7. Varieties of Single-Factor Experimental Designs
Goals

- Be able to identify 4 different types of single-factor experimental / quasi-experimental designs
- Understand the advantages of multi-level designs compared to two-level designs.
- Understand the different kinds of control groups
- Understand when to use graphs, tables and text to describe statistical results.
- Understand which procedures to use in analyzing multi-level designs
7.1 The Simplest Experiment

Single Factor Designs with 2 Levels
Vocabulary

- Methodology: Complete and detailed description of the way the experiment is carried out (subjects tested, instructions, stimuli, etc.)

- Design: Short-hand method of designating different general experimental set-ups. Describes:
  - IVs and their levels
  - Whether each IV is within or between
  - “We used a mixed 2 x 2 factorial design, with gender (male or female) as a between-subjects factor and navigation cues (landmarks vs. none) as a within-subjects factor.”
**Single Factor Designs**

The most basic experiment has a single IV (factor). There are four basic variants on this design

<table>
<thead>
<tr>
<th>Type of Design</th>
<th>Between Ss or Within?</th>
<th>Experimental Design?</th>
<th>Method for assigning Ss to conditions</th>
<th>IV Manipulated or Subject?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent groups</td>
<td>Between</td>
<td>Yes</td>
<td>Random Assign.</td>
<td>Manip</td>
</tr>
<tr>
<td>Matched groups</td>
<td>Between</td>
<td>Yes</td>
<td>Matching</td>
<td>Manip</td>
</tr>
<tr>
<td>Non-equivalent groups</td>
<td>Between</td>
<td>No (quasi)</td>
<td>n/a</td>
<td>Subject</td>
</tr>
<tr>
<td>Repeated measures</td>
<td>Within</td>
<td>Yes</td>
<td>n/a</td>
<td>Manip</td>
</tr>
</tbody>
</table>
2-level designs

• The simplest form of experiment has 1 IV (or factor), which has only 2 levels.

• We will now look at 4 examples of 2 level designs, illustrating the different types of single factor designs.
Example Experiment 1: Independent Groups

• Question: Last-minute cramming (LC) vs. distributed studying (DS), which is best?

• Conventional wisdom suggests that LC is bad for your grades, but is it?

• First step is to look at previous research to see what it says. Let’s say you find just two studies...
# Previous Study #1

<table>
<thead>
<tr>
<th></th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thr</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC</td>
<td></td>
<td></td>
<td></td>
<td>3 hrs.</td>
<td>Test</td>
</tr>
<tr>
<td>DS</td>
<td>3 hrs.</td>
<td>3 hrs.</td>
<td>3 hrs.</td>
<td></td>
<td>Test</td>
</tr>
</tbody>
</table>
## Previous Study #2

<table>
<thead>
<tr>
<th></th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thr</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC</td>
<td></td>
<td></td>
<td>3 hrs.</td>
<td></td>
<td>Test</td>
</tr>
<tr>
<td>DS</td>
<td>1 hrs.</td>
<td>1 hrs.</td>
<td>1 hrs.</td>
<td></td>
<td>Test</td>
</tr>
</tbody>
</table>
We will use a method that has more internal validity than Previous Study #1, and more external validity than Previous Study #2.

Method: 60 participants, divided into two groups of 30 (LC and DS) by random assignment. Each group studies a text.

- LC group studies text day before test for 3 hours
- DS group studies text for 3 days, 1 hour each day.
- Both groups then tested on knowledge of text.
**Example Experiment 1: Independent Groups**

<table>
<thead>
<tr>
<th></th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thr</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LC (N=30)</strong></td>
<td></td>
<td></td>
<td></td>
<td>3 hrs.</td>
<td>Test</td>
</tr>
<tr>
<td><strong>DS (N=30)</strong></td>
<td>1 hrs.</td>
<td>1 hrs.</td>
<td>1 hrs.</td>
<td>Test</td>
<td></td>
</tr>
</tbody>
</table>
Example Experiment 1: Independent Groups

- Expected results: Conventional wisdom suggests that the mean score of the LC group should be lower than that of the DS group.

- Implications: Assuming we see the expected results, LC should be discouraged. If we do not, it may indicate that the traditional studying advice is incorrect.

