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 1st, 2nd, and 3rd 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

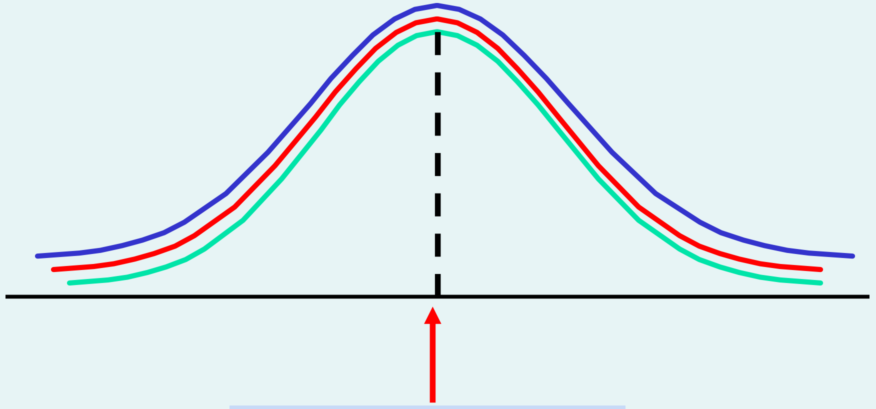
- $H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$
  - All population means are equal
  - i.e., no factor effect (no variation in means among groups)
- $H_1 : \text{Not all of the population means are equal}$
  - At least one population mean is different
  - i.e., there is a factor effect
  - Does not 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

(continued)

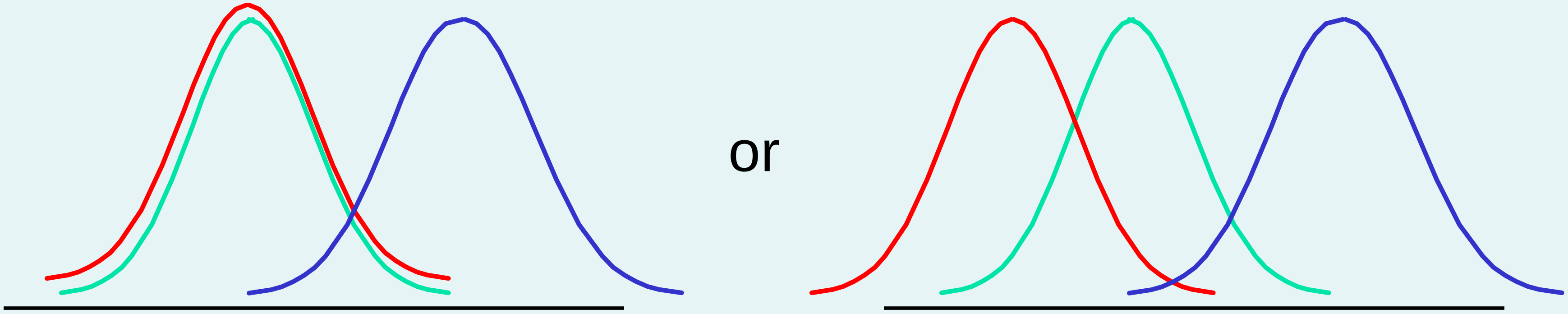
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$$

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

The Null Hypothesis is NOT true

At least one of the means is different

(Factor Effect is present)



$$\mu_1 = \mu_2 \neq \mu_3 \qquad \mu_1 \neq \mu_2 \neq \mu_3$$

Method: Partitioning the Variation

- Total variation can be split into two parts:

$$\boxed{SST = SSB + SSW}$$

SST = Total Sum of Squares
(Total variation)
SSB = Sum of Squares Between Groups
(Between-group variation)
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

$$\boxed{SST} = SSB + SSW$$

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

Where:

SST = Total sum of squares

c = number of groups or levels

n_j = number of observations in group j

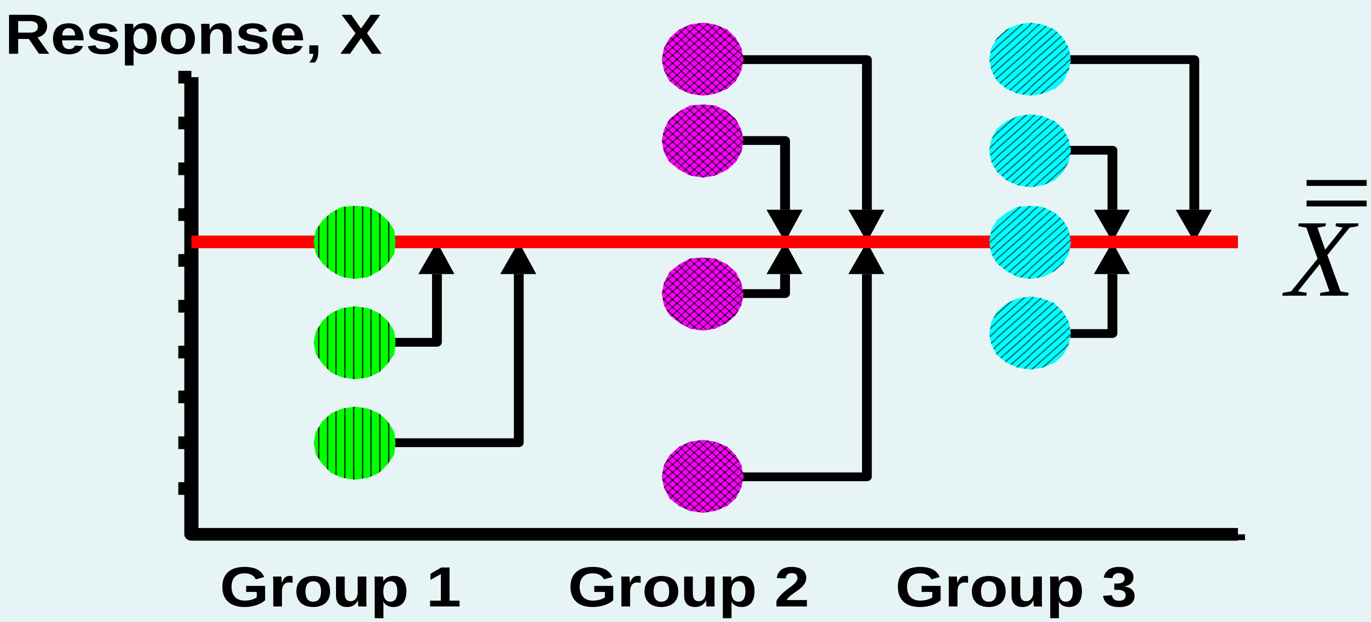
X_{ij} = i^{th} observation from group j

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

Total Variation

(continued)

$$SST = (X_{11} - \bar{\bar{X}})^2 + (X_{12} - \bar{\bar{X}})^2 + \cdots + (X_{cn_c} - \bar{\bar{X}})^2$$



Between-Group Variation

$$SST = \boxed{SSB} + SSW$$

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

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

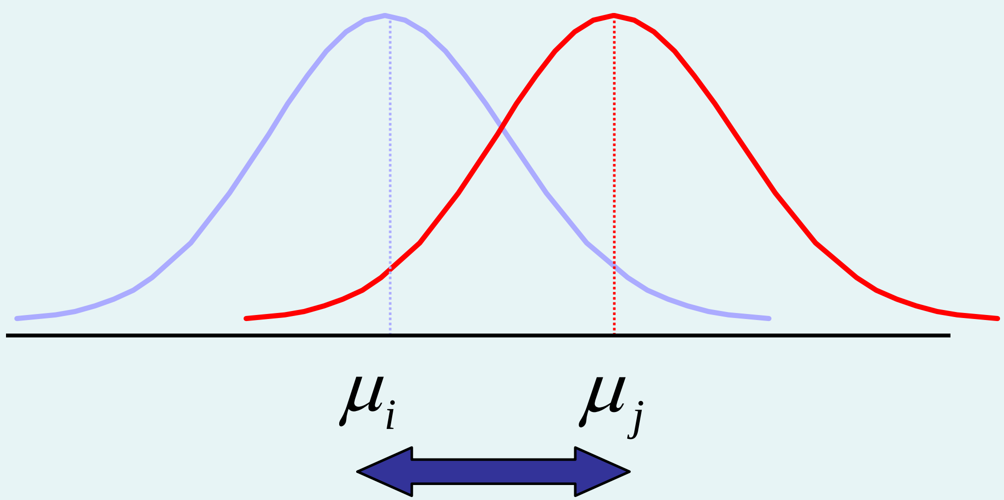
$\bar{\bar{X}}$ = grand mean (mean of all data values)

