Univariate Analysis of Variance (ANOVA)
General ANOVA Setting

- Investigator controls one or more factors of interest
  - Each factor contains two or more levels
  - Levels can be numerical or categorical
  - Different levels produce different groups
  - Think of each group as a sample from a different population

- Observe effects on the dependent variable
  - Are the groups the same?
One-Way Analysis of Variance

- Evaluate the difference among the means of three or more groups

**Examples:** Number of accidents for 1\textsuperscript{st}, 2\textsuperscript{nd}, and 3\textsuperscript{rd} shift
Expected mileage for five brands of tires

- Assumptions
  - Populations are normally distributed
  - Populations have equal variances
  - Samples are randomly and independently drawn
Hypotheses of One-Way ANOVA

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$
  - All population means are equal
  - i.e., no factor effect (no variation in means among groups)

- $H_1 : \text{Not all of the population means are equal}$
  - At least one population mean is different
  - i.e., there is a factor effect
  - Does not mean that all population means are different (some pairs may be the same)
One-Way ANOVA

$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$

$H_1 : \text{Not all } \mu_j \text{ are equal}$

The Null Hypothesis is True
All Means are the same:
(No Factor Effect)

$\mu_1 = \mu_2 = \mu_3$
One-Way ANOVA

H₀: μ₁ = μ₂ = μ₃ = … = μₖ

H₁: Not all μⱼ are equal

The Null Hypothesis is NOT true
At least one of the means is different
(Factor Effect is present)

or

μ₁ = μ₂ ≠ μ₃

μ₁ ≠ μ₂ ≠ μ₃
Total variation can be split into two parts:

\[ \text{SST} = \text{SSB} + \text{SSW} \]

\( \text{SST} = \) Total Sum of Squares (Total variation)
\( \text{SSB} = \) Sum of Squares Between Groups (Between-group variation)
\( \text{SSW} = \) Sum of Squares Within Groups (Within-group variation)
Partition of Total Variation

Total Variation (SST) = Variation Due to Factor (SSA) + Variation Due to Random Error (SSW)

SSA = variation explained by different levels of the factor
SSW = variation not explained by the factor
Total Sum of Squares

\[ \text{SST} = \text{SSB} + \text{SSW} \]

Where:

\[ \text{SST} = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2 \]

SST = Total sum of squares

c = number of groups or levels

\( n_j \) = number of observations in group \( j \)

\( X_{ij} \) = \( i^{th} \) observation from group \( j \)

\( \bar{X} \) = grand mean (mean of all data values)
Total Variation

\[
SST = (X_{11} - \bar{X})^2 + (X_{12} - \bar{X})^2 + \cdots + (X_{cn_c} - \bar{X})^2
\]
Between-Group Variation

$$SSB = \sum_{j=1}^{c} n_j \left( X_j - \bar{X} \right)^2$$

Where:

- $SSB$ = Sum of squares between groups
- $c$ = number of groups
- $n_j$ = sample size from group $j$
- $\bar{X}_j$ = sample mean from group $j$
- $\bar{X}$ = grand mean (mean of all data values)

$$SST = SSB + SSW$$
Between-Group Variation

(continued)

\[ SSB = \sum_{j=1}^{c} n_j (\bar{X}_j - \bar{X})^2 \]

Variation Due to Differences Between Groups

\[ MSB = \frac{SSB}{c - 1} \]

Mean Square Among = SSA/degrees of freedom
Between-Group Variation

\[ SSB = n_1 (\bar{X}_1 - \bar{X})^2 + n_2 (\bar{X}_2 - \bar{X})^2 + \cdots + n_c (\bar{X}_c - \bar{X})^2 \]
Within-Group Variation

\[ \text{SST} = \text{SSB} + \text{SSW} \]

\[ \text{SSW} = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 \]

Where:

- \( \text{SSW} \) = Sum of squares within groups
- \( c \) = number of groups
- \( n_j \) = sample size from group \( j \)
- \( \bar{X}_j \) = sample mean from group \( j \)
- \( X_{ij} \) = \( i^{th} \) observation in group \( j \)
Within-Group Variation

SSW = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2

Summing the variation within each group and then adding over all groups

Mean Square Within = \frac{SSW}{n - c}

\mu_j
Within-Group Variation

\[
SSW = (X_{11} - \bar{X}_1)^2 + (X_{12} - \bar{X}_2)^2 + \cdots + (X_{cn_c} - \bar{X}_c)^2
\]

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

(continued)
Obtaining the Mean Squares

The Mean Squares are obtained by dividing the various sum of squares by their associated degrees of freedom.

\[
MSB = \frac{SSB}{c - 1}
\]

Mean Square Between (d.f. = c-1)

\[
MSW = \frac{SSW}{n - c}
\]

Mean Square Within (d.f. = n-c)

\[
MST = \frac{SST}{n - 1}
\]