- [Question your question: Are grades really what we want to know about?]
Example Experiment 2: Matched Groups

- Same as ExEx1, but let’s say we can only get 10 participants, for some reason.
- Random assignment will not work, so create two groups based on matching via GPA.
- Methodology is otherwise the same: One group crams, other does distributed study, then both are tested.
### Step 1: Order Values

<table>
<thead>
<tr>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
</tr>
<tr>
<td>9.0</td>
</tr>
<tr>
<td>8.5</td>
</tr>
<tr>
<td>8.0</td>
</tr>
<tr>
<td>7.3</td>
</tr>
<tr>
<td>7.1</td>
</tr>
<tr>
<td>7.0</td>
</tr>
<tr>
<td>6.8</td>
</tr>
<tr>
<td>6.5</td>
</tr>
<tr>
<td>5.5</td>
</tr>
</tbody>
</table>

### Step 2: Create pairs of adjacent values

<table>
<thead>
<tr>
<th>Adjacent Values</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>8.5</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>7.3</td>
<td>7.1</td>
<td></td>
</tr>
<tr>
<td>7.0</td>
<td></td>
<td>6.8</td>
</tr>
<tr>
<td>6.5</td>
<td></td>
<td>5.5</td>
</tr>
</tbody>
</table>

### Step 3: From each pair, randomly assign one to each group

<table>
<thead>
<tr>
<th>Values</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.0</td>
<td></td>
<td>9.1</td>
</tr>
<tr>
<td>8.5</td>
<td></td>
<td>8.0</td>
</tr>
<tr>
<td>7.3</td>
<td></td>
<td>7.1</td>
</tr>
<tr>
<td>7.0</td>
<td></td>
<td>6.8</td>
</tr>
<tr>
<td>6.8</td>
<td></td>
<td>7.0</td>
</tr>
<tr>
<td>5.5</td>
<td></td>
<td>6.5</td>
</tr>
</tbody>
</table>

**Average**

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu = 7.42$</td>
<td>$\mu = 7.54$</td>
</tr>
</tbody>
</table>
Which kind of 2-level design is this?

In a study on the effects of motivation on maze-running time, a researcher puts two groups of rats through a training exercise whereby they learn that there is food at the end of an alley-maze. She then has one group deprived of food for 48 hours before running the two groups through the same maze again. She measures the time it takes for the rats to reach the end of the maze. Rats are assigned to groups (motivated or not) such that mean body weight is the same for the two groups.
Nonequivalent Groups

• Quasi-experimental design that uses subject variables (male/female, depressed/non-depressed, etc.)
• a.k.a. “Ex post facto” (after the fact), or “natural groups” design
• Cannot establish causality because the IV is not manipulated
Example Experiment 3: Nonequivalent Groups

• Same as ExEx1 & ExEx 2, but instead of assigning a study style to people at random, we look at self-described “crammers” vs. “distributers”

• Method would change slightly: Give Ss text and tell them to use their own preferred strategy

• No random assignment, so can’t be sure what is causing any differences we see

• Also, the IV is less well-defined and controlled here. How individuals define LC and DS might differ

• Test after a week of using own strategy.
Example Experiment 3: Nonequivalent Groups

• Expected results: Again, LC should do worse than DS, according to traditional advice.

• Implications: Expected results would again support distributed study. But, if expected results not obtained, it may indicate that people choose a style because it works for them. Perhaps there are different cognitive styles that benefit from one study strategy or the other.

• Future directions: Do a two-factor study with one IV being preferred style and another IV being assigned style.
Example Experiment 4: Repeated Measures

- Methods: Ps study two different texts (equated for difficulty), one after the other.
- One text is studied by LC, the other by DS.
- Order of conditions must be counterbalanced:
  - Half of Ps do LC 1st (text A), DS 2nd (text B)
  - Half of Ps do DS 1st (text A), LC 2nd (text B)
- This helps control for practice or fatigue effects.
### Example Experiment 4: Repeated Measures

<table>
<thead>
<tr>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thr</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Week 1 (text A)**

Half of participants do this

### Week 2 (text B)

<table>
<thead>
<tr>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thr</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other half do this
## Analysis of 2-Level Data

<table>
<thead>
<tr>
<th>Relation Between Groups</th>
<th>Data Type</th>
<th>Normal Numerical</th>
<th>Non-normal Numerical</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>t-test, independent groups</td>
<td>Mann-Whitney U test</td>
<td></td>
<td>Chi-square test, two samples</td>
</tr>
<tr>
<td>Dependent</td>
<td>t-test, dependent groups</td>
<td>Wilcoxon signed-rank test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Independent* = Randomly-assigned or non-equivalent groups  
*Dependent* = Matched or repeated-measures groups
Which kind of 2-level design is this?

