Analysis of Variance

Learning Objectives

In this chapter, you learn:

- The basic concepts of experimental design
- How to use one-way analysis of variance to test for differences among the means of several groups
- How to use two-way analysis of variance and interpret the interaction effect
- How to perform multiple comparisons in a one-way analysis of variance and a two-way analysis of variance

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?
- Experimental design: the plan used to collect the data

Completely Randomized Design

- Experimental units (subjects) are assigned randomly to groups
  - Subjects are assumed homogeneous
- Only one factor or independent variable
  - With two or more levels
- Analyzed by one-factor analysis of variance (ANOVA)

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)

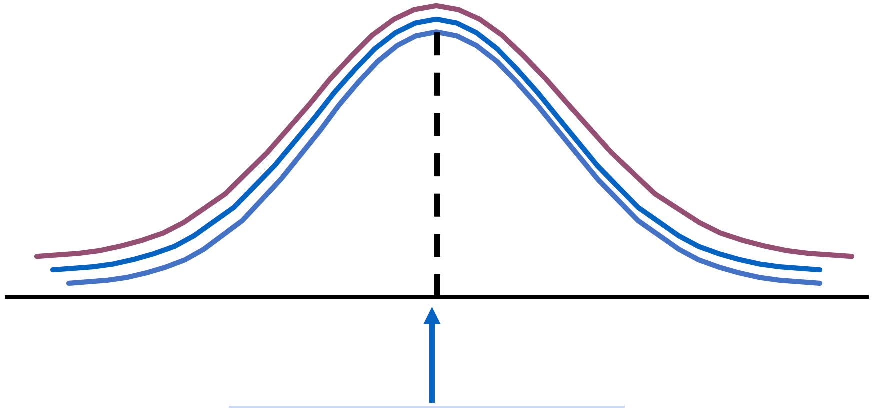
- $H_1 : \text{Not all of the population means are equal}$
  - i.e., there is a factor effect
  - Does not mean that all population means are different (some pairs may be the same)

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)



$$\mu_1 = \mu_2 = \mu_3$$

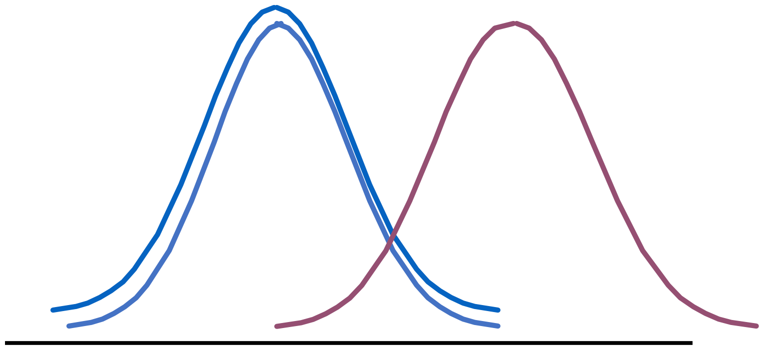
One-Way ANOVA

(continued)

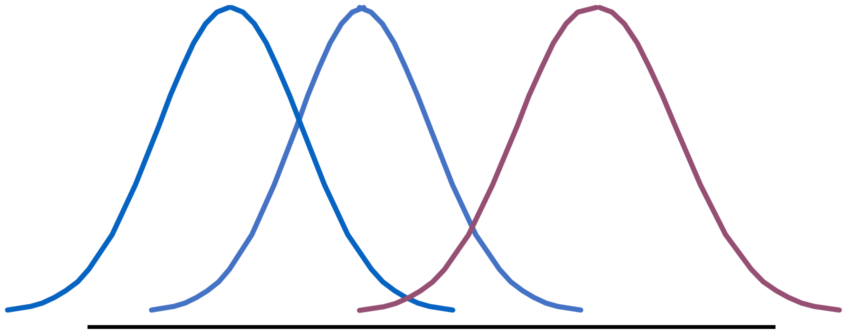
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$$

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

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



or



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

Partitioning the Variation

- Total variation can be split into two parts:

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

SST = Total Sum of Squares
(Total variation)
SSA = Sum of Squares Among Groups
(Among-group variation)
SSW = Sum of Squares Within Groups
(Within-group variation)

Partitioning the Variation

$$SST = SSA + SSW$$

Total Variation = the aggregate variation of the individual data values across the various factor levels (SST)

Among-Group Variation = variation among the factor sample means (SSA)

Within-Group Variation = variation that exists among the data values within a particular factor level (SSW)

Partition of Total Variation

Total Variation (SST)

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

Total Sum of Squares

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

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

Where:

SST = Total sum of squares

c = number of groups or levels

n_j = number of observations in group j

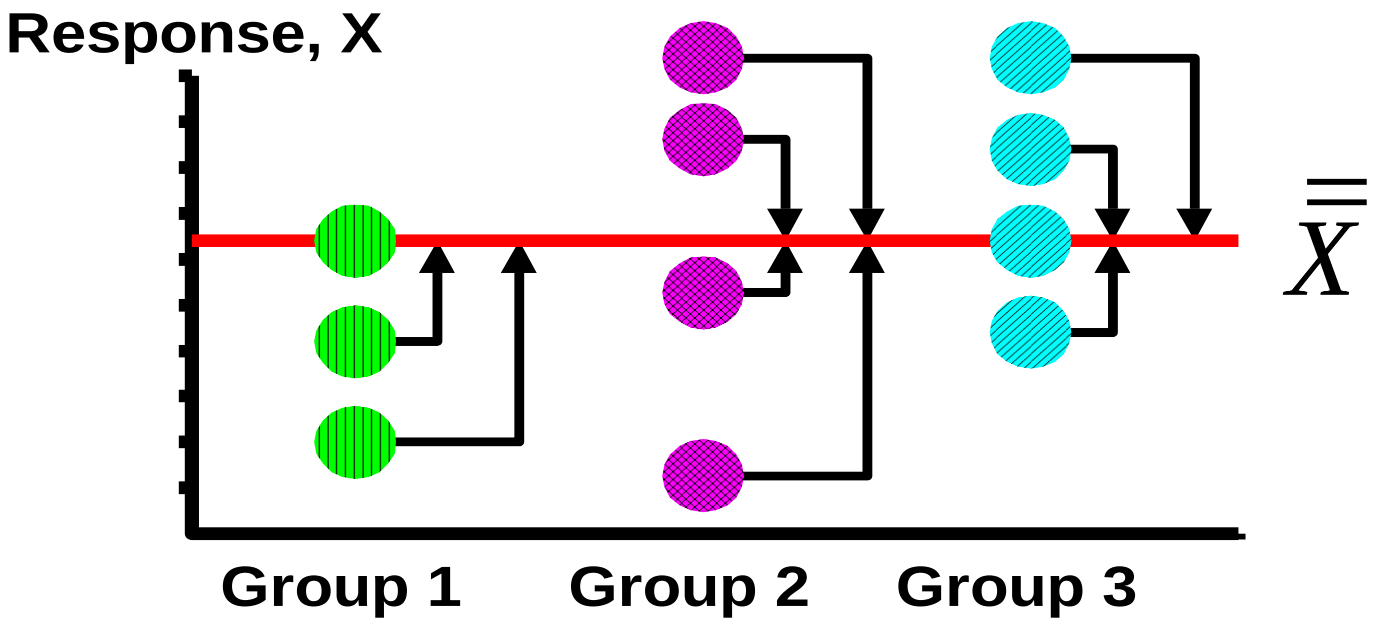
X_{ij} = i^{th} observation from group j

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

Total Variation

(continued)

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



Among-Group Variation

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

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

Where:

SSA = Sum of squares among groups

c = number of groups

n_j = sample size from group j

$\bar{X}_j$ = sample mean from group j

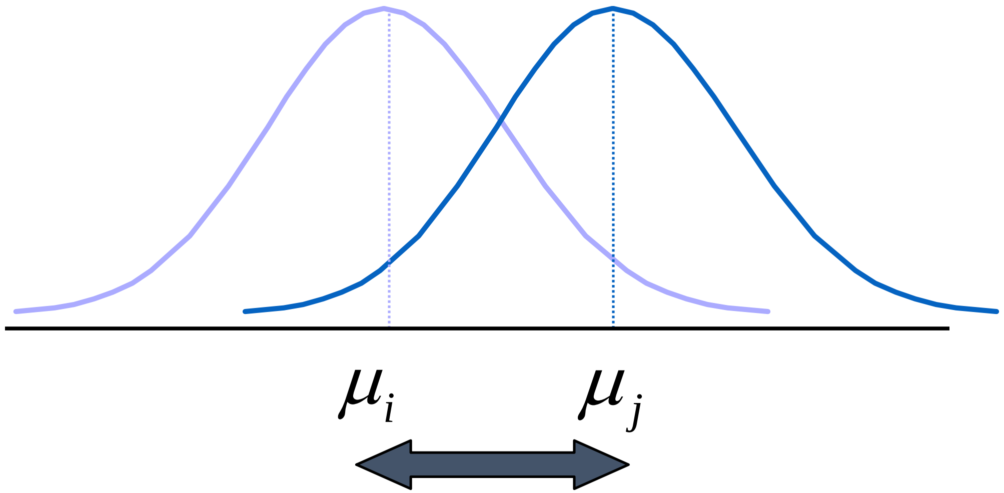
$\bar{\bar{X}}$ = grand mean (mean of all data values)

