Randomized Block Designs I

Sean Hellingman ©

Design for Data Science (ADSC2030)

shellingman@tru.ca

Winter 2025

THOMPSON RIVERS UNIVERSITY

Topics

Introduction

- It is advisable to keep experimental units as homogeneous as possible
  - Reduces the variance (σ^2) of the experimental error.
  - Allows for better detection of treatment factor effects.
- In practice, it may be difficult to select such experimental units.
  - Especially if you would like to generalize your results.
- *Blocking may help to solve this problem.*

Illustrative Example 1

- Suppose we would like to examine the impact of various workout routines on stress and anxiety levels.
- Based on standardized test scores, we would see a wide variance in the stress and anxiety levels of the general population.
  - Which would make it difficult to detect any significant differences (high variance).
- How do we reduce the variance and examine homogeneous experimental units?

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.

Road Map

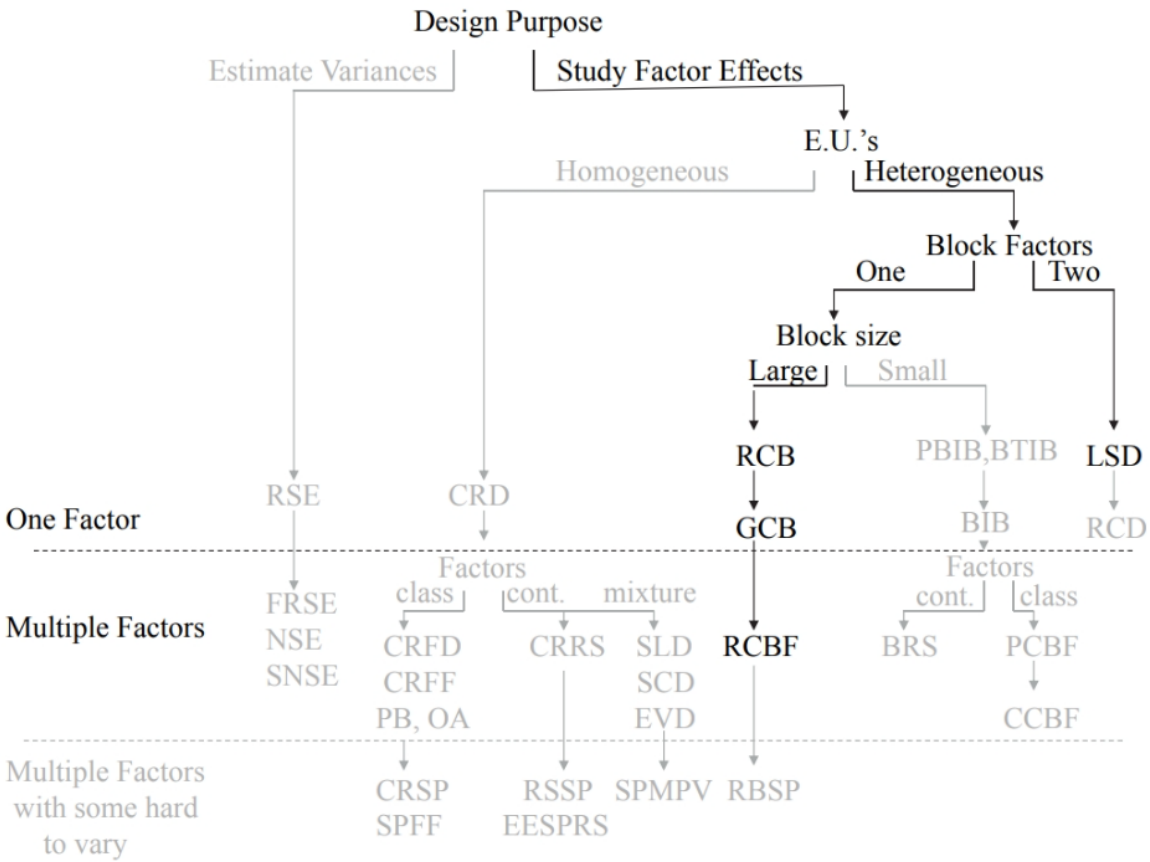

Figure: Source: (1)

Assigning Blocks Basics

- Physical entities: block (group) by similar physical characteristics.
  - Plots of land in agricultural experiments are usually blocked by proximity (similar soil characteristics)
- Animals: Genetically similar animals (same litter).
- Trials/treatments: Usually blocked by the same time period.
  - Lurking variables may change over time.

