Planning Experiments

Davood Tofighi, Ph.D.

Planning Experiments

Experimental design forms the foundation of scientific inquiry whenever the goal is to understand how certain controllable inputs (factors) influence outcomes (responses). By carefully planning how to collect data and assigning treatments to experimental units, we can draw valid causal inferences about these relationships. Unlike observational studies, experiments involve actively manipulating factors to reveal their effects while controlling for unwanted variability.

This set of lecture notes provides a comprehensive overview of experimental design concepts and applications. We will address how to define clear objectives, classify sources of variation, choose appropriate experimental designs, apply randomization, incorporate blocking, specify statistical models, and determine appropriate sample sizes. Real-world examples and R code snippets will illustrate key ideas. Additionally, we will integrate exercises, class activities, summaries, challenges, and visual aids to reinforce understanding.

Objectives

1. Understand how to identify and classify sources of variation in an experiment.
2. Define treatment factors, factor levels, and experimental units.
3. Explore blocking, nuisance factors, and strategies to manage them.

4. Grasp the principles of randomization and why it underpins statistical inference.
5. Examine standard experimental designs (CRD, RBD, Factorial, Latin Square, Split-Plot) and when to use each.
6. Understand the distinction between fixed and random effects in modeling.
7. Learn how to plan data collection, run pilot experiments, and determine the number of observations (sample size).
8. Gain hands-on experience using R code for randomization, exploratory data analysis, and basic ANOVA.

Readings

- Dean et al. (2017, Ch. 3)

Checklist for Experimental Design

Steps in Planning an Experiment

It is very tempting to jump into data collection without a clear plan. However, a well-thought-out experimental design is crucial for obtaining valid and interpretable results. Here is a checklist of steps to guide you through the planning process:

1. Define the Objectives of the Experiment

- Clearly state the goals (e.g., comparing treatments, optimizing factors, studying interactions).
- Write objectives as specific questions or hypotheses to guide the design.

2. Identify All Sources of Variation

1. **Treatment Factors and Their Levels**
    - Factors actively manipulated (e.g., temperature, dosage) and their specific levels.

- Example: Dosage levels: 10 mg, 20 mg, 30 mg.

2. **Experimental Units**
    - Smallest division of material receiving a treatment independently.
    - Example: Individual plants in a field trial.
3. **Blocking Factors, Noise Factors, and Covariates**
    - Blocking: Group similar units (e.g., day of testing).
    - Noise: Uncontrollable variability (e.g., weather conditions).
    - Covariates: Measureable properties (e.g., baseline performance).

3. Choose a Rule for Assigning Experimental Units to Treatments

- Select randomization techniques:
    - Completely Randomized Design (CRD).
    - Randomized Block Design (RBD).
    - Split-Plot Design.

4. Specify the Measurements, Procedure, and Anticipated Difficulties

1. **Measurements**
    - Specify precision, units, and frequency of measurement.
    - Example: Growth in cm, tensile strength in MPa.
2. **Procedure**
    - Step-by-step instructions for consistent data collection.
3. **Anticipated Difficulties**
    - Identify challenges (e.g., equipment failure) and mitigation strategies.

5. Run a Pilot Experiment

- Test feasibility, validate instruments, and refine factor levels or treatment combinations.

6. Specify the Statistical Model

- Example for one-way ANOVA:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

where:

- Y_{ij}: Observed response.
- μ: Overall mean.
- τ_i: Treatment effect.
- ε_{ij}: Random error term.

7. Outline the Analysis

- Plan descriptive and inferential methods (e.g., ANOVA, regression).
- Include diagnostic checks for model assumptions.

8. Calculate Sample Size

- Perform power analysis based on variability estimates and desired precision.

9. Review and Revise

- Revisit and refine all decisions based on pilot results and practical constraints.