Among-Group Variation

(continued)

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

Variation Due to
Differences Among Groups



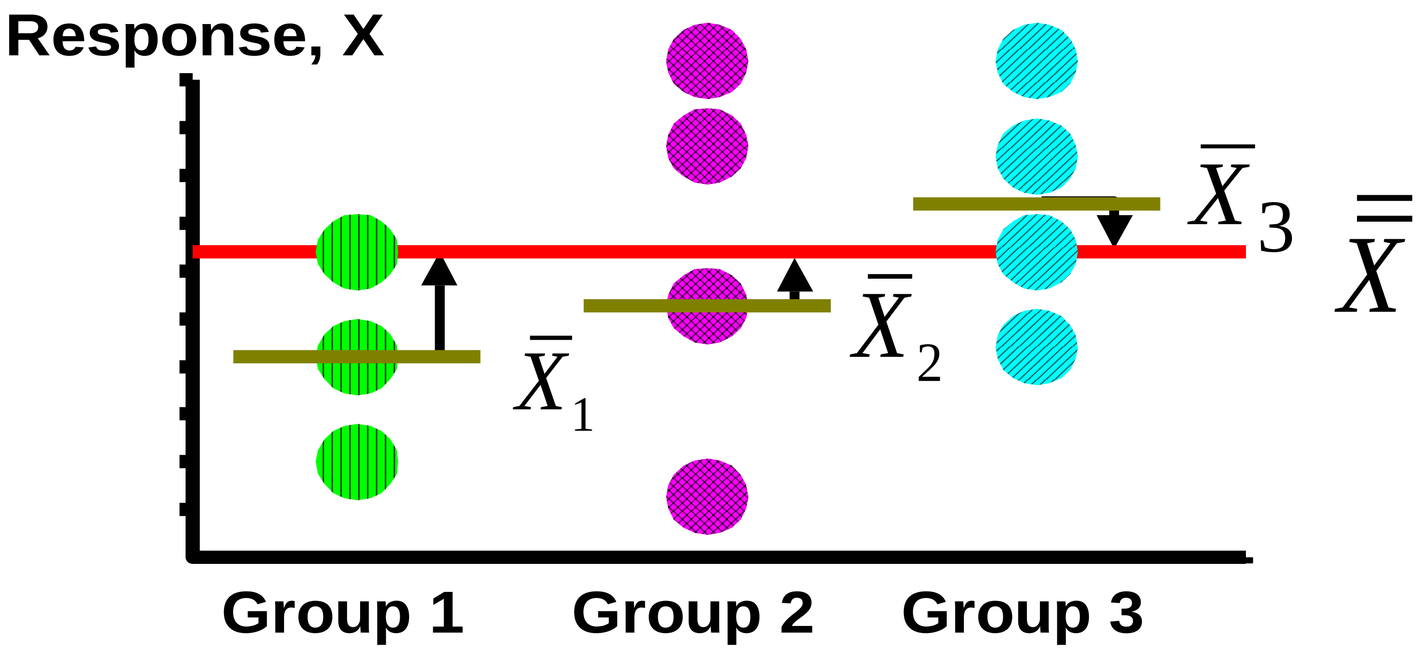
$$MSA = \frac{SSA}{c-1}$$

Mean Square Among =
SSA/degrees of freedom

Among-Group Variation

(continued)

