

Source	df	Sum Sq	Mean Sq	F	p
Subjects	5	0.95	0.19		
Treatments	3	14.37	4.79	17.44	$< .0001$
T $\times$ S	15	4.12	0.27	39.81	$< .0001$
Within	48	15.84	0.33	0.01	

Alfalfa Variety Trial

		Blocks			
		1	2	$\cdots$	6
Varieties	1	y_{111}	y_{121}	$\cdots$	y_{161}
		y_{112}	y_{122}	$\cdots$	y_{162}
		y_{113}	y_{123}	$\cdots$	y_{163}
		$\vdots$	$\vdots$	$\vdots$	$\vdots$
	4	y_{411}	y_{421}	$\cdots$	y_{461}
		y_{412}	y_{422}	$\cdots$	y_{462}
		y_{413}	y_{423}	$\cdots$	y_{463}

Source	df	Sum Sq	Mean Sq	F	p
Block	5	3.98	1.33		
Variety	3	37.20	12.40	26.07	$< .0001$
V $\times$ B	15	4.28	0.48	1.88	0.050
Within	48	8.10	0.25		

Microarray Stem Cell

		Subjects			
		1	2	$\cdots$	6
Treatments	1	y_{111}	y_{121}	$\cdots$	y_{161}
		y_{112}	y_{122}	$\cdots$	y_{162}
		y_{113}	y_{123}	$\cdots$	y_{163}
		$\vdots$	$\vdots$	$\vdots$	$\vdots$
	4	y_{411}	y_{421}	$\cdots$	y_{461}
		y_{412}	y_{422}	$\cdots$	y_{462}
		y_{413}	y_{423}	$\cdots$	y_{463}

Source	df	Sum Sq	Mean Sq	F	p
Subjects	5	0.95	0.19		
Treatments	3	14.37	4.79	17.44	$< .0001$
T $\times$ S	15	4.12	0.27	39.81	$< .0001$
Within	48	15.84	0.33	0.01	

Revisiting the Field and the Lab

Source	df	Sum Sq	Mean Sq	F	p
Block	5	3.98	1.33		
Variety	3	37.20	12.40	26.07	$< .0001$
V $\times$ B	15	4.28	0.48	1.88	0.050
Within	48	8.10	0.25		

Source	df	Sum Sq	Mean Sq	F	p
Subjects	5	0.95	0.19		
Treatments	3	14.37	4.79	17.44	$< .0001$
T $\times$ S	15	4.12	0.27	39.81	$< .0001$
Within	48	15.84	0.33	0.01	

What to do with all of those Within Degrees of Freedom?

- Better Design Strategy: Increase the Number of Blocks
  - subjects
  - biological reps
  - arrays

Revisiting the Field and the Lab

Source	df	Sum Sq	Mean Sq	F	p
Block	5	3.98	1.33		
Variety	3	37.20	12.40	26.07	$< .0001$
V $\times$ B	15	4.28	0.48	1.88	0.050
Within	48	8.10	0.25		

Source	df	Sum Sq	Mean Sq	F	p
Subjects	5	0.95	0.19		
Treatments	3	14.37	4.79	17.44	$< .0001$
T $\times$ S	15	4.12	0.27	39.81	$< .0001$
Within	48	15.84	0.33	0.01	

Testing $H_0 : \sigma^2_{\tau\beta} = 0$?

Pooling? – Better to avoid

- Maybe
  - Type II error $\downarrow$
  - Type I error $\uparrow$

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

- For the RCB model

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk},$$

  - we have

$$\text{Corr}(\varepsilon_{ijk}, \varepsilon_{i'jk'}) = \begin{cases} \rho_\varepsilon & \text{for technical replication} \\ 0 & \text{for true replication.} \end{cases}$$

  - which affects the error term

- Look at the Expected Mean Squares

- For the RCB model

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk},$$

  - we have

$$\text{Corr}(\varepsilon_{ijk}, \varepsilon_{i'jk'}) = \begin{cases} \rho_\varepsilon & \text{for technical replication} \\ 0 & \text{for true replication.} \end{cases}$$

  - which effects the error term

- Look at the Expected Mean Squares

- RCB anova - technical replication

Source	df	EMS
Blocks	$b-1$	$\sigma^2_\varepsilon[1+(r-1)\rho_\varepsilon]+r\sigma^2_{\tau\beta}+rt\sigma^2_\beta$
Treatments	$t-1$	$\sigma^2_\varepsilon[1+(r-1)\rho_\varepsilon]+r\sigma^2_{\tau\beta}+\frac{rt}{t-1}\sum_i(\tau_i-\bar{\tau})^2$
T $\times$ B	$(t-1)(b-1)$	$\sigma^2_\varepsilon[1+(r-1)\rho_\varepsilon]+r\sigma^2_{\tau\beta}$
Within	$bt(r-1)$	$(1-\rho_\varepsilon)\sigma^2_\varepsilon$

- Treatment Test OK

- RCB anova - technical replication

Source	df	EMS
Blocks	$b-1$	$\sigma^2_\varepsilon[1+(r-1)\rho_\varepsilon]+r\sigma^2_{\tau\beta}+rt\sigma^2_\beta$
Treatments	$t-1$	$\sigma^2_\varepsilon[1+(r-1)\rho_\varepsilon]+r\sigma^2_{\tau\beta}+\frac{rt}{t-1}\sum_i(\tau_i-\bar{\tau})^2$
T $\times$ B	$(t-1)(b-1)$	$\sigma^2_\varepsilon[1+(r-1)\rho_\varepsilon]+r\sigma^2_{\tau\beta}$
Within	$bt(r-1)$	$(1-\rho_\varepsilon)\sigma^2_\varepsilon$

- Cannot test the interaction - often an important test

Example

- Microarray Experiment on Unirradiated vs. Irradiated Cells
- Twoway crossed treatment design

	Treatment U	Treatment I
Line 1	x	x
	x	x
Line 2	x	x
	x	x

Independent Replication?