Mean Square Total (d.f. = n-1)
One-Way ANOVA Table

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum Of Squares</th>
<th>Mean Square (Variance)</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>c - 1</td>
<td>SSB</td>
<td>MSB = \frac{SSB}{c - 1}</td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td>n - c</td>
<td>SSW</td>
<td>MSW = \frac{SSW}{n - c}</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>n - 1</td>
<td>SST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F_{STAT} = \frac{MSB}{MSW}

c = number of groups
n = sum of the sample sizes from all groups
df = degrees of freedom
One-Way ANOVA
F Test Statistic

H₀: \( \mu_1 = \mu_2 = \cdots = \mu_c \)

H₁: At least two population means are different

Test statistic

\[ F_{STAT} = \frac{MSB}{MSW} \]

MSB is mean squares between groups
MSW is mean squares within groups

Degrees of freedom

- \( df_1 = c - 1 \)  (c = number of groups)
- \( df_2 = n - c \)  (n = sum of sample sizes from all populations)
Interpreting One-Way ANOVA

F Statistic

- The F statistic is the ratio of the among estimate of variance and the within estimate of variance.
  - The ratio must always be positive.
  - \( df_1 = c - 1 \) will typically be small.
  - \( df_2 = n - c \) will typically be large.

Decision Rule:
- Reject \( H_0 \) if \( F_{\text{STAT}} > F_\alpha \), otherwise do not reject \( H_0 \).
One-Way ANOVA F Test Example

You want to see if three different golf clubs yield different distances. You randomly select five measurements from trials on an automated driving machine for each club. At the 0.05 significance level, is there a difference in mean distance?

<table>
<thead>
<tr>
<th>Club 1</th>
<th>Club 2</th>
<th>Club 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>254</td>
<td>234</td>
<td>200</td>
</tr>
<tr>
<td>263</td>
<td>218</td>
<td>222</td>
</tr>
<tr>
<td>241</td>
<td>235</td>
<td>197</td>
</tr>
<tr>
<td>237</td>
<td>227</td>
<td>206</td>
</tr>
<tr>
<td>251</td>
<td>216</td>
<td>204</td>
</tr>
</tbody>
</table>
One-Way ANOVA Example: Scatter Plot

<table>
<thead>
<tr>
<th>Club 1</th>
<th>Club 2</th>
<th>Club 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>254</td>
<td>234</td>
<td>200</td>
</tr>
<tr>
<td>263</td>
<td>218</td>
<td>222</td>
</tr>
<tr>
<td>241</td>
<td>235</td>
<td>197</td>
</tr>
<tr>
<td>237</td>
<td>227</td>
<td>206</td>
</tr>
<tr>
<td>251</td>
<td>216</td>
<td>204</td>
</tr>
</tbody>
</table>

\[
\bar{x}_1 = 249.2 \quad \bar{x}_2 = 226.0 \quad \bar{x}_3 = 205.8
\]

\[
\bar{x} = 227.0
\]
### One-Way ANOVA Example Computations

<table>
<thead>
<tr>
<th>Club 1</th>
<th>Club 2</th>
<th>Club 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>254</td>
<td>234</td>
<td>200</td>
</tr>
<tr>
<td>263</td>
<td>218</td>
<td>222</td>
</tr>
<tr>
<td>241</td>
<td>235</td>
<td>197</td>
</tr>
<tr>
<td>237</td>
<td>227</td>
<td>206</td>
</tr>
<tr>
<td>251</td>
<td>216</td>
<td>204</td>
</tr>
</tbody>
</table>

- $\bar{X}_1 = 249.2$, $n_1 = 5$
- $\bar{X}_2 = 226.0$, $n_2 = 5$
- $\bar{X}_3 = 205.8$, $n_3 = 5$
- $\bar{X} = 227.0$, $n = 15$
- $c = 3$

---

SSB = 5 \((249.2 - 227)^2\) + 5 \((226 - 227)^2\) + 5 \((205.8 - 227)^2\) = 4716.4

SSW = (254 - 249.2)$^2$ + (263 - 249.2)$^2$ + ... + (204 - 205.8)$^2$ = 1119.6

MSB = $\frac{4716.4}{(3-1)} = 2358.2$

MSW = $\frac{1119.6}{(15-3)} = 93.3$

$F_{STAT} = \frac{2358.2}{93.3} = 25.275$
One-Way ANOVA Example Solution

H₀: μ₁ = μ₂ = μ₃
H₁: μⱼ not all equal

α = 0.05

df₁ = 2  \hspace{0.2cm} df₂ = 12

\[ \text{Test Statistic:} \quad F_{\text{STAT}} = \frac{MSB}{MSW} = \frac{2358.2}{93.3} = 25.275 \]

\[ F_{\alpha} = 3.89 \]

Decision: Reject H₀ at α = 0.05

Conclusion: There is evidence that at least one μⱼ differs from the rest