

C2M4: Peer Reviewed Assignment

Outline:

The objectives for this assignment:

1. Get a better understanding of Experimental design patterns.
2. Prove some of the background intuition in blocking and interblock interactions.
3. Understand how and when to apply different model structures for different experimental designs.

General tips:

1. Read the questions carefully to understand what is being asked.
2. This work will be reviewed by another human, so make sure that you are clear and concise in what your explanations and answers.

Problem 1: Experimental Design

This problem is to get you thinking about how experiments are designed and how data is collected, because those influence what models we end up using.

1. (a)

In your own words, define experimental design. Describe some negative effects of making an incorrect experimental design decision.

Experimental Design refers to the process of planning an experiment to ensure that the data collected can provide valid and objective conclusions. It involves determining how the experiment will be conducted, including the selection of experimental units, the allocation of treatments, and the method of data collection. The key elements of experimental design include randomization, replication, and blocking to control for variability and ensure the results are reliable and unbiased.

Negative Effects of Incorrect Experimental Design Decisions:

- **Confounding Variables:** Poor design can lead to confounding, where the effect of the treatment is mixed with other variables, making it difficult to determine the actual cause of observed effects.
- **Bias:** Non-random allocation of treatments can introduce bias, affecting the validity of the conclusions.
- **Reduced Precision:** Insufficient replication or improper blocking can increase variability, reducing the precision of the results.
- **Inability to Generalize:** If the experimental units are not representative of the population, the findings may not be generalizable.
- **Ethical Issues:** Inappropriate design can lead to unethical practices, such as exposing subjects to unnecessary risks or providing insufficient treatment.

1. (b)

In your own words, describe the difference between an experimental unit and a treatment unit. Why does this distinction matter?

Experimental Unit: An experimental unit is the smallest division of the experimental material such that any two units can receive different treatments. It is the entity to which a treatment is applied.

Treatment Unit: A treatment unit refers to the different levels or types of treatment applied to the experimental units. It represents the specific interventions or conditions imposed in the experiment.

Importance of Distinction:

- The distinction matters because it affects the analysis and interpretation of results. Incorrectly identifying experimental units can lead to incorrect conclusions about treatment effects.
- Understanding this distinction helps in proper randomization and replication, ensuring the validity and reliability of the experiment.
- It ensures that the statistical analysis matches the experimental design, leading to correct estimation of treatment effects and variability.

Problem 2: Proving the Intuition

Show that, for the randomized complete block design:

$$SS_{total} = SS_{treat} + SS_{block} + SS_R$$

```

In [1]: # Load necessary Library
library(dplyr)

# Example data for RCBD
# Assume we have 3 treatments (A, B, C) and 4 blocks (1, 2, 3, 4)
data <- data.frame(
  block = rep(1:4, each = 3),
  treatment = rep(c("A", "B", "C"), times = 4),
  response = c(15, 19, 22, 17, 21, 23, 18, 20, 24, 20, 22, 25)
)

# Calculate the grand mean
grand_mean <- mean(data$response)

# Calculate the treatment means
treatment_means <- data %>%
  group_by(treatment) %>%
  summarize(mean_response_treat = mean(response))

# Calculate the block means
block_means <- data %>%
  group_by(block) %>%
  summarize(mean_response_block = mean(response))

# Calculate SS_total
SS_total <- sum((data$response - grand_mean)^2)

# Calculate SS_treat
SS_treat <- sum((treatment_means$mean_response_treat - grand_mean)^2) * 4 # Mult

# Calculate SS_block
SS_block <- sum((block_means$mean_response_block - grand_mean)^2) * 3 # Multiply

# Calculate the residuals
data <- data %>%
  left_join(treatment_means, by = "treatment") %>%
  left_join(block_means, by = "block")

# Residuals
data <- data %>%
  mutate(residual = response - mean_response_treat - mean_response_block + grand_mean)

# Calculate SS_R
SS_R <- sum(data$residual^2)

# Print the results
cat("SS_total:", SS_total, "\n")
cat("SS_treat:", SS_treat, "\n")
cat("SS_block:", SS_block, "\n")
cat("SS_R:", SS_R, "\n")
cat("SS_total == SS_treat + SS_block + SS_R:", all.equal(SS_total, SS_treat + SS_block + SS_R), "\n")