Technical Replication?

Cell Line Anova

- Technical Replication $\Rightarrow$ No interaction test in RCB

	Treatment	
	U	I
	x	x
Line 1	x	x
	x	x
Line 2	x	x

True Rep

Source	df
Blocks(Lines)	1
Treatments(U/I)	1
L $\times$ T	1
Within	4
Total	7

Technical Rep

Source	df
Blocks(Lines)	1
Treatments(U/I)	1
L $\times$ T	1
Subsampling	4
Total	7

- Subsampling = Pseudoreplication = Technical Replication

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

- Balanced Incomplete Block Designs (BIBD)
  - Are needed when all treatments cannot fit in one block
  - Arise naturally in microarray experiments
- Some Examples
  - Two-Dye System $\Rightarrow$ Block of Size 2
  - GCP Example - Treatment applied to cells - Subject is block
  - Other Technological advances

- Balanced Incomplete Block Designs (BIBD)
  - Are needed when all treatments cannot fit in one block
  - Arise naturally in microarray experiments
- Some Examples
  - Two-Dye System $\Rightarrow$ Block of Size 2
  - GCP Example - Treatment applied to cells - Subject is block
  - Other Technological advances

- The new Agilent arrays

- "Mulitplex formats enable hybridization of multiple samples on a single chip"
- Available formats
  - 1x244K
  - 2x105K
  - 4x44K
  - 8x15K
- A Statistical Design Nightmare?

- The new Agilent arrays

- "Mulitplex formats enable hybridization of multiple samples on a single chip"
- Available formats
  - 1x244K
  - 2x105K
  - 4x44K
  - 8x15K
- A Statistical Design Nightmare?

Properties of Balanced Incomplete Block Designs

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

▷ $\varepsilon_{ij} \sim$ iid $\mathsf{N}(0, \sigma^2_\varepsilon)$.

▷ $\beta_1, \ldots, \beta_b$, are iid $\mathsf{N}(0, \sigma^2_\beta)$

- Treatment contrasts are free of block effects

$$\mathrm{Var}\left(\sum_{i=1}^{t} a_i \hat{\tau}_i\right) = \frac{k}{\lambda t}\sigma^2_\varepsilon \sum_{i=1}^{t} a_i^2.$$

▷ no σ^2_β

Properties of Balanced Incomplete Block Designs

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

▷ $\varepsilon_{ij} \sim$ iid $\mathsf{N}(0, \sigma^2_\varepsilon)$.

▷ $\beta_1, \ldots, \beta_b$, are iid $\mathsf{N}(0, \sigma^2_\beta)$

- Treatment contrasts are free of block effects

$$\mathrm{Var}\left(\sum_{i=1}^{t} a_i \hat{\tau}_i\right) = \frac{k}{\lambda t}\sigma^2_\varepsilon \sum_{i=1}^{t} a_i^2.$$

▷ no σ^2_β

What can go wrong?

- NASA Experiment on Substantiality of Crops (Potatoes)
  - Two Factors: Photoperiod (P), and bioactive Tuber Inducing Factor solution (TIF)
  - Each at two levels = Four Treatments
  - `Agilent` two-dye microarray chip $\Rightarrow$ BIBD
- "I ran all four pairs"
- But $\binom{4}{2} = 6$!!!

What can go wrong?

- NASA Experiment on Substantiality of Crops (Potatoes)
  - Two Factors: Photoperiod (P), and bioactive Tuber Inducing Factor solution (TIF)
  - Each at two levels = Four Treatments
  - `Agilent` two-dye microarray chip $\Rightarrow$ BIBD
- "I ran all four pairs"
- But $\binom{4}{2} = 6$!!!

What can go wrong?

- NASA Experiment on Substantiality of Crops (Potatoes)
  - Two Factors: Photoperiod (P), and bioactive Tuber Inducing Factor solution (TIF)
  - Each at two levels = Four Treatments
  - `Agilent` two-dye microarray chip $\Rightarrow$ BIBD
- "I ran all four pairs"
- But $\binom{4}{2} = 6$!!!

