1-Way ANOVA

MATH 143

Department of Mathematics and Statistics
Calvin College

Spring 2010
The basic ANOVA situation

- Two variables: 1 Categorical, 1 Quantitative
- Main Question: Do the (means of) the quantitative variables depend on which group (given by categorical variable) the individual is in?

- If categorical variable has only 2 values:
  - 2-sample t-test
  - ANOVA allows for 3 or more groups (sub-populations)
The basic ANOVA situation

- Two variables: 1 Categorical, 1 Quantitative
- Main Question: Do the (means of) the quantitative variables depend on which group (given by categorical variable) the individual is in?
- If categorical variable has only 2 values:
The basic ANOVA situation

- Two variables: 1 Categorical, 1 Quantitative
- Main Question: Do the (means of) the quantitative variables depend on which group (given by categorical variable) the individual is in?
- If categorical variable has only 2 values:
  - 2-sample t-test
- ANOVA allows for 3 or more groups (sub-populations)
An example ANOVA situation

Example (Treating Blisters)

- Subjects: 25 patients with blisters
- Treatments: Treatment A, Treatment B, Placebo (P)
- Measurement: # of days until blisters heal
- Data [and means]:
  - A: 5,6,6,7,7,8,9,10 [7.25]
  - B: 7,7,8,9,9,10,10,11 [8.875]
  - P: 7,9,9,10,10,10,11,12,13 [10.11]

Question: Are these differences significant?
(Or would we expect differences this large by random chance?)
Are these differences significant?

Whether differences between the groups are significant depends on

- the difference in the means
- the amount of variation within each group
- the sample sizes
Are these differences significant?

Whether differences between the groups are significant depends on

- the difference in the means
- the amount of variation within each group
- the sample sizes
Some notation for ANOVA

Entire data set:

- $N =$ number of individuals all together
- $k =$ number of groups
- $x_{ij} =$ value for individual $j$ in group $i$
- $\bar{x} =$ sample mean of quant. variable for entire data set
- $s =$ sample s.d. of quant. variable for entire data set

Group $i$:

- $n_i =$ # of individuals in group $i$
- $\mu_i =$ population mean for group $i$
- $\bar{x}_i =$ sample mean for group $i$
- $\sigma_i =$ population standard deviation for group $i$
- $s_i =$ sample standard deviation for group $i$
The ANOVA model

\[ x_{ij} = \mu_i + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma) \]

\[ \text{DATA} = \text{FIT} + \text{ERROR} \]

Relationship to Regression:

- FIT gives mean for each level of categorical variable, but the means need not fall in a pattern as in regression.
- ANOVA can be thought of as regression with a categorical predictor.
The ANOVA model

\[ \hat{x}_{ij} = \mu_i + \epsilon_{ij} \]
\[ \text{DATA} \quad \text{FIT} \quad \text{ERROR} \]

\[ \epsilon_{ij} \sim N(0, \sigma) \]

Relationship to Regression:

- FIT gives mean for each level of categorical variable, but the means need not fall in a pattern as in regression.
- ANOVA can be thought of as regression with a categorical predictor.

Assumptions of this model: Each group . . .

- is normally distributed about its group mean (\( \mu_i \))
- has the same standard deviation
Checking the assumptions

Once again we look at residuals:

\[ \text{residual}_{ij} = e_{ij} = x_{ij} - \bar{x}_i \]

- Normality Check:
  - normal quantile plots of residuals (boxplots, histograms)
  - ANOVA can handle some skewness, but is quite sensitive to outliers

- Equal Variance Check:
  - \( s_i \)'s should be approximately equal
  - Rule of Thumb: \( \leq 2:1 \) ratio between largest and smallest of group sample standard deviations
    - 4:1 ratio between variances is equivalent
  - Remember that the assumptions are about the population, not the sample.
    - These are hard to check when the data sets are small.
Checking the assumptions in the blister example

- Equal Variances Check:

<table>
<thead>
<tr>
<th>treatment</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>days</td>
<td>A</td>
<td>8</td>
<td>7.250</td>
<td>7.000</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>8</td>
<td>8.875</td>
<td>9.000</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>9</td>
<td>10.111</td>
<td>10.000</td>
</tr>
</tbody>
</table>

- largest sd: 1.764; smallest sd : 1.458
- $1.458 \times 2 = 2.916 > 1.764$, so this passes our rule of thumb.