A researcher is interested in the effects of political orientation (conservative vs. liberal) on tolerance for ambiguity. She divides participants into conservative and liberal groups based on responses to a questionnaire measuring political orientation, then has each group go through an exercise involving difficult ambiguous situations. She then has them fill out a questionnaire measuring their frustration levels.
7.2 Taking it to the Next Level

Multilevel (3+ level) designs
Multilevel Designs

- Design having one IV with 3+ levels: *Single factor multilevel design*

- Main advantage: Can detect *non-linear effects* when the IV is continuous

- Also, can be more efficient

  **Example:** testing many different therapies at once vs. a single control group is more efficient than testing each therapy on its own vs. its own control group.
Non-Linear Effect: PTSD Reduction Over Time

- Source: Australian Centre for Post-traumatic Mental Health (www.acpmh.unimelb.edu.au)

- Measured clients on intake to the centre and after 3, 9 and 21 months of therapy.

- One factor (length of therapy), 4 levels (0, 3, 9, 21 months).
PTSD Levels

PTSD Checklist Scores over Months After Intake
Non-linear Effects

![Graph showing the relationship between Mean Grade, Hours of Charles's Cups of Coffee, Mean Lecture Clarity Rating, and Performance, as well as Task Difficulty.](image-url)
Example
Multilevel Experiment #1

- Question: Therapy X has been proven effective for depression, but no one is sure how many hours per week are needed.
- As usual, first step is to examine the literature. Let’s say you find just 3 studies, each reporting a 2-level exp:
  - One says 1 hour/week is effective relative to a control
  - One says 2 hours/week is effective relative to a control
  - One says 3 hours/week is effective relative to a control
- How do these different therapy intensities compare to one another? Is 1 hour/week enough? Is 3 hours/week overkill?
Example Multilevel Experiment #1

- Will use method allowing comparison of all three intensities to one another, as well as a control group.
- Methodology: Let’s say you are part of a large-scale research project wherein Ss have agreed to be assigned at random to one of the three intensities or to a control group (therapy to be delivered later, thus this is a waiting list control group).
- This gives a rare opportunity to explore therapy effectiveness using a true experiment design.
- OpDef: Measure change in depression pre- vs. post-therapy using two administrations of the BDI.
Example
Multilevel Experiment #1

- Expected results: Therapy groups expected to do better than control. But previous research is not clear as to what the pattern of effectiveness will be across the different intensities.

- Thus this work has both hypothesis testing and exploratory aspects, as is often the case.

- Implications: Complex. For example, if the 3 hr/week intensity is no better than the lower intensities, this suggests that (for most patients, at least), this is excessive. If the 1 hr/week intensity is worse than the higher ones, it suggests this is not enough.
A Non-linear Effect

Therapy Intensity (hrs./week)

- Control
- CBT
- ECT
- Prozac

Mean Change in BDI
Example Multilevel Experiment #2

• Question: Many different kinds of therapy are attempted with depression. Do any of them work, and which, if any, is the best?

• As usual, first step is to examine the literature. Let’s say you find just 3 studies, each reporting a 2-level exp:
  • One says that CBT is effective relative to a control
  • One says that EFT is effective relative to a control
  • One says that Prozac is effective relative to control

• No one seems to have compared these therapies to one another! Plus, this piecemeal approach is inefficient!
Example Multilevel Experiment #2

- Will use method allowing comparison of all three therapies to one another, as well as a control group.
- Methodology: Let’s say you work at a private practice where there are three psychologists: One does CBT, one EFT, the last just prescribes Prozac for depression.
- Can’t assign people to a therapist at random, that’s their choice, so there is no random assignment or matching. Therefore this is a non-equivalent groups design.
- OpDef: Measure change in depression pre- vs. post-therapy using two administrations of the BDI.
Example Multilevel Experiment #2

• Who will act as the controls? Often, there is a backlog of patients waiting for therapy. They can be recruited as control participants. This is a waiting list control.