- He ran all four pairs

P-1/TIF1	P-2/TIF-1

P-1/TIF-1	P-1/TIF-2

P-1/TIF-1	P-2/TIF-2

P-2/TIF-1	P-2/TIF-2

The case of the missing pairs

- Damage Control - confounding - contrasts have block effects

- He ran all four pairs

P-1/TIF1	P-1/TIF-1
P-2/TIF-1	P-1/TIF-2

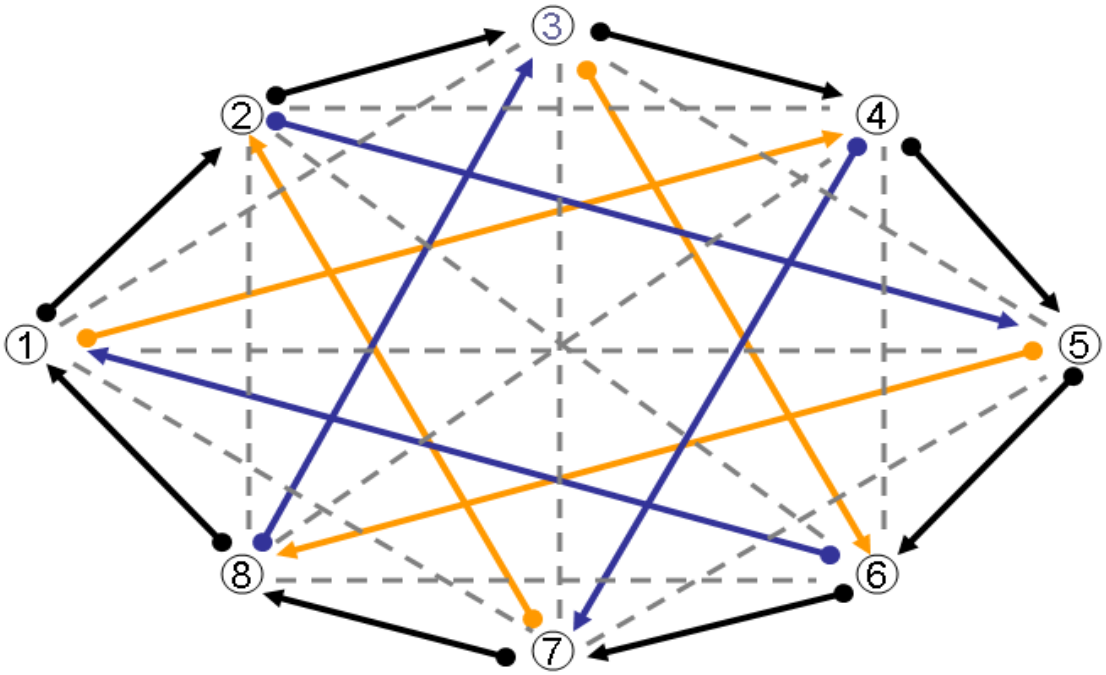
P-1/TIF-1	P-2/TIF-1
P-2/TIF-2	P-2/TIF-2

The case of the missing pairs

- Damage Control - confounding - contrasts have block effects

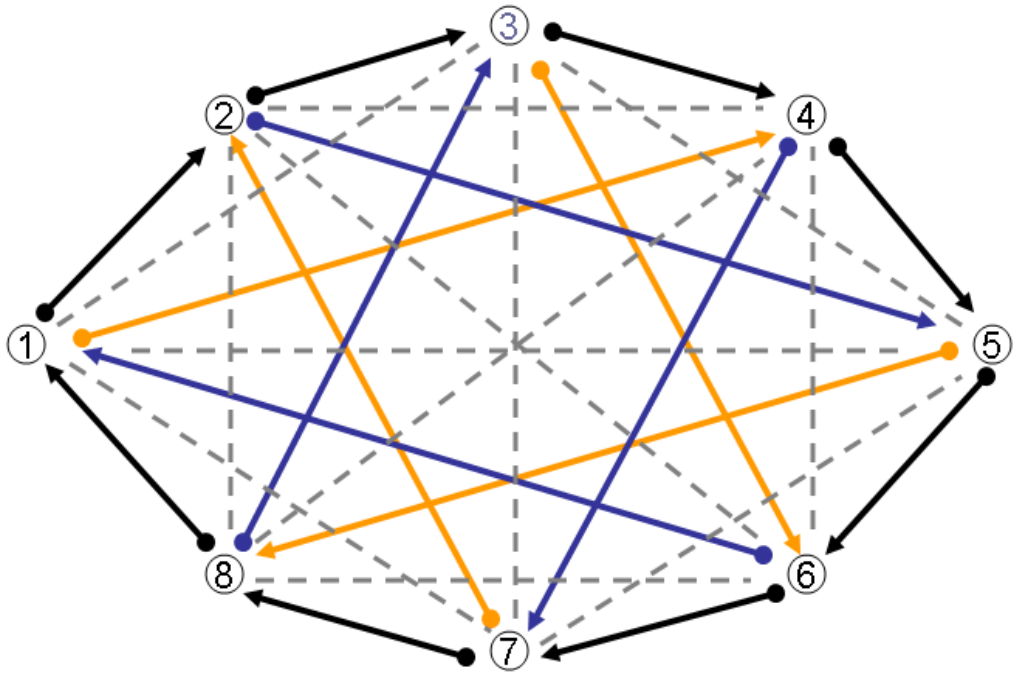
Loops, Augmented Loops, and BIBDs

▷ *Persea* (Avocado) Experiment-Eight Different Tissues



▷ Reducing Some Variances

Persea (Avocado) Experiment

