Giancola & Corman (2007)

- Effects of alcohol & attention on aggressive behaviour
- Hypothesis: Alcohol facilitates aggression by focusing attention on salient provocative cues
  - Presenting Ss with a task should reduce attention on provocative cues and therefore reduce aggression
  - But if task is too complex, Ss will stop focusing on task and aggressive behaviour will resume/increase
- Ss were given alcohol to raise blood-alcohol level to 0.10%
- Different groups of Ss performed a memory task that varied in difficulty
- All Ss first received shocks from partner based on task performance
  - Later, all Ss delivered shocks to partner
    - Aggression measured as intensity & duration of shocks
- What is/are the dependent & independent variables?

Giancola & Corman (2007)

- Subjects assigned randomly to 5 groups
  - Groups varied in task difficulty (independent variable)
  - 12 Ss per group
- Shock “level” (Intensity & duration) was dependent variable
- Compare mean shock levels in each group
  - Are observed differences across groups due to chance?
- Null Hypothesis:
  - H0: Mean shock level was the same in all groups
- Alternative/Research Hypothesis:
  - H1: Mean shock level was not the same in all groups
- Cannot evaluate H0 & H1 with a t test because we have 5 (not 2) groups
- Instead, evaluate H0 & H1 using the Analysis of Variance (ANOVA)
Why evaluate differences in group means with the analysis of variance?

By hypothesis, groups differ only in terms of mean aggression score.

When null hypothesis is true, what causes variation within each group?
- population error variance $\sigma_e$
- Variance within each group is an estimate of $\sigma_e$.
- average within-group variance is best estimate of $\sigma_e$.

\[
\hat{\sigma}_e^2 = s_j^2 = \frac{s_1^2 + s_2^2 + s_3^2 + s_4^2 + s_5^2}{5}
\]

When $H_0$ is true, what causes variation among group means?
- sampling error
- When $H_0$ is true, 5 groups are 5 independent samples from same population of aggression scores
- Between-group variance related to $\sigma_e$ and sample size ($n$)

\[
\sigma_X^2 = \frac{\sigma_e^2}{n}
\]

When $H_0$ is true, variation within and between groups is related to $\sigma_e$.
- $\sigma_e$ is estimated by average of within-group variances [MSwithin, MSerror]
- $\sigma_e$ is estimated by product of $n$ and between-group variance [MSgroup]
- Hence, within- & between-group variances provide 2 independent estimates of $\sigma_e$.
- and therefore the MSgroup & MSerror should be similar

\[
\sigma_e^2 = \frac{s_j^2}{s_j^2} = \frac{s_1^2 + s_2^2 + s_3^2 + s_4^2 + s_5^2}{5}
\]

\[
\hat{\sigma}_X^2 = \frac{\sigma_e^2}{n} = \frac{n \times \sigma_x^2}{n} = \sigma_e^2
\]
When \( H_0 \) is false:
- within-group variance still determined only by \( \sigma_e \)
- but variation between groups is related to \( \sigma_e \) and to effects of the independent variable
- hence, variation between group means should be greater than variation predicted solely by \( \sigma_e \)
- so \( MS_{\text{group}} \) should be greater than \( MS_{\text{error}} \)

\[
\sigma^2 = \frac{s^2}{n}
\]

\[
v_X = n \times \sigma^2
\]

\[
Variance \ among \ group \ means: \quad MS_{\text{group}} = \frac{\sum (\bar{X}_i - \bar{X})^2}{k-1} = 1.000
\]

\[
Average \ of \ group \ variances: \quad MS_{\text{error}} = \frac{\sum s_i^2}{k} = \frac{100.00 + 10.00 + 7.75}{3} = 39.175
\]

\[
MS_{\text{group}} = \frac{MS_{\text{group}}}{MS_{\text{error}}} = 9.25
\]

ANOVA (general strategy)
- compute within-group and between-group estimates of \( \sigma^2_e \)
- when \( \text{H}_0 \) is true, the within- and between-group estimates should be similar
- when \( \text{H}_0 \) is false, the between-group estimate should be larger than the within-group estimate
- so, we will evaluate \( \text{H}_0 \) by determining if the ratio of between- and within-group estimates of \( \sigma^2_e \) is unusually large given that \( \text{H}_0 \) is true
ANOVA calculations

