Recall two-sample \( t \) statistics

- used to compare the means of two populations
- if
  - the two populations are assumed to have equal but unknown standard deviations \( \sigma \)
  - the sample sizes are both equal to \( n \)
  - \( s_p^2 \) is the pooled sample estimate of the common variance \( \sigma^2 \)
- then the \( t \) statistic is

\[
t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{n}}} = \frac{\sqrt{\frac{n}{n}(\bar{x} - \bar{y})}}{s_p}
\]

The One-way ANOVA model

- For testing the hypothesis that the means are equal in \( I \) different populations
  - take random samples from each of \( I \) populations
  - \( n_i \) is sample size in \( i \)th population
  - \( y_{ij} \) is \( j \)th observation from \( i \)th population
  - \( \mu_i \) is unknown population mean of \( i \)th population
  - \( \sigma^2 \) is unknown common variance of all \( I \) populations
- then the one-way ANOVA model is

\[
y_{ij} = \mu_i + e_{ij}
\]

- for \( i = 1, \ldots, I \) and \( j = 1, \ldots, n_i \)
- \( e_{ij} \) assumed to be from \( N(0, \sigma^2) \) distribution
- unknown model parameters to be estimated:
  \( \mu_1, \mu_2, \ldots, \mu_I, \sigma^2 \)
Assessing the assumptions of one-way ANOVA

- normal distribution of response variable in all populations of interest
  - histograms, boxplots for sample data from each population
  - normal qq plot for sample data from each population
- same standard deviation $\sigma$ in all populations
  - possible to use Levene’s test for homogeneity of variance, but it is less robust to non-normality than the ANOVA procedure!
  - rule of thumb: if largest sample standard deviation isn’t more than twice as large as smallest sample standard deviation, assumption probably is met closely enough that ANOVA results will be OK
  - also, ANOVA is less sensitive to violation of equality of variance assumption if $n_i$’s are close to equal
- If one or both assumptions are violated, often a transformation of the response variable will improve both
  - or nonparametric procedure might be used

Estimating the population parameters

- For each population $i$, the estimate of $\mu_i$ is
  \[ \bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \]
- pooled estimator of the common variance $\sigma^2$
  \[ s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_I - 1)} \]
  where $s_i^2$ is the sample variance in the $i$th sample
- pooled estimator of the common standard deviation $\sigma$
  \[ s_p = \sqrt{s_p^2} \]

Hypothesis testing in one-way ANOVA

The null and alternative hypothesis for one-way ANOVA are:

\[ H_0 : \mu_1 = \mu_2 = \cdots = \mu_I \]
\[ H_a : \text{not all the } \mu_i \text{ are equal} \]
### The ANOVA Table

<table>
<thead>
<tr>
<th>Source of freedom</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td>I - 1</td>
<td>$\sum_{\text{groups}} n_i (\bar{y}_i - \bar{y})^2$</td>
<td>GSS/GDF</td>
<td>MSG/RMS</td>
</tr>
<tr>
<td>Residuals</td>
<td>N - I</td>
<td>$\sum_{\text{groups}} (n_i - 1)s_i^2$</td>
<td>RSS/RDF</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>N - 1</td>
<td>$\sum_{\text{obs}} (y_{ijk} - \bar{y})^2$</td>
<td>TSS/TDF</td>
<td></td>
</tr>
</tbody>
</table>

where

- $\bar{y}$ is the overall mean from all samples combined
- GSS, GMS, and GDF are sum of squares, mean square, and degrees of freedom for groups
  - SAS calls this for “Model”
- RSS, RMS, and RDF are sum of squares, mean square, and degrees of freedom for residuals
- TSS, TMS, and TDF are total sum of squares, mean square, and degrees of freedom

### Coefficient of determination in ANOVA

$$R^2 = \frac{\text{GSS}}{\text{TSS}}$$

This is the proportion of the total variation in the data that is accounted for by the model

- (that is, by the differences among the group means)

### Two-way ANOVA

- used to compare population means when the populations are classified according to *two* factors
- advantages compared to doing two one-way ANOVAs
  - more efficient to study two factors simultaneously rather than separately
  - can reduce residual variation in a model by including a second factor believed to affect the response
  - can study interaction between factors

### Assumptions of two-way ANOVA

- But first some notation:
  - first categorical variable: factor A with I levels
  - second categorical variable: factor B with J levels
  - $n_{ij}$: sample size for level $i$ of A and level $j$ of B
  - total number of obs $N = \sum_{i} \sum_{j} n_{ij}$
  - $\mu_{ij}$: unknown mean of response variable in population defined by level $i$ of A and level $j$ of B
  - $y_{ijk}$ is $k$th observation from population having factor A at level $i$ and factor B at level $j$
- assumptions
  - independent simple random samples of size $n_{ij}$ from each of $I \times J$ normal populations
- all these populations have same standard deviation $\sigma$ (unknown)
- model
  
  $y_{ijk} = \mu_{ij} + \epsilon_{ijk}$
  
  $\epsilon_{ijk}$’s follow a $N(0, \sigma^2)$ distribution

The ANOVA table for two-way ANOVA

As in one-way ANOVA, both the sums of squares and the degrees of freedom add.

- total sum of squares = model sum of squares + residual sum of squares
  
or
  TSS = MSS + RSS
- total degrees of freedom = model degrees of freedom + residual degrees of freedom
  
or
  TDF = MDF + RDF

In two-way ANOVA, however, MSS and MDF are each split up 3 ways:

- SSM = SSA + SSB + SSAB
- DFM = DFA + DBF + DFAB

producing 3 distinct F statistics: for factor A, for factor B, and for the interaction.

