Incomplete Block Designs I

Sean Hellingman ©

Design for Data Science (ADSC2030)

shellingman@tru.ca

Winter 2025

THOMPSON RIVERS UNIVERSITY

Topics

Recall: Blocking

- **Blocking** is an error control technique that allows for heterogeneous experimental units to be studied to arrive at general conclusions.

- In a **randomized block design**, the heterogeneous experimental units are grouped into homogeneous sub-groups (blocks) before randomization.

- Treatment factors are then randomly assigned to the units within the smaller homogeneous blocks.

Introduction

- In randomized block designs, the number of EUs in the block must be greater than or equal to the number of factor levels.
  - RCB: Number of EUs per block $= t$.
  - GCB: Number of EUs per block $= tr$.

- When larger block sizes are needed, the blocks become less homogeneous.

- One solution is to construct block designs where each block only takes a subset of the possible factor levels.
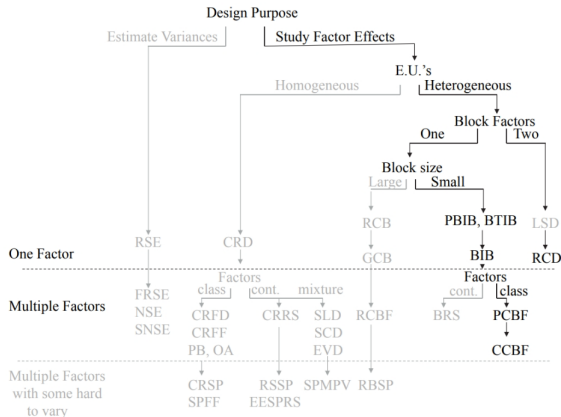 - *Incomplete block designs*

Road Map

Figure: Source: (1)

Balanced Incomplete Block (BIB) Designs

- **Balanced incomplete block** (BIB) designs: Each treatment level is equally replicated and appears within a block with every other treatment level an equal number of times.

- All pairwise differences have the same standard error and the power of detection will be the same.

- *If one treatment factor level is left out of every block, it cannot be compared to the other treatment levels.*

- *If different treatment levels are unequally represented, some comparisons of factor levels will have more precision than others.*

Illustrative Example 1

- In food science, taste panels are used to test palatability of new recipes and food products.

- In this situation, the different food products represent the different factor levels.

- A random sample of subjects (EUs) is used to represent the consumer market.
  - Individual tastes are different therefore subjects represent the different blocks.

- It is not reasonable to make the subjects taste and rate every single recipe (lose their ability to discriminate).
- **Solution:** Each subject (block) only tastes a subset of the recipes.

Constructing Balanced Incomplete Block Designs I

- The simplest way to construct a BIB ($k < t$) is to form all possible subsets of k treatment levels chosen from t.

- For Example:
    - $t = 6$ recipes to be tested.
    - Each subject can taste at most $k = 3$ recipes without losing discriminatory power
    - $\binom{6}{3} = 20$ subjects (EUs) are required.
    - *All possible combinations are needed.*
    - Each recipe is (factor level) is replicated $r = 10$ times.
    - Each pair of treatment levels occurs in the same block $\lambda = 4$ times.

- This is only one way to set this experiment up, the number of blocks and replicates can change depending on the required practical difference.

Constructing Balanced Incomplete Block Designs

- The following requirements must be met in order for the design to be BIB:
    1. $b \geq t$
    2. $tr = bk$
    3. $\lambda(t-1) = r(k-1)$
        - r is the number of times each treatment level is replicated.
        - λ is the number of times each treatment level occurs with every other treatment level in the same block.
        - t is the total number of levels of the treatment factor.
        - k is the number of EUs **in a block**.
        - b is the number of blocks.

- These inequalities can be used to determine the minimum number of blocks required for a BIB design.

Constructing Balanced Incomplete Block Designs in R I

- R can be used to suggest values of λ and r to satisfy the inequalities:
  - library(daewr)
  - BIBsize(t,k)

- *Even if the inequalities are satisfied, a corresponding BIB design may not exist.*

Example 1

- Use the `BIBsize()` function to identify a possible design for the food science example ($t = 6, k = 3$).

Constructing Balanced Incomplete Block Designs in R II

- The following R code can be used to find BIB designs:
  - ```
    library(AlgDesign)
    ```

  - ```
    BIB <- optBlock( ~., withinData = factor(1:t),
    blocksizes = (k,b))
    ```
  - ```
    des <- BIB$rows
    ```
  - ```
    dim(des) <- NULL
    ```
  - ```
    des <- matrix(des, nrow = b, ncol = k, byrow = TRUE,
    dimnames = list(c("Block1", ..., "Blockb"), c("unit1",
    ..., "unitk")))
    ```
  - ```
    des
    ```

- Once a BIB design is found, the levels of the treatment factor should be randomized within each block (like was done with RCB).

Example 2

- Use the optBlock() function to find a BIB design for the food science example ($t = 6, k = 3, b = 10$).

- What are the values of r and λ?

Incomplete Block Design Model

- The model for the incomplete block design:

$$y_{ij} = \mu + b_i + \tau_j + \epsilon_{ij} \tag{1}$$

- It is the same model used for the randomized complete block design.

- The analysis is slightly different due to the missing observations.
    - May bias the estimates.

Analysis in R I

- May use the same methods used for analyzing data from factorial designs with an unequal number of replicates per cell (type III sum of squares).

  OR

- If the differences in blocks are not of interest can use the following method in R:
    - model <- aov(response $\sim$ block + factor, data = data)
    - summary(model)

Analysis in R II

- Can use the adjusted sum of squares to compare the adjusted means
    - library(lsmeans)
    - lsmeans(model, pairwise $\sim$ factor, adjust = ("tukey"))

- Directly compares the adjusted means and performs hypothesis tests.

Example 3

1. Load the *taste* dataset from daewr package and take a moment to familiarize yourself with the data.

2. Construct an ANOVA table to examine the significance of the blocks and the recipe treatment.

3. Using the appropriate method, compare the adjusted means to determine if any of the recipes are *better* than any of the others.

4. Comment on the results.

5. *Do you notice anything about the standard errors of the differences?*

Determining the Number of Replicates

- Rough estimate for the number of replicates of each treatment level: $r = \frac{bk}{t}$
- Rough estimate for the number of blocks: $b = \frac{tr}{k}$

- If an estimate of the variance of the heterogeneous EUs (σ^2) and the practical difference (Δ) is known:
    - Number of replicates needed (small blocks): $r = r_{crd}/RE$
    - Number of blocks of size k in an incomplete block design: $b = tr/k$

- *RE is the relative efficiency (how much blocking is expected to reduce the variance)*

Example 4

- Suppose:
  - $\hat{\sigma}^2_{crd} = 0.040865$.
  - The number of replicates of each treatment required for a CRD design to achieve adequate power for detecting a practical difference in means was $r_{crd} = 20$.
  - Blocking is expected to reduce the variance by 90% ($\hat{\sigma}^2_{rcb} = 10 \cdot \hat{\sigma}^2_{crd}$ or $RE = 10.0$)

1. Determine the number of replicates (r) needed for for the small blocks to achieve the same power.

2. How many blocks (b) are needed if the block size is $k = 2$ and there are $t = 4$ treatment levels.

References & Resources

1. Lawson, J. (2014). *Design and Analysis of Experiments with R (Vol. 115)*. CRC press.

- BIBsize()
- optBlock()