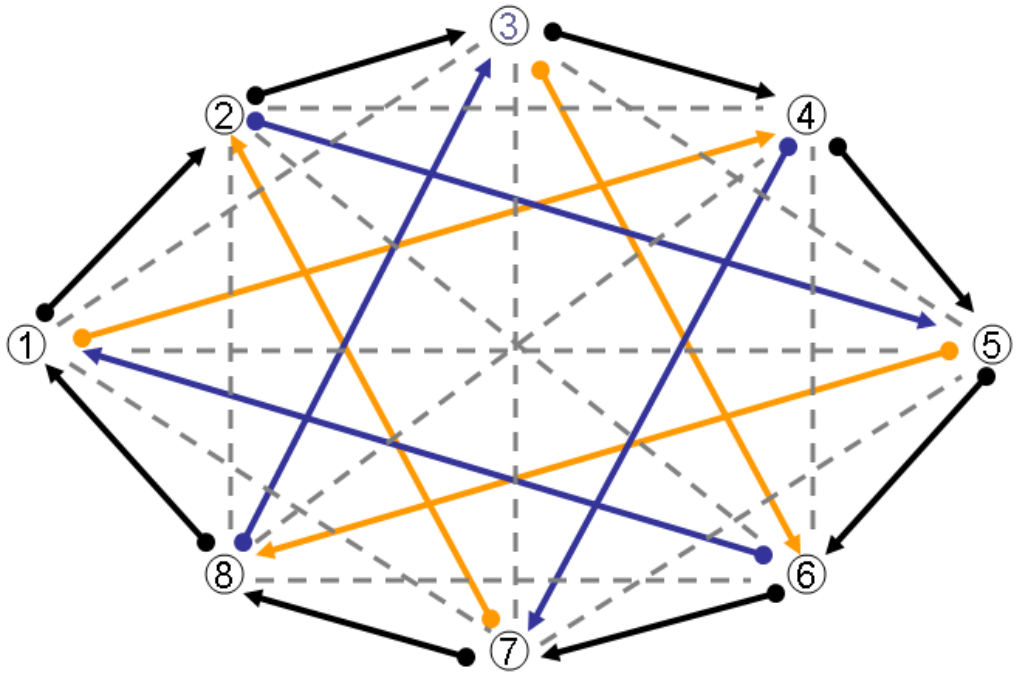


Outer Black Lines
=Loop Design

Outer Black Lines
+ Orange and Blue
=Augmented Loop

All Lines
=BIBD

Persea (Avocado) Experiment

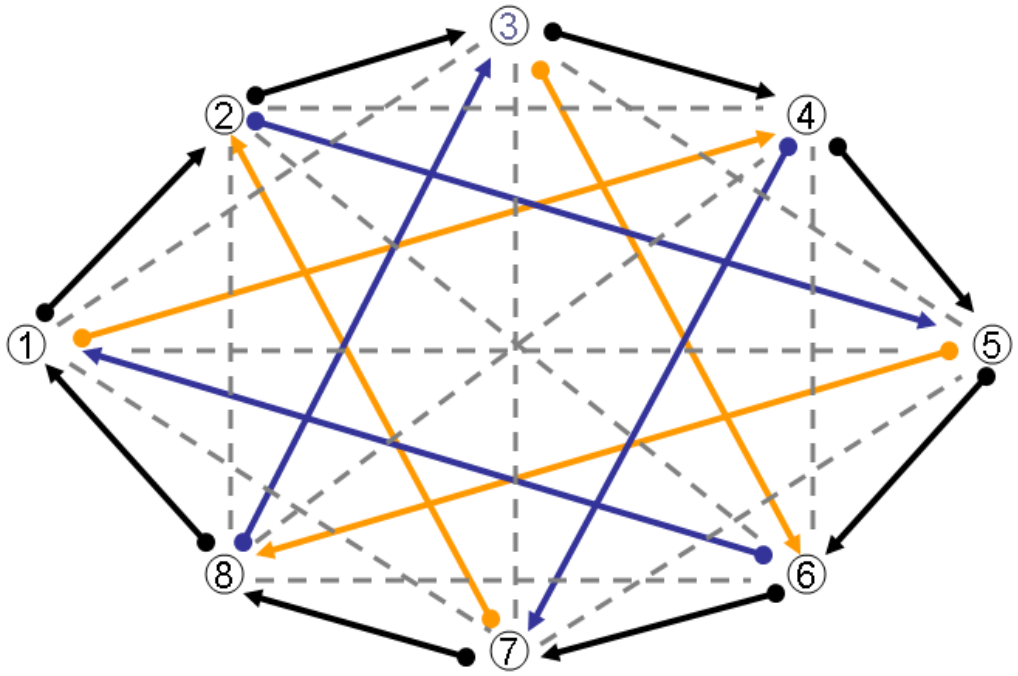


Outer Black Lines
=Loop Design

Outer Black Lines
+ Orange and Blue
=Augmented Loop

All Lines
=BIBD

Persea (Avocado) Experiment

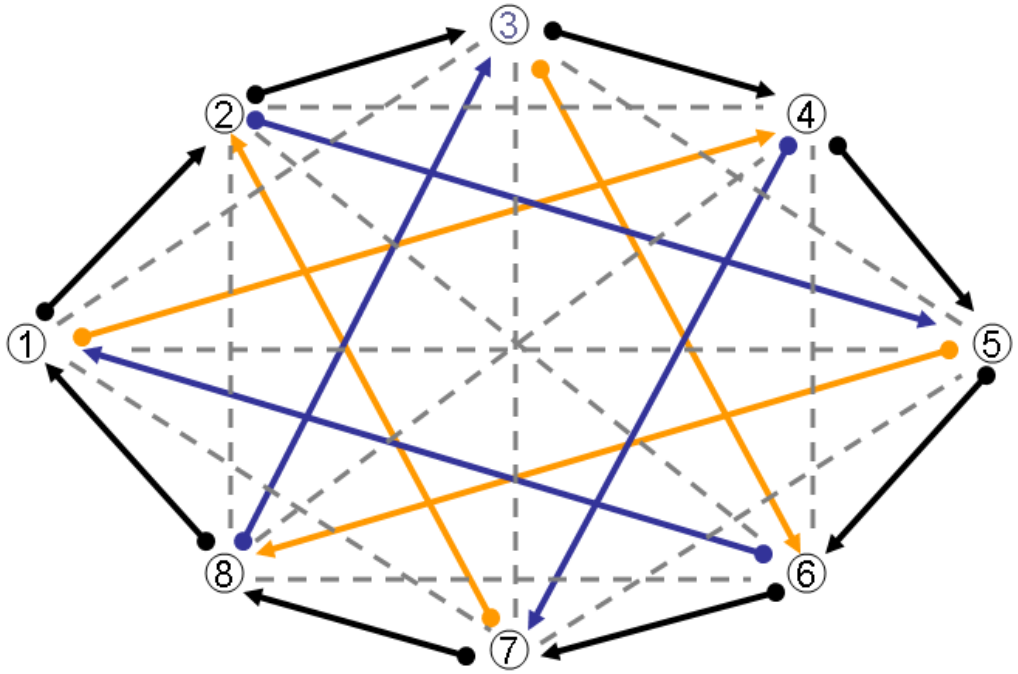


Outer Black Lines
=Loop Design

Outer Black Lines
+ Orange and Blue
=Augmented Loop

All Lines
=BIBD

Loop Design Variances



Adjacent Treatments

$$\text{Var}(\hat{\tau}_i - \hat{\tau}_{i\prime}) = \sigma^2_\varepsilon + \tfrac{1}{2}\sigma^2_\beta$$