```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
SS_total: 95
SS_treat: 72
SS_block: 20.33333
SS_R: 2.666667
SS_total == SS_treat + SS_block + SS_R: TRUE
```

- 1. Data Preparation: We create a sample dataset with block, treatment, and response values.**
- 2. Grand Mean Calculation: Calculate the overall mean of the responses.**
- 3. Treatment Means Calculation: Calculate the mean response for each treatment.**
- 4. Block Means Calculation: Calculate the mean response for each block.**
- 5. Sum of Squares Calculations:**
 - **SS_total:** Total sum of squares of the responses from the grand mean.
 - **SS_treat:** Sum of squares of the treatment means from the grand mean, multiplied by the number of blocks.
 - **SS_block:** Sum of squares of the block means from the grand mean, multiplied by the number of treatments.
 - **SS_R:** Sum of squares of the residuals (differences between observed and expected values).

- 1. Validation: Print the sum of squares and check if**

$$SS_{total} = SS_{treat} + SS_{block} + SS_R$$

Problem 3: Interblock Interactions

Describe why, in a randomized complete block design (RCBD), it is not possible to test whether interactions exist between the treatment and blocks.

In a Randomized Complete Block Design (RCBD), it is not possible to test for interactions between treatments and blocks because the primary purpose of blocking is to control for variability among experimental units by grouping similar units together within blocks. The focus in RCBD is to remove block effects to get a more accurate estimate of treatment effects.

Reasons for Not Testing Interactions in RCBD:

- **Homogeneity Within Blocks:** Blocks are created to be as homogeneous as possible, meaning the variability within blocks is minimized. This reduces the likelihood of detecting interactions between treatments and blocks because the blocks themselves are designed to be consistent.
- **Confounding Effects:** The block effects are considered nuisance variables and are not of primary interest in the analysis. They are controlled to reduce variability, not to be examined for interactions. Testing for interactions would confound the main treatment effects with the block effects, making it difficult to interpret the results.
- **Model Simplicity:** The RCBD model typically includes main effects for treatments and blocks but does not include interaction terms. Including interaction terms would complicate the model and potentially reduce the power to detect the main treatment effects.
- **Blocking Objective:** The main objective of blocking is to reduce experimental error by accounting for known sources of variability. It is not intended to explore potential interactions between treatments and blocks.

Problem 4: 99 Designs for 99 Problems

For each of the following design patterns, give an example (that wasn't given in the videos) for an experiment that would best lend itself to the specified design pattern. Make sure to explain why the specified design is more applicable for your experiment than the other design patterns.

1. Complete Randomized Design (CRD)
2. Complete Randomized Block Design (CRBD)
3. Factorial Design

1. Complete Randomized Design (CRD)

Example: Testing the Effect of Different Fertilizers on Plant Growth

Explanation:

- **Design:** A group of plants is randomly assigned to different fertilizer treatments without considering any other factors. Each plant receives only one type of fertilizer, and the growth is measured after a specified period.
- **Reason:** CRD is suitable here because it allows for a straightforward comparison of fertilizer effects. It is assumed that all other conditions (e.g., sunlight, water, soil) are uniform or have minimal impact. Random assignment helps to ensure that any other variations are evenly distributed across treatments.

1. Complete Randomized Block Design (CRBD)

Example: Evaluating the Effect of Different Teaching Methods on Student Performance Across Different Schools

Explanation:

- **Design:** Students within each school are grouped into blocks. Within each block (school), students are randomly assigned to different teaching methods. The performance of students is then measured.
- **Reason:** CRBD is appropriate here because students within the same school are likely to have similar characteristics (e.g., socio-economic background, school resources). Blocking by school helps to control for this variability, allowing a more accurate comparison of the teaching methods.

1. Factorial Design

Example: Studying the Combined Effects of Exercise and Diet on Weight Loss

Explanation:

- **Design:** Participants are randomly assigned to one of several combinations of exercise levels (e.g., no exercise, moderate exercise, intense exercise) and diet types (e.g., low-carb, low-fat, balanced diet). Each combination is a treatment group, and weight loss is measured over a period.
- **Reason:** Factorial design is suitable here because it allows for the investigation of both main effects and interactions between exercise and diet. It helps in understanding whether the effect of one factor (e.g., exercise) depends on the level of the other factor (e.g., diet).