Comparing Several Groups
Do the group means differ?

Naive approach
Independent-samples t-tests of all pairs
Each test doesn't use all data → Less power
10 total tests → Greater chance of Type I error

Analysis of Variance (ANOVA)
Single test for any group differences
Null hypothesis: All means are equal
Works using variance of the sample means
Also based on separating explained and unexplained variance

Variance of Sample Means
Say we have \( k \) groups of \( n \) subjects each
Want to test whether \( \mu_i \) are all equal for \( i = 1, \ldots, k \)
Even if all population means are equal, sample means will vary

Central limit theorem tells how much: \( \sigma^2_M = \frac{\sigma^2}{n} \)

Compare actual variance of sample means to amount expected by chance

\[
\frac{\text{var}(M)}{\sigma^2/n} \quad \text{or} \quad \frac{n \cdot \text{var}(M)}{\sigma^2}
\]

Estimate \( \sigma^2 \) using \( \text{MS}_{\text{residual}} \)

Test statistic: \( F = \frac{n \cdot \text{var}(M)}{\text{MS}_{\text{residual}}} \)

If \( F \) is large enough
Sample means vary more than expected by chance
Reject null hypothesis

Residual Mean Square
Best estimate of population variance, \( \sigma^2 \)
Based on remaining variability after removing group differences
Extends MS for independent-samples t-test

For \( 2 \) groups:
\[
\text{MS}_{\text{residual}} = \frac{\sum_{\text{group } 1} (X_1 - M_1)^2 + \sum_{\text{group } 2} (X_2 - M_2)^2}{n_1 + n_2 - 2}
\]

For \( k \) groups:
\[
\text{MS}_{\text{residual}} = \frac{\sum_{\text{group } 1} (X_1 - M_1)^2 + \ldots + \sum_{\text{group } k} (X_k - M_k)^2}{n_1 + \ldots + n_k - k}
\]
Example: Lab Sections

Group means: \( M = [76.6 \ 82.5 \ 83.7 \ 79.5 \ 82.9] \)

Variance of group means:

\[
\text{mean}(M) = \frac{\sum_i M_i}{5} = 81.0
\]

\[
\text{var}(M) = \frac{\sum_i (M_i - 81.0)^2}{5 - 1} = 8.69
\]

\[\text{var}(M) \text{ expected by chance: } \frac{22}{22}\]

\[\text{var}(M) \text{ vs. } \frac{22}{22}\]

Instead: \( 22 \cdot \text{var}(M) \text{ vs. } \frac{22}{22}\)

191.1 vs. \( \frac{22}{22}\)

Estimate of \( \sigma^2\):

\[
\text{MS} = \frac{\sum \left( X - M_1 \right)^2 + \sum \left( X - M_2 \right)^2 + \sum \left( X - M_3 \right)^2 + \sum \left( X - M_4 \right)^2 + \sum \left( X - M_5 \right)^2}{110 - 5} = 199.1
\]

Ratio of between-group variance to amount expected by chance: \( F_{4,105} = \frac{191.1}{199.1} = .96\)

\( F_{\text{crit}} = 2.46\)

\( p = .44\)

Partitioning Sum of Squares

General approach to ANOVA uses sums of squares

Works with unequal sample sizes and other complications

Breaks total variability into explained and unexplained parts

\( SS_{\text{total}}\)

Sum of squares for all data, treated as a single sample

Based on differences from grand mean, \( \overline{M} \)

\[
SS_{\text{total}} = \sum_{\text{all groups}} \left( X - \overline{M} \right)^2
\]

\( SS_{\text{residual}} \) (\( SS_{\text{within}} \))

Variability within groups, just like with independent-samples t-test

Not explainable by group differences

\[
SS_{\text{residual}} = \sum_{i=1 \text{ to } k} \left( \sum_{\text{group } i} \left( X_i - M_i \right)^2 \right) = \sum_{\text{group } 1} \left( X_1 - M_1 \right)^2 + \ldots + \sum_{\text{group } k} \left( X_k - M_k \right)^2
\]