Between-Group Variation

(continued)

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

Variation Due to
Differences Between Groups



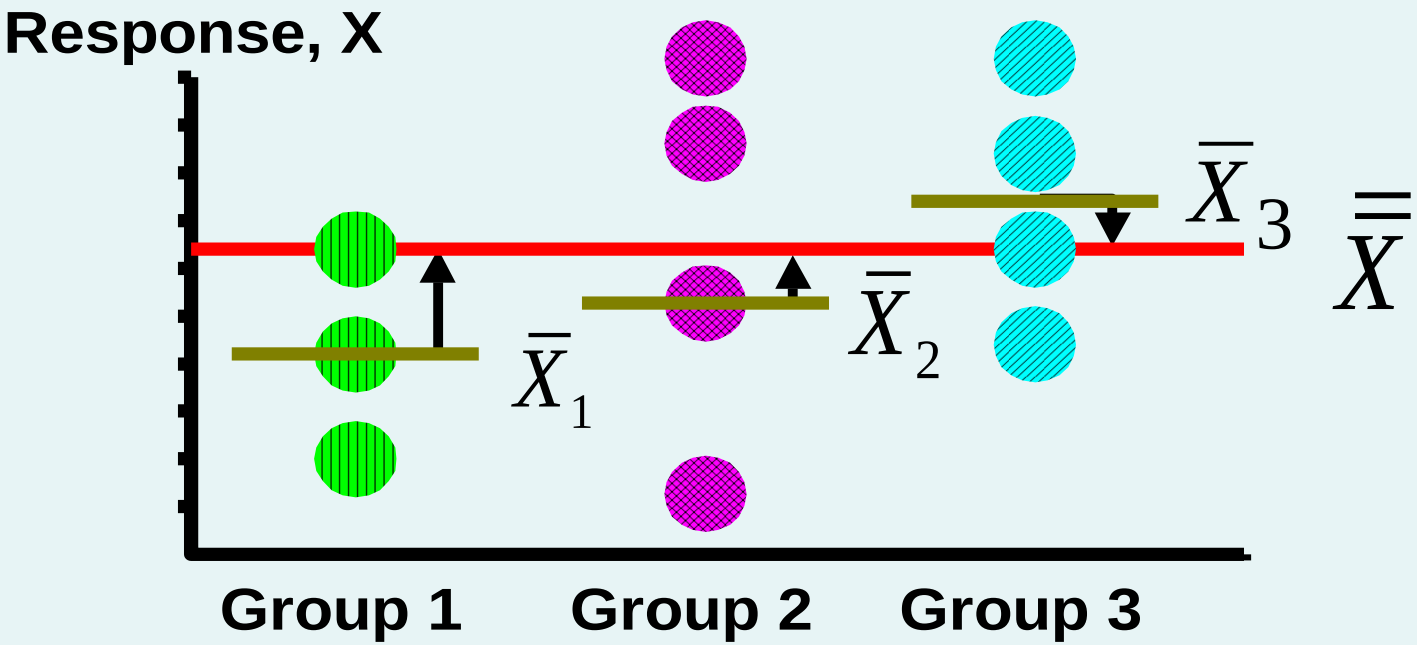
$$\mathrm{MSB} = \frac{\mathrm{SSB}}{\mathrm{c}-1}$$

Mean Square Among =
SSA/degrees of freedom

Between-Group Variation

(continued)

$$SSB = n_1(\bar{X}_1 - \bar{\bar{X}})^2 + n_2(\bar{X}_2 - \bar{\bar{X}})^2 + \cdots + n_c(\bar{X}_c - \bar{\bar{X}})^2$$



Within-Group Variation

$$SST = SSB + \boxed{SSW}$$

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

Where:

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

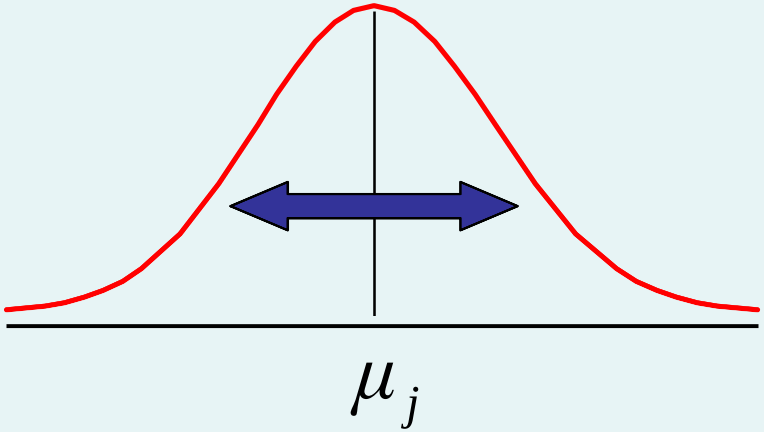
X_{ij} = i^{th} observation in group j

Within-Group Variation

(continued)

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

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



$$MSW = \frac{SSW}{n - c}$$

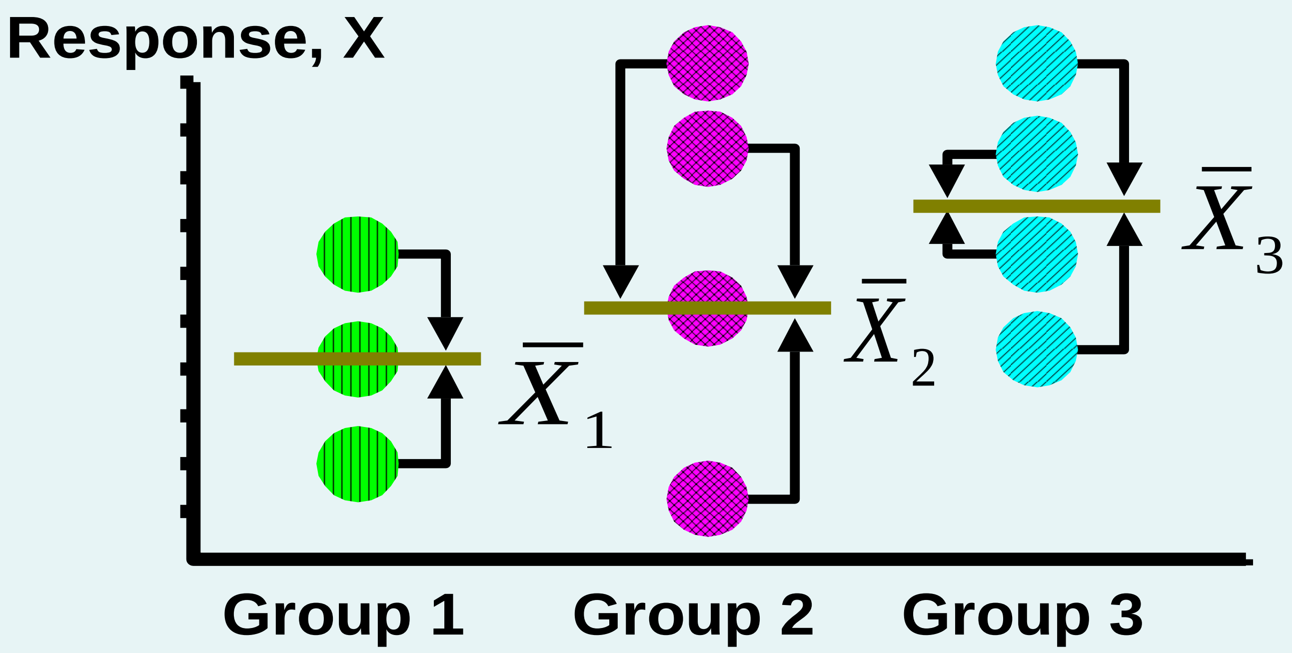
Mean Square Within =

SSW/degrees of freedom

Within-Group Variation

(continued)

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



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)

$$\text{MSW} = \frac{\text{SSW}}{\text{n} - \text{c}}$$

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

$$\text{MST} = \frac{\text{SST}}{\text{n} - 1}$$

Mean Square Total
(d.f. = n-1)

