INTRO TO INCOMPLETE BLOCK DESIGNS

Chapter 11

LEARNING OBJECTIVES

- Explain advantages and disadvantages of using an incomplete block design (IBD)
- Describe how to properly randomize IBD's
- Define and check for connected-ness
- Write the statistical model and explain why treatment means are not used for contrast estimators

EYE DROP EXAMPLE

- There are 4 eye drop medications to compare
- Randomly select 12 subjects to test
- Can only test a person once (can't flush out eye and apply different treatment)
- Q1: How would you design this using a CRD? How many reps per treatment?
- Q2: Describe a design that could take into account person-to-person variability.

WHY USE INCOMPLETE BLOCKS

- RCBD: apply and observe each treatment one time in each block
- Requires the block size (k) to equal # treatments (t)
- EUs in block should be homogeneous, otherwise blocking has minimal effect
- Problem: The more EUs we put in a block, the less likely this will be true
  - Brought up this concern when discussing GRCBD

- Incomplete block designs (IBD) used when the number of treatments is larger than block size ($k < t$)

ADVANTAGES AND DISADVANTAGES

- Advantages:
  - Reduces experimental error and increases representativeness
  - Better justification of homogeneity of EU's for many treatments than RCBD or GRCBD
  - Accommodating when we can't get large block sizes

- Disadvantages:
  - IBDs are not equivalent in their analysis properties!
  - The "best" IBD's are difficult to construct or may not exist
  - Analysis becomes even trickier and need to make assumptions

IBD SET UP AND INCIDENCE MATRIX

- Have *b* blocks, *t* treatments, block sizes $k < t$
- IBD is binary if each treatment appears at most one time for any block (some will not appear)
- Let $n_{hi} = 1$ if treatment *i* is in block *h*, otherwise 0
- Describe IBD using block-treatment incidence matrix

$$N_{b \times t} = (n_{hi})$$

  - Columns correspond to treatments; row correspond to blocks
- For this matrix
  - RCBD: $n_{hi} = 1$
  - GRCBD: $n_{hi} = r$
  - IBD: : $n_{hi} = ?$

IBD SET UP AND INCIDENCE MATRIX

- **Group:** For the following IBD plan, write out the block/treatment incidence matrix

Block	1	2	3	4	5
	1	2	3	4	5
	2	3	4	5	1
	3	4	5	1	2
	4	5	1	2	3

RANDOMIZING IBD

- IBD plan determines the treatment sets that will be applied to EU's in the blocks
- Treatments obviously need to be randomly assigned to the EUs within the blocks
- Also randomize treatment sets to the blocks

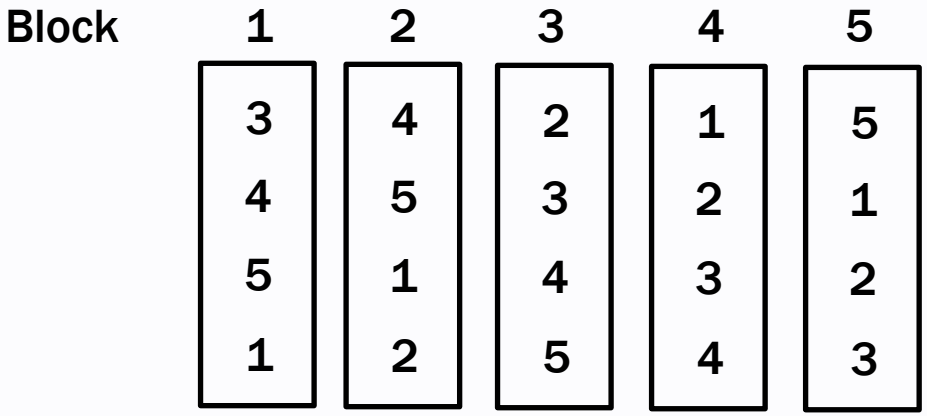
Block	1	2	3	4	5
	1	2	3	4	5
	2	3	4	5	1
	3	4	5	1	2
	4	5	1	2	3

Randomize sets to blocks

- Why didn't we care about the treatment set randomization for RCBDs?

RANDOMIZING IBD

- IBD plan determines the treatment sets that will be applied to EU's in the blocks
- Treatments obviously need to be randomly assigned to the EUs within the blocks
- Also randomize treatment sets to the blocks



STATISTICAL MODEL – IBD OVERSIMPLIFIED

- Following model is used to analyze data

$$Y_{hi} = \mu + \beta_h + \tau_i + E_{hi} \qquad \begin{matrix} h = 1, \ldots, b \\ i = 1, \ldots, t \end{matrix}$$

- Remember the index issues we had with categorical ANCOVA? The same issues come up here
  - We do not observe every combination of *h* and *i*

- Instead, let *u* = 1,…,*k* index the EU in block *h*
- Define $d[h, u] =$ treatment assigned to EU *u* in block *h* (post randomization)

STATISTICAL MODEL – IBD EXPLICIT

- More explicit to write the model as

$$Y_{hu} = \mu + \beta_h + \tau_{d[h,u]} + E_{hu} \qquad \begin{matrix} h = 1, \ldots, b \\ u = 1, \ldots, k \end{matrix}$$

- Again, Rwill make the appropriate adjustments so nothing explicit needs to be done on your end

- This model does help to explicitly show that treatment contrasts are estimated differently than before

CONNECTED DESIGNS

- Basic requirement when choosing an IBD is that we can estimate all treatment contrasts
- Such designs are said to be connected
  - Not every IBD is connected! Be careful!

- A sufficient condition is that every treatment pair occurs in some block but not always necessary

- Checking for connectedness using $N_{b\times t}$:
  - For treatment pair, find path between corresponding columns
  - Cannot move diagonally
  - Free to move across rows or columns only if the elements are 1

TREATMENT MEANS AND LSMEANS

- Expected value of treatment means **depends on set of blocks a treatment appears** in

$$E(\bar{Y}_{.i}) = \mu + \tau_i + \boxed{\frac{\sum_h n_{hi}\beta_h}{\sum_h n_{hi}}} = \mu + \tau_i + \boxed{\bar{\beta}_i}$$

=Block average for treatment *i*

- When we ask JMP for LSmeans, it gives the BLUE for

$$\mu + \tau_i + \boxed{\bar{\beta}_.}$$

=Overall block average

- **Conclusion:** Treatment mean does not equal Lsmean
- **Advice:** be careful when using treatment means to summarize data, they may be biased!

TREATMENT MEANS AND CONTRAST BLUES

- This bias carries over if we use the treatment means to estimate contrasts

- For CRDs, RCBDs, and LSDs, BLUEs for any contrast was same contrast of the treatment means, but now

$$E\left(\sum_i c_i \bar{Y}_{.i}\right) = \sum_i c_i \tau_i + \boxed{\sum_i c_i \bar{\beta}_i}$$

Bias!

- Unequal block means will potentially bias estimates

- Unbiased estimates exist, let R calculate BLUEs

LEARNING OBJECTIVES REVIEW

- Explain advantages and disadvantages of using an incomplete block design (IBD)
- Describe how to properly randomize IBD's
- Define and check for connected-ness
- Write the statistical model and explain why treatment means are not used for contrast estimators