ANOVA

Analysis of Variance

Chapter 16

• A procedure for comparing more than two groups
  – independent variable: smoking status
    • non-smoking
    • one pack a day
    • > two packs a day
  – dependent variable: number of coughs per day

• $k =$ number of conditions (in this case, 3)

One-Way ANOVA

• One-Way ANOVA has one independent variable (1 factor) with > 2 conditions
  – conditions = levels = treatments
  – e.g., for a brand of cola factor, the levels are:
    • Coke, Pepsi, RC Cola

• Independent variables = factors

Two-Way ANOVA

• Two-Way ANOVA has 2 independent variables (factors)
  – each can have multiple conditions

Example
• Two Independent Variables (IV’s)
  – IV1: Brand; and IV2: Calories
  – Three levels of Brand:
    • Coke, Pepsi, RC Cola
  – Two levels of Calories:
    • Regular, Diet

When to use ANOVA

• One-way ANOVA: you have more than two levels (conditions) of a single IV
  – EXAMPLE: studying effectiveness of three types of pain reliever
    • aspirin vs. tylenol vs. ibuprofen
• Two-way ANOVA: you have more than one IV (factor)
  – EXAMPLE: studying pain relief based on pain reliever and type of pain
    • Factor A: Pain reliever (aspirin vs. tylenol)
    • Factor B: type of pain (headache vs. back pain)

ANOVA

• When a factor uses independent samples in all conditions, it is called a between-subjects factor
  – between-subjects ANOVA
• When a factor uses related samples in all conditions, it is called a within-subjects factor
  – within-subjects ANOVA
  – PASW: referred to as repeated measures
ANOVA & PASW

<table>
<thead>
<tr>
<th></th>
<th>2 samples</th>
<th>2 or more samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>t-test</td>
<td>Between Subjects</td>
</tr>
<tr>
<td>Samples t-test</td>
<td></td>
<td>ANOVA</td>
</tr>
<tr>
<td>Related</td>
<td>t-test</td>
<td>Repeated Measures</td>
</tr>
<tr>
<td>Samples</td>
<td></td>
<td>ANOVA</td>
</tr>
</tbody>
</table>

Why bother with ANOVA?

Would require six t-tests, each with an associated Type I (false alarm) error rate.

Familywise error rate

- Overall probability of making a Type I (false alarm) error somewhere in an experiment
  - One t-test, familywise error rate is equal to $\alpha$ (e.g., .05)
  - Multiple t-tests result in a familywise error rate much larger than the $\alpha$ we selected
- ANOVA keeps the familywise error rate equal to $\alpha$

Post-hoc Tests

- If the ANOVA is significant at least one significant difference between conditions
  - In that case, we follow the ANOVA with post-hoc tests that compare two conditions at a time
    - post-hoc comparisons identify the specific significant differences between each pair

ANOVA Assumptions

- Homogeneity of variance
  - $\sigma^2_1 = \sigma^2_2 = \sigma^2_3 = \sigma^2_4 = \sigma^2_5$
- Normality scores in each population are normally distributed

Partitioning Variance
Partitioning Variance

Total Variance in Scores

Variance Within Groups (error)  Variance Between Groups

(chance variance)  (systematic variance: treatment effect)

- $MS_{\text{error}}$ is an estimate of the variability as measured by differences within the conditions
  - sometimes called the within group variance or the error term
  - chance variance (random error + individual differences)

- $MS_{\text{group}}$ is an estimate of the differences in scores that occurs between the levels in a factor
  - also called $MS_{\text{between}}$
  - Treatment effect (systematic variance)