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



Within-Group Variation

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

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

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

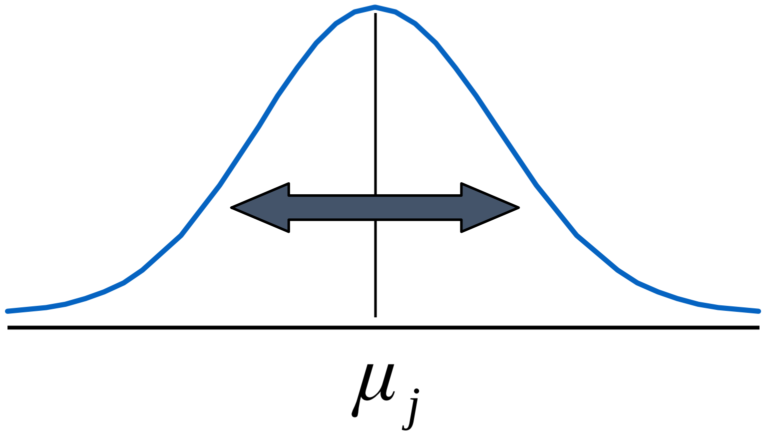
X_{ij} = i^{th} observation in group j

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



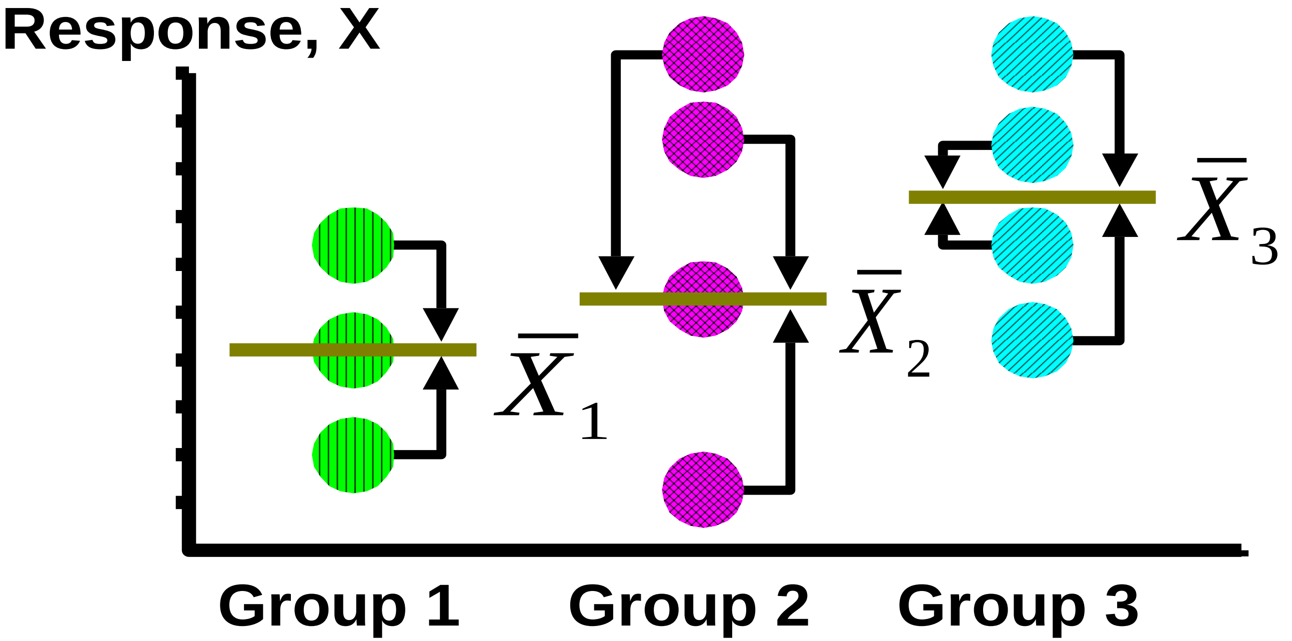
$$MSW = \frac{SSW}{n - c}$$

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

$$MSA = \frac{SSA}{c-1}$$

Mean Square Among
(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

Source of Variation	Degrees of Freedom	Sum Of Squares	Mean Square (Variance)	F
Among Groups	c - 1	SSA	$MSA = \frac{SSA}{c - 1}$	$F_{STAT} = \frac{MSA}{MSW}$
Within Groups	n - c	SSW	$MSW = \frac{SSW}{n - c}$	
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

$$F_{STAT} = \frac{MSA}{MSW}$$

MSA is mean squares among 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)

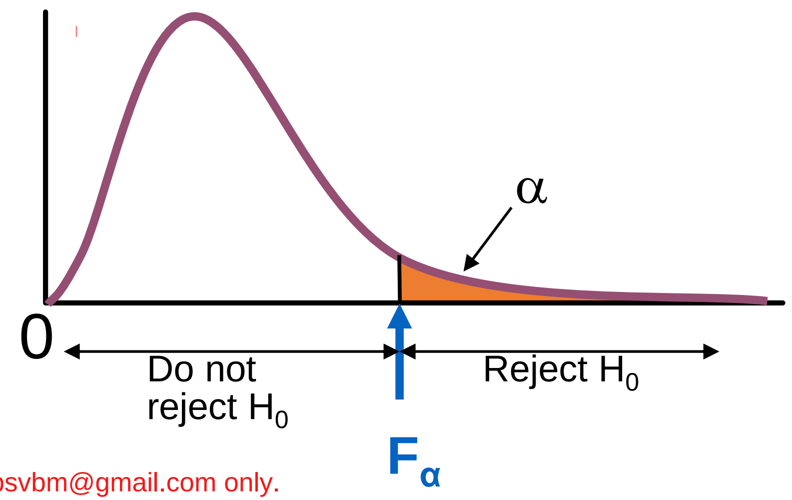
Interpreting One-Way ANOVA F Statistic

greatlearning

- 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?