- Sums of Squares: sum of squared deviations from mean
  - $SS_{total} = \Sigma(X - \overline{X}_{gm})^2$  ($\overline{X}_{gm}$ is grand mean (mean of all scores))
  - $SS_{group} = n \times \Sigma((\text{Mean Group } j) - \overline{X}_{gm})^2$  ($n$ = scores per group)
  - $SS_{total} = SS_{group} + SS_{error}$
    - so: $SS_{error} = SS_{total} - SS_{group}$
- degrees of freedom:
  - $df_{total} = N - 1$  ($N$ is the total number of observations)
  - $df_{group} = k - 1$  ($k$ is the number of groups)
  - $df_{error} = df_{total} - df_{group}$

- Mean Squares:
  - $MS_{error} = SS_{error}/df_{error}$  [estimate of $\sigma_e^2$ when H0 is true]
  - $MS_{group} = SS_{group}/df_{group}$  [estimate of $\sigma_e^2$ when H0 is true]
- F statistic: $MS_{group}/MS_{error}$

ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>4</td>
<td>62.460</td>
<td>15.615</td>
<td>6.90</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>124.458</td>
<td>2.263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>186.918</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F distribution

- ratio of 2 variance estimates follows F distribution
- F distribution is a theoretical distribution (like t distribution)
- family of curves
  - defined by 2 parameters:
    - df in numerator & df in denominator
- large values of F are unusual when numerator & denominator are estimates of same population variance

F test

- observed $F(4,55) = 6.90$
- critical $F(4,55) = 2.54$  ($\alpha=.05$)
- $F_{observed} > F_{critical}$
  - reject H0 (all means equal),
    $F(4,55)=6.90$,  $p<.05$
- note that 2-tailed test does not make sense here because H1 always predicts
  $MS_{group} > MS_{error}$
ANOVA Assumptions

- data in each group are distributed normally
- groups have the same variance
- observations are independent
- violations of assumptions mean F’s p-value will be inaccurate
- ANOVA is somewhat robust to violations of normality and constant-variance assumptions

Non-normality & non-constant variance

- ANOVA reasonably robust to deviations from normality
  - if deviations are similar in all groups
  - robustness declines if n is not equal across groups
  - also declines if deviations differ across groups
    - e.g., positive skew in 1 group, negative skew in others
- ANOVA is reasonably robust to 3-4 fold differences in variances
  - if scores are normally distributed and equal n per group

Bartlett test for constant variance

```r
bartlett.test(mood.data$mood,mood.data$group)
## Bartlett test of homogeneity of variances
## data: mood.data$mood and mood.data$group
## Bartlett's K-squared = 2.6, df = 2, p-value = 0.3
```

Bartlett test is a common procedure for evaluating the null hypothesis that variance is constant across groups

Common tests for non-normality

- Kolmogorov-Smirnov test
- Shapiro-Wilk’s test

```r
shapiro.test(residuals(mood.full) )
## Shapiro-Wilk normality test
## data: residuals(mood.full)
## W = 0.85, p-value = 5e-04
```

Null hypothesis is that within-group scores are distributed normally.
When assumptions are violated

• perform ANOVA on transformed data
  - square-root, log, & inverse-sine transformations common
  - conclusions apply to transformed data
• Welch df correction for non-constant variance [assumes normality]
• Kruskal-Wallis test for group differences
  - does not assume normality or constant variance
  - KW test evaluates null hypothesis that means of ranked data are the same in each group
  - if we assume that distributions for each group have same shape (not necessarily normal), then KW test evaluates null hypothesis that group MEDIANs are equal

Measures of association strength & effect size

• Association Strength:
  - proportion of total variation in scores associated with grouping variable
  - varies between 0 & 1
  - similar to $R^2$ measure in linear regression
  - $\eta^2$
    - $SS_{\text{group}}/SS_{\text{total}} = SS_{\text{group}}/(SS_{\text{group}}+SS_{\text{error}})$
    - biased estimate of association score-group association in population
• $\omega^2$
  - like $\eta^2$, $\omega^2$ measures association between scores and grouping variable
  - less biased than $\eta^2$
• Effect Size
  - measure of difference between means relative to within-group standard deviation
  - Cohen's $d$: could compute this for each pair of groups
  - Cohen's $f$: generalization of Cohen's $d$ to more than 2 groups
    - similar to average Cohen's $d$ for each group:
      - e.g., average of (group mean - grand mean)/ $\sqrt{MS_{\text{error}}}$

Association strength & effect size

Guidelines from Cohen (1988)

<table>
<thead>
<tr>
<th>Magnitude of association or effect</th>
<th>Omega-squared</th>
<th>Cohen's $f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>small</td>
<td>0.01</td>
<td>0.1</td>
</tr>
<tr>
<td>medium</td>
<td>0.06</td>
<td>0.25</td>
</tr>
<tr>
<td>large</td>
<td>0.14</td>
<td>0.4</td>
</tr>
</tbody>
</table>
low association strength