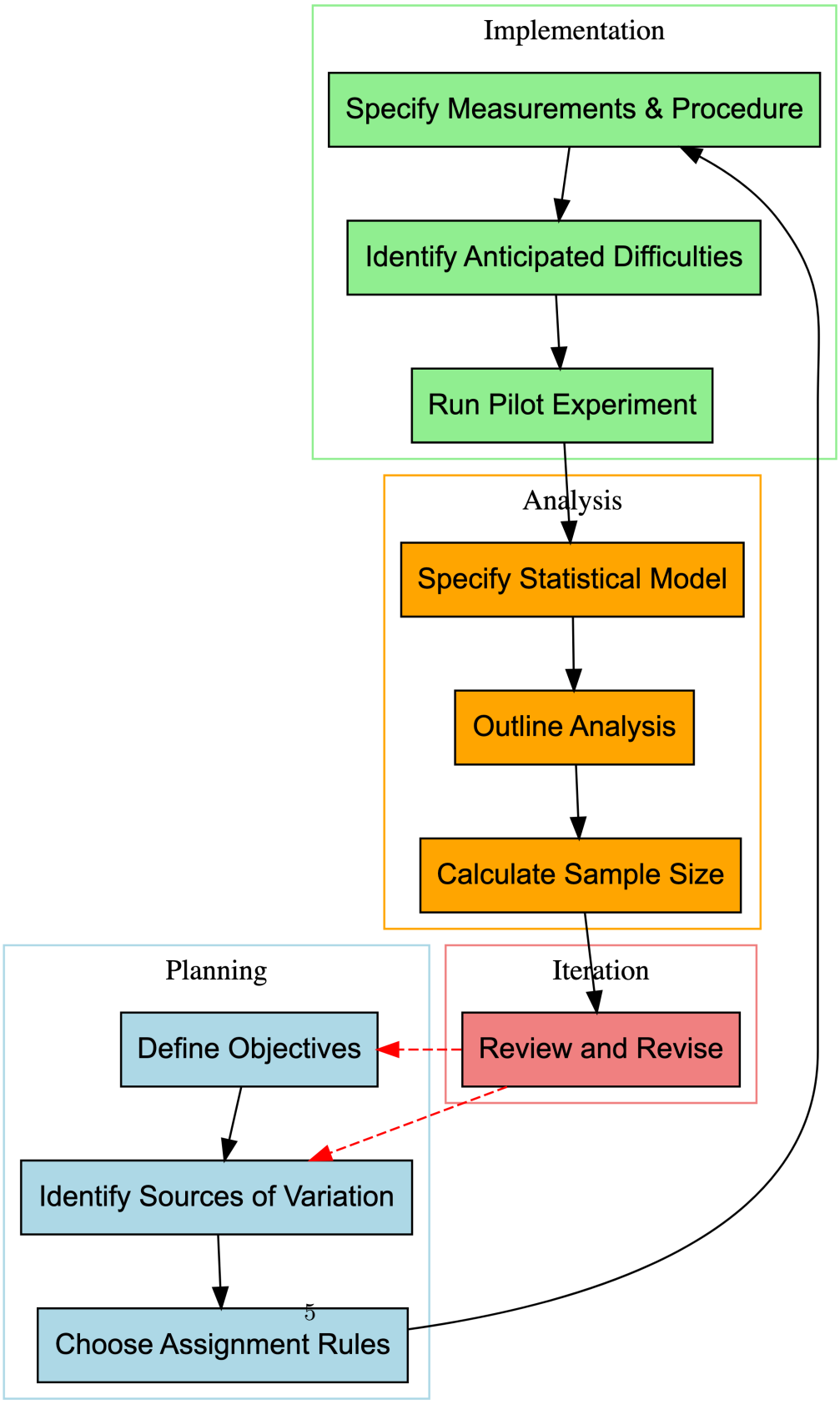

Figure 1: Experimental Design Planning Flowchart

Flowchart of Experimental Design Steps

Explaining Experimental Design Concepts

Identifying Experiment Objectives and Variation Sources

Setting Objectives

Every experiment must start with a clear research objective. Ask: *What do we hope to learn or achieve?* Objectives may include:

- Determining which manufacturing process leads to the highest product quality.
- Comparing new teaching methods for improved student performance.
- Finding the optimal settings of machine parameters to reduce defects.