### Data Table

<table>
<thead>
<tr>
<th>Tylenol</th>
<th>Aspirin</th>
<th>Ibuprofen</th>
<th>Gin</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.6</td>
<td>2.1</td>
<td>1.3</td>
</tr>
<tr>
<td>2.3</td>
<td>2.7</td>
<td>1.6</td>
<td>3.3</td>
</tr>
<tr>
<td>1.7</td>
<td>2.4</td>
<td>2.4</td>
<td>1.9</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1.9</td>
<td>2.7</td>
</tr>
<tr>
<td>2</td>
<td>3.3</td>
<td>1</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Mean: 2.1  Overall $\overline{X} \approx 2.1$

### Calculations

- $F_{\text{obs}} = \frac{X_1 - X_2}{S_{X_1 - X_2}}$
  - difference between sample means
  - difference expected by chance (standard error)

- $F_{\text{obs}} = \frac{\text{variance between sample means}}{\text{variance expected by chance (error)}}$
  - systematic variance
  - chance variance

- $MS_{\text{error}} = \text{error variance (within groups)}$
  - $MS_{\text{error}} = \text{mean squared deviation}$
  - $MS_{\text{error}}$ is similar to variance ($s^2$)

- $MS_{\text{group}} = \text{variance between groups}$
  - $MS_{\text{group}}$ is an estimate of the differences in scores that occurs between the levels in a factor
  - also called $MS_{\text{between}}$
  - Treatment effect (systematic variance)

### Total Variance

Between Groups Variance
1. Treatment effect (systematic)
2. Chance (random error + individual differences)

Error Variance (within groups)
1. Chance (random error + individual differences)
When $H_0$ is TRUE (there is no treatment effect):

\[ F = \frac{0 + \text{Chance}}{\text{Chance}} \cong 1 \]

When $H_0$ is FALSE (there is a treatment effect):

\[ F = \frac{\text{Treatment effect} + \text{Chance}}{\text{Chance}} > 1 \]

**Signal-to Noise Ratio**

- ANOVA is about looking at the *signal* relative to *noise*
  - $MS_{\text{group}}$ is the *signal*
  - $MS_{\text{error}}$ is the *noise*
  - We want to see if the between-group variance (signal), is comparable to the within-group variance (noise)

**Logic Behind ANOVA**

- If there is no true difference between groups at the population level:
  - the only differences we get between groups in the sample should be due to error.
  - if that’s the case, differences between groups should be about the same as differences among individual scores within groups (error).
  - $MS_{\text{group}}$ and $MS_{\text{error}}$ will be about the same.

- If there are true differences between groups:
  - variance between groups will exceed error variance (within groups)
  - $F$ will be much greater than 1
  - $F_{\text{obt}}$ can also deviate from 1 by chance alone
  - we need a sampling distribution to tell how high $F_{\text{obt}}$ needs to be before we reject the $H_0$
  - compare $F_{\text{obt}}$ to critical value (e.g., $F_{.05}$)

- The critical value ($F_{.05}$) depends on
  - degrees of freedom
    - between groups: $df_{\text{group}} = k - 1$
    - error (within): $df_{\text{error}} = k (n - 1)$
    - Total: $df_{\text{total}} = N - 1$
  - alpha ($\alpha$)
    - e.g., .05, .01

- In ANOVA, variance = Mean Square ($MS$)

\[ F_{\text{Ratio}} = \frac{\text{between group variance}}{\text{error variance (within groups)}} = \frac{MS_{\text{group}}}{MS_{\text{error}}} \]
ANOVA Example: Cell phones

Research Question:
• Is your reaction time when driving slowed by a cell phone? Does it matter if it’s a hands-free phone?
• Twelve participants went into a driving simulator.
  1. A random subset of 4 drove while listening to the radio (control group).
  2. Another 4 drove while talking on a cell phone.
  3. Remaining 4 drove while talking on a hands-free cell phone.
• Every so often, participants would approach a traffic light that was turning red. The time it took for participants to hit the breaks was measured.

