A Brief Introduction to Experimental Design

BAE 815
Dr. Zifei Liu
• Originate with a question or problem
• Require a clear articulation of an objective
• Follow a specific plan or procedure (a method)
• Require collection and interpretation of data

• Empirical research consists of:
  – Experimentation
  – Interpretation of results
  – Presentation of results

Philosophy of empirical research
Without proper experimental design/method

- Restricted to mainly exploratory experimentation
- Can gain intuition, but no real answers
- Can be used to help generate hypotheses
- Difficult to justify results to others
- May never appear in published materials

With proper experimental design/method

- Allow full range of types of experimentation
- Can be used to determine conditional answers
- Convince an audience of probable fact
- Facilitates justification of results

Why we need experimental design?
• Validity (Avoid systematic error/bias)
  – Comparison/control
  – Randomization
  – Blocking
  – Blinding

• Reliability (Reduce random error/bias)
  – Replication
  – Balance

• Maximum information from minimum effort/cost

Experimental design is a careful balancing of several features:
• statistical power
• generalizability
• various forms of “validity”
• reliability
• practicality
• cost.

Goals of experimental design
An experiment has treatments, experimental units, and a method to assign treatments to units.

The selection of experimental units and assignment of treatments is called “experiment design.”

- Types of variables
- Comparison/control
- Blocking, randomization, and blinding
- Replication
- Factorial experiments

**Basic concepts in experimental design**
• Explores relationship between variables
  – Independent variable is the variable (factor) that is purposely changed. It is the manipulated variable.
  – Dependent variable changes in response to the independent variable. It is the responding variable.
  – Confounding/nuisance variables are the variables that mask or distort the association between measured variables in a study. They are undesired sources of variation that can affect the dependent variable.

• Constant variables
  – Factors that are kept the same and not allowed to change

• Uncontrollable variables
  – Variables that are known to exist, but conditions prevent them from being manipulated, or it is very difficult to measure them.

Identify the types of variables
• What is a control?
  – The part of the experiment that serves as the standard of comparison.

• Why is a control necessary?
  – It is the unchanged part of the experiment that detects the effects of hidden variables.

• Internal controls
  – It can be useful to use the subjects themselves as their own controls (e.g., consider the response after vs. before treatment).

The control in an experiment
• Include the confounding variable as one of the factors in the experiment.
• If you can, fix the confounding variable (make it a constant).
• If you can’t fix the confounding variable, use blocking.
• If you can neither fix nor block the confounding variable, use randomization.

Minimize problems from extraneous sources of variability
Experimental units

Treatments (Red vs green)

Without blocking or randomization, the confounding variable differs among treatments
A block of units is a set of units that are homogeneous in some sense.

Blocking is an experimental procedure for isolating variation attributable to a confounding/nuisance variable.
Randomization is the random assignment of treatments to units.

With randomization, the confounding variable does not differ among treatments, such that we can assume the differences seen may be largely due to the treatment.
• Randomization protects against confounding.
  – Breaks the association between potential confounding variables and the dependent variables

• Randomization can form the basis for inference
  – Randomization allows the later use of probability theory, and so gives a solid foundation for statistical analysis.

Why randomize?
• Are the subjects you are studying really representative of the population you want to study?
• Ideally, your study material is a random sample from the population of interest.

Representativeness and generalizability
Randomized Block Design
A Latin Square Design

- A Latin Square is a design that blocks for two sources of variation.
- There are two blocking factors in a Latin Square, and these are represented by the rows and columns of the square.
- Each treatment occurs once in each row and once in each column.

<table>
<thead>
<tr>
<th>Test session</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  2  3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject 1</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 2</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>Subject 3</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>
• Analysis of covariance is another general approach to control confounding/nuisance variables.
• It involves measuring one or more confounding/nuisance variables.
• Through analysis of covariance, the dependent variable can be statistically adjusted to remove the effects of the uncontrolled source of variation.

Analysis of covariance
• Blinding is the concealment of information from the participants and/or researchers about which subjects are receiving which treatments.

• Measurements made by people can be influenced by unconscious biases.

• Two treatments: drug and placebo
  – Single blind: the patients don’t know which group they are in, but the doctors do.
  – Double blind: neither the patients nor the doctors administering the drug know which group the patients are in.

**Why blinding?**
• Reduce the effect of uncontrolled variation (i.e., increase precision). Larger $n \Rightarrow$ Smaller SE

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

• Quantify uncertainty.
  – An estimate is of no value without some statement of the uncertainty in the estimate

**Why replicate?**
• In a balanced experimental design, all treatments have equal sample size
• This maximizes power and makes tests more robust to violating assumptions

Why balance?
Types of design

• Single factor (one-way):
  – Studies *one independent variable*
  – Pretest-posttest (one-group)
  – Pretest-posttest (control group)
  – Posttest-only (control group)

• Multi-factor
  – Studies *multiple independent variables*
  – Each with a number of levels (L)
  – Two-way (e.g., 2 x 2)
  – Three-way (e.g., 2 x 2 x 2)

• Time-series
  – Dependent measure is continuous
  – Establish baseline
  – Measure treatment effect over time
• Factorial design
  • Investigates all treatment combinations of two or more variables
    – Permits study of interactions between treatment variables
• Analysis
  – ANOVA

Example: 2 x 3

Two-way factorial design
Interactions
Characteristics of good experiments

• Unbiased
  – Randomization
  – Blinding

• High precision
  – Uniform material
  – Replication
  – Stratification/blocking

• Able to estimate uncertainty
  – Replication
  – Randomization
• A description of what you used for your experiment, and how you did it.
• Usually should include:
  – Levels of the independent variable
  – Repeated trials
  – Statistics methods

Materials and methods section of your paper
• Data exploration and analysis
• Inductive inference with probability
• Quantification of uncertainty

Why statistics?
• Two possible errors (Type I and Type II)
• Confidence intervals
• Significance level, P-value
• Statistical power
• Determining sample size

Basic concepts in statistics
Use the observed data to answer a yes/no question, “Does X have an effect on Y?”