\( SS_{\text{treatment}} \) (\( SS_{\text{between}} \))

Variability explainable by difference among groups

Treat each datum as its group mean and take difference from grand mean

\[
SS_{\text{treatment}} = \sum_{i=1 \text{ to } k} \left( n_i \cdot \left( M_i - \overline{M} \right)^2 \right)
\]
Magic of Squares

\[ SS_{\text{total}} = SS_{\text{treatment}} + SS_{\text{residual}} \]

\[ SS_{\text{total}} = \sum (X - \bar{M})^2 \]

\[ SS_{\text{residual}} = \sum (\sum (X_i - M_i))^2 \]

\[ SS_{\text{treatment}} = \sum (n_i (M_i - \bar{M})^2) \]

Test Statistic for ANOVA

Goal: Determine whether group differences account for more variability than expected by chance

Same approach as with regression

Calculate Mean Square: \[ MS_{\text{treatment}} = \frac{SS_{\text{treatment}}}{df_{\text{treatment}}} \]

\[ H_0: MS_{\text{treatment}} \approx \sigma^2 \]

Estimate \( \sigma^2 \) using \[ MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} \]

Test statistic: \[ F = \frac{MS_{\text{treatment}}}{MS_{\text{residual}}} \]

Lab Sections Revisited

Two approaches

Variance of section means
Mean Square for treatment (lab section)

Variance of group means

\[ \text{Var}(76.6, 82.5, 83.7, 79.5, 82.9) = 8.69 \]

Expected by chance: \( \sigma^2 / 22 \)

\[ \text{Var}(M) \text{ vs. } \sigma^2 / 22, \text{ or } 22 \cdot \text{var}(M) \text{ vs. } \sigma^2 \]

Mean Square for treatment

\[ MS_{\text{treatment}} = \frac{\sum_i n_i (M_i - \bar{M})^2}{5 - 1} \]

\[ = \frac{22 \cdot (76.6 - 81.0)^2 + 22 \cdot (82.5 - 81.0)^2 + 22 \cdot (83.7 - 81.0)^2 + 22 \cdot (79.5 - 81.0)^2 + 22 \cdot (82.9 - 81.0)^2}{4} \]

\[ = 191.1 \]

Because equal group sizes:

\[ n_i = 22 \]

\[ \bar{M} = \text{mean}(M) \]

\[ MS_{\text{treatment}} = 22 \cdot \text{var}(M) \]

\[ F = \frac{191.1}{MS_{\text{residual}}} = \frac{191.1}{199.1} = .96 \]
Two Views of ANOVA

Simple: Equal sample sizes
\[ F = \frac{\text{var}(M)}{\frac{\sigma^2}{n}} = \frac{n \cdot \text{var}(M)}{\sigma^2} = \frac{n \cdot \text{var}(M)}{\text{MS}_{\text{residual}}} \]

General: Using sums of squares
\[ F = \frac{\text{MS}_{\text{treatment}}}{\text{MS}_{\text{residual}}} \]
\[ \text{MS}_{\text{treatment}} = \frac{SS_{\text{treatment}}}{df_{\text{treatment}}} = \frac{\sum_i n_i (M_i - \bar{M})^2}{k-1} \]

If sample sizes equal: \[ \text{MS}_{\text{treatment}} = \frac{n \cdot \sum_i (M_i - \bar{M})^2}{k-1} = n \cdot \text{var}(M) \]

Degrees of Freedom
\[ SS_{\text{total}} = \sum (X - \bar{M})^2 \]
\[ df_{\text{total}} = \sum n_i - 1 \]
\[ SS_{\text{treatment}} = SS(M) \]
\[ df_{\text{treatment}} = k - 1 \]
\[ SS_{\text{residual}} = \sum_i SS(X_i) \]
\[ df_{\text{residual}} = \sum (n_i - 1) = \sum n_i - k \]