Chapter 13

Introduction to Analysis of Variance (ANOVA)

University of Guelph
Psychology 3320 — Dr. K. Hennig
Winter 2003 Term

Figure 13-2 (p. 397)
A typical situation in which ANOVA would be used. Three separate samples are obtained to evaluate the mean differences among three populations (or treatments) with unknown means.
A research design with two factors. The research study uses two factors: One factor uses two levels of therapy technique (I versus II), and the second factor uses three levels of time (before, after, and 6 months after). Also notice that the therapy factor uses two separate groups (independent measures) and the time factor uses the same group for all three levels (repeated measures).
The test statistics

\[ t = \frac{\text{difference between sample means}}{\text{difference expected by chance (error)}} \]

\[ t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{(M_1-M_2)}} \]

\[ F = \frac{\text{variance (differences) between sample means}}{\text{variance (differences) expected by chance (error)}} \]

- E.g., \( M_1 = 20 \) \( M_2 = 30 \); can find the difference.
- but if there are three means? Use variance to measure the size of the difference between groups.

Logic of ANOVA

- Goal: measure the variability and determine where it comes from.
- Determine the total variability of the data
  - break into parts (analyze) the variability/variance, i.e., ANOVA
Logic (contd.)

- Between-treatment variance (note 50° vs. 70°)
- Within-treatment variance (e.g., 70° - not all the same)
- When you see “variance,” think “differences”

<table>
<thead>
<tr>
<th>TREATMENT 1 50° (SAMPLE 1)</th>
<th>TREATMENT 2 70° (SAMPLE 2)</th>
<th>TREATMENT 3 90° (SAMPLE 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ M = 1 \quad M = 4 \quad M = 1 \]

Logic (contd.)

- \( H_0 \): the differences between treatments are simply due to chance
- \( H_1 \): the differences are significantly greater than can be explained by chance, i.e., the differences are caused by treatment effects
- Two primary sources of chance differences:
  - individual differences - people are different
  - experimental error - measurement always involves
Logic (contd.)

- Compare with chance - how big are the differences when there is no treatment effect?
- Within-treatments variance:
  - e.g., in 70° condition individuals were all tested in same condition, yet different scores
  - within-treatment variance is a measure of difference expected just by chance

**Figure 13-4** (p. 403)
The independent-measures analysis of variance partitions, or analyzes, the total variability into two components: variance between treatments and variance within treatments.

- Total variability
  - Between-treatments variance
    - Measures differences due to
      1. Treatment effects
      2. Chance
  - Within-treatments variance
    - Measures differences due to
      1. Chance
The $F$-ratio: The test statistic for ANOVA

$$F = \frac{\text{variance between treatments}}{\text{variance within treatments}}$$

$$= \frac{\text{treatment effect} + \text{differences due to chance}}{\text{differences due to chance}}$$

$$= \frac{0 + \text{differences due to chance}}{\text{differences due to chance}}$$

- An $F$-ratio near 1.00 means differences are due to chance
- In ANOVA the denominator is the error term; the same as the numerator when the treatment effect is zero

Steps

1. Analysis of the SS
   - $SS_{\text{within}}$
   - $SS_{\text{between}}$
2. Analysis of the degrees of freedom ($df$)
   - $df_{\text{within}}$
   - $df_{\text{between}}$
3. Calculation of variances
1. Analysis of the SS

<table>
<thead>
<tr>
<th>TREATMENT 1 50° (SAMPLE 1)</th>
<th>TREATMENT 2 70° (SAMPLE 2)</th>
<th>TREATMENT 3 90° (SAMPLE 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>M = 1</td>
<td>M = 4</td>
<td>M = 1</td>
</tr>
</tbody>
</table>

- Calculate $T_{1,2,3}$ ($X=\bar{T} \pm S$) for each of the $k=3$ treatment conditions.
- Calculate $G$ (grand total) = $T_1 + T_2 + T_3$
- $N = n_1 + n_2 + n_3$ Calculate $X^2$ for entire $N = 15$.

---

Table 13-2 (p. 406)
Hypothetical data from an experiment examining learning performance under three temperature conditions.