- Null hypothesis ($H_0$): X has no effect on Y.
- Alt. hypothesis ($H_a$): X does have an effect on Y.

Significance tests
<table>
<thead>
<tr>
<th>True situation</th>
<th>Our conclusion</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No effect</td>
<td>Not significant</td>
<td>Correct</td>
</tr>
<tr>
<td></td>
<td>Significant (Reject H₀)</td>
<td>Type I error</td>
</tr>
<tr>
<td>Has an effect</td>
<td>Significant (Reject H₀)</td>
<td>Correct</td>
</tr>
<tr>
<td></td>
<td>Not significant</td>
<td>Type II error</td>
</tr>
</tbody>
</table>

**Type I and Type II Errors**
• A false positive result is known as a **Type I Error**.

• We control for Type I errors explicitly by selecting an appropriate confidence level.

• The confidence is the probability of not getting a false positive result.

• Confidence level is the probability of accepting the null hypothesis when the null hypothesis is true.

**Confidence level**
• Plausible values for the true population average or treatment effect, given the observed data.

95% CI for population average (93.4 to 113.8)

Confidence interval (CI)
• A P-value is the probability of obtaining data as extreme as was observed, if the null hypothesis were true (i.e., if the treatment has no effect).

• If your P-value is smaller than your chosen significance level $\alpha$, you reject the null hypothesis.

• P-value
  – Summarizes the result of the significance test.
  – Small P-value conclude that there is an effect.

**P-values**
• A false negative result is known as a **Type II Error**.

• We control the power implicitly via the confidence level and the experimental design.

• The power is the probability of not getting a false negative result.

• Statistical power is the probability of rejecting the null hypothesis when the null hypothesis is false.

---

**Statistical power**
Statistical power

Dist’n of D when $\Delta = 0$

Dist’n of D when $\Delta = 8.5$

Power = 70%
• The structure of the experiment
• The method for analyzing the data
• The size of the true underlying effect
• The variability in the measurements
• The chosen significance level $\alpha$
• The sample size

Note: We usually try to determine the sample size to give a particular power (e.g. 80%).

**Power depends on**
• Method for analysis
• Chosen significance level, $\alpha$ (e.g. 5%)
  – More stringent of statistical test (Smaller $\alpha$) $\rightarrow$ larger sample size
• Desired statistical power (e.g. 80%)
  – More statistical power $\rightarrow$ larger sample size
• Variability in the measurements
  – Larger variability $\rightarrow$ larger sample size
• The smallest meaningful effect
  – Smaller treatment effect $\rightarrow$ larger sample size

Determining sample size
• Testing the differences between two or more groups on one dependent variable.
  – One way or factorial
• Variation due to treatment and random variation (error)
• Typical breakdown in a between groups design
  \[ SS_{tot} = SS_{b/t} + SS_e \]
• The F statistic is a ratio of these variances
  \[ F = MS_{b}/MS_e \]

Analysis of variances (ANOVA)
• Experiments should be designed.
• Good design and good analysis can lead to reduced sample sizes.
• Consult an expert on both the analysis and the design of your experiment.
• If the results do not make sense, it may suggest a problem in methodology or experimental design

Final words
# Sample Data Table

**Title:** The Effect of the independent variable on the dependent variable

<table>
<thead>
<tr>
<th>Column for independent variable</th>
<th>Column for dependent variable</th>
<th>Column for derived quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Label – with units if necessary</strong></td>
<td>Label – with units if necessary – multiple trials included</td>
<td>Label – with units if necessary. Example = average of trials</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>
Example

- 20 male mice and 20 female mice.
- Half to be treated; the other half left untreated.
- Can only work with 4 mice per day.

Question: How to assign individuals to treatment groups and to days?
An extremely bad design

<table>
<thead>
<tr>
<th>Week One</th>
<th>Week Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Tu</td>
<td>Tu</td>
</tr>
<tr>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>Th</td>
<td>Th</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

C = control, T = treated, pink = female, blue = male
Randomized

Week One

M | Tu | W | Th | F
---|----|---|----|---
T | T  | T | T  | T
C | T  | T | T  | T
C | C  | C | T  | T
T | C  | C | C  | C

Week Two

M | Tu | W | Th | F
---|----|---|----|---
C | T  | T | T  | T
C | C  | C | T  | T
C | T  | C | T  | T
T | T  | C | T  | T

T = treated, C = control, pink = female, blue = male
A stratified/block design

Week One

M  Tu  W  Th  F
C  T  T  C  T
T  T  C  C  C
C  C  T  T  C
T  C  C  T  T

Week Two

M  Tu  W  Th  F
C  C  T  C  T
T  T  T  C  C
C  T  C  T  C
T  C  C  T  T

T = treated, C = control, pink = female, blue = male