A 6 Step Program for Hypothesis Testing

1. State your research question
   • Is your reaction time when driving influenced by cell-phone usage?
2. Choose a statistical test
   • three levels of a single independent variable (cell; hands-free; control)
     → One-Way ANOVA, between subjects

3. Select \( \alpha \), which determines the critical value
   • \( \alpha = .05 \) in this case
   • See \( F \)-tables (page 543 in the Appendix)
     • \( df_{group} = k – 1 = 3 – 1 = 2 \) (numerator)
     • \( df_{error} = k (n - 1) = 3(4-1) = 9 \) (denominator)
     • \( F_{.05} = ? \)
     4.26

4. State Hypotheses
   - referred to as the omnibus null hypothesis
   \( H_0: \mu_1 = \mu_2 = \mu_3 \)
   \( H_1: \) not all \( \mu \)'s are equal
   • When rejecting the Null in ANOVA, we can only conclude that there is at least one significant difference among conditions.
   • If ANOVA significant – pinpoint the actual difference(s), with post-hoc comparisons

\( F \) Distribution critical values (alpha = .05)

<table>
<thead>
<tr>
<th>df</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>denom.</td>
<td>163.1</td>
<td>199.5</td>
<td>213.8</td>
<td>224.8</td>
<td>230.0</td>
<td>233.8</td>
<td>236.5</td>
<td>238.6</td>
<td>240.1</td>
<td>242.1</td>
</tr>
<tr>
<td>3</td>
<td>10.1</td>
<td>9.5</td>
<td>9.28</td>
<td>9.12</td>
<td>9.01</td>
<td>8.94</td>
<td>8.89</td>
<td>8.85</td>
<td>8.81</td>
<td>8.79</td>
</tr>
<tr>
<td>4</td>
<td>7.1</td>
<td>6.9</td>
<td>6.59</td>
<td>6.39</td>
<td>6.26</td>
<td>6.16</td>
<td>6.09</td>
<td>6.04</td>
<td>6.00</td>
<td>5.96</td>
</tr>
<tr>
<td>5</td>
<td>6.1</td>
<td>5.7</td>
<td>5.41</td>
<td>5.19</td>
<td>5.05</td>
<td>4.95</td>
<td>4.88</td>
<td>4.82</td>
<td>4.77</td>
<td>4.74</td>
</tr>
<tr>
<td>6</td>
<td>5.9</td>
<td>5.14</td>
<td>4.75</td>
<td>4.53</td>
<td>4.39</td>
<td>4.28</td>
<td>4.21</td>
<td>4.15</td>
<td>4.10</td>
<td>4.06</td>
</tr>
<tr>
<td>7</td>
<td>5.9</td>
<td>4.74</td>
<td>4.35</td>
<td>4.12</td>
<td>3.97</td>
<td>3.87</td>
<td>3.79</td>
<td>3.73</td>
<td>3.68</td>
<td>3.64</td>
</tr>
<tr>
<td>8</td>
<td>5.3</td>
<td>4.46</td>
<td>4.07</td>
<td>3.84</td>
<td>3.69</td>
<td>3.56</td>
<td>3.50</td>
<td>3.44</td>
<td>3.39</td>
<td>3.35</td>
</tr>
<tr>
<td>9</td>
<td>5.1</td>
<td>4.26</td>
<td>3.90</td>
<td>3.63</td>
<td>3.48</td>
<td>3.37</td>
<td>3.29</td>
<td>3.23</td>
<td>3.18</td>
<td>3.14</td>
</tr>
<tr>
<td>10</td>
<td>4.9</td>
<td>4.10</td>
<td>3.71</td>
<td>3.48</td>
<td>3.33</td>
<td>3.22</td>
<td>3.14</td>
<td>3.07</td>
<td>3.02</td>
<td>2.98</td>
</tr>
<tr>
<td>11</td>
<td>4.8</td>
<td>3.98</td>
<td>3.59</td>
<td>3.36</td>
<td>3.27</td>
<td>3.19</td>
<td>3.12</td>
<td>3.05</td>
<td>3.00</td>
<td>2.95</td>
</tr>
</tbody>
</table>
Examine Data and Calculate $F_{obt}$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Normal Cell</td>
<td>Hands-free Cell</td>
</tr>
<tr>
<td>.50</td>
<td>.75</td>
<td>.65</td>
</tr>
<tr>
<td>.55</td>
<td>.65</td>
<td>.50</td>
</tr>
<tr>
<td>.45</td>
<td>.60</td>
<td>.65</td>
</tr>
<tr>
<td>.40</td>
<td>.60</td>
<td>.70</td>
</tr>
</tbody>
</table>

