

# Statistik Bisnis

Week 12

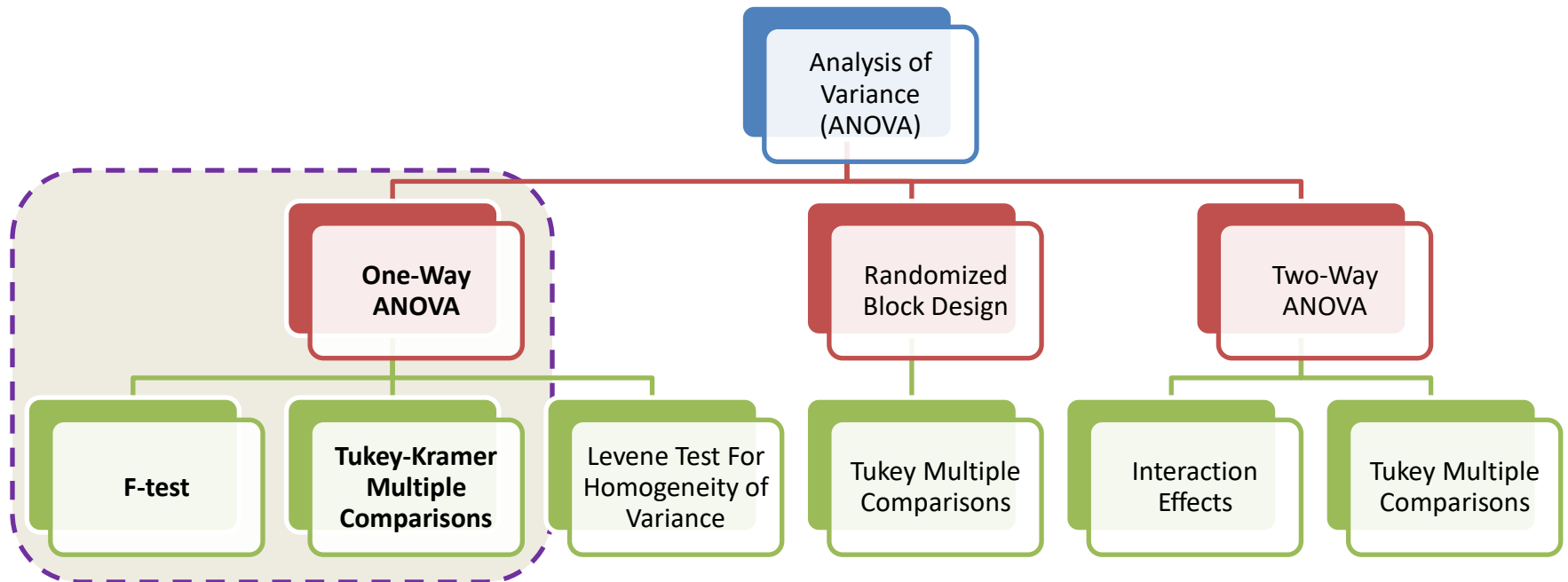
Analysis of Variance

# Learning Objectives

## **This week, you learn:**

- How to use one-way analysis of variance to test for differences among the means of several populations (also referred to as “groups” in this chapter)

# Chapter Overview



# **F TEST OF ANOVA**

# One-Way Analysis of Variance

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

**Examples:** Accident rates for 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> 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

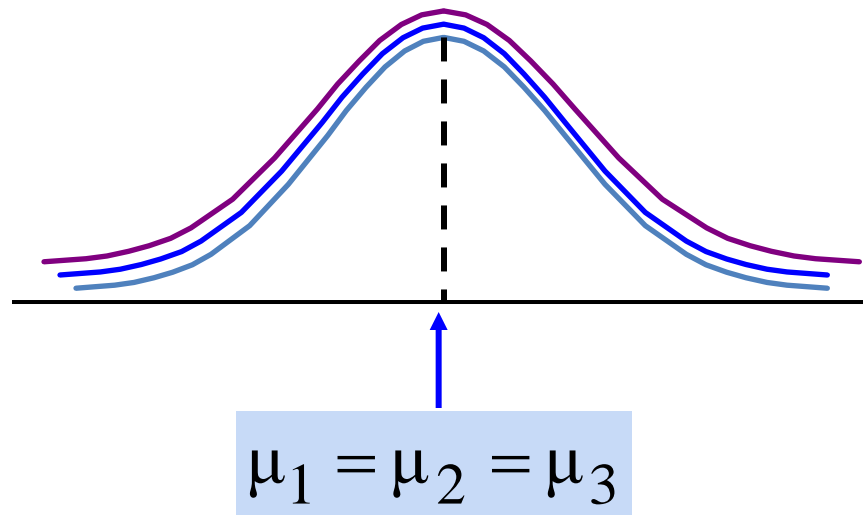
- $H_0 : \mu_1 = \mu_2 = \mu_3 = \Lambda = \mu_c$ 
  - All population means are equal
  - i.e., no factor effect (no variation in means among groups)
- $H_1 : \text{Not all of the population means are the same}$ 
  - At least one population mean is different
  - i.e., there is a factor effect
  - Does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

