

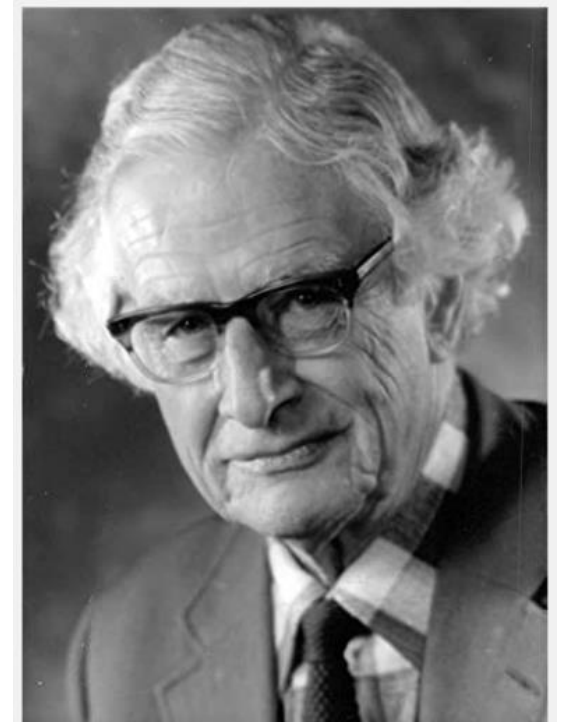
L12 – Incomplete Block Design

Michele Jorge Silva Siqueira

2023

Introduction

- Frank Yates (1902-1994) was appointed as assistant statistician in 1931 at Rothamsted Experimental Station by Ronald Fisher
- He became the head of the Statistics Department on Fisher's departure in 1933 until his retirement in 1968
- He originated the idea of balanced incomplete block design



Introduction

- The designs like CRD and RBD are the complete block designs

a	b	c	d
A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

We now discuss the Incomplete Block Design (IBD)

Introduction

- In some situations, it can be difficult to obtain enough homogeneous experimental units to construct complete blocks, for example, when the number of treatments is numerous, physical constraints, costs or time
- In other situations we might start with a complete block design but lose some observations during the study due to contamination, false recording, accidents, etc

So we have an **incomplete block design** because not every treatment appears in each block

Introduction

- Incomplete Block Design (IBD) was introduced by Yates in order to **eliminate heterogeneity** to a greater extent when **the number of treatments is large**
- IBD mitigates heterogeneity between experimental units, so experimental units within each incomplete block are more homogeneous

Introduction

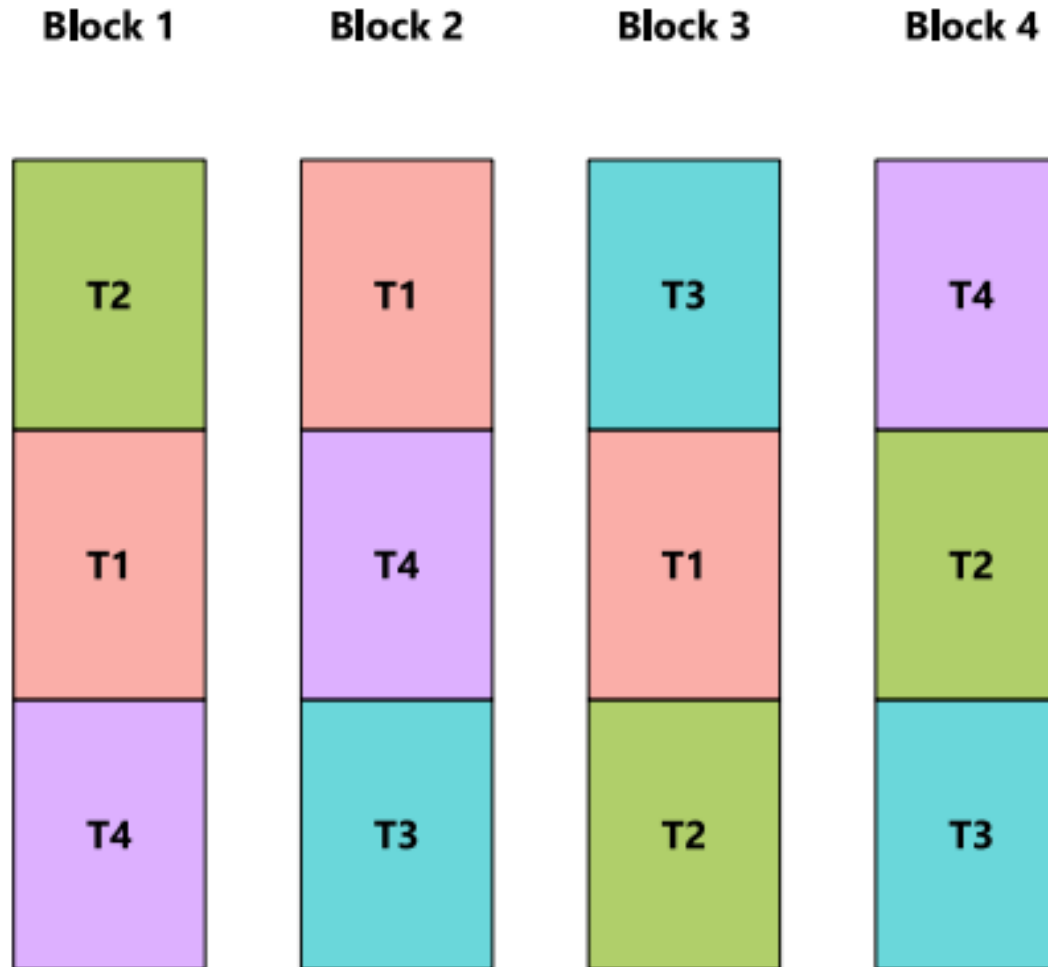
- We can use **incomplete block** designs