The general ANOVA table for two-way ANOVA

<table>
<thead>
<tr>
<th>Source of freedom</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>I - 1</td>
<td>SSA</td>
<td>SSA / DFA</td>
<td>MSA / MSE</td>
</tr>
<tr>
<td>B</td>
<td>J - 1</td>
<td>SSB</td>
<td>SSB / DBF</td>
<td>MSB / MSE</td>
</tr>
<tr>
<td>AB</td>
<td>(I - 1)(J-1)</td>
<td>SSAB</td>
<td>SSAB / DFAB</td>
<td>MSAB / MSE</td>
</tr>
<tr>
<td>Error</td>
<td>N - IJ</td>
<td>SSE</td>
<td>SSE / DF E</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>N - 1</td>
<td>TSS</td>
<td>TSS / TDF</td>
<td></td>
</tr>
</tbody>
</table>
Examples of what a significant interaction can mean

Example: the depression data Without interaction term

```plaintext
data depress;
infile '/group/ftp/pub/snowles/datasets/depress.dat';
input eff age trt $;
trta = 0; trtb = 0;
if trt eq 'A' then trta = 1;
else if trt eq 'B' then trtb = 1;
if trt eq 'C' then do;
  trta = .;
  trtb = .;
end;
age40 = 0;
if age > 40 then age40 = 1;
run;

proc glm;
class trt age40;
model eff = trt age40;
means trt age40 / bon;
run;
```

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>3</td>
<td>A B C</td>
</tr>
<tr>
<td>age40</td>
<td>2</td>
<td>0 1</td>
</tr>
</tbody>
</table>

Number of observations 36

The GLM Procedure

Dependent Variable: eff

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>3524.461601</td>
<td>1174.820534</td>
<td>20.10</td>
</tr>
<tr>
<td>Error</td>
<td>32</td>
<td>1870.583399</td>
<td>58.454325</td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>35</td>
<td>5395.000000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source | Pr > F |
-------|--------|
Model   | <.0001 |
Error   |        |
Corrected Total | <.0001 |

R-Square Coeff Var Room MSE eff Mean

- R-Square: 0.63283
- Coeff Var: 13.85899
- Room MSE: 7.645543
- eff Mean: 55.16667

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>2</td>
<td>927.16667</td>
<td>463.583333</td>
<td>7.93</td>
</tr>
</tbody>
</table>
The GLM Procedure

Bonferroni (Gum) t Tests for eff

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Degrees of Freedom</td>
<td>32</td>
</tr>
<tr>
<td>Error Mean Square</td>
<td>58.45432</td>
</tr>
<tr>
<td>Critical Value of t</td>
<td>2.03693</td>
</tr>
<tr>
<td>Minimum Significant Difference</td>
<td>5.3243</td>
</tr>
<tr>
<td>Harmonic Mean of Cell Sizes</td>
<td>17.1111</td>
</tr>
</tbody>
</table>

Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>Bon Grouping</th>
<th>Mean</th>
<th>N</th>
<th>trt</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>62.333</td>
<td>12</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>51.917</td>
<td>12</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>51.250</td>
<td>12</td>
<td>C</td>
</tr>
</tbody>
</table>

With interaction

proc glm;
class trt ag3gt40 ;
model eff = trt ag3gt40 trt * ag3gt40 ;
mean trt ag3gt40 trt * ag3gt40 / bon ;
run ;

Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>3</td>
<td>A B C</td>
</tr>
<tr>
<td>ag3gt40</td>
<td>2</td>
<td>0 1</td>
</tr>
</tbody>
</table>

Dependent Variable: eff

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4224.667857</td>
<td>844.933571</td>
<td>21.66</td>
</tr>
<tr>
<td>Error</td>
<td>1170.332143</td>
<td>39.011071</td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>5395.000000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source | Pr > F |
-------|--------|
trt    | <.0001 |
ag3gt40| <.0001 |
trt*ag3gt40 | <.0001 |

Bonferroni (Gum) t Tests for eff

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.
### Bonferroni (Quantile) t Tests for eff

**Means with the same letter are not significantly different.**

<table>
<thead>
<tr>
<th>Bonferroni Grouping</th>
<th>Mean</th>
<th>N</th>
<th>eff</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>61.773</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>44.786</td>
<td>14</td>
<td>0</td>
</tr>
</tbody>
</table>

The GLM Procedure

<table>
<thead>
<tr>
<th>Level of</th>
<th>Level of</th>
<th>Bonferroni Grouping</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>ageq40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>5</td>
<td>56.8000000</td>
<td>4.08665336</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>7</td>
<td>66.2857143</td>
<td>4.46147531</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>4</td>
<td>43.5000000</td>
<td>3.81884582</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>8</td>
<td>66.1260000</td>
<td>6.62112420</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>5</td>
<td>33.8000000</td>
<td>8.13633224</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>7</td>
<td>63.7142857</td>
<td>7.91021040</td>
</tr>
</tbody>
</table>

**NOTE:** This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

---

### Summary Statistics

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Error Degrees of Freedom</th>
<th>Error Mean Square</th>
<th>Critical Value of t</th>
<th>Minimum Significant Difference</th>
<th>Harmonic Mean of Cell Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>30</td>
<td>39.01107</td>
<td>2.53274</td>
<td>6.4688</td>
<td>17.1111</td>
</tr>
</tbody>
</table>

**Means with the same letter are not significantly different.**

<table>
<thead>
<tr>
<th>Bonferroni Grouping</th>
<th>Mean</th>
<th>N</th>
<th>ageq40</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>61.773</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>44.786</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** Cell sizes are not equal.