One-Way ANOVA Table

Source of Variation	Degrees of Freedom	Sum Of Squares	Mean Square (Variance)	F
Between Groups	c - 1	SSB	$MSB = \dfrac{SSB}{c - 1}$	$F_{STAT} = \dfrac{MSB}{MSW}$
Within Groups	n - c	SSW	$MSW = \dfrac{SSW}{n - c}$	
Total	n – 1	SST		

c = number of groups
n = sum of the sample sizes from all groups
df = degrees of freedom

One-Way ANOVA F Test Statistic

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

H_1: 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

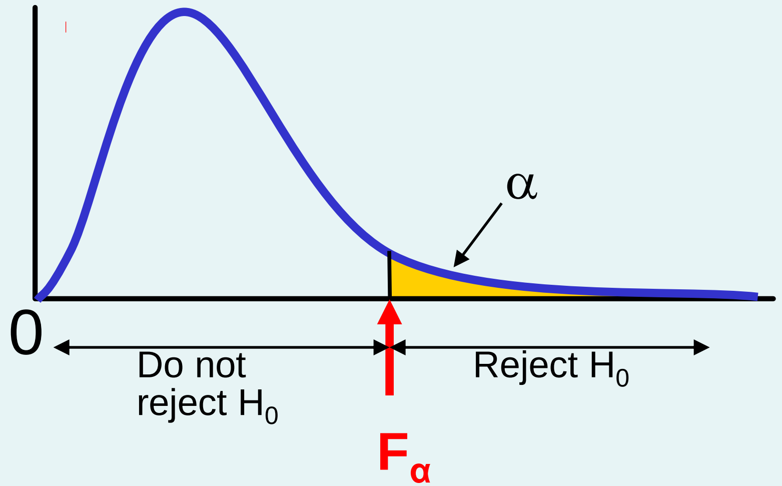
- Degrees of freedom
  - $df_1 = c - 1$ (c = number of groups)
  - $df_2 = n - c$ (n = sum of sample sizes from all populations)

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 $F_{STAT} > F_\alpha$, 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?

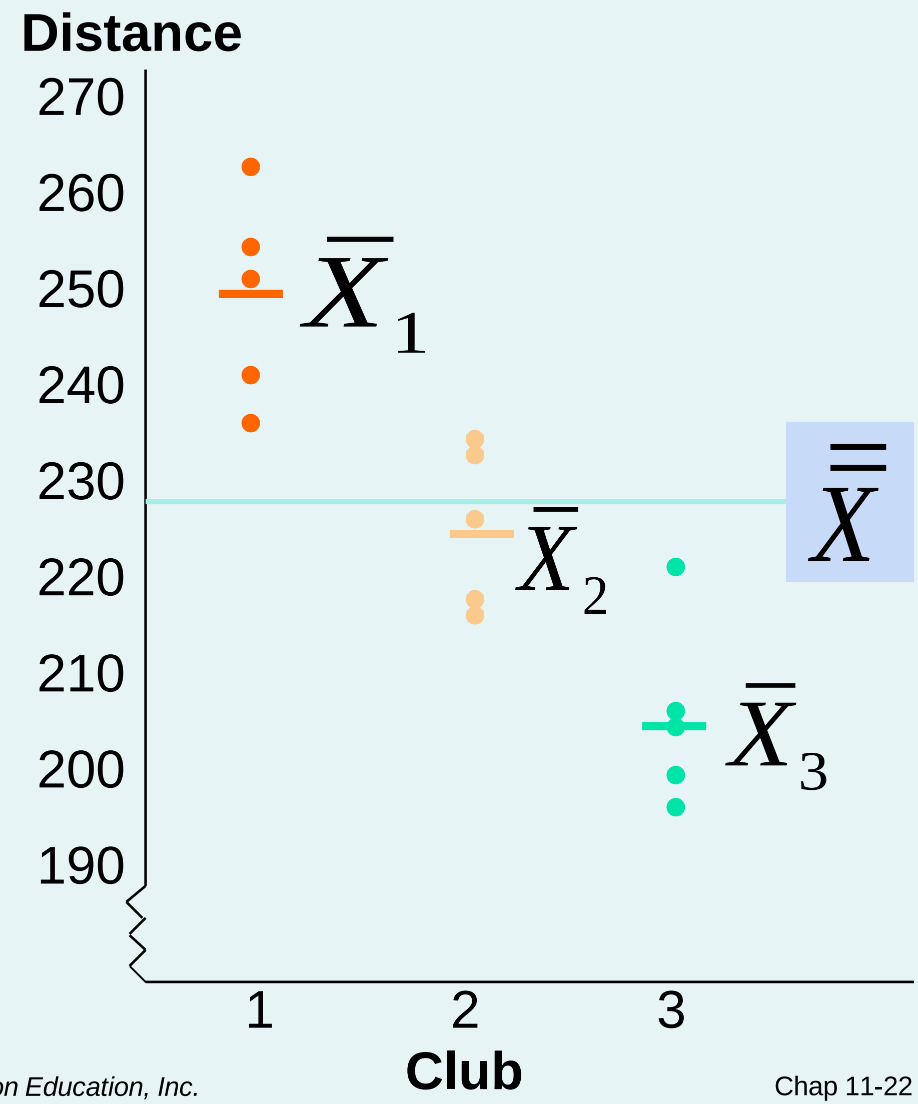
Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

