Statistical Design for Animal Research

Dr. Dhafer Malouche

Department of Mathematics and Statistics,
College of Arts and Sciences,
Qatar University.

Outline

Fundamental Concepts in Statistical Design

Definition

Definition

Statistical design in research refers to the structured planning of an experiment or study to ensure that the data collected can be analyzed to yield valid, objective, and reliable conclusions.

- Involves careful selection of animal models.
- Guides allocation of animals to treatment groups.
- Determines methods for measuring and analyzing outcomes.

Purpose

- Maximize precision of the study's findings by ensuring consistent and reliable measurements.
- Minimize variability by controlling factors that introduce noise, such as random errors or biological differences.
- **Balance between precision and variability**:
  - Reducing variability too much can limit generalizability of the findings.
  - The goal is to find a balance that increases precision while keeping the study feasible, ethical, and broadly applicable.
- A well-designed experiment enables researchers to:
  - Draw clear inferences about the relationships between variables.
  - Assess the effect of a treatment on a biological response.

Experiemental Design: Randomization, Replication, and Blocking

Early Steps in Animal Research: Choosing the Best Design

- Once the research question and objectives are defined, choosing the best experimental design is critical.
- A well-chosen design ensures accuracy, reduces variability, and helps ensure efficient use of animals, supporting both ethical and scientific standards.
- **Experimental Design Methods:** Randomization, Replication, Blocking, and others.

What is Randomization?

Randomization:

- Randomization is the process of **randomly assigning** subjects (animals) to different groups in an experiment.
- This ensures that the groups are similar and that no hidden factors (like age or health) unfairly affect the results.
- **Example:** By randomly assigning some rats to the ”drug” group and others to the ”no drug” group, we ensure that:
  - Any differences in response are likely due to the drug.
  - Other characteristics like age, weight, or health status do not unfairly affect the results.

Why Do We Randomize?

- **To make fair comparisons:** Randomization ensures that differences between groups are due to the treatment, not because of other factors.
- **To prevent bias:** It reduces the risk of the experiment unintentionally favoring one group over the other.

Other Methods

- **Replication:**
  - Replication means repeating the experiment under the same conditions to ensure the results are consistent and reproducible.
  - Helps confirm that your findings are reliable and not just due to random chance.
- **Blocking:**
  - Blocking groups similar subjects together (e.g., by age, weight) before randomly assigning them to treatment groups.
  - Reduces the effect of variation between subjects, allowing the focus to be on the treatment's real impact.
- Stratification, Matched-Pairs Design, Crossover Design, Factorial Design, Sequential Design...
- **RandoMice** (van Eenige et al., 2020), User-friendly tool designed for animal research (Downloadable from Github).

Ethical Considerations in Statistical Design

The 3Rs Principle: Replacement, Reduction, and Refinement

Introduction:

- The 3Rs Principle—Replacement, Reduction, and Refinement—was first introduced by Russell and Burch in their 1959 book *The Principles of Humane Experimental Technique*.
- This framework promotes humane animal research by minimizing the use of animals and improving their welfare.
- The 3Rs have since become a cornerstone of ethical animal research, widely adopted in scientific and regulatory practices globally (Vitale et al., 2022; Lewis, 2019).

The 3Rs in Practice

- **Replacement:** Use non-animal alternatives like cell cultures or computer simulations wherever possible.
- **Reduction:** Minimize the number of animals used through careful planning, statistical design (e.g., power analysis), and data sharing to avoid repetition.
- **Refinement:** Improve experimental techniques to reduce animal suffering, using better housing, handling, and less invasive procedures.

Ethical Challenges: Balancing Scientific Rigor with Ethical Treatment

- **Scientific Rigor:** Reliable and reproducible results often require specific sample sizes and methods, which may involve more animals or invasive techniques.
- **Ethical Treatment:** Researchers must minimize harm and distress to animals, which can limit the scope or scale of experiments.
- **The Challenge:** Balancing scientific needs with ethical responsibilities to ensure studies are both valid and humane.