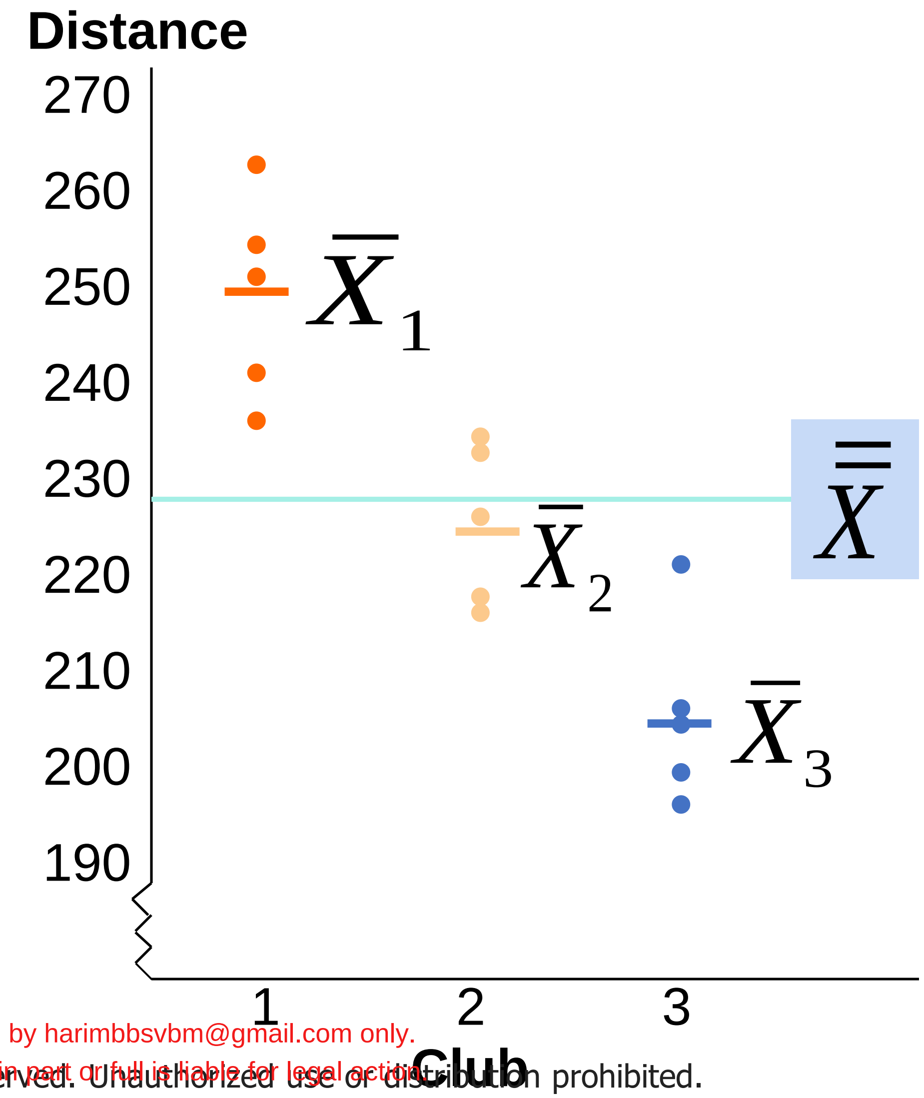
Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

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

greatlearning

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

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

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

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

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

$c = 3$

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

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

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

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

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

One-Way ANOVA Example Solution

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

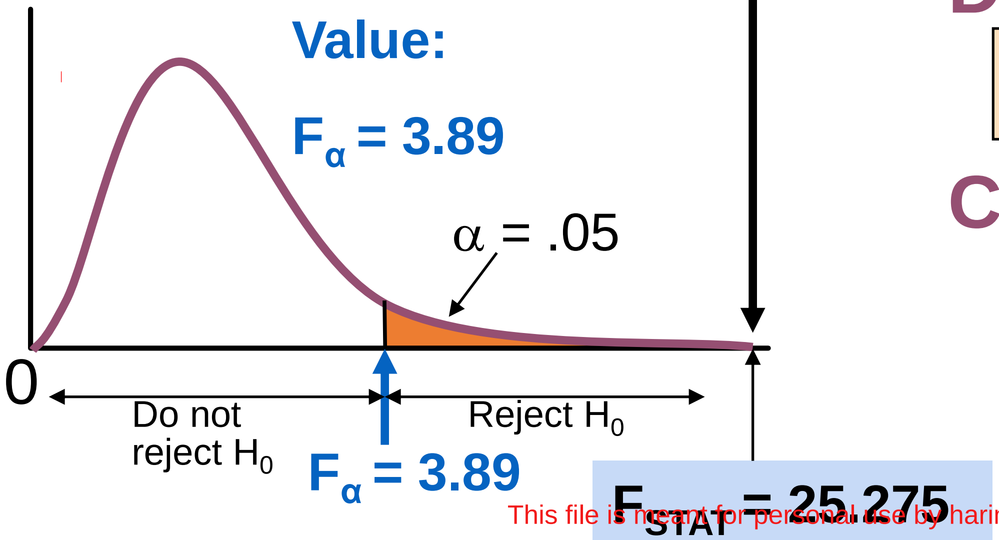
H_1: μ_j not all equal

$\alpha = 0.05$

$df_1 = 2$ $df_2 = 12$

Critical Value:

$F_\alpha = 3.89$



Test Statistic:

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

Decision:

Reject H_0 at $\alpha = 0.05$

Conclusion:

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

One-Way ANOVA Excel Output

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Club 1	5	1246	249.2	108.2		
Club 2	5	1130	226	77.5		
Club 3	5	1029	205.8	94.2		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	**4716.4**	**2**	**2358.2**	**25.275**	0.0000	**3.89**
Within Groups	**1119.6**	**12**	**93.3**			
Total	5836.0	14				

ANOVA Assumptions

- Randomness and Independence
  - Select random samples from the c groups (or randomly assign the levels)
- Normality
  - The sample values for each group are from a normal population
- Homogeneity of Variance
  - All populations sampled from have the same variance
  - Can be tested with Levene’s Test

Factorial Design: Two-Way ANOVA

- Examines the effect of
  - Two factors of interest on the dependent variable
    - e.g., Percent carbonation and line speed on soft drink bottling process
  - Interaction between the different levels of these two factors
    - e.g., Does the effect of one particular carbonation level depend on which level the line speed is set?

Two-Way ANOVA

(continued)

- Assumptions
  - Populations are normally distributed
  - Populations have equal variances
  - Independent random samples are drawn

Two-Way ANOVA
Sources of Variation

Two Factors of interest: A and B

r = number of levels of factor A

c = number of levels of factor B

n′ = number of replications for each cell

n = total number of observations in all cells

$$n = (r)(c)(n')$$

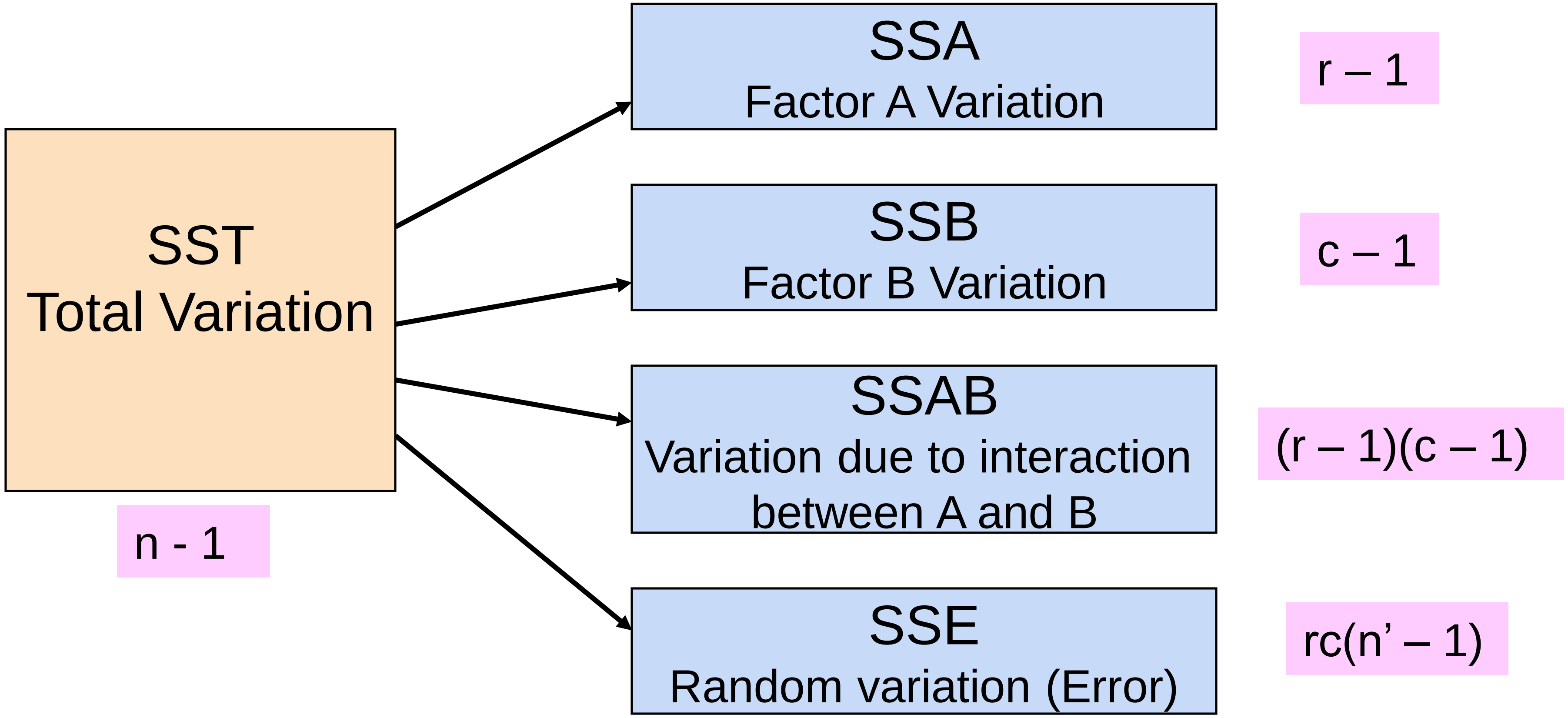
X_{ijk} = value of the k^{th} observation of level i of factor A and level j of factor B