One-Way ANOVA Example: Scatter Plot

Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{x}_1 = 249.2$ $\bar{x}_2 = 226.0$ $\bar{x}_3 = 205.8$

$\bar{\bar{x}} = 227.0$



One-Way ANOVA Example Computations

Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{X}_1 = 249.2$ $\quad n_1 = 5$

$\bar{X}_2 = 226.0$ $\quad n_2 = 5$

$\bar{X}_3 = 205.8$ $\quad n_3 = 5$

$\bar{\bar{X}} = 227.0$ $\quad n = 15$

$c = 3$

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

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

$$MSB = 4716.4 \,/\, (3-1) = 2358.2$$

$$MSW = 1119.6 \,/\, (15-3) = 93.3$$

$$F_{STAT} = \frac{2358.2}{93.3} = 25.275$$

One-Way ANOVA Example Solution

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

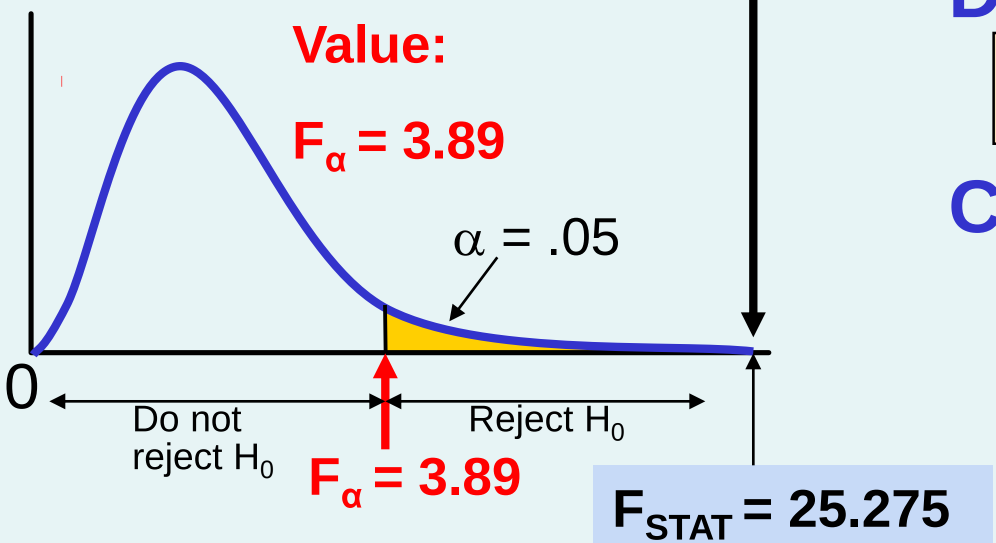
$H_1: \mu_j$ not all equal

$\alpha = 0.05$

$df_1 = 2 \qquad df_2 = 12$

Critical Value:

$F_\alpha = 3.89$



Test Statistic:

$$F_{STAT} = \frac{MSB}{MSW} = \frac{2358.2}{93.3} = 25.275$$

Decision:

Reject H_0 at $\alpha = 0.05$

Conclusion:

There is evidence that at least one μ_j differs from the rest