DV: Response time (seconds)

Is there at least one significant difference between conditions?

A Reminder about Variance

- **Variance**: the average squared deviation from the mean

\[ X = 2, 4, 6 \quad \bar{X} = 4 \]

Sample Variance Definitional Formula

\[
s^2_X = \frac{\sum (X - \bar{X})^2}{N-I}
\]

\[
\begin{array}{c|c|c|c|c}
\hline
X & X - \bar{X} & (X - \bar{X})^2 \\
\hline
2 & -2 & 4 \\
4 & 0 & 0 \\
6 & 2 & 4 \\
\hline
\end{array}
\]

\[
\sum (X - \bar{X})^2 = 8
\]

\[
s^2_X = \frac{8}{2} = 4
\]

ANOVA Summary Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>$df$</th>
<th>Mean Squares</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>$SS_{group}$</td>
<td>$df_{group}$</td>
<td>$MS_{group}$</td>
<td>$F_{obt}$</td>
</tr>
<tr>
<td>Error</td>
<td>$SS_{error}$</td>
<td>$df_{error}$</td>
<td>$MS_{error}$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SS_{total}$</td>
<td>$df_{total}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- $SS = $ Sum of squared deviations or “Sum of Squares”

\[
SS_{total} = SS_{error} + SS_{group}
\]

\[
SS_{total} = .072 + .050 = .122
\]

Computing the Mean Squares

\[
MS_{group} = \frac{SS_{group}}{df_{group}} = \frac{.072}{2} = .0360
\]

\[
MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{.050}{9} = .0056
\]

Examine Data and Calculate $F_{obt}$

- $df$ between groups = $k - 1$
- $df$ error (within groups) = $k (n - 1)$
- $df$ total = $N - 1$
  (the sum of $df_{group}$ and $df_{error}$)
Examine Data and Calculate $F_{\text{obt}}$

- Compute $F_{\text{obt}}$
  
  \[
  F_{\text{obt}} = \frac{MS_{\text{group}}}{MS_{\text{error}}} = \frac{.0360}{.0056} = 6.45
  \]

**ANOVA Summary Table**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>.072</td>
<td>2</td>
<td>.0360</td>
<td>6.45</td>
</tr>
<tr>
<td>Error</td>
<td>.050</td>
<td>9</td>
<td>.0056</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.122</td>
<td>11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ANOVA Example: Cell phones**

- Interpret results in terms of hypothesis
  
  $6.45 > 4.26$; Reject $H_0$ and accept $H_1$

- Report results
  
  $F(2, 9) = 6.45, \ p < .05$

- Explain in plain language
  
  – Among those three groups, there is at least one significant difference

**Post-hoc Comparisons**

- Fisher’s Least Significant Difference Test (LSD)
  
  – uses $t$-tests to perform all pairwise comparisons between group means.
  