- Overall Normality Check:

![Normal Q–Q Plot for Blister Residuals](image-url)
What does ANOVA do?

At its simplest (there are extensions) ANOVA tests the following hypotheses:

- $H_0$: The means of all the groups are equal.
  
  
  \[ \mu_1 = \mu_2 = \cdots = \mu_I \]

- $H_a$: Not all the means are equal.
  
  - doesn’t say how or which ones differ
  
  - can follow up with “multiple comparisons” if we reject $H_0$
A quick look at the ANOVA table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F-stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>2</td>
<td>34.74</td>
<td>17.37</td>
<td>6.45</td>
<td>0.0063</td>
</tr>
<tr>
<td>Residuals</td>
<td>22</td>
<td>59.26</td>
<td>2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>94.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What is the conclusion here?
The ANOVA test statistic (ingredients)

ANOVA measures two sources of variation in the data and compares their relative sizes

- **variation BETWEEN** groups
  - for each data value look at the difference between its group mean and the overall mean

\[
SSG = \sum (\bar{x}_i - \bar{x})^2 = \sum (\text{response} - \text{group mean})^2
\]

\[
MSG = \frac{SSG}{k - 1}
\]

- **variation WITHIN** groups
  - for each data value we look at the difference between that value and the mean of its group

\[
SSE = \sum (x_{ij} - \bar{x}_i)^2 = \sum (\text{residual})^2
\]

\[
MSE = \frac{SSE}{n - k}
\]
The ANOVA test statistic

The ANOVA F-statistic is a ratio of the Between Group Variation (explained by FIT) divided by the Within Group Variation (Residuals):

\[ F = \frac{\text{Between}}{\text{Within}} = \frac{MSG}{MSE} \]

- A large value of \( F \) is evidence against \( H_0 \), since it indicates that there is more difference between groups than within groups.

- The \( F \) distribution has two degrees of freedom parameters
  - numerator: \( k - 1 \) (one less than number of groups)
  - denominator: \( n - k \) (observations minus number of groups)
The ANOVA test statistic (example)

We want to measure and compare the amount of variation due to **BETWEEN** group variation and **WITHIN** group variation:

So for each data value, we calculate its contribution to:

- **BETWEEN** group variation: \((\bar{x}_i - \bar{x})^2\)
- **WITHIN** group variation: \((x_{ij} - \bar{x}_i)^2\)

**Example (An Even Smaller Dataset)**

Suppose we have three groups

- Group 1: 5.3, 6.0, 6.7 \[\bar{x}_1 = 6.00\]
- Group 2: 5.5, 6.2, 6.4, 5.7 \[\bar{x}_2 = 5.95\]
- Group 3: 7.5, 7.2, 7.9 \[\bar{x}_3 = 7.53\]
Computing $F$

<table>
<thead>
<tr>
<th>data</th>
<th>group</th>
<th>mean</th>
<th>WITHIN</th>
<th>BETWEEN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>difference:</td>
<td>group mean - overall mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>data - group mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>plain</td>
<td>squared</td>
</tr>
<tr>
<td>5.3</td>
<td>1</td>
<td>6.00</td>
<td>-0.70</td>
<td>0.490</td>
</tr>
<tr>
<td>6.0</td>
<td>1</td>
<td>6.00</td>
<td>0.00</td>
<td>0.000</td>
</tr>
<tr>
<td>6.7</td>
<td>1</td>
<td>6.00</td>
<td>0.70</td>
<td>0.490</td>
</tr>
<tr>
<td>5.5</td>
<td>2</td>
<td>5.95</td>
<td>-0.45</td>
<td>0.203</td>
</tr>
<tr>
<td>6.2</td>
<td>2</td>
<td>5.95</td>
<td>0.25</td>
<td>0.063</td>
</tr>
<tr>
<td>6.4</td>
<td>2</td>
<td>5.95</td>
<td>0.45</td>
<td>0.203</td>
</tr>
<tr>
<td>5.7</td>
<td>2</td>
<td>5.95</td>
<td>-0.25</td>
<td>0.063</td>
</tr>
<tr>
<td>7.5</td>
<td>3</td>
<td>7.53</td>
<td>-0.03</td>
<td>0.001</td>
</tr>
<tr>
<td>7.2</td>
<td>3</td>
<td>7.53</td>
<td>-0.33</td>
<td>0.109</td>
</tr>
<tr>
<td>7.9</td>
<td>3</td>
<td>7.53</td>
<td>0.37</td>
<td>0.137</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td>1.757</td>
<td></td>
</tr>
<tr>
<td>TOTAL/df</td>
<td></td>
<td></td>
<td>0.25095714</td>
<td></td>
</tr>
</tbody>
</table>

overall mean: 6.44; $F = \frac{2.5528}{0.25025} = 10.21575$
The ANOVA table revisited