Without well-defined objectives, the experiment risks being unfocused, making the results difficult to interpret.

Variation Sources

Variation in responses can arise from multiple sources:

- **Major Variation (Treatment Effects):** Due to the factors of interest (e.g., different fertilizers or drug dosages).
- **Nuisance Factors (noise):** Variables that influence the response but are not of primary interest (e.g., room temperature, operator skill). Some nuisance factors can be controlled or incorporated as blocking factors, while others must be accepted as noise.

By identifying and classifying sources of variation, we can decide which factors to manipulate as treatments, which to fix or control, and which to block or treat as covariates.

Example

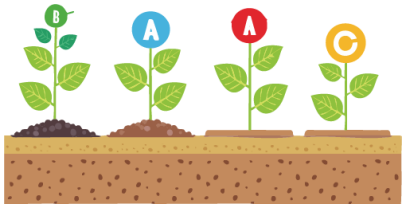

Figure 2: Variation Sources in an Experiment

Consider an agricultural experiment testing different fertilizer types (A, B, C). The main factors are the fertilizer types. Soil heterogeneity could be a nuisance factor. If the field has varying soil quality, it might mask the fertilizer effect. Recognizing this early allows us to incorporate blocking or other strategies to handle this nuisance variation.

Activity (Work Out Experiment): Think-Pair-Share

1. **Think:** Imagine you are testing two workout plans (high-intensity vs. moderate-intensity) to improve muscle strength.
2. **Pair:** Discuss with a partner what the main objective is and what factors could cause unwanted variation (e.g., baseline fitness level, diet).
3. **Share:** Volunteer pairs share their classification of factors and whether they could be controlled or blocked.

Challenges and Common Mistakes

- Starting without a clear objective can lead to inconclusive results.
- Confusing major variation with nuisance variation can misdirect the design.

Exercises

1. Why must objectives be clear before starting an experiment?

2. Given three fertilizer types and two soil conditions, identify which factors should be considered treatments and which might be nuisance factors.
3. Explain how failing to control a nuisance factor increases the variance of treatment effect estimates.

Key Takeaways: - Well-defined objectives shape the experiment. - Classifying factors into treatment and nuisance categories enhances design efficiency.

Treatment Factors and Levels

Defining Treatment Factors

A **treatment factor** is any variable **deliberately** manipulated by the experimenter to study its effect on the response.

Example: In a detergent study, water temperature and detergent concentration are treatment factors.

Factor Levels

Factor levels are the specific settings or categories of a factor. For instance, twist levels in a cotton-spinning experiment might be 1.63, 1.69, 1.78, and 1.90 turns per inch. Chosen based on subject matter knowledge, factor levels must be realistic and relevant.

Activity (Baking Experiment): Design Your Own Factorial Experiment

Individually pick a factor (e.g., cooking time) and propose three levels. Pair up with another student who has chosen a different factor (e.g., oven temperature), and combine them into a factorial design. Discuss the complexity and the number of total treatment combinations.

Challenges and Common Mistakes

- Choosing impractical or irrelevant factor levels can undermine the experiment's usefulness.
- Too many factor levels complicate analysis without necessarily adding insight.

Exercises

1. What criteria guide the selection of factor levels?
2. For a factor with 4 levels and another with 3 levels, determine how many treatment combinations exist.
3. Explain how replicates per factor level (with total resources fixed) decreases the variance of treatment effect estimates.

Key Takeaways: - Treatments factors are deliberately manipulated variables. - Factor levels must be meaningful, informed by real-world considerations or pilot data.

