Multiple Group Experiments

Experimental Psychology

Arlo Clark-Foos
Two Group Experiments

• Design: Between-Subjects vs. Within-Subjects

• $\alpha$ (Type I) vs. $\beta$ (Type II)

• Logic of Significance Testing: Rare occurrences

• Directional Testing...
Multiple Groups

• Two Groups
  – Experimental + Control
  – Treatment + Placebo

• Multiple Groups:
  – 2 x Experimental + Control
  – 3 Experimental
Which design works best?

How many IVs will my experiment have?
- One (Chs. 10, 11)
- Two or more (Ch. 12)

How many groups (levels) will my IV have?
- Two (Ch. 10)
- Three or more (Ch. 11)

What type of participant groups do I have?
- Independent (between subjects)
  - Two independent groups design (Ch. 10)
  - t test for independent samples
- Correlated (within subjects)
  - Two correlated groups design (Ch. 10)
  - t test for correlated samples
- Multiple independent groups design (Ch. 11)
  - One-way ANOVA for independent groups
- Correlated (within subjects)
  - Multiple correlated groups design (Ch. 11)
  - One-way ANOVA for correlated groups
Multiple Groups Example

- **IV**: Type of Role Model
- **DV**: Likelihood of purchasing product

None | Peer | Celebrity | Parent
Participant Assignment

• **Independent Samples** *(Between-Subjects)*
  – Completely Random Assignment
  – Controls for unknowns
Participant Assignment

- Correlated Samples
  - Nonrandom Assignment
    - Matched Sets
    - Repeated Measures
    - Natural Sets
  - Goal of all 3 is to reduce error!
Analyzing Data from Multiple Groups

• Two: \( t \) test

• Multiple: **Analysis Of VAriance (ANOVA)**
  – Also used for designs with multiple IVs
  – For now, One IV \( \rightarrow \) One Way ANOVA

  – We still have 2 types of One Way ANOVA
    • Independent Samples (Between-Subjects)
    • Repeated Measures (Within-Subjects)
A Familiar Example

• Old Two Group Example: Room Lighting
  – Brightly Lit vs. Dimly Lit
    • Sig. difference between groups
    • Are there differences between those levels?
Lighting Conditions Example

• Operational Definition(s)
  – What constitutes Bright, Moderate, & Dim?

• Experimental Design
  – Repeated Measures?
  – Independent Samples?

• How do we analyze the data?
Analysis of Variance

- **Rationale of ANOVA**

- **Between Groups Variability**
  - Variability in DV that is due to IV

- **Within Groups Variability**
  - Variability in DV that is due to something other than IV (individual differences, measurement error, extraneous variables)
Analysis of Variance

\[ F = \frac{\text{variability due to IV} + \text{error variability}}{\text{error variability}} \]

• When the IV has a significant effect on the DV, the F ratio will be large.
• When the IV has no effect or only a small effect, the F ratio will be small (near 1).

\[ F = \frac{\text{between-groups variability}}{\text{within-groups variability}} \]
Interpreting Output

- **Source Table**
  - Specifies the source of variability
- **Sum of Squares**
  - A sum of squared deviations
- **Mean Square**
  - Divide SS by df to put on equal footing

\[
F = \frac{\text{mean square between groups}}{\text{mean square within groups}}
\]
Determining Significance

• If computing $F$ ratio by hand...
  – Just as with $t$ test, check a table
    • Use both types of $df$ to look up critical value

• If computing $F$ ratio by computer...
  – Look for one of the following words
    • PROB, SIGNIFICANCE, P-VALUE
## Interpreting Differences

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bright Light</td>
<td>86.6</td>
<td>6.43</td>
</tr>
<tr>
<td>Moderate</td>
<td>81.9</td>
<td>6.97</td>
</tr>
<tr>
<td>Dim Light</td>
<td>58.4</td>
<td>5.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>DF</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>2</td>
<td>570.85</td>
</tr>
<tr>
<td>Within Groups</td>
<td>21</td>
<td>121.25</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

\[ F \text{ ratio} = 4.71 \quad \text{PROB} = .02 \]
Interpreting Differences

• What do learn from a significant difference?
  – A difference between groups
  – Where is the difference?
  – We need to test for differences between each of our groups...after we compute the $F$ statistic

• post hoc tests (after the fact tests)
Post Hoc Tests

• Many types of Post Hoc tests
  – Tukey’s HSD, Bonferroni, Duncan, Dunn, Fisher’s LSD, Dunnet, Scheffe, Student-Neuman-Keuls, etc.
Brief Digression on Probability

- Let’s say we put the numbers 1-6 on slips of paper and put them all in a hat...

- Probability of 1 AND 3: $P(A \cap B) = P(A) \times P(B)$
- Probability of 1 OR 3: $P(A \cup B) = P(A) + P(B)$
Brief Digression on Probability

- Planned Comparisons (*a priori*)
- Inflation of $\alpha$
- Impact on Post Hoc testing
  - More stringent $\alpha$ in many post hoc tests
Translating Stats Into Words

• “Lighting conditions affected learning”
  – Not enough information

“The effect of different lighting on students’ learning was significant $F(2, 21) = 4.71, p = .02$. Tukey tests indicated ($p < .05$) that learning in brightly lit rooms ($M = 86.6, SD = 6.43$) was better than learning in dimly lit rooms ($M = 58.4, SD = 5.74$). Learning in brightly lit and moderately lit rooms did not differ from each other.”

Lay person should be able to understand it.
Repeated Measures ANOVA

• **Example:** What type of words do people remember the best?

• **IV:** Type of Word (Noun, Adjective, Adverb)

• **DV:** Recognition or Recall

• **Groups:** Same participants studying all three word types
A Note on Computer Output

What does it mean when $p = .000$?

Let’s look at the normal distribution.

Tails are asymptotic = They never touch the baseline

$p$ is NEVER equal to 0 (probability of chance is NEVER 0)

Just say that $p < .001$
Translating Stats Into Words

• “People remember some word types better than others”
  – Still need post hoc tests

“The effect of word type on participants’ later recall was significant, $F(2, 14) = 19.71, p < .001$. Tukey tests showed ($p < .01$) that people remembered nouns better ($M = 63.25, SD = 11.73$) than either adjectives ($M = 48.38, SD = 9.46$) or adverbs ($M = 48.88, SD = 9.55$). Recall did not differ between adjectives and adverbs.”
What comes next?

• Now we know how to design and analyze data from multiple groups
  – Where can we go from here?
  – Multiple IVs