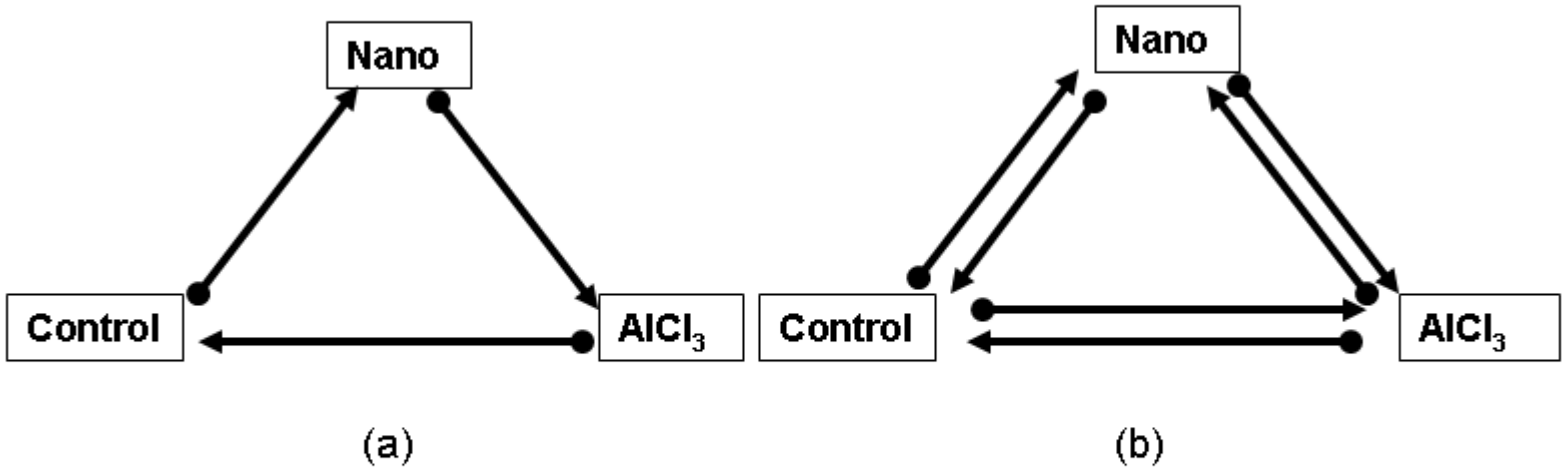
Non-Adjacent Treatments

$$\text{Var}(\hat{\tau}_i - \hat{\tau}_{i\prime}) = \sigma^2_\varepsilon + \sigma^2_\beta$$

BIBD: No σ^2_β

A Word About Reference Designs

- Effect of Aluminum on Zebrafish



- Recommended: Loop (left) or Double Loop (right)
- Experimenter chose Reference Design

Reference or Loop?

Dye	Reference			Loop		
Green	Ref.	Ref.	Ref.	Cont.	Nano	$AlCl_3$
Red	Cont.	$AlCl_3$	Nano	Nano	$AlCl_3$	Cont.

Reference Design : $\mathrm{Var}(\hat{\tau}_i - \hat{\tau}_{i'}) = 2\sigma^2_\varepsilon + 2\sigma^2_\beta$

Loop Design : $\mathrm{Var}(\hat{\tau}_i - \hat{\tau}_{i'}) = \sigma^2_\varepsilon + \frac{1}{2}\sigma^2_\beta$

- "Half of your observations are on a treatment that you don't care about"

Reference or Loop?

Dye	Reference			Loop		
Green	Ref.	Ref.	Ref.	Cont.	Nano	$AlCl_3$
Red	Cont.	$AlCl_3$	Nano	Nano	$AlCl_3$	Cont.

$$\text{Reference Design} : \text{Var}(\hat{\tau}_i - \hat{\tau}_{i'}) = 2\sigma^2_\varepsilon + 2\sigma^2_\beta$$

$$\text{Loop Design} : \text{Var}(\hat{\tau}_i - \hat{\tau}_{i'}) = \sigma^2_\varepsilon + \frac{1}{2}\sigma^2_\beta$$

▷ "Half of your observations are on a treatment that you don't care about"

Reference or Loop?

Dye	Reference			Loop		
Green	Ref.	Ref.	Ref.	Cont.	Nano	$AlCl_3$
Red	Cont.	$AlCl_3$	Nano	Nano	$AlCl_3$	Cont.