Experimental Units, Blocking, and Nuisance Factors

Experimental Units

The **experimental unit** is the smallest entity to which treatments are independently assigned. This concept is crucial for valid statistical inference.

Examples of Experimental Units

- Plants in a field trial if each plant gets a different fertilizer.
- Engine test benches if each engine is run under a unique setup.
- Students in a classroom if each student receives a different teaching method.
- Patients in a clinical trial if each patient gets a different drug.

Blocking Factors

When a known nuisance factor could confound results, use **blocking**. Blocking forms groups of similar experimental units, each block receiving all treatments. This controls for block-to-block variation, increasing *precision*.

Equation (Block Design):

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

{#eq:block-design}

Here, τ_i is the treatment effect and β_j the block effect.

Noise Factors and Covariates

- **Noise Factors:** Uncontrollable variations like daily temperature fluctuations.
- **Covariates:** Measured continuous variables (e.g., initial weight) that help explain variation when included in the model.

Factor Classification Flowchart

Activity: Blocking Brainstorm

Consider a baking experiment where ovens differ in temperature calibration. Could ovens form blocks? Discuss how blocking might improve detection of differences in recipes.

Challenges and Common Mistakes

- Confusing the experimental unit with observational units leads to incorrect analysis.
- Ignoring a known source of variation that could be blocked may inflate residual error.