Two-Way ANOVA
Sources of Variation

(continued)

SST = SSA + SSB + SSAB + SSE

Degrees of Freedom:



Two-Way ANOVA Equations

Total Variation:

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n'}(X_{ijk} - \bar{\bar{X}})^2$$

Factor A Variation:

$$SSA = cn'\sum_{i=1}^{r}(\bar{X}_{i..} - \bar{\bar{X}})^2$$

Factor B Variation:

$$SSB = rn'\sum_{j=1}^{c}(\bar{X}_{.j.} - \bar{\bar{X}})^2$$

Two-Way ANOVA Equations

(continued)

Interaction Variation:

$$SSAB = n'\sum_{i=1}^{r}\sum_{j=1}^{c}(\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{\bar{X}})^2$$

Sum of Squares Error:

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n'}(X_{ijk} - \bar{X}_{ij.})^2$$

Two-Way ANOVA Equations

(continued)

where:

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n'} X_{ijk}}{rcn'} = \text{Grand Mean}$$

$$\bar{X}_{i..} = \frac{\sum_{j=1}^{c}\sum_{k=1}^{n'} X_{ijk}}{cn'} = \text{Mean of } i^{th} \text{ level of factor A } \ (i = 1, 2, ..., r)$$

$$\bar{X}_{.j.} = \frac{\sum_{i=1}^{r}\sum_{k=1}^{n'} X_{ijk}}{rn'} = \text{Mean of } j^{th} \text{ level of factor B } \ (j = 1, 2, ..., c)$$

$$\bar{X}_{ij.} = \sum_{k=1}^{n'} \frac{X_{ijk}}{n'} = \text{Mean of cell } ij$$

r = number of levels of factor A

c = number of levels of factor B

n' = number of replications in each cell

Mean Square Calculations

$$\text{MSA} = \text{Mean square factor A} = \frac{\text{SSA}}{r-1}$$

$$\text{MSB} = \text{Mean square factor B} = \frac{\text{SSB}}{c-1}$$

$$\text{MSAB} = \text{Mean square interaction} = \frac{\text{SSAB}}{(r-1)(c-1)}$$