# One-Way ANOVA

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

$H_1$  : Not all  $\mu_j$  are the same

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

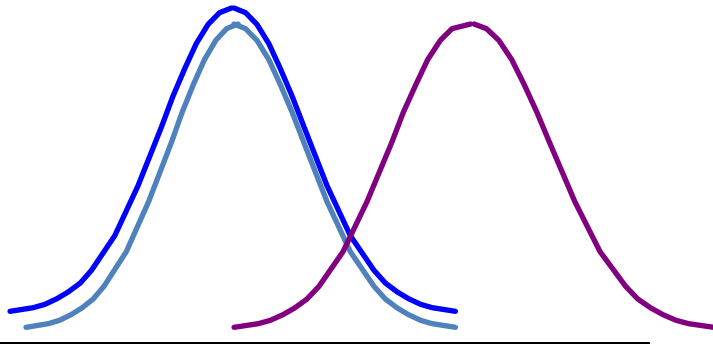


# One-Way ANOVA

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

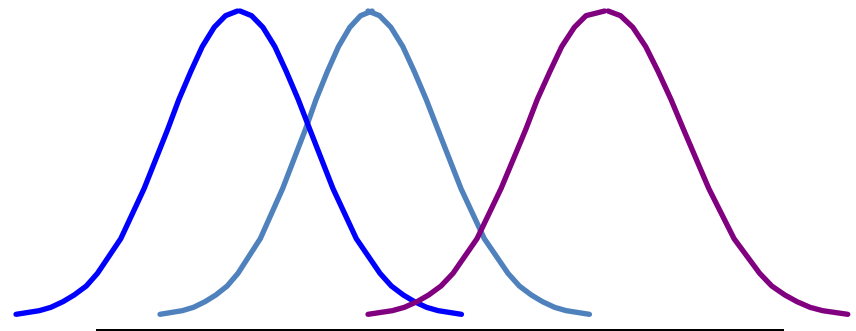
$H_1$  : Not all  $\mu_j$  are the same

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



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

or



$$\mu_1 \neq \mu_2 \neq \mu_3$$



# Partitioning the Variation

- Total variation can be split into two parts:

$$SST = SSA + SSW$$

$SST$  = 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

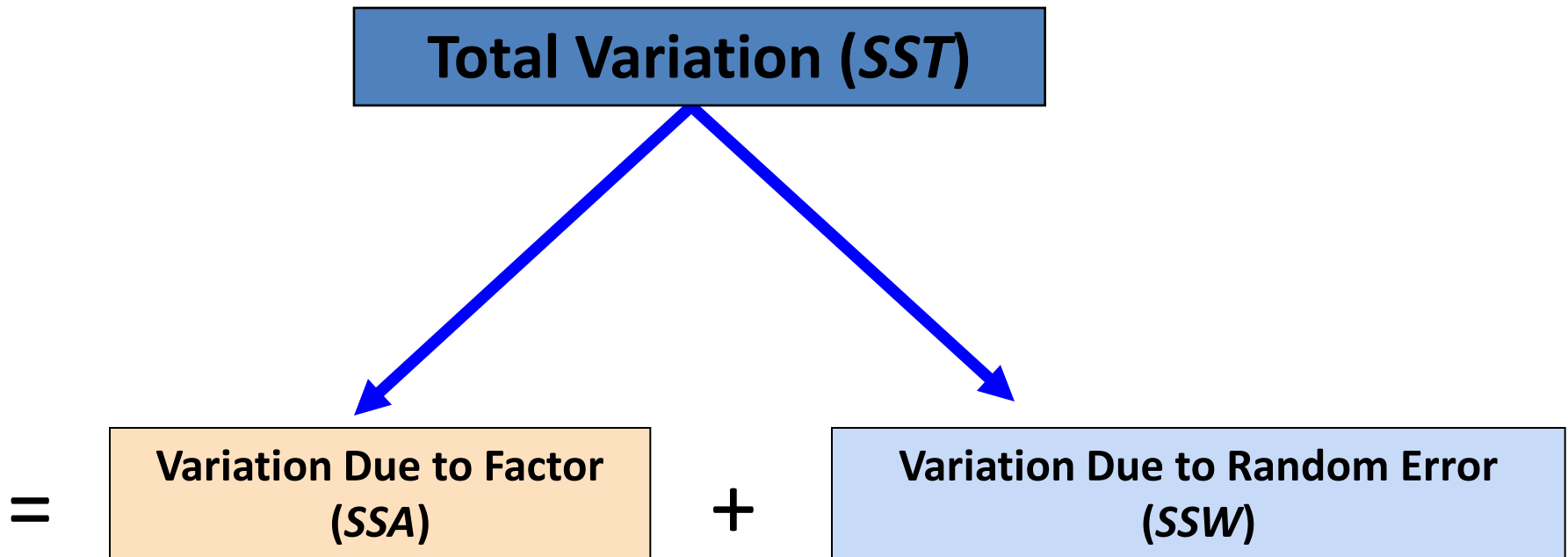
$$SST = SSA + SSW$$

**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 Sum of Squares

$$SST = SSA + SSW$$

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

Where:

$SST$  = Total sum of squares

$c$  = number of groups or levels

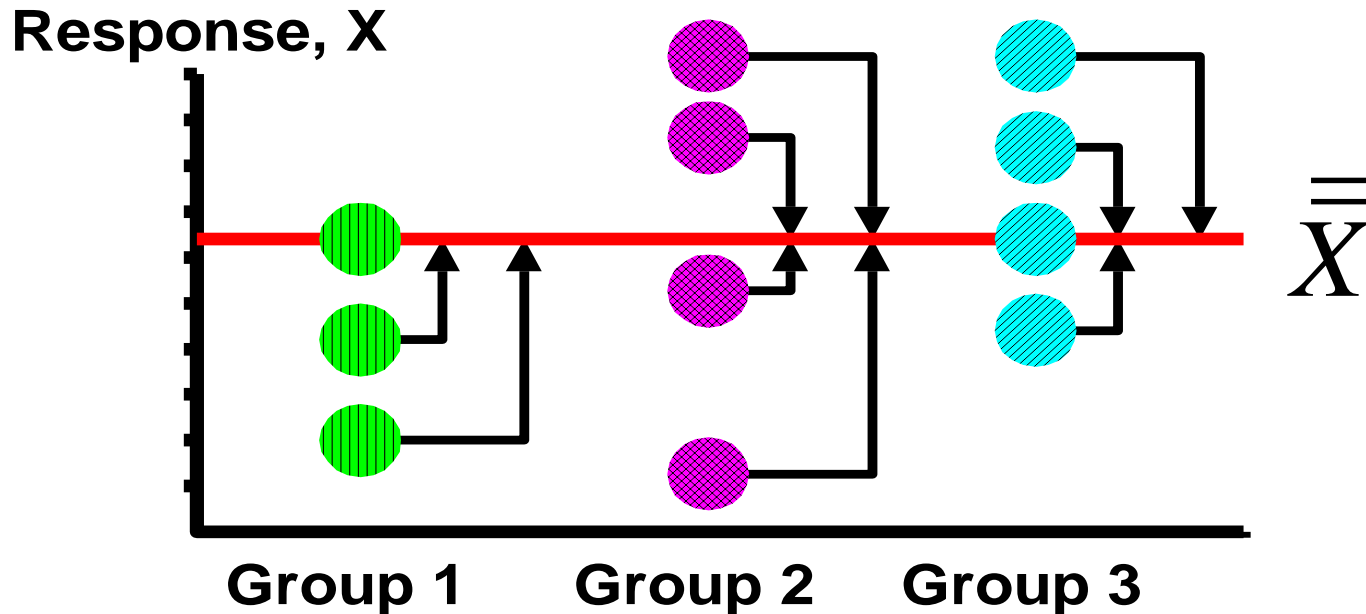
$n_j$  = number of observations in group  $j$

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

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

# Total Variation

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



# Among-Group Variation

$$SST = SSA + SSW$$

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

Where:

$SSA$  = Sum of squares among groups

$c$  = number of groups

$n_j$  = sample size from group  $j$

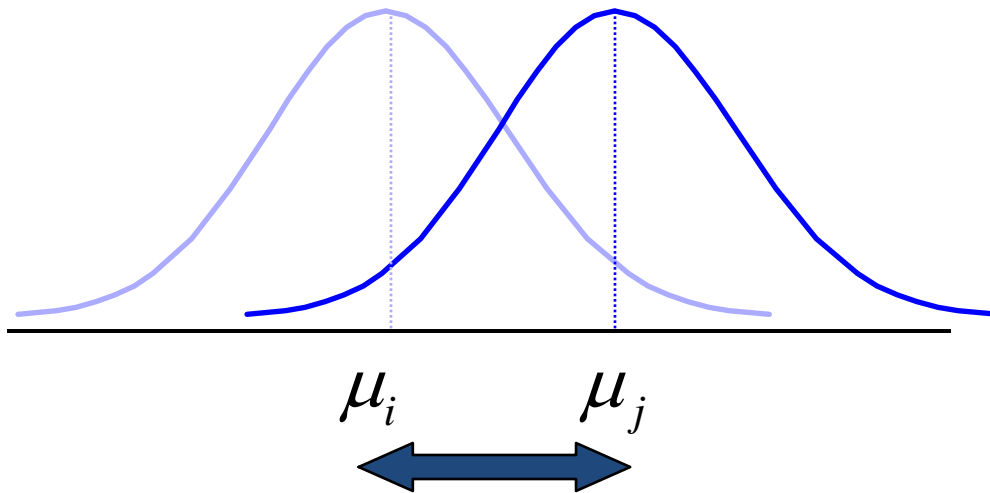
$\bar{X}_j$  = sample mean from group  $j$

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

# Among-Group Variation

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

Variation Due to  
Differences Among Groups

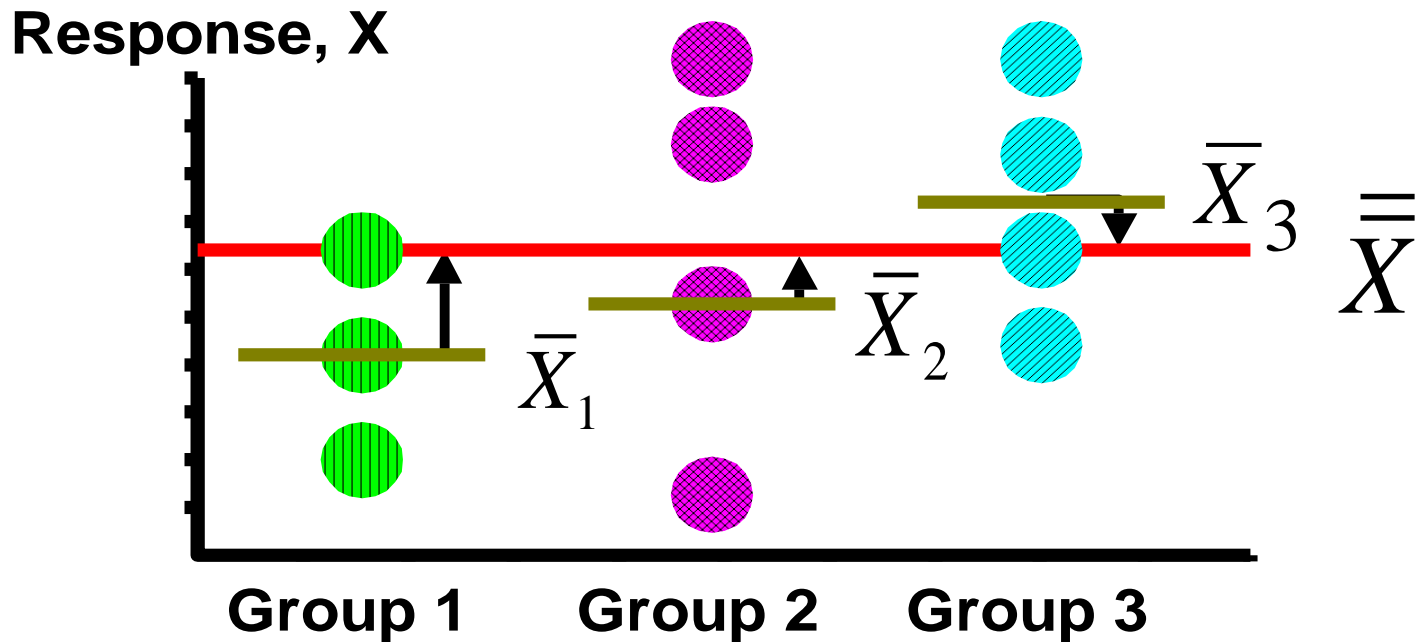


$$MSA = \frac{SSA}{c - 1}$$

Mean Square Among =  
SSA/degrees of freedom

# Among-Group Variation

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





# Within-Group Variation

$$SST = SSA + SSW$$

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

Where:

$SSW$  = Sum of squares within groups

$c$  = number of groups

$n_j$  = sample size from group  $j$

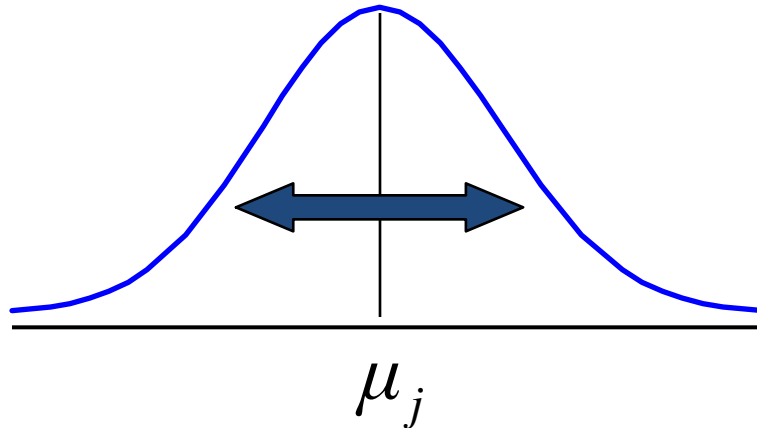
$\bar{X}_j$  = sample mean from group  $j$

$X_{ij}$  =  $i^{\text{th}}$  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

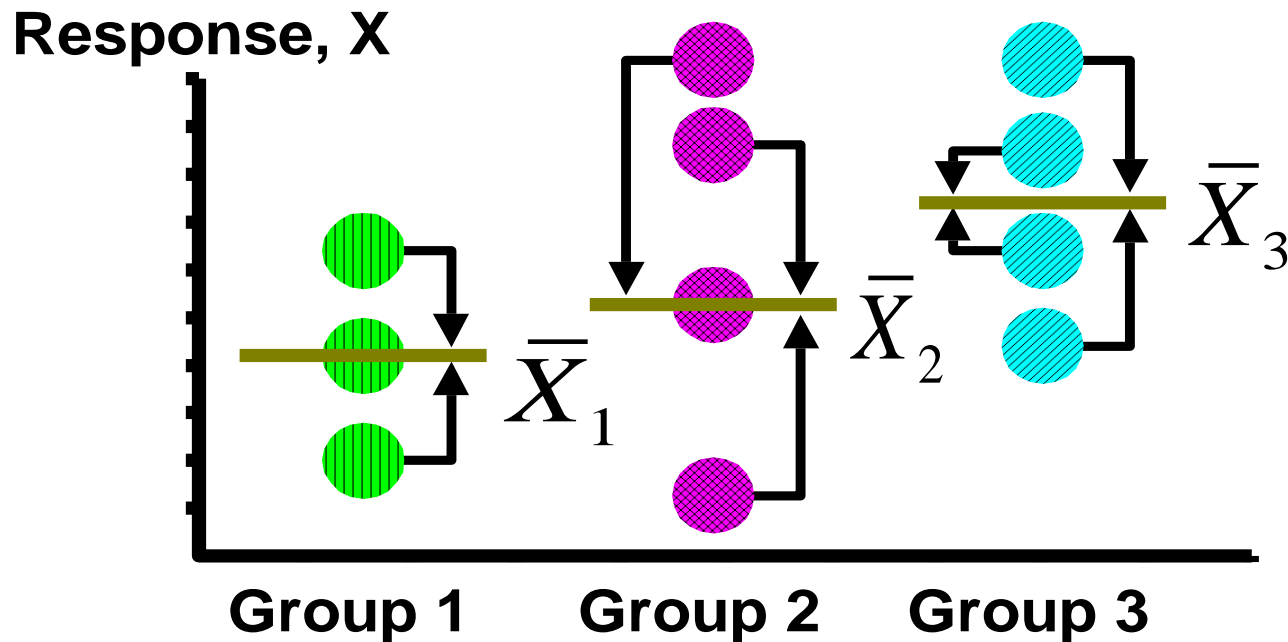


$$MSW = \frac{SSW}{n - c}$$

Mean Square Within =  
SSW/degrees of freedom

# Within-Group Variation

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



# 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

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

$H_1$ : 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)