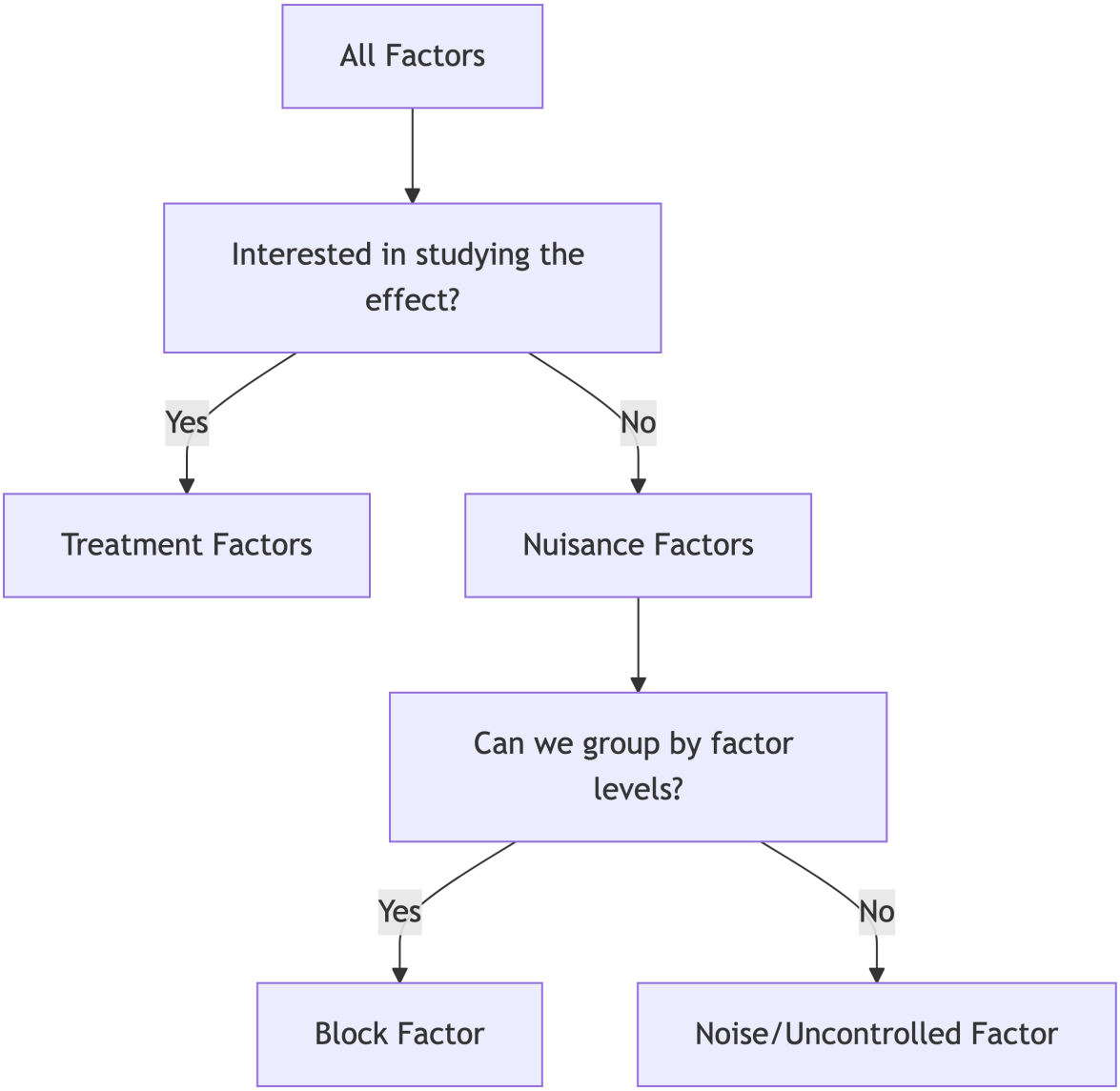

Figure 3: Factor Classification Flowchart

Exercises

1. **Conceptual:** Differentiate between experimental and observational units.
2. **Numerical:** If you have 4 treatments and 6 blocks (each block gets all 4 treatments), how many total experimental units are there?
3. **Proof/Derivation:** Show that including a block factor in the model reduces the residual variance compared to an unblocked design.

Key Takeaways:

- Correctly identifying experimental units is essential.
- Blocking controls known nuisance factors, improving the experiment's power to detect treatment effects.

Principles of Randomization

Randomization Concept

Randomization ensures that each experimental unit has an equal chance of receiving any treatment. This prevents systematic bias and justifies the assumptions underlying standard statistical tests.

Example: Use a random number generator to assign treatments to plots, ensuring no predetermined pattern.

How to Perform Randomization

Randomization ensures unbiased allocation of experimental units to treatments and reduces the risk of systematic error. Here's how you can perform randomization in practice:

Procedure

1. **List Experimental Units**:
    - Create a list of all experimental units (e.g., subjects, plots of land).
2. **List Treatments**:
    - Specify the treatments or factor levels to be assigned.
3. **Determine Design Type**:
    - Choose a design type: Completely Randomized Design (CRD), Randomized Block Design (RBD), or Split-Plot Design.
4. **Generate Random Numbers**:
    - Use random number generation to assign treatments to units.
5. **Assign Treatments**:
    - Pair random numbers with treatments to allocate them to experimental units.
6. **Verify**:
    - Ensure assignments are correct and check for any unintended patterns.

Pseudocode for Randomization

Completely Randomized Design

```
INPUT: List of experimental units, List of treatments
OUTPUT: Randomized assignments of treatments to units

1. Define N = Total number of experimental units
2. Define T = List of treatments
3. Repeat T to match the number of experimental units if unbalanced
4. Shuffle the list of treatments randomly
5. Assign shuffled treatments to the experimental units
6. Return randomized assignments
```

Randomized Block Design

```
INPUT: List of blocks, List of treatments
OUTPUT: Randomized assignments within each block

1. For each block in blocks:
    a. Shuffle the list of treatments randomly
    b. Assign shuffled treatments to experimental units in the block
2. Combine all block-level assignments
3. Return randomized assignments
```

R Implementation: Randomization Examples

1. Completely Randomized Design

2. Randomized Block Design

Explanation in R

- `rep()`: Repeats the treatment list to match the total number of experimental units.
- `sample()`: Randomizes the order of the treatments.
- `set.seed()`: Ensures reproducibility of the randomization process.
- **`for` loop (RBD)**: Iterates through blocks, randomizing treatments for each.