Power Analysis

- Power analysis helps determine the minimum number of animals required to detect a significant effect.
- It avoids underpowered studies (wasting resources) and overpowered studies (using too many animals).

Tools for Power Analysis:

- **pwr:** General-purpose package for power calculations (t-tests, ANOVA, etc.).
- **sspower:** Designed for sample size and power calculations in animal studies and clinical trials.
- **WSSPAS:** Interactive web application for power and sample size analysis using R.
- **BayesPPD:** Bayesian sample size determination using historical data for generalized linear models.

Statistical Models for Animal Studies

Linear Models (LM)

Introduction:

- A Linear Model (LM) is used when the dependent variable is continuous and normally distributed, and the relationship between the independent variables and the dependent variable is assumed to be linear.

- The model can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon$$

  where:

  - Y is the outcome variable (dependent variable).
  - $X_1, X_2, \ldots$ are the predictors (independent variables).
  - $\beta_0, \beta_1, \ldots$ are the <u>unknown</u> coefficients .
  - ϵ is the error term.

Application in Animal Research

Application:

- Linear models are widely used to assess the effect of various treatments (e.g., drugs, diets) on continuous outcomes.
- Common continuous outcomes include body weight, blood pressure, or enzyme activity.
- The assumption is that changes in the outcome are proportional to changes in the independent variables (e.g., dose of drug or type of diet).

Case Study: Effect of Dietary Supplement on Weight Gain in Rabbits

- **Objective:** Investigate whether a new dietary supplement increases weight gain in rabbits.
- **Model:** A linear model can be applied to analyze the effect of the supplement (independent variable) on the rabbits' weight gain (dependent variable).
- **Result:** The linear model can show that the supplement group had a significant impact on weight gain compared to the control group.

Linear Mixed Models (LMM)

- A Linear Mixed Model (LMM) is an extension of the linear model that includes both:
  - **Fixed effects**: parameters associated with the entire population.
  - **Random effects**: parameters associated with individual units, such as animals.
- This model is useful for repeated measures or hierarchical data structures, such as data collected from the same animals over time.

Model Equation:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + \epsilon_{ij}$$

where:

- Y_{ij} is the outcome for the i-th time point on the j-th animal.
- X_{ij} is the value of a covariate for the i-th time point on the j-th animal (e.g., time point, dosage, or any characteristic being measured).
- u_j is the random effect for animal j.
- ϵ_{ij} is the error term.

Application in Animal Research

- Mixed models are often used in studies where data are collected from the same animals at different time points.
- They help manage the fact that measurements from the same animal are not independent.
- These models also handle variations between different animals, making them more effective for analyzing complex data.
- This allows researchers to study both individual differences and overall trends across the population.

Case Study: Drug Effect on Heart Rate in Rodents

- **Objective:** To evaluate how a drug impacts heart rate in rodents over time.
- **Model:** A Linear Mixed Model (LMM) is used to analyze heart rate measurements taken at multiple time points for each rodent. The model includes a random effect to capture variability between individual animals.
- **Result:** The LMM shows that the drug leads to a significant reduction in heart rate over time, while accounting for differences between individual rodents.

Generalized Linear Models (GLM)

- Generalized Linear Models (GLMs) extend linear models to accommodate non-normally distributed response variables, such as binary outcomes or count data.
- GLMs consist of three components: **Random component:** the distribution of the outcome variable (e.g., binomial, Poisson). **Systematic component:** the linear predictor. **Link function:** transforms the expected value of the outcome to a linear scale.

Example - Poisson Regression:

$$\log(\mu_i) = \beta_0 + \beta_1 X_i$$

where:

- μ_i is the expected count for subject i.
- X_i is the predictor variable.

Application in Animal Research

- GLMs are used when the outcome variable is not continuous. Examples include modeling:
  - The number of occurrences (e.g., tumors, lesions).
  - Binary outcomes (e.g., survival vs. death).
- GLMs are flexible and can handle various data types commonly found in animal studies, such as count and binary data.

