C2M4_peer_reviewed

September 22, 2024

1 C2M4: Peer Reviewed Assignment

1.0.1 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 strucutres 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.

2 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.

2.0.1 1. (a)

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

3 My Answer:

Experimental design is the systematic framework through which researchers plan and conduct experiments to investigate the relationships between variables. It involves carefully selecting and manipulating independent variables, establishing control and experimental groups, and ensuring that the conditions are consistent and unbiased across all trials. A well-crafted experimental design aims to isolate the effects of the variables of interest, thereby enabling accurate and reliable conclusions about cause-and-effect relationships.

Making incorrect decisions in experimental design can have several detrimental effects. For instance, flawed designs may introduce confounding variables that obscure the true relationship between the studied factors, leading to invalid or misleading results. This can undermine the credibility of the research and render the findings unusable for further scientific inquiry or practical application. Additionally, poor experimental design often results in increased variability and reduced statistical power, making it difficult to detect significant effects even if they exist. Inefficient use of resources, such as time, funding, and materials, is another negative consequence, as flawed experiments may need to be repeated or abandoned altogether. Furthermore, incorrect design choices can lead to ethical issues, especially in studies involving human or animal subjects, where inadequate controls or biases can cause harm or distress. Ultimately, these negative effects not only compromise the integrity and usefulness of individual studies but also hinder the overall progress and trustworthiness of scientific research.

3.0.1 1. (b)

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

4 My Answer:

In experimental design, an experimental unit refers to the smallest entity to which a treatment is directly applied. This could be an individual organism, a specific plot of land, a laboratory sample, or an entire environment, depending on the study's context. For example, in a clinical trial testing a new medication, each patient receiving the treatment would be considered an experimental unit. Conversely, a treatment unit is essentially synonymous with the experimental unit, representing the entity that undergoes the treatment or intervention being studied. This distinction is vital because accurately identifying the experimental or treatment units ensures that treatments are administered consistently and appropriately across all subjects or samples. Misidentifying these units can lead to confounded results, where the effects of different treatments become indistinguishable or where variability within the data is improperly attributed. Moreover, understanding the difference between experimental units and the units from which data is collected (often referred to as sampling units) is crucial for valid statistical analysis. It ensures that the variability is correctly partitioned, preventing issues such as inflated sample sizes or underestimated standard errors, which can compromise the study's reliability and the validity of its conclusions. Therefore, clearly distinguishing between experimental and treatment units is fundamental to designing robust experiments that yield accurate and meaningful results.

5 Problem 2: Proving the Intuition

Show that, for the randomized complete block design:

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

6 My Answer:

In the context of a **Randomized Complete Block Design (RCBD)**, it's essential to understand how the **Total Sum of Squares ($SS_{\textbf{total}}$)** is partitioned into its constituent components: **Treatment Sum of Squares ($SS_{\textbf{treat}}$)**, **Block Sum of Squares ($SS_{\textbf{block}}$)**, and **Residual Sum of Squares ($SS_{\textbf{R}}$)**. This partitioning helps in assessing the variability attributable to treatments, blocks, and random error, respectively.

Mathematically, the relationship is expressed as:

$$SS_{\text{total}} = SS_{\text{treat}} + SS_{\text{block}} + SS_{\text{R}}$$

Let's delve into the proof of this equation.

Definitions and Notations

- **Observations:** Let y_{ij} denote the observation in the i^{th} treatment and j^{th} block.
- **Overall Mean:** $\bar{y}_{..} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij}$
- **Treatment Mean:** $\bar{y}_{i.} = \frac{1}{b} \sum_{j=1}^{b} y_{ij}$
- **Block Mean:** $\bar{y}_{.j} = \frac{1}{a} \sum_{i=1}^{a} y_{ij}$
- **Number of Treatments:** a
- **Number of Blocks:** b

Total Sum of Squares (SS_{total})

The Total Sum of Squares measures the total variability in the data and is given by:

$$SS_{\text{total}} = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{..})^2$$

Treatment Sum of Squares (SS_{treat})

This measures the variability due to the treatment effects:

$$SS_{\text{treat}} = b \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2$$

Block Sum of Squares (SS_{block})

This captures the variability due to differences between blocks:

$$SS_{\text{block}} = a \sum_{j=1}^{b} (\bar{y}_{.j} - \bar{y}_{..})^2$$

Residual Sum of Squares (SS_{R})

Also known as the Error Sum of Squares, it accounts for the variability within treatments and blocks:

$$SS_{\mathrm{R}} = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

Proof of Partitioning

To prove that $SS_{\text{total}} = SS_{\text{treat}} + SS_{\text{block}} + SS_{\mathrm{R}}$, we'll expand and simplify the expressions.

Expanding $SS_{\textbf{total}}$:

$$SS_{\text{total}} = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2$$

Expanding $SS_{\textbf{treat}} + SS_{\textbf{block}}$:

$$SS_{\text{treat}} + SS_{\text{block}} = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2$$

Adding $SS_{\mathbf{R}}$:

$$SS_{\text{treat}} + SS_{\text{block}} + SS_{\mathrm{R}} = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

Expanding the Residual Sum of Squares:

$$SS_{\mathrm{R}} = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

Notice that this term represents the variation that is not explained by the treatment or block effects.

Combining All the Terms:

When you sum SS_{treat}, SS_{block}, and SS_{R}, you account for all possible sources of variability in the data:

- **Between-Treatment Variability ($SS_{\textbf{treat}}$):** Differences due to different treatments.
- **Between-Block Variability ($SS_{\textbf{block}}$):** Differences due to different blocks.
- **Within-Treatment and Within-Block Variability ($SS_{\mathbf{R}}$):** Random error or unexplained variability.

Tthe total variability in the data (SS_{total}) is fully partitioned into these three components.

7 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 treatment and blocks because blocks are not considered as random variables that can interact with treatments. Instead, blocks are treated as a way to control for variability among experimental units, and the primary focus is on isolating the treatment effects. Since the design assumes that block effects are additive and do not interact with treatments, testing for such interactions is outside the scope of RCBD."

8 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)

Testing the effect of different fertilizers on plant growth in a uniform greenhouse environment. Why CRD?: In this case, all plants are subjected to the same environmental conditions, and the only variable is the type of fertilizer. A completely randomized design is appropriate as it randomly assigns treatments without the need for blocking. 2. Complete Randomized Block Design (CRBD)

Studying the effect of different diets on the weight gain of animals across different farms. Why CRBD?: Different farms (blocks) may have varying environmental conditions, which could influence weight gain. Blocking by farm controls for this variability, while allowing the effects of the diets to be isolated. 3. Factorial Design

Investigating the effect of temperature and humidity on the shelf life of a product. Why Factorial Design?: A factorial design allows the experimenter to study the effect of each factor (temperature and humidity) independently, as well as their interaction, on the shelf life. This design is more efficient than conducting separate experiments for each factor.

[ ]:
```
```