These R scripts provide easy-to-follow approaches to randomization, ensuring balanced and unbiased treatment assignment. Let me know if you'd like further assistance with advanced designs!

R Example: Simple Random Assignment

- See the help page for rep() in the Appendix for more details.
- See the help page for sample() in the Appendix for more details.

Table 1: Randomized Assignments in CRD

```r
n <- 10 # Number of experimental units
# Define experimental units and treatments
experimental_units <- 1:n # 10 experimental units
treatments <- c("A", "B", "C") # 3 treatments

# Repeat treatments to match the number of units
treatments <- rep(treatments, length.out = n)

# Shuffle treatments randomly
set.seed(123) # For reproducibility
randomized_treatments <- sample(treatments)

# Create a data frame with randomized assignments
randomization_crd <- data.frame(Unit = experimental_units, Treatment = randomized_treatments)
print(randomization_crd)
```

```
   Unit Treatment
1     1         C
2     2         A
3     3         B
4     4         B
5     5         C
6     6         C
7     7         A
8     8         A
9     9         B
10   10         A
```

Table 2: Randomized Assignments in RBD

```r
# Define blocks and treatments
blocks <- 1:4 # 4 blocks
treatments <- c("A", "B", "C") # 3 treatments

# Create a data frame to store results
randomization_rbd <- data.frame(Block = integer(), Unit = integer(), Treatment = character())

# Randomize within each block
set.seed(123)
for (block in blocks) {
    # Randomize treatments
    randomized_treatments <- sample(treatments)

    # Assign treatments to units within the block
    block_data <- data.frame(
        Block = block,
        Unit = seq_len(length(treatments)),
        Treatment = randomized_treatments
    )
    randomization_rbd <- rbind(randomization_rbd, block_data)
}

print(randomization_rbd)
```

```
   Block Unit Treatment
1      1    1         C
2      1    2         A
3      1    3         B
4      2    1         B
5      2    2         A
6      2    3         C
7      3    1         B
8      3    2         C
9      3    3         A
10     4    1         A
11     4    2         B
12     4    3         C
```

```r
set.seed(123)
treatments <- rep(c("A", "B", "C"), each = 4)
assignments <- sample(treatments)
assignments
```

```
 [1] "A" "C" "C" "A" "B" "C" "B" "A" "C" "B" "A" "B"
```

Activity: Randomization Drill

Write down three treatments on slips of paper and randomly assign them to hypothetical units. Discuss how randomization prevents biased placement of favorable treatments.

Challenges and Common Mistakes

- Assigning treatments alphabetically or in a patterned manner is not randomization.
- Randomization must be deliberate, not haphazard.

Exercises

1. Why is randomization crucial for valid statistical inference?
2. With 12 units and 3 treatments, show one method of random assignment.
3. Under randomization and the null hypothesis, show that the expected difference between treatment means is zero.

Key Takeaways:

- Randomization removes systematic bias.
- It underpins the validity of ANOVA and other inferential methods.

Standard Experimental Designs

Completely Randomized Design (CRD)

- No blocking, treatments assigned entirely at random.
- Suitable when units are homogeneous.

Randomized Block Design (RBD)

- Introduce a blocking factor to control known variation.
- Each block receives all treatments, improving the precision of treatment comparisons.

Row-Column and Latin Squares

- Control for two perpendicular nuisance factors simultaneously.

Factorial Designs

- Study multiple factors and their interactions simultaneously.
- Efficiently explore how factors work together.

Split-Plot Designs

- Useful when some factors are harder to change than others.
- Creates a hierarchy of experimental units (e.g., fields as main plots and subplots within fields).

Activity: Design Match-Up

Given three scenarios (no nuisance factor, known nuisance factor, and multiple factors with a complex structure), choose the appropriate design (CRD, RBD, Factorial, Latin Square).