Back to our blister example:

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F-stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>2</td>
<td>34.74</td>
<td>17.37</td>
<td>6.45</td>
<td>0.0063</td>
</tr>
<tr>
<td>Residuals</td>
<td>22</td>
<td>59.26</td>
<td>2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>94.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The ANOVA table revisited

Back to our blister example:

<table>
<thead>
<tr>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F-stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>2</td>
<td>34.74</td>
<td>17.37</td>
<td>6.45</td>
</tr>
<tr>
<td>Residuals</td>
<td>22</td>
<td>59.26</td>
<td>2.69</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>94.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Degrees of Freedom:

- $2 = k - 1 = 1$ less than $\#$ of groups
- $24 = N - 1 = 1$ less than $\#$ of observations
- $22 = N - k = 25 - 3$
  - also equals $24 - 2$
  - also equals sum of degrees of freedom for each group:

$$ (n_1 - 1) + (n_2 - 1) + (n_3 - 1) = (8 - 1) + (8 - 1) + (9 - 1) $$
The ANOVA table revisited

Back to our blister example:

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F-stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>2</td>
<td>34.74</td>
<td>17.37</td>
<td>6.45</td>
<td>0.0063</td>
</tr>
<tr>
<td>Residuals</td>
<td>22</td>
<td>59.26</td>
<td>2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>94.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sum of Squares (SS):

- $34.74 = SSG = \sum_{obs} (\bar{x}_i - \bar{x})^2$
- $59.26 = SSE = \sum_{obs} (x_{ij} - \bar{x}_i)^2$
- $34.74 + 59.26 = 94.0 = SST = \sum_{obs} (x_{ij} - \bar{x})^2$
The ANOVA table revisited

Back to our blister example:

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F-stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>2</td>
<td>34.74</td>
<td>17.37</td>
<td>6.45</td>
<td>0.0063</td>
</tr>
<tr>
<td>Residuals</td>
<td>22</td>
<td>59.26</td>
<td>2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>94.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean Squares (MS):
- $MS = SS/DF$
The ANOVA table revisited

Back to our blister example:

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F-stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>2</td>
<td>34.74</td>
<td>17.37</td>
<td>6.45</td>
<td>0.0063</td>
</tr>
<tr>
<td>Residuals</td>
<td>22</td>
<td>59.26</td>
<td>2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>94.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean Squares (MS):

- $MS = SS/DF$

$F = MSG/MSE$

- $DFG = I - 1 = 2$ degrees of freedom for numerator
- $DFE = n - I = 22$ degrees of freedom for denominator
Connections between SST, MST, and standard deviation

Ignoring groups for a moment:

\[
SST = \frac{\sum(x_{ij} - \bar{x})^2}{n - 1}
\]

\[
= \frac{SST}{DFT} = MST
\]

So \( MST = s_x^2 \) and \( SST = (n - 1)s_x^2 \).

These measure the TOTAL variation (of \( x \)).
Connections between SSE, MSE, and standard deviation

If we just look at the data from one of the groups:

\[ s_i^2 = \frac{\sum_j (x_{ij} - \bar{x}_i)^2}{n_i - 1} = \frac{SS[\text{within group } i]}{DF_i} \]

So \( SS[\text{within Group } i] = (s_i^2)(DF_i) \)

This means

\[ SSE = SS[\text{within}] = \sum_i SS[\text{within group } i] \]
\[ = \sum_i s_i^2 (n_i - 1) \]
\[ = \sum_i s_i^2 \cdot DF_i \]
Pooled estimate for $\sigma$

One of the ANOVA assumptions is that all groups have the same standard deviation. We can estimate this with a weighted average:

$$s_p^2 = \frac{DF_1 \cdot s_1^2 + DF_2 \cdot s_2^2 + \cdots + DF_I \cdot s_I^2}{DF_1 + DF_2 + \cdots + DF_I}$$

$$s_p^2 = \frac{SSE}{DFE} = MSE$$

So MSE is the pooled estimate of variance parameter $\sigma^2$.