- variation between groups is very small compared to variation within groups
  - $\eta^2 = \frac{55}{(55+4238)} = 0.013$
  - $\omega^2 \approx 0$
  - Cohen’s $f \approx 0$
- average difference between group means and overall means is approximately zero

Giancola & Corman (2007) [Aggression Study]

- omnibus $F=6.901$, $p<.01$
- reject null hypothesis of no group difference
- $\eta^2 = \frac{62.46}{(62.46+124.46)} = 0.33$
- $\omega^2 = 0.28$
- Cohen’s $f = 0.68$
- big effect of group

Multiple Comparisons
Giancola & Corman (2007) [Aggression Study]

* significant omnibus F suggests H0 is false
  - group means are not all equal
* but how do groups differ?
  - H1 can take many forms
* to answer this question, researchers often perform multiple comparisons or contrasts between groups
  - e.g., which pairwise group differences are significant?
  - does mean of g3 differ from mean of groups g1 & g5?

---

Multiple Comparisons & Type I Error

* Consider case of evaluating each pairwise difference between 5 groups
  - total of 10 pairwise tests:
    - g1 vs g2, g1 vs g3, g1 vs g4, g1 vs g5
    - g2 vs g3, g2 vs g4, g2 vs g5
    - g3 vs g4, g3 vs g5
    - g4 vs g5
  - For each comparison, we use a t test with $\alpha = 0.05$
  - Assume complete null hypothesis is true: all group means are the same
  - For our 10 tests, what is probability of making at least one Type I error?
    - $p(\text{at least 1 Type I error}) = 1 - (1-0.05)^{10} = 0.40$ [this value differs from textbook]
    - familywise $\alpha$ [$\alpha$ for entire set/family of tests] is 0.40, not .05

---

Giancola & Corman (2007) [Aggression Study]

* Using Tukey HSD to evaluate all pairwise differences between group means:

| Tukey multiple comparisons of means 95% family-wise confidence level |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                 | diff            | lwr             | upr             | p-adj           |
| g2-g1                           | -1.1275         | -2.860          | 0.605           | 0.364           |
| g3-g1                           | -2.7150         | -4.447          | -0.983          | 0.000           |
| g4-g1                           | -1.2583         | -2.990          | 0.474           | 0.257           |
| g5-g1                           | 0.0867          | -1.645          | 1.819           | 1.000           |
| g3-g2                           | -1.5875         | -3.320          | 0.145           | 0.087           |
| g4-g2                           | -0.1308         | -1.863          | 1.601           | 1.000           |
| g5-g2                           | 1.2142          | -0.518          | 2.946           | 0.291           |
| g4-g3                           | 1.4587          | -0.275          | 3.189           | 0.139           |
| g5-g3                           | 2.8017          | 1.070           | 4.534           | 0.000           |
| g5-g4                           | 1.3450          | -0.387          | 3.077           | 0.199           |
Giancola & Corman (2007) [Aggression Study]

- Other procedures exist to do more complex comparisons:
  - e.g., compare the mean of g3 to the mean of g1 & g2 (combined)
  - e.g., compare difference between g1 & g2 to difference between g5 & g4

- These comparisons are linear contrasts
  - often more appropriate & powerful test of experimental hypothesis that omnibus F test
  - important difference between planned vs post-hoc contrasts
  - post-hoc contrasts use Scheffé’s method to control Type I error rate

Omnibus F & Multiple Comparisons

- Omnibus F and multiple comparison procedures often yield different results
  - e.g., omnibus F may be significant, but Tukey HSD may fail to find any significant pairwise differences
  - e.g., Tukey HSD may find significant pairwise differences but omnibus F may not be significant
  - these procedures evaluate different null hypotheses and therefore different results are not unexpected

- One exception: Omnibus F & Scheffe test are consistent:
  - if omnibus F is significant there is at least one linear contrast that is significant with Scheffe method
  - if omnibus F is not significant then no linear contrast will be significant with Scheffe method

One-way ANOVA summary

- One-way ANOVA: 1 independent variable with more than 2 groups
- Assumptions:
  - normality, homogeneity of variance, independent observations
  - Bartlett, Kolmogorov-Smirnov, & Shapiro-Wilks tests
  - alternative analyses: data transformations, Welch df correction, & Kruskal-Wallis test
- Sums-of-squares, degrees-of-freedom, Mean Squares
- omnibus F test:
  - \( F = \frac{MS_{group}}{MS_{error}} \)
  - F distribution
  - null hypothesis of no difference among group means
- Association Strength & Effect Size: eta-squared, omega-squared, Cohen’s f
- Multiple Comparisons & Linear Contrasts