  – good with three groups, risky with $> 3$
  
  – this is a liberal test; i.e., it gives you high **power** to detect a difference if there is one, but at an increased risk of a Type I error
Post-hoc Comparisons

• Bonferroni procedure
  – uses \( t \)-tests to perform pairwise comparisons between group means,
  – but controls overall error rate by setting the error rate for each test to the familywise error rate divided by the total number of tests.
  – Hence, the observed significance level is adjusted for the fact that multiple comparisons are being made.
  – e.g., if six comparisons are being made (all possibilities for four groups), then alpha = \( .05/6 = .0083 \)

• Tukey HSD (Honestly Significant Difference)
  – sets the familywise error rate at the error rate for the collection for all pairwise comparisons.
  – very common test

• Other post-hoc tests also seen:
  – e.g., Newman-Keuls, Duncan, Scheffe’…

Effect Size: Partial Eta Squared

• Partial Eta squared (\( \eta^2 \)) indicates the proportion of variance attributable to a factor
  – 0.20 small effect
  – 0.50 medium effect
  – 0.80 large effect

• Calculation: PASW

Effect Size: Omega Squared

• A less biased indicator of variance explained in the population by a predictor variable

\[
\omega^2 = \frac{SS_{group} - (k-1)MS_{error}}{SS_{total} + MS_{error}}
\]

\[
\omega^2 = \frac{.072 - (3-1)(.0056)}{.122 + .0056} = .48
\]

• 48% of the variability in response times can be attributed to group membership (medium effect)

PASW: One-Way ANOVA (Between Subjects)

• Setup a one-way between subjects ANOVA as you would an independent samples \( t \)-test:
• Create two variables
  – one variable contains levels of your independent variable (here called “group”).
    • there are three groups in this case numbered 1-3.
  – second variable contains the scores of your dependent variable (here called “time”)

PASW : One-Way ANOVA (Between Subjects)
Label the numbers you used to differentiate groups:

Go to “Variable View”, then click on the “Values” box, then the gray box labeled “…”

Enter Value (in this case 1, 2 or 3) and the Value Label (in this case: control, cell, hands)

Click “Add”, and then add the next two variables.

Performing Test

Select from Menu: Analyze -> General Linear Model -> Univariate

Select your dependent variable (here: “time”) in the “Dependent Variable” box

Select your independent variable (here: “group”) in the “Fixed Factor(s)” box

Click “Options” button,
– check Descriptives (this will print the means for each of your levels)
– check Estimates of effect size for Partial Eta Squared

Click the Post Hoc button for post hoc comparisons; move factor to “Post Hoc Tests for” box; then check “LSD, Bonferroni, or Tukey”

Click OK

PASW and Effect Size

Click Options menu;
then check Estimates of effect size box

This option produces partial eta squared
PASW Data Example

- Three groups with three in each group ($N = 9$)

<table>
<thead>
<tr>
<th></th>
<th>Fast</th>
<th>Medium</th>
<th>Slow</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>44.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>30.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
</tbody>
</table>

$\bar{x} = 31.3 \quad 8.7 \quad 2.0$

ANOVA Summary

<table>
<thead>
<tr>
<th>Source</th>
<th>Type I Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Error</td>
<td>14.18447</td>
<td>2</td>
<td>7.09221</td>
<td>7.424</td>
<td>.022</td>
</tr>
<tr>
<td>Intercept</td>
<td>255.000</td>
<td>1</td>
<td>255.000</td>
<td>18.655</td>
<td>.001</td>
</tr>
<tr>
<td>Group (between)</td>
<td>18.655</td>
<td>2</td>
<td>9.32751</td>
<td>7.424</td>
<td>.022</td>
</tr>
<tr>
<td>Error</td>
<td>161.187</td>
<td>6</td>
<td>26.86467</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>391.037</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- if $p < .05$, then significant effect

Post Hoc Comparisons

- slow and medium groups are not significantly different
- slow and fast groups are significantly different


- A Sarcastic Version Item
  - Joe came to work, and instead of beginning to work, he sat down to rest.
  - His boss noticed his behavior and said, “Joe, don’t work too hard!”

- A Neutral Version Item
  - Joe came to work and immediately began to work. His boss noticed his behavior and said, “Joe, don’t work too hard!”

- Following each story, participants were asked:
  - Did the manager believe Joe was working hard?