Challenges and Common Mistakes

- Using a CRD when a known nuisance factor should be blocked.
- Ignoring interactions in factorial experiments.

Exercises

1. **Conceptual:** Under what conditions would you choose a CRD over an RBD?
2. **Numerical:** Compute the number of runs in a 2x4 factorial with 3 replicates.
3. **Proof/Derivation:** Show how the variance decomposition changes from CRD to RBD, highlighting the block effect.

Key Takeaways:

- Different designs address different research needs.
- Factorial designs detect interactions, blocking improves precision, and split-plots handle difficult-to-change factors.

Model Specification and Effect Types

Linear Models and ANOVA

Experiments often use a linear model to relate responses to treatments and blocks:

$$Y = \mu + \text{treatment effects} + \text{block effects} + \text{error}$$

{#eq:linear-model}

ANOVA decomposes total variation into components attributable to treatments, blocks, and error, enabling inference on treatment effects.

Fixed vs. Random Effects

- **Fixed Effects:** Chosen levels of interest; inferences apply only to those tested levels.
- **Random Effects:** Levels are a random sample from a broader population; inferences generalize beyond observed levels.

Activity: Fixed or Random?

Consider an experiment with 3 specific fertilizer brands (likely fixed) vs. an experiment using 3 randomly chosen brands from a large market (random). Discuss how interpretation changes.

Challenges and Common Mistakes

- Treating random factors as fixed or vice versa leads to incorrect conclusions.
- Confusing the meaning of random effects with randomization.

Exercises

1. **Conceptual:** Why is it important to distinguish fixed and random effects?
2. **Numerical:** In an RBD, show how expected mean squares differ for a random block factor vs. a fixed block factor.
3. **Proof/Derivation:** Derive the expected mean squares for a two-way ANOVA with one random factor.

Key Takeaways:

- Correct model specification ensures valid inference.
- Fixed vs. random distinction affects interpretation and generalized conclusions.

Planning Data Collection and Pilot Experiments

Measurement and Procedure

Careful data collection planning ensures relevance and quality:

- Choose appropriate measurement scales and ensure instruments are calibrated.
- Train personnel to reduce measurement errors.

Pilot Experiments

A small-scale pilot run can identify unforeseen difficulties, refine factor levels, and confirm data collection procedures before launching the main study.

Activity: Pilot Study Discussion

Discuss what a pilot experiment might reveal in a crop study (e.g., unexpected soil pests) and how that influences the main experiment's design.

Challenges and Common Mistakes

- Omitting a pilot study can lead to costly mistakes in the main experiment.
- Relying on untested procedures risks invalid data collection.

Exercises

1. Why are pilot experiments beneficial?
2. If pilot data suggest variance $= 4$ and you want a margin of error $= 1$, how many samples per treatment are needed (assuming normality)?
3. Show how a pilot-based variance estimate informs sample size calculations.

Key Takeaways:

- Planning and pilot testing prevent wasted resources.
- Well-designed procedures and preliminary runs ensure high-quality, interpretable data.

Determining the Number of Observations

Sample Size and Power

The number of observations affects the experiment's ability to detect true effects:

- **Power:** Probability of detecting a true effect if it exists.
- Larger sample size generally increases power but also increases cost.

Power calculations balance desired precision, variance estimates, and available resources.

Activity: Sample Size Calculation

Given an estimated variance and a desired effect size, estimate the required sample size. Discuss trade-offs between resource constraints and statistical power.

Challenges and Common Mistakes

- Choosing sample size arbitrarily can lead to low power or wasted resources.
- Overly large samples may be unnecessary and expensive.

Exercises

1. **Conceptual:** Explain why sample size must be justified.
2. **Numerical:** Given effect size = 2 (SD=3) and desired power = 0.8, calculate required sample size per group.
3. **Proof/Derivation:** Derive a basic formula relating power, effect size, and sample size for a one-way ANOVA.