• So, participants are drawn from four groups (note, NOT assigned to groups)
  • Waiting list control
  • Cognitive-behavioural therapy (CBT)
  • Emotionally-focussed therapy (EFT)
  • Prozac
Example Multilevel Experiment #2

- Expected results: Based on previous work, therapy groups expected to do better than control. But previous research is not clear as to which therapy will be best, if any.

- Implications: If one therapy is superior to the others, this recommends it for use. If therapies have similar effectiveness, then all can be recommended.
NOT a Non-linear Effect

Mean Change in BDI

<table>
<thead>
<tr>
<th>Therapy Type</th>
<th>Control</th>
<th>CBT</th>
<th>EFT</th>
<th>Prozac</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>10</td>
<td>15</td>
<td>12</td>
</tr>
</tbody>
</table>
Discussion / Questions

- What would make our example multilevel design an Independent Groups Multilevel design?
- What would make it a Matched groups design?
- What would make it a repeated measures design?
- What do you think of the validity and practicality of these different options?
7.3 Presenting & Analyzing Single-Factor Data
Presenting Multilevel Data

**GOLD**

**BAD**

**Therapy Type**

Mean Change in BDI
Presenting Multilevel Data

Okay Better
Presenting Multilevel Data

### Table 1. Mean residual BDI scores following 3 months of therapy. All therapy groups show significant improvement compared to controls

<table>
<thead>
<tr>
<th>Therapy Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>6.2</td>
<td>0.15</td>
</tr>
<tr>
<td>CBT</td>
<td>8.8</td>
<td>0.12</td>
</tr>
<tr>
<td>EFT</td>
<td>9.0</td>
<td>0.15</td>
</tr>
<tr>
<td>Prozac</td>
<td>8.5</td>
<td>0.11</td>
</tr>
</tbody>
</table>
“Mean residual BDI scores (pre-test minus post-test) were lowest for the control group, at 6.2 ± .12. All the therapy groups showed significantly more improvement. For the CBT group, mean improvement was 8.8 ± .15. For the EFT group, mean improvement was 9.0 ± .15. For the Prozac group, mean improvement was 8.5 ± .11.”
Analysis of Multi-Level Data

• Can’t use multiple 2-level tests due to *alpha inflation*

• Each test you run has a chance of Type I error (i.e., saying there is a difference when there isn’t)

• Multiple tests means an overall chance of at least one Type I error of $1 - (1-\alpha)^n$, where $n$ is the number of tests

  **Example:** For a study with 4 groups (ABCD), you would run 6 tests (A-B, A-C, A-D, B-C, B-D, C-D). Overall chance of Type I error is $1 - (1-.05)^6 = 26.5\%$

• Could reduce base alpha to compensate (e.g., to .01), but this also reduces statistical *power* (i.e., increases chances of missing real fx)
Analysis of Multi-Level Data

• Instead use One-way ANOVA (Analysis of Variance) plus post-hoc tests.

• ANOVA tests the null hypothesis that $\mu_1 = \mu_2 = \mu_3$ ...

• Typically, posthoc tests (Fisher’s PLSD, Tukey’s HSD, etc) used to figure out which pairs differ from one another. Or planned comparisons can be used instead.

• ANOVA relies on several assumptions about the data, which should be tested before ANOVA is carried out
Analysis of Multi-Level Data

- MANOVA (multivariate ANOVA): For multiple DVs
- ANCOVA (analysis of covariance): Factoring out certain variables (e.g., factoring age out of driving ability)
- MANCOVA: Combination of above
- Krukal-Wallis: Equivalent of a one-way ANOVA for non-parametric data (i.e., doesn’t meet assumptions of ANOVA)
Discussion / Questions

• What would be the appropriate analysis for Example Multilevel Experiment #1? What about #2?
THE END