Incomplete Block Designs

- The incomplete block design is so called because a block does not contain all treatments
- In other words, the block size (number of experimental units) is smaller than the number of treatments

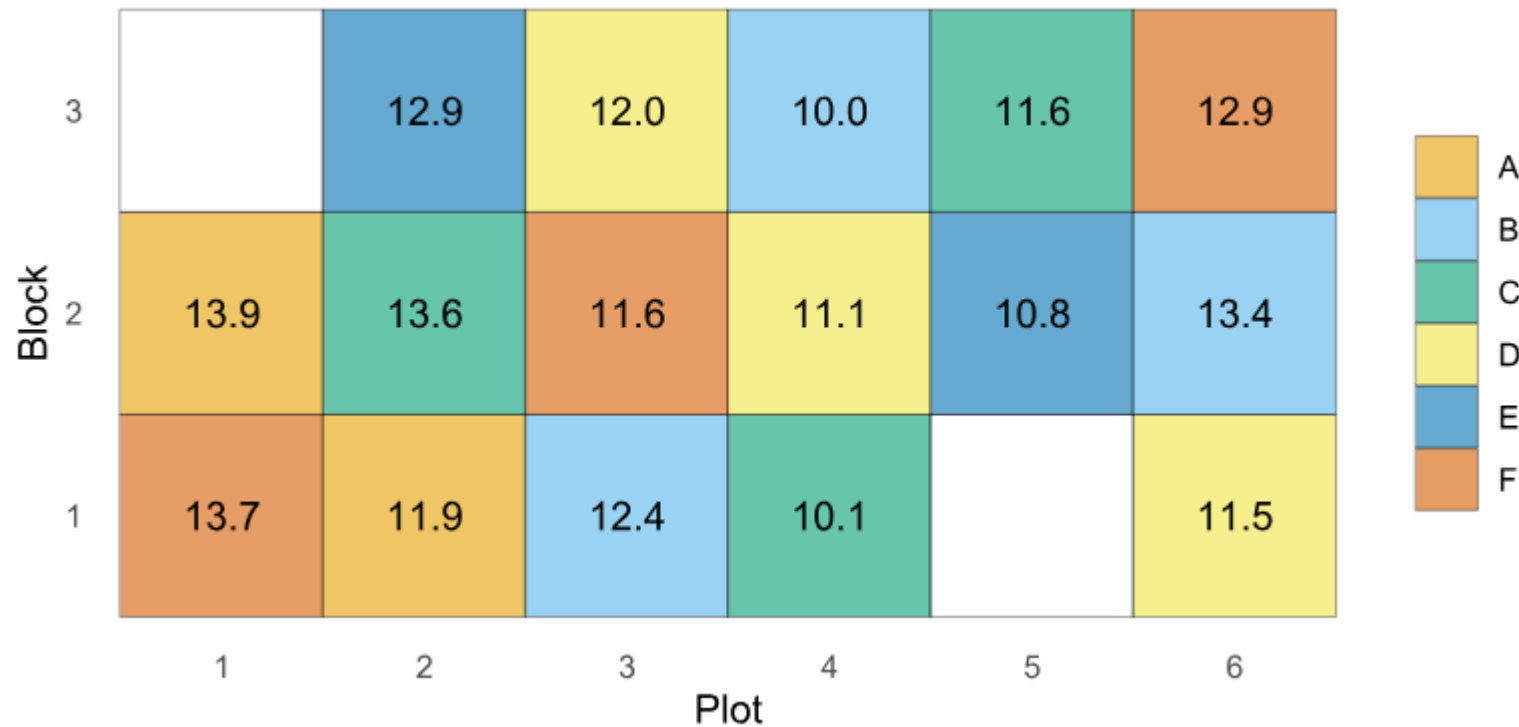
Experimental Design Layout

Incomplete Block Designs



Experimental Design Layout

Incomplete Block Designs



Example - Potato yield

dat

	Y	Block	Trt	plot
5	13.7	1	F	1
1	11.9	1	A	2
2	12.4	1	B	3
3	10.1	1	C	4
4	11.5	1	D	6
6	13.9	2	A	1
8	13.6	2	C	2
11	11.6	2	F	3
9	11.1	2	D	4
10	10.8	2	E	5
7	13.4	2	B	6
15	12.9	3	E	2
14	12.0	3	D	3
12	10.0	3	B	4
13	11.6	3	C	5
16	12.9	3	F	6
17		3	<NA>	1
21		1	<NA>	5

Experimental Design Layout

Incomplete Block Designs

- Let t represent the number of treatments and r the number of replications of each treatment
- Each of the t treatments appears in r incomplete blocks
- Experimental units are divided into b blocks of k units
 - b is the *total* number of blocks in the experiment
 - k represents the size (number of experimental units) of each incomplete block

Experimental Design Layout

- Let λ represent the number of times that pairs of treatments occur in the same incomplete block

$$\lambda = \frac{r(k-1)}{t-1}$$

Experimental Design Layout

Balanced Incomplete Block Designs (BIBD)

- Each treatment occurs at most once in a block
- Each treatment occurs the same number of times r
- Each pair of treatments occur together in a block the same number of times λ

In that case, we have that $tr = bk$ is the total number of experimental units

Experimental Design Layout

Balanced Incomplete Block Designs (BIBD)

- In some cases, particularly when the number of treatments is large, achieving balance may be impractical
- Exceedingly large numbers of replicates
- Not all pairs of treatments are equally represented within the incomplete blocks
- Some pairs may never even occur in the same block
- Comparisons of treatment means are made with equivalent precision for all pairs of treatments

Experimental Design Layout

Example 01 - BIBD

- A BIB design for $b = 5$, $r = k = 4$ and $\lambda = 3$ in the following:

Blocks	(I)	1	2	3	4
	(II)	1	2	3	5
	(III)	1	2	4	5
	(IV)	1	3	4	5
	(V)	2	3	4	5

λ represent the number of times that pairs of treatments occur in the same incomplete block

- Not all treatments occur in every block, such that some i, j combinations do not exist

Experimental Design Layout

Example 02 - BIBD

- A BIB design for $b = 10$, $r = k = 5$ and $\lambda = 2$ in the following:

Blocks Treatments

B_1	T_1	T_2	T_5
B_2	T_1	T_2	T_6
B_3	T_1	T_3	T_4
B_4	T_1	T_3	T_6
B_5	T_1	T_4	T_5
B_6	T_2	T_3	T_4
B_7	T_2	T_3	T_5
B_8	T_2	T_4	T_6
B_9	T_3	T_5	T_6
B_{10}	T_4	T_5	T_6

Experimental Design Layout

Partially Balanced Incomplete Block Designs

- In some cases, particularly when the number of treatments is large, achieving balance may be impractical
- Not all pairs of treatments are equally represented within the incomplete blocks
- Some pairs may never even occur in the same block

Experimental Design Layout

Partially Balanced Incomplete Block Designs

- If each pair of treatments appears λ_1 or λ_2 times in the same block, the design is called a Partially Balanced Incomplete Block Design

Statistical Model

- Let y_i represent the observed response of the i th treatment ($i = 1, \dots, t$), in the j th block ($j = 1, \dots, r$)
- For a BIB design (Balanced Incomplete Block), the corresponding model can be written as:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

Where μ is the overall mean, τ_i are treatment effects, β_j are block effects and ε_{ij} are the associated error random variables

- Not all treatments occur in every block, such that some i, j combinations do not exist

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F -Statistic
Blocks	$b - 1$	SS_{Blocks}	$MS_{\text{Blocks}} = \frac{SS_{\text{Blocks}}}{b - 1}$	$F = \frac{MS_{\text{TrtAdj}}}{MS_E}$
Trt _{Adj}	$t - 1$	SS_{TrtAdj}	$MS_{\text{TrtAdj}} = \frac{SS_{\text{TrtAdj}}}{t - 1}$	
Error	$bk - b - t + 1$	SS_E	$MS_E = \frac{SS_E}{bk - b - t + 1}$	
Total	$bk - 1$	SS_{Total}		

Analysis of Variance

- The treatment sum of squares needs to be adjusted, because not all treatments are present in each block
- For each treatment, we sum over different blocks

Analysis of Variance

- If we model block *effects as fixed*, pairwise differences between treatments can be estimated in the blocks in which both treatments occur
- This is called the **intrablock** analysis of variance: examines variation in treatment outcomes within each block, without considering differences between blocks
- Differences between treatments in the same block are free from block effects

Analysis of Variance

- If block effects are treated *as random*, their totals may carry information about the treatment combinations
- Differences between treatments can be extracted from blocks that do not contain both treatments
- This results in the **interblock** analysis of variance
- The effective variance includes both the intrablock and interblock sources of variation

Lattice designs

- Lattice designs form an important class of useful incomplete block designs
- These designs were originally introduced by Yates (1936)

Lattice designs

- Square lattice
- Circular lattice
- Cubic lattice
- Rectangular lattice

Let's Practice 01!

- The example is from a soybean balanced incomplete block experiment
- Yield measurements (in bushels per acre) were made for 30 varieties
- There are 31 genotype levels, because one variety was duplicated (G07 and G14)
- The experiment consisted of 31 blocks of six plots each
- This data set can be found in the R package agridat and is originally from Weiss & Cox (1939) *Iowa State Agricultural Research Bulletin* 257. Ames, Iowa. p. 28.



Let's Practice 01!



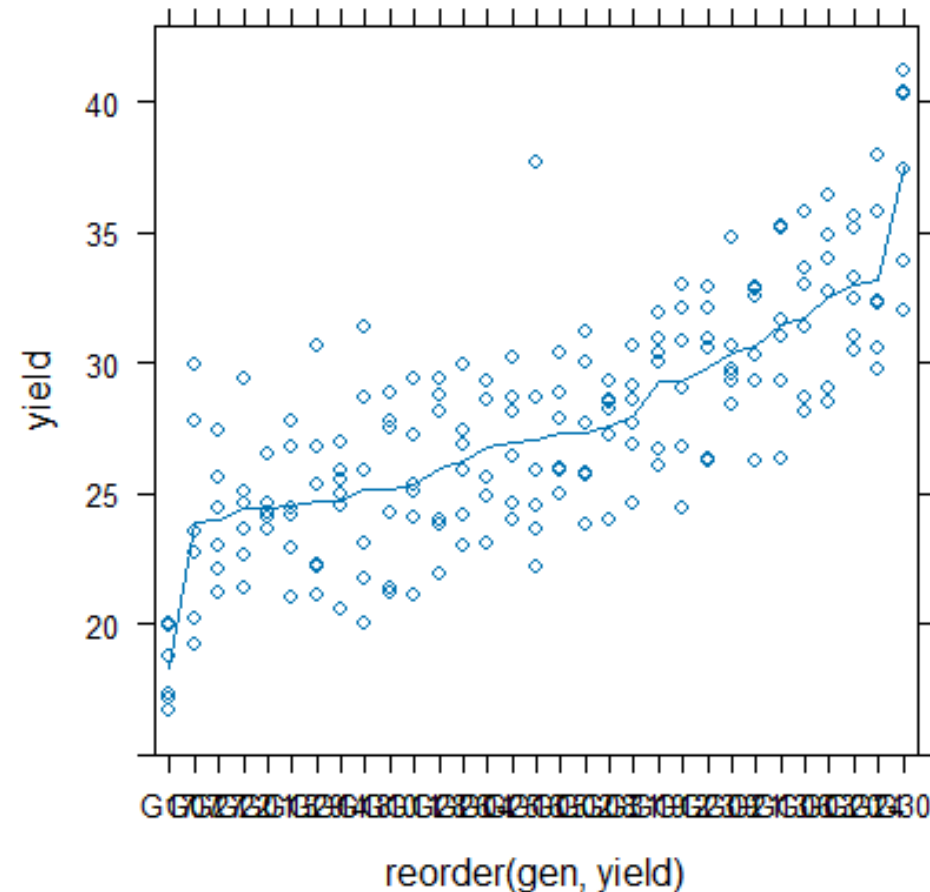
- Use the R function `read.csv` to import the data
- Fit the model with fixed (and/or random) effects
- Check if model assumptions are met
- Build the ANOVA table and test the null hypothesis of no difference between genotypes
- Use multiple comparisons to assess pairwise differences

Let's Practice 01!

	block	gen	yield
1	B01	G24	29.8
2	B02	G15	24.2
3	B03	G20	30.5
4	B04	G18	20.0
5	B05	G20	35.2
6	B06	G05	25.0
7	B07	G22	23.6
8	B08	G01	23.6
9	B09	G09	29.3
10	B10	G14	25.5
11	B11	G10	21.4
12	B12	G17	16.7
13	B13	G03	32.7
14	B14	G24	32.3
15	B15	G21	32.6
16	B16	G02	30.0
17	B17	G17	17.1
18	B18	G18	21.7
19	B19	G01	24.3
20	B20	G28	23.8
21	B21	G25	26.4
22	B22	G08	27.2
23	B23	G10	27.8
24	B24	G05	30.4

Let's Practice 01!

```
library(lattice)
xyplot(yield ~ reorder(gen, yield), data = dados, type = c("p", "a"))
xyplot(yield ~ reorder(block, yield), data = dados, type = c("p", "a"))
```



Let's Practice 01!

```
# Fixed Model

fm <- lm(yield ~ block + gen, data = dados)
anova(fm)
```

Analysis of Variance Table

```
Response: yield
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
block	30	1642.61	54.754	15.272	< 2.2e-16	***
gen	30	1841.28	61.376	17.119	< 2.2e-16	***
Residuals	125	448.16	3.585			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Let's Practice 01!

```
# Random Model
```

```
library(nlme)
```

```
fme <- lme(yield ~ gen, random = ~1 |block, data = dados)
```

```
anova(fme)
```

```
> anova(fme)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	125	4042.016	<.0001
gen	30	125	17.675	<.0001

Let's Practice 01!

```
# Multiple Comparisons
```

```
library(emmeans)
fm_means1 <- emmeans(fm, "gen")
fm_means1
pairs(fm_means1)
```

```
> pairs(fm_means1)
```

contrast	estimate	SE	df	t.ratio	p.value
G01 - G02	-2.33548	1.18	125	-1.982	0.9796
G01 - G03	-8.02903	1.18	125	-6.815	<.0001
G01 - G04	-2.37742	1.18	125	-2.018	0.9745
G01 - G05	-1.42903	1.18	125	-1.213	1.0000
G01 - G06	-7.40000	1.18	125	-6.281	<.0001
G01 - G07	0.40000	1.18	125	0.340	1.0000
G01 - G08	-3.01936	1.18	125	-2.563	0.7441
G01 - G09	-4.68710	1.18	125	-3.979	0.0339
G01 - G10	0.14516	1.18	125	0.123	1.0000
G01 - G11	-2.69032	1.18	125	-2.284	0.9010
G01 - G12	-4.65806	1.18	125	-3.954	0.0367
G01 - G13	-5.15161	1.18	125	-4.373	0.0087
G01 - G14	0.40968	1.18	125	0.348	1.0000
G01 - G15	-1.70645	1.18	125	-1.449	0.9999
G01 - G16	-1.18387	1.18	125	-1.005	1.0000
G01 - G17	4.70645	1.18	125	3.995	0.0321
G01 - G18	-1.21290	1.18	125	-1.030	1.0000
G01 - G19	-4.41290	1.18	125	-3.746	0.0696
G01 - G20	-8.58387	1.18	125	-7.286	<.0001
G01 - G21	-6.55161	1.18	125	-5.561	0.0001
G01 - G22	-0.67742	1.18	125	-0.575	1.0000
G01 - G23	-5.22581	1.18	125	-4.436	0.0069
G01 - G24	-9.11613	1.18	125	-7.738	<.0001
G01 - G25	-2.40323	1.18	125	-2.040	0.9709

References

1. Casella, G. Statistical Design. (2008).
2. Patterson, H. D. & Williams, E. R. A new class of resolvable incomplete block designs. *Biometrika* **63**, 83–92 (1976).