<table>
<thead>
<tr>
<th>TEMPERATURE CONDITIONS 1 50°</th>
<th>2 70°</th>
<th>3 90°</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

$\Sigma X^2 = 106$
$G = 30$
$N = 15$
$k = 3$

$T_1 = 5$
$T_2 = 20$
$T_3 = 5$
$SS_1 = 6$
$SS_2 = 6$
$SS_3 = 4$
$n_1 = 5$
$n_2 = 5$
$n_3 = 5$
$M_1 = 1$
$M_2 = 4$
$M_3 = 1$

*Summary values and notation for an analysis of variance are also presented.*
1a. SS within treatments
Partitioning the sum of squares (SS) for the independent-measures analysis of variance.

\[ SS = \sum X^2 - \frac{(\sum X)^2}{N} \]

\[ SS_{\text{Total}} = \sum X^2 - \frac{G^2}{N} \]

\[ SS_{\text{within}} = SS_1 + SS_2 + SS_3 \]

1b. SS between treatments

<table>
<thead>
<tr>
<th>MEAN (X)</th>
<th>( X^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
</tr>
</tbody>
</table>

\[ SS = \sum X^2 - \frac{(\sum X)^2}{n} \]

\[ = 18 - \frac{(6)^2}{3} \]

\[ = 18 - 12 = 6 \]

But, only works if samples are all same size (\( ns \) are equal), thus use a computational formula:

\[ SS_{\text{between}} = \sum \frac{T^2}{n} - \frac{G^2}{N} \]
Step 2: Analysis of the degrees of freedom (df)
Partitioning degrees of freedom (df) for the independent-measures analysis of variance.

\[ df_{\text{total}} = N - 1 = 15 - 1 = 14 \]

\[ df_{\text{within}} = N - k = 15 - 3 = 12 \]

\[ df_{\text{between}} = k - 1 = 3 - 1 = 2 \]

Step 3: Analysis of the variances
The structure and sequence of calculations for the analysis of variance. (Recall: \( s^2 = \frac{SS}{df} \))
Results are organized into a table

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>( F = 11.28 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between treatments</td>
<td>30</td>
<td>2</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Within treatments</td>
<td>16</td>
<td>12</td>
<td>1.33</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
<td>14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( F \) ratios with \( df = 2.12 \). Of all the values in the distribution, only 5% are larger than \( F = 3.88 \), and only 1% are larger than \( F = 6.93 \).
### Table 13-3 (p. 414)
A portion of the \( F \) distribution table. Entries in roman type are critical values for the .05 level of significance, and bold type values are for the .01 level of significance. The critical values for \( df = 2.12 \) have been highlighted (see text).

<table>
<thead>
<tr>
<th>Degrees of Freedom:</th>
<th>Degrees of Freedom: Numerator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denominator</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>4.96</td>
</tr>
<tr>
<td>11</td>
<td>10.04</td>
</tr>
<tr>
<td>12</td>
<td>4.84</td>
</tr>
<tr>
<td>9.65</td>
<td>7.20</td>
</tr>
<tr>
<td>13</td>
<td>4.75</td>
</tr>
<tr>
<td>14</td>
<td>9.33</td>
</tr>
<tr>
<td>21</td>
<td>4.67</td>
</tr>
<tr>
<td>14</td>
<td>9.07</td>
</tr>
<tr>
<td>8.86</td>
<td>6.51</td>
</tr>
</tbody>
</table>

### Figure 13-15 (p. 421)
A visual representation of the between-treatments variability and the within-treatments variability that form the numerator and denominator, respectively, of the \( F \)-ratio. In (a), the difference between treatments is relatively large and easy to see. In (b), the same 4-point difference between treatments is relatively small and is overwhelmed by the within-treatments variability.
The distribution of $t$ statistics with $df = 18$ and the corresponding distribution of $F$-ratios with $df = 1.18$. Notice that the critical values for $\alpha = .05$ are $t = \pm 2.101$ and that $F = 2.101^2 = 4.41$

**Figure 13-11** (p. 431)

Post hoc tests

- E.g., $M_1 = 3$  $M_2 = 5$  $M_3 = 10$
- 2-pt difference $M_1$ and $M_2$
- experimentwise alpha level
- planned vs. unplanned comparisons
- Tukey (HSD)

$$HSD = q\sqrt{\frac{MS_{within}}{n}}$$

- Scheffe