$$\text{MSE} = \text{Mean square error} = \frac{\text{SSE}}{rc(n'-1)}$$

Two-Way ANOVA: The F Test Statistics

F Test for Factor A Effect

$H_0: \mu_{1..} = \mu_{2..} = \mu_{3..} = \cdots = \mu_{r..}$

H_1: Not all $\mu_{i..}$ are equal

$$F_{STAT} = \frac{MSA}{MSE}$$

Reject H_0 if $F_{STAT} > F_\alpha$

F Test for Factor B Effect

$H_0: \mu_{.1.} = \mu_{.2.} = \mu_{.3.} = \cdots = \mu_{.c.}$

H_1: Not all $\mu_{.j.}$ are equal

$$F_{STAT} = \frac{MSB}{MSE}$$

Reject H_0 if $F_{STAT} > F_\alpha$

F Test for Interaction Effect

H_0: the interaction of A and B is equal to zero

H_1: interaction of A and B is not zero

$$F_{STAT} = \frac{MSAB}{MSE}$$

Reject H_0 if $F_{STAT} > F_\alpha$

Two-Way ANOVA Summary Table

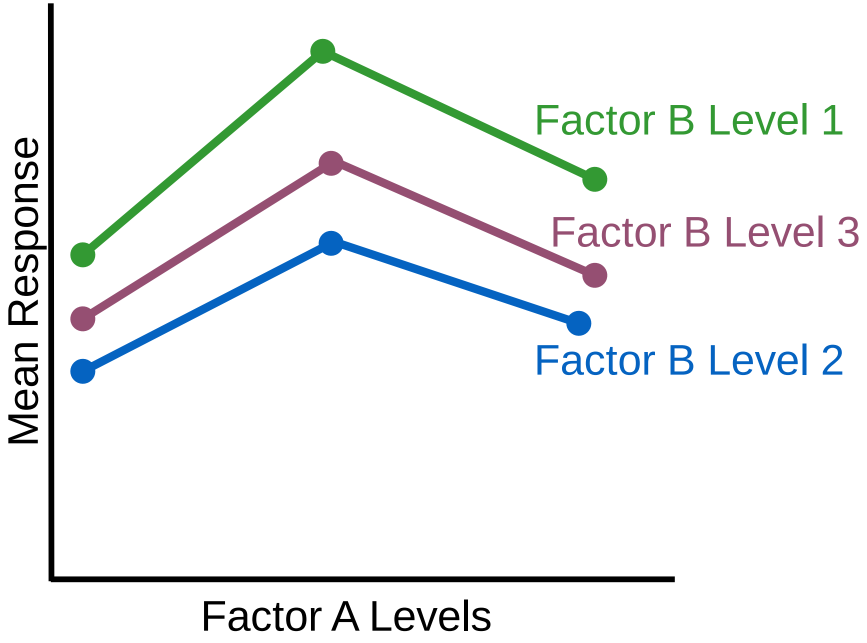
Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F
Factor A	**SSA**	**r – 1**	**MSA** = SSA /(r – 1)	$\frac{\textbf{MSA}}{\textbf{MSE}}$
Factor B	**SSB**	**c – 1**	**MSB** = SSB /(c – 1)	$\frac{\textbf{MSB}}{\textbf{MSE}}$
AB (Interaction)	**SSAB**	**(r – 1)(c – 1)**	**MSAB** = SSAB / (r – 1)(c – 1)	$\frac{\textbf{MSAB}}{\textbf{MSE}}$
Error	**SSE**	**rc(n′ – 1)**	**MSE =** SSE/rc(n’ – 1)	
Total	**SST**	**n – 1**		

Features of Two-Way ANOVA F Test

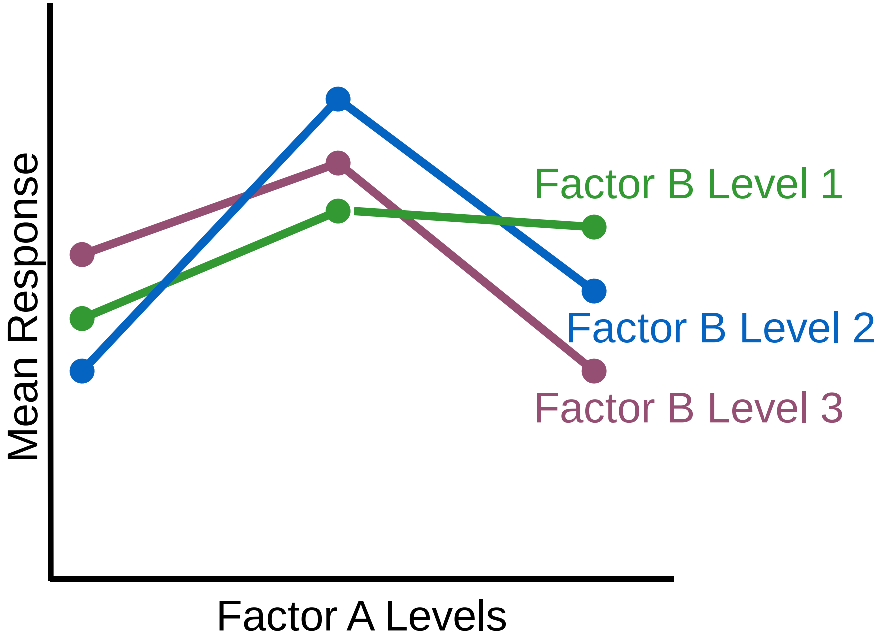
- Degrees of freedom always add up
  - $n-1 = rc(n'-1) + (r-1) + (c-1) + (r-1)(c-1)$
  - Total = error + factor A + factor B + interaction
- The denominators of the F Test are always the same but the numerators are different
- The sums of squares always add up
  - SST = SSE + SSA + SSB + SSAB
  - Total = error + factor A + factor B + interaction

Examples:
Interaction vs. No Interaction

- No interaction: line segments are parallel



- Interaction is present: some line segments not parallel



Summary

In this chapter we discussed

- The one-way analysis of variance
  - The logic of ANOVA
  - ANOVA assumptions
  - F test for difference in c means
- The two-way analysis of variance
  - Examined effects of multiple factors
  - Examined interaction between factors