- This is a direct generalization of our pooled estimate from two-sample $t$ tests.
- Also just like it was for linear regression
For both regression and ANOVA, $R^2$ can be used as a measure of the portion of variation explained by the model.

In fact, if we let $\hat{x} = \bar{x}_i$ be the group mean (because our model predicts a mean for each group), we see that they are really the same thing:

For ANOVA:

$$R^2 = \frac{SSG}{SST} = \frac{\sum (\bar{x}_i - \bar{x})^2}{\sum (x_{ij} - \bar{x})^2} = \frac{\sum (\hat{x} - \bar{x})^2}{\sum (x - \bar{x})^2}$$

For regression:

$$R^2 = \frac{SSM}{SST} = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y - \bar{y})^2} = \frac{\sum (\text{residual})^2}{\sum (y - \bar{y})^2}$$
Where’s the Difference?

Once ANOVA indicates that the groups do not all have the same means, we can compare them two by two using the 2-sample $t$ test.

- This is $\binom{I}{2} = \frac{I(I-1)}{2} \approx \frac{l^2}{2}$ different tests.
- We need to adjust our p-value threshold because we are doing multiple tests with the same data.
- There are several methods for doing this.
- If we really just want to test the difference between one pair of treatments, we should set the study up that way.
Tukey’s Pairwise Comparisons

One method of adjusting the simultaneous confidence intervals for all the pairs is called Tukey’s method.

Tukey 95% Simultaneous Confidence Intervals

<table>
<thead>
<tr>
<th></th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>-0.4365</td>
<td>3.6865</td>
</tr>
<tr>
<td>P</td>
<td>0.8577</td>
<td>4.8645</td>
</tr>
</tbody>
</table>

- This corresponds to 98.01% CIs for each pairwise difference.
- Only P vs A is significant in our example (lower and upper limits have the same sign).
Other views of Multiple Comparisons

Tukey multiple comparisons of means

<table>
<thead>
<tr>
<th></th>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-A</td>
<td>1.6250</td>
<td>-0.43650</td>
<td>3.6865</td>
</tr>
<tr>
<td>P-A</td>
<td>2.8611</td>
<td>0.85769</td>
<td>4.8645</td>
</tr>
<tr>
<td>P-B</td>
<td>1.2361</td>
<td>-0.76731</td>
<td>3.2395</td>
</tr>
</tbody>
</table>

Individual 95% CIs For Mean
Based on Pooled StDev

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8</td>
<td>7.250</td>
<td>1.669</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>8.875</td>
<td>1.458</td>
</tr>
<tr>
<td>P</td>
<td>9</td>
<td>10.111</td>
<td>1.764</td>
</tr>
</tbody>
</table>

Pooled StDev = 1.641

95% family-wise confidence level

Differences in mean levels of treatment

---

---
Another example

Example

Coagulation Experiment In a 1978 study of blood coagulation, 24 animals were randomly assigned one of four different diets. Blood coagulation times for the 24 animals were then measured to see if diet has an effect on coagulation time.

<table>
<thead>
<tr>
<th>group</th>
<th>count</th>
<th>mean</th>
<th>st dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>61</td>
<td>1.8257</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>66</td>
<td>2.8284</td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>68</td>
<td>1.6733</td>
</tr>
<tr>
<td>D</td>
<td>8</td>
<td>61</td>
<td>2.6186</td>
</tr>
</tbody>
</table>
### ANOVA for coagulation example

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>diet</td>
<td>3</td>
<td>228.00</td>
<td>76.00</td>
<td>13.57</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residuals</td>
<td>20</td>
<td>112.00</td>
<td>5.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tukey multiple comparisons of means (95% family-wise confidence level)

<table>
<thead>
<tr>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-A</td>
<td>5.72455</td>
<td>9.2754</td>
</tr>
<tr>
<td>C-A</td>
<td>2.72455</td>
<td>11.2754</td>
</tr>
<tr>
<td>D-A</td>
<td>4.05604</td>
<td>4.0560</td>
</tr>
<tr>
<td>C-B</td>
<td>1.82407</td>
<td>5.8241</td>
</tr>
<tr>
<td>D-B</td>
<td>8.57709</td>
<td>1.4229</td>
</tr>
<tr>
<td>D-C</td>
<td>10.57709</td>
<td>3.4229</td>
</tr>
</tbody>
</table>

95% family-wise confidence level

Differences in mean levels of diet