Case Study: Tumors in Mice

- **Objective:** Investigate the effect of a new drug on the number of cancerous tumors in mice.
- **Model:** A Poisson regression (a type of GLM) was used to model the number of tumors in mice as a function of drug treatment. The model accounts for the count nature of the outcome.
- **Result:** The Poisson regression showed a statistically significant reduction in the number of tumors in the treated group compared to the control group.

Bayesian version of the Models

- In animal research, we often deal with small sample sizes, whimaking difficult to obtain reliable estimates using traditional (frequentist) methods.
- **Bayesian Methods:** offer a powerful alternative by incorporating prior knowledge or previous data into the analysis, improving the robustness of models, especially when sample sizes are small.

Advantages of Bayesian Methods in Animal Research

- **Incorporation of Prior Knowledge**: Bayesian methods allow researchers to integrate prior data or knowledge (e.g., from previous experiments) to strengthen analysis.
- **Flexible Framework for Small Samples**: Bayesian methods are particularly helpful when dealing with small sample sizes, as they provide more reliable estimates by combining prior information with current data.
- **Probabilistic Interpretation**: Unlike traditional methods that provide point estimates, Bayesian methods yield probability distributions for parameters, offering a richer interpretation of the uncertainty in estimates.

Key R Packages for Bayesian Modeling

- **rstanarm:**
  - Fits Bayesian versions of GLMs, LMs, and LMMs using Stan.
  - Easy to use with syntax similar to base R functions like `lm()`, `glm()`, and `lmer()`.
- **brms:**
  - Provides a flexible interface to fit Bayesian models using Stan.
  - Supports a wide range of models including GLMs, LMMs, and hierarchical models.
- **BayesFactor:**
  - Specializes in Bayesian hypothesis testing and model comparison.
  - Supports models like t-tests, ANOVA, and simple regression.

Additional R Packages for Bayesian Models

- **blme:**
  - Extends the `lme4` package to fit Bayesian linear mixed models.
  - Compatible with `lmer()` and `glmer()` functions for familiar syntax.
- **INLA:**
  - Uses Integrated Nested Laplace Approximations (INLA) to fit Bayesian models efficiently.
  - Best for spatial and temporal models with large datasets.
- **MCMCglmm:**
  - Fits Bayesian generalized linear mixed models (GLMMs) using Markov Chain Monte Carlo (MCMC) methods.
  - Suitable for hierarchical models and mixed effects.

Practical Guidelines for Robust Statistical Design

Best Practices for Planning and Designing Experiments

- **Define Clear Objectives:** Outline the goals (e.g., testing treatment, comparing groups) to guide the design and models.
- **Identify Variables:** Choose relevant independent and dependent variables (treatments, measurements, outcomes).
- **Randomization and Blocking:** Randomly assign animals to groups and use blocking to control variability (e.g., by age or weight).
- **Replication:** Ensure enough replicates to detect the treatment effect and reduce random noise.

Software Tools for Design and Analysis

- **R and Python:** Widely used for statistical analysis and visualization.
  - **R:** Packages like `lme4` (linear mixed models), `pwr` (power analysis), and `blockrand` (randomization).
  - **Python:** Libraries like `statsmodels` and `scikit-learn` for statistical analysis and machine learning.
- **SPSS, STATA, and SAS:** User-friendly tools for complex statistical analysis and Design of Experiments (DoE) functionality.
- **JMP:** A graphical tool for planning, randomizing, and analyzing experiments, especially suited for complex factorial designs.

Collaboration with a Statistician During the Planning Phase

- **Choosing the Right Design:** Statisticians help select the best experimental design (e.g., factorial, crossover, split-plot) for the research question.
- **Avoiding Pitfalls:** A statistician can help prevent common issues such as improper randomization, insufficient sample size, or incorrect data handling.
- **Interpreting Results:** Provides guidance on result interpretation, ensuring conclusions are data-driven and addressing any limitations.
- **Long-Term Planning:** Helps in data collection, analysis, and reporting, ensuring adherence to ethical guidelines and the 3Rs (Replacement, Reduction, and Refinement).


Thank You!
Questions?