Table 3: ANOVA Table

```r
set.seed(123)
treatments <- rep(c("A", "B", "C"), each = 4)
assignments <- sample(treatments)

# Simulate some response data
response <- rnorm(12, mean = 50, sd = 5)

# Fit ANOVA model
model <- aov(response ~ factor(assignments))
summary(model)
```

```
                    Df Sum Sq Mean Sq F value Pr(>F)
factor(assignments)  2   14.2    7.08   0.163  0.852
Residuals            9  391.5   43.50
```

Key Takeaways:

- Adequate sample size ensures meaningful, reliable conclusions.
- Power analysis guides optimal resource allocation.

Practical Example in R

Example: A Simple CRD

Compare three battery types (A, B, C) with 4 replicates each.

R Code

Interpret the ANOVA table:

- If the p-value for factor(assignments) is small, it suggests a difference among battery types.

Activity: Interpret R Output

Examine the ANOVA table. Identify the F-statistic and p-value, and discuss whether treatments differ.

Challenges and Common Mistakes

- Misinterpreting p-values without context.
- Ignoring model assumptions such as normality and equal variances.

Exercises

1. **Conceptual:** Explain why randomization is essential before running ANOVA.
2. **Numerical:** Modify the code for 2 treatments with 5 replicates and interpret results.
3. **Proof/Derivation:** Show how the F-statistic relates to the ratio of variances in ANOVA.

Key Takeaways:

- R facilitates practical application of design concepts.
- Interpreting software output requires understanding design principles and model assumptions.

Summary of Key Takeaways

- **Objectives:** Clear objectives guide the entire design and analysis.
- **Factors and Levels:** Treatment factors and meaningful factor levels ensure relevance.
- **Experimental Units and Blocking:** Correct identification of units and use of blocking increases precision.
- **Randomization:** Ensures unbiased estimates and valid statistical inference.
- **Designs (CRD, RBD, Factorial, etc.):** Different designs solve different problems; factorial designs explore interactions, blocking controls known variation.

- **Fixed vs. Random Effects:** Proper classification determines the scope of inferences.
- **Planning and Pilots:** Thoughtful planning and preliminary trials prevent costly errors.
- **Sample Size and Power:** Adequate sample size ensures detectable effects without wasting resources.
- **R Implementation:** Practical coding examples reinforce theoretical principles.

Definitions of Key Terms

- **Objective:** The main research question or goal of the experiment.
- **Nuisance Factor:** A variable that affects the response but is not of primary interest.
- **Treatment Factor:** A variable manipulated by the experimenter.
- **Factor Levels:** The specific settings or categories of a factor.
- **Experimental Unit:** The entity to which a treatment is applied independently.
- **Block Factor:** A factor used to group units into homogeneous sets.
- **Randomization:** Assigning treatments randomly to avoid bias.
- **CRD (Completely Randomized Design):** A design without blocking; treatments are assigned randomly.
- **RBD (Randomized Block Design):** A design that uses blocks to control known nuisance factors.
- **Factorial Design:** A design that includes multiple factors simultaneously.
- **Fixed Effects:** Effects of chosen factor levels of specific interest.
- **Random Effects:** Effects of factor levels considered as a random sample from a population.
- **Pilot Experiment:** A small preliminary study to refine methods before the main experiment.
- **Power:** The probability of detecting a true treatment effect.

- **ANOVA:** A statistical method for comparing means across groups.

References

Christensen, R. (2018). *Analysis of variance, design, and regression: Linear modeling for unbalanced data* (2nd ed.). Chapman and Hall/CRC.

Dean, A., Voss, D., & Draguljić, D. (2017). *Design and analysis of experiments* (2nd ed.). Springer International Publishing. https://doi.org/10.1007/978-3-319-52250-0

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2020). *Introduction to linear regression analysis* (Fifth edition). Wiley.

Scheffé, H. (1959). *The analysis of variance*. Wiley.