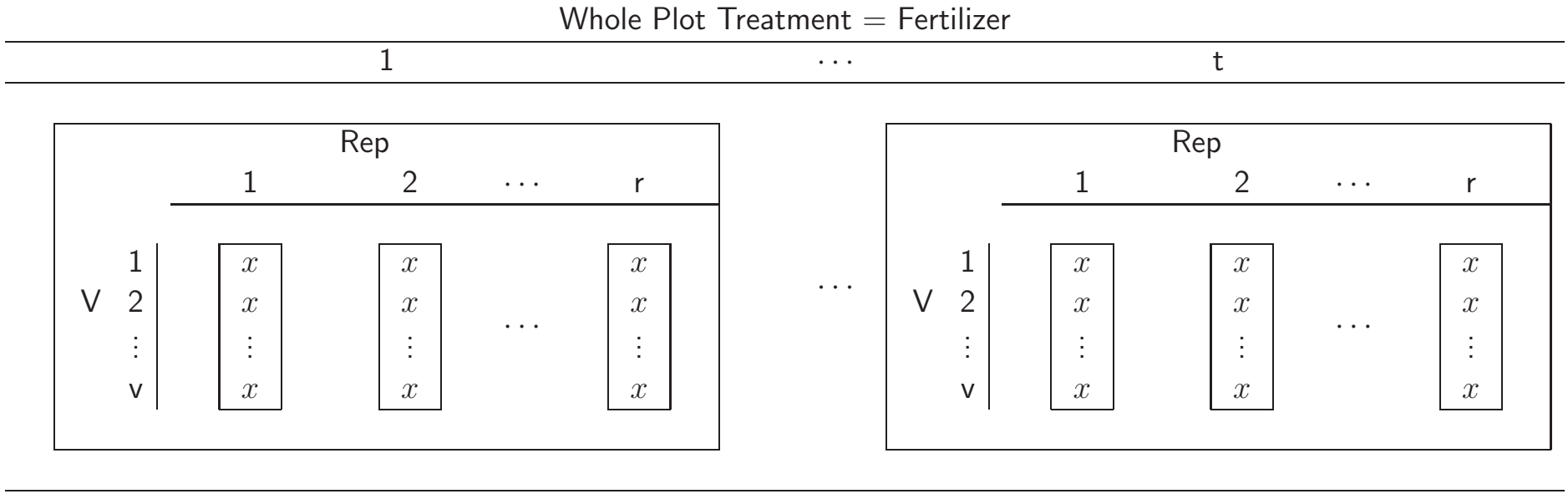
Reference Design : $\mathrm{Var}(\hat{\tau}_i - \hat{\tau}_{i'}) = 2\sigma^2_\varepsilon + 2\sigma^2_\beta$

Loop Design : $\mathrm{Var}(\hat{\tau}_i - \hat{\tau}_{i'}) = \sigma^2_\varepsilon + \frac{1}{2}\sigma^2_\beta$

▷ "Half of your observations are on a treatment that you don't care about"

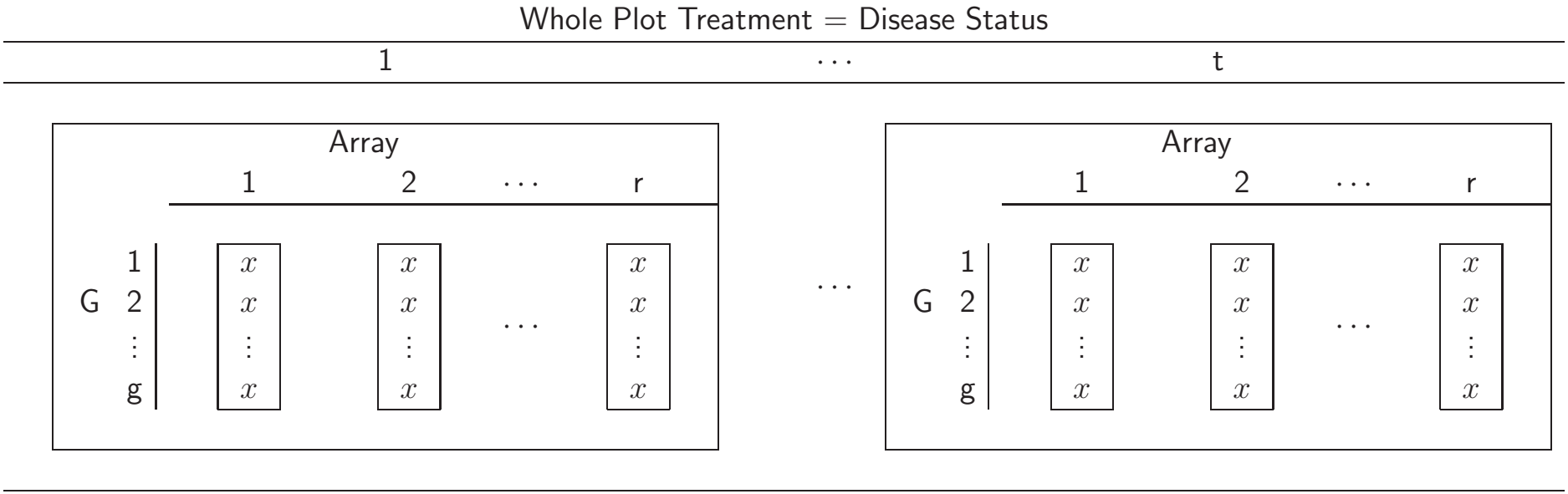
- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

Split Plots - Agricultural Experiment



- ▷ Fertilizer applied to Whole Plots
- ▷ Varieties within Whole Plots = Split Plots

Split Plots - Microarray Experiment



- ▷ Arrays= Whole Plots
- ▷ Gene within Arrays = Split Plots

Split Plot Anova

Source	df	EMS
Whole Plot Trt	t-1	$\sigma^2_\delta + g\sigma^2_\varepsilon + \frac{rg}{t-1}\sum_i \tau_i^2$
Arrays (in Whole Plots)	t(r-1)	$\sigma^2_\delta + g\sigma^2_\varepsilon$
Genes (Split Plot Trt)	g-1	$\sigma^2_\delta + \frac{rt}{g-1}\sum_k \gamma_k^2$
Genes $\times$ Whole Plot Trt	(g-1)(t-1)	$\sigma^2_\delta + \frac{r}{(g-1)(t-1)}\sum_{ik}(\tau\gamma)^2_{ik}$
Error (Split Plot Trt $\times$ Replication in Whole Plots)	t(g-1)(r-1)	σ^2_δ
Total	grt-1	

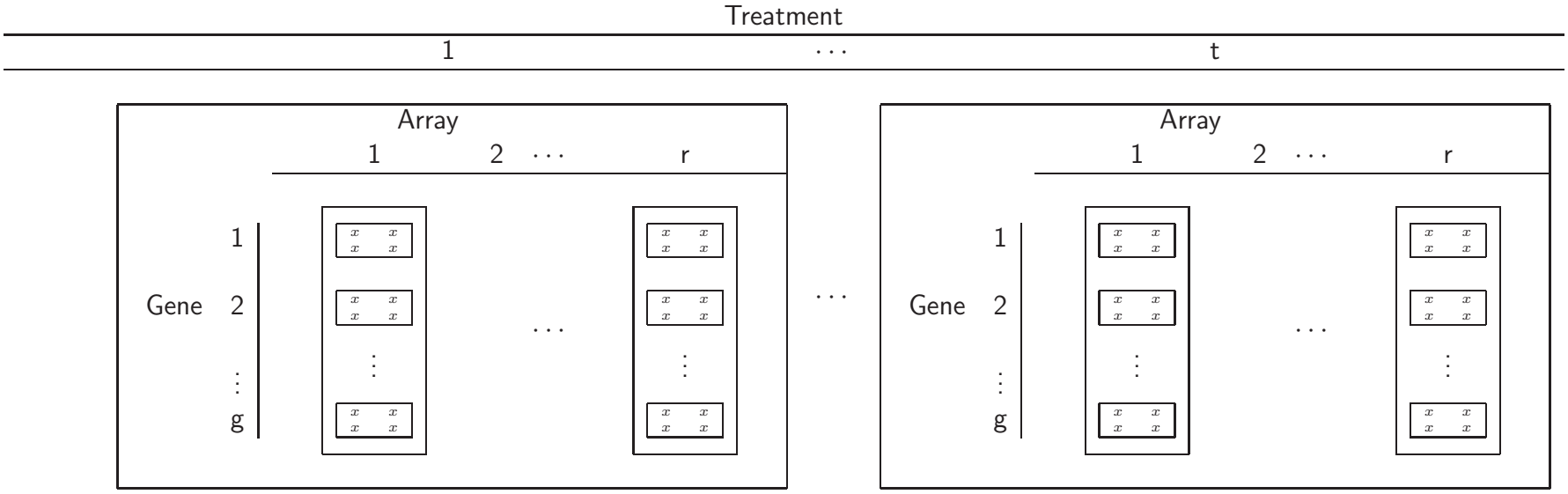
- ▷ Interest is at Split Plot level
- ▷ Genes $\times$ Whole Plot Trt = Differential Expression
- ▷ Split Plot comparisons have smaller error
- ▷ Pooling Genes - Analysis Question

Split Plot Anova

Source	df	EMS
Whole Plot Trt	t-1	$\sigma^2_\delta + g\sigma^2_\varepsilon + \frac{rg}{t-1}\sum_i \tau_i^2$
Arrays (in Whole Plots)	t(r-1)	$\sigma^2_\delta + g\sigma^2_\varepsilon$
Genes (Split Plot Trt)	g-1	$\sigma^2_\delta + \frac{rt}{g-1}\sum_k \gamma_k^2$
Genes $\times$ Whole Plot Trt	(g-1)(t-1)	$\sigma^2_\delta + \frac{r}{(g-1)(t-1)}\sum_{ik}(\tau\gamma)^2_{ik}$
Error (Split Plot Trt $\times$ Replication in Whole Plots)	t(g-1)(r-1)	σ^2_δ
Total	grt-1	

- Interest is at Split Plot level
- Genes $\times$ Whole Plot Trt = Differential Expression
- Split Plot comparisons have smaller error
- Pooling Genes - Analysis Question

Splitting Even More



- ▷ Probes nested within Genes
- ▷ Probe × Gene Interaction ⇒ SNP?
- ▷ Split Split Plot comparisons

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

Pooling

- Should we pool RNA of Subjects?

$$\sigma^2 = \sigma^2_B + \sigma^2_W \qquad \text{Between + Within Variance}$$

- Variance of Treatment Mean

$$\text{Var}(\bar{Y}_i) = \frac{1}{rp}\left(\sigma^2_B + \frac{\sigma^2_W}{s}\right).$$

  - r=True Rep, p=Pooling, s=Technical Rep

Pooling

- Should we pool RNA of Subjects?

$$\sigma^2 = \sigma^2_B + \sigma^2_W \qquad \text{Between + Within Variance}$$

- Variance of Treatment Mean

$$\text{Var}(\bar{Y}_i) = \frac{1}{rp}\left(\sigma^2_B + \frac{\sigma^2_W}{s}\right).$$

▷ r=True Rep, p=Pooling, s=Technical Rep

Conservative Tests

- Completely Randomized Design

Source	df	EMS
Treatment T	$t-1$	$\sigma^2 + \frac{rg}{t-1}\sum_i \tau_i^2$
Treatment G	$g-1$	$\sigma^2 + \frac{rt}{g-1}\sum_j \gamma_i^2$
T $\times$ G	$(t-1)(g-1)$	$\sigma^2 + \frac{r}{(t-1)(g-1)}\sum_{ij}(\tau\gamma)_{ij}^2$
Within	$tg(r-1)$	σ^2

- Pooling $\Rightarrow$
  - Lower Power
  - Lower Type I Error

Conservative Tests

- Completely Randomized Design

Source	df	EMS
Treatment T	$t-1$	$\sigma^2 + \frac{rg}{t-1}\sum_i \tau_i^2$
Treatment G	$g-1$	$\sigma^2 + \frac{rt}{g-1}\sum_j \gamma_i^2$
T × G	$(t-1)(g-1)$	$\sigma^2 + \frac{r}{(t-1)(g-1)}\sum_{ij}(\tau\gamma)_{ij}^2$
Within	$tg(r-1)$	σ^2

- Pooling $\Rightarrow$
  - ▷ Lower Power
  - ▷ Lower Type I Error

Anti-Conservative Tests

- Randomized Complete Block Design

Source	df	EMS
Blocks	$b-1$	$\sigma^2_\varepsilon + r\sigma^2_{\tau\beta} + rt\sigma^2_\beta$
Treatments	$t-1$	$\sigma^2_\varepsilon + r\sigma^2_{\tau\beta} + \frac{rt}{t-1}\sum_i(\tau_i - \bar{\tau})^2$
T $\times$ B	$(t-1)(b-1)$	$\sigma^2_\varepsilon + r\sigma^2_{\tau\beta}$
Within	$bt(r-1)$	σ^2_ε

- Pooling $\Rightarrow$
  - Higher Power
  - Higher Type I Error

Anti-Conservative Tests

- Randomized Complete Block Design

Source	df	EMS
Blocks	$b-1$	$\sigma^2_\varepsilon + r\sigma^2_{\tau\beta} + rt\sigma^2_\beta$
Treatments	$t-1$	$\sigma^2_\varepsilon + r\sigma^2_{\tau\beta} + \frac{rt}{t-1}\sum_i(\tau_i - \bar{\tau})^2$
T $\times$ B	$(t-1)(b-1)$	$\sigma^2_\varepsilon + r\sigma^2_{\tau\beta}$
Within	$bt(r-1)$	σ^2_ε

- Pooling $\Rightarrow$
  - Higher Power
  - Higher Type I Error

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

Summary

- From the Field to the Lab ▷Check the Design
- Getting the Errors Correct ▷ Check the Denominator
- Revisiting the Field and the Lab ▷ Ag Design = Microarray Design
- Replication - True and Technical ▷ Identify the Experimental Unit
- BIBDs and Their Variations ▷ Balance the Blocks
- Splitting the Plot ▷ Using Split Plot Errors
- Lightning Round ▷ Useful Tidbits
- Conclusions ▷ One More Slide

Summary

Summary

- From the Field to the Lab
  - ▷Check the Design
- Getting the Errors Correct
  - ▷ Check the Denominator
- Revisiting the Field and the Lab
  - ▷ Ag Design = Microarray Design
- Replication - True and Technical
  - ▷ Identify the Experimental Unit
- BIBDs and Their Variations
  - ▷ Balance the Blocks
- Splitting the Plot
  - ▷ Using Split Plot Errors
- Lightning Round
  - ▷ Useful Tidbits
- Conclusions
  - ▷ One More Slide

Summary

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

▷Check the Design
▷ Check the Denominator
▷ Ag Design = Microarray Design
▷ Identify the Experimental Unit
▷ Balance the Blocks
▷ Using Split Plot Errors
▷ Useful Tidbits
▷ One More Slide

Summary

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

▷Check the Design
▷ Check the Denominator
▷ Ag Design = Microarray Design
▷ Identify the Experimental Unit
▷ Balance the Blocks
▷ Using Split Plot Errors
▷ Useful Tidbits
▷ One More Slide

Summary

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

▷Check the Design
▷ Check the Denominator
▷ Ag Design = Microarray Design
▷ Identify the Experimental Unit
▷ Balance the Blocks
▷ Using Split Plot Errors
▷ Useful Tidbits
▷ One More Slide

Summary

- From the Field to the Lab
- Getting the Errors Correct
- Revisiting the Field and the Lab
- Replication - True and Technical
- BIBDs and Their Variations
- Splitting the Plot
- Lightning Round
- Conclusions

▷Check the Design
▷ Check the Denominator
▷ Ag Design = Microarray Design
▷ Identify the Experimental Unit
▷ Balance the Blocks
▷ Using Split Plot Errors
▷ Useful Tidbits
▷ One More Slide

Summary

- From the Field to the Lab
  - ▷Check the Design
- Getting the Errors Correct
  - ▷ Check the Denominator
- Revisiting the Field and the Lab
  - ▷ Ag Design = Microarray Design
- Replication - True and Technical
  - ▷ Identify the Experimental Unit
- BIBDs and Their Variations
  - ▷ Balance the Blocks
- Splitting the Plot
  - ▷ Using Split Plot Errors
- Lightning Round
  - ▷ Useful Tidbits
- Conclusions
  - ▷ One More Slide

Summary

- From the Field to the Lab ▷Check the Design
- Getting the Errors Correct ▷ Check the Denominator
- Revisiting the Field and the Lab ▷ Ag Design = Microarray Design
- Replication - True and Technical ▷ Identify the Experimental Unit
- BIBDs and Their Variations ▷ Balance the Blocks
- Splitting the Plot ▷ Using Split Plot Errors
- Lightning Round ▷ Useful Tidbits
- Conclusions ▷ One More Slide

Thank You for Your Attention

casella@ufl.edu

University of Florida Gators!