Randomized Complete Block Design

- In a **randomized complete block design (RCB)** with one treatment factor of t levels there will be b blocks.
- Each block will contain exactly t experimental units.
  - $n = t \cdot b$
- The t EUs in each block are as similar as possible.
- The randomization is performed within each block.

General Complete Block Design

- A **general complete block design** occurs if there are replicates within each block.
- Resulting in a total of $n = r \cdot t \cdot b$ experimental units.
- Generally the **randomized complete block design (RCB) is preferred over the general complete block design.**
  - Smaller blocks mean greater homogeneity within the blocks.
  - Smaller experimental error leading to more precise tests and estimates.

Creating an RBC in R I

- This can be done using base R:
  - `set.seed(2020)`
  - `f <- factor(c(1,2,3,...,t))` *Levels in your factor*
  - `b1t <- sample(f,t)` *randomize* ***each*** *block*
  - ⋮
  - `bbt <- sample(f,t)`
  - `t <- c(b1t, ..., bbt)` *Combine treatment vectors*
  - `block <- factor(rep(c(``block.1", ..., ``block.b"), each = t))` *label the block names*
  - `fnum <- rep(f,t)`
  - `plan <- data.frame(Block.Col.Name = block, EU.Number = fnum, treatment = t)`

- *You can save you design if you want*

Creating an RCB in R II

- This can be done using the *agricolae* package:
  - `treat <- c(1,2,3,...,t)` *Levels in your factor*
  - `outdesign <- design.rcbd(treat, b, seed = 2030)` *randomize* ***each*** *block*
  - `rcb <- outdesign$book`
  - `levels(rcb$block) <- c('Block.1',...,'block.b')`

- *You can save you design if you want*
- **The plots column represents the EUs.**

Example 1

- Suppose we are designing an experiment to investigate methods for extending the life of cut flowers.

- The treatment factor is the liquid in the vase:
  1. Tap water
  2. Tap water with one spoon full of sugar
  3. Tap water with one cup of carbonated water
  4. Tap water with one up of cola

- We have four different kinds of cut flowers (blocks):
  1. Rose
  2. Carnation
  3. Daisy
  4. Tulip

- **Use both methods in R to design the experiment.**

Mathematical Model (RCB) I

- The mathematical model for a randomized complete block design (RCB) with one factor can be written as:

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

- b_i represents the block effects.
- τ_j represents the treatment effets.

- **The usual assumptions of normality of experimental error and homogeneity of variance of experimental error across levels of the treatment factors and blocks are required for this model**.
  - *Can use the methods from earlier on in the course to check these assumptions.*

Mathematical Model (RCB) II

- This is an *additive* model.
  - If interactions are considered, there are not enough degrees of freedom to estimate the variance of the error term.
- The difference in treatment effects among blocks is exactly what the interaction measures and is therefore the correct error term.
- By leaving the interaction out of the model, the *ssE* becomes identical to the interaction sums of squares.

Analysis of Variance Table (ANOVA)

- When there are b blocks, t levels for the treatment, and 1 replication for each group, the ANOVA table for a randomized complete block design (RCB) with one factor:

Source	df	Sum of Squares (ss)	Mean Squares	F-ratio
Blocks	$b-1$	$ssBlk$ $\mathrm{R}(b\|\mu)$	$\frac{ssBlk}{(b-1)}$	
Treatments	$t-1$	ssT $\mathrm{R}(\tau\|b,\mu)$	$\frac{ssT}{(t-1)}$	$F=\frac{msT}{msE}$
Error	$(b-1)(t-1)$	ssE	$\frac{ssE}{(b-1)(t-1)}$	

Variance Estimates I

- The estimate of the variance of the homogeneous experimental units within each block:

$$\hat{\sigma}^2_{rcb} = \frac{ssE}{(b-1)(t-1)} \tag{2}$$

- An estimate of the variance of the entire group of heterogeneous experimental units (from RCB ANOVA):

$$\hat{\sigma}^2_{crd} = \frac{ssBlk + ssE}{t(b-1)} \tag{3}$$

Variance Estimates II

- The ratio of $\hat{\sigma}^2_{crd}$ and $\hat{\sigma}^2_{rcb}$ is a measure of the efficacy of the blocking.

- Relative efficiency (RE):

$$RE = \frac{(\nu_{rcb} + 1)(\nu_{crd} + 3)\hat{\sigma}^2_{crd}}{(\nu_{rcb} + 3)(\nu_{crd} + 1)\hat{\sigma}^2_{rcb}} \tag{4}$$

- Where $\nu_{rcb} = (b-1)(t-1)$ (RCB error degrees of freedom) and $\nu_{crd} = t(b-1)$ (CRD error degrees of freedom).

- RE can be used to determine how many observations would be needed in a CRD with heterogeneous EUs:

$$EUs = RE \cdot (b \cdot t)$$

ANOVA Table in R

- To perform this analysis in R we need to examine the additive relationship:
  - `model <- aov(response ~ block + factor, data = data)`
  - `summary(model)`

- *If the factor is significant, we can examine the nature of the relationship.*

Example 2

- Import the *drug* data from the *daewr* package.
- Take some time to get to know the data.
- What do you think the blocks should be for this experiment?
- Use R to construct an ANOVA table to determine any significance of the treatment factor.

Comparisons of Means

- If the factor levels are significant and quantitative, we can examine the nature of the relationship.
  - *Is there is a polynomial relationship?*

- In R we can use the `contr.poly()` function:
  - `contrasts(data$factor) <- contr.poly(5)` *check up to* 5^{th} *order polynomial*
  - `model <- aov(response ~ block + factor, data = data)`
  - `summary.aov(model, split = list(factor = list('Linear' = 1, 'Quadratic' = 2, 'Cubic' = 3, 'Quartic' = 4)))`

- The results will indicate which trends are significant.

Example 3

- Using the data from Example 2, examine the nature of the relationship between the *dose* and the *rate*.

Visualize the relationship

- In R you can visualize the relationship:
  - `R <-do.call("cbind", split(data$response, data$block))`
  - `y <-apply(R, 1, mean )`
  - `x <-as.double( levels(data$factor) )`
  - `plot( x, y, xlab = "factor", ylab = "response" )`
  - `xx <-seq( 0.0, 2.0, .1 )` *Numeric sequence of factor levels*
  - `rate.quad <-lm( y  poly( x, 2) )` *Change 2 to change polynomial level*
  - `lines(xx, predict( rate.quad, data.frame( x = xx) ))` *Draw relationship*

Example 4

- Change the code to plot the highest order polynomial relationship you found in Example 3.

Example 5

- How effective is the blocking in this example?
- Hint:
  - Relative efficiency (RE):

$$RE = \frac{(\nu_{rcb} + 1)(\nu_{crd} + 3)\hat{\sigma}^2_{crd}}{(\nu_{rcb} + 3)(\nu_{crd} + 1)\hat{\sigma}^2_{rcb}}$$

  - Where $\nu_{rcb} = (b-1)(t-1)$ (RCB error degrees of freedom) and $\nu_{crd} = t(b-1)$ (CRD error degrees of freedom).

Number of Blocks

- To identify the number of blocks needed for a certain power (usually 0.8 - 0.9) using R:
  - `bmin <- min`
  - `bmax <- max`
  - `alpha <- 0.05`
  - `sigma2 <-` σ^2_{rcb}
  - `css <-` $\sum_j \tau_j^2$ (can get it from ANOVA table)
  - `nu1 <-` $t - 1$
  - `blocks <- c(bmin:bmax)`
  - `nu2 <- (blocks - 1) * nu1`
  - `nc <- (blocks * css) / sigma2`
  - `Power <- Fpower(alpha, nu1, nu2, nc )`
  - `data.frame(blocks, nu1, nu2, nc, Power)`

Example 6

- Assuming an RCB is used in the *drugs* data how many blocks do we need to achieve a power of 0.9 or greater?

- Assume the following information (from our first ANOVA table):
  - $\sigma^2_{rcb} = 0.0083487$
  - $\sum_j \tau_j^2 = 0.4602$

Exercise 1

- Using the *RCB.csv* data conduct the following analysis:
  1. Construct an ANOVA table with the response variable being y and interpret the results.
  2. Is there a polynomial relationship?
  3. Visualize the highest order polynomial (if it exists).
  4. How effective is the blocking in this example?
  5. How many blocks do we need to achieve a power of 0.9 or greater?

References & Resources

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

- `design.rcbd()`
- `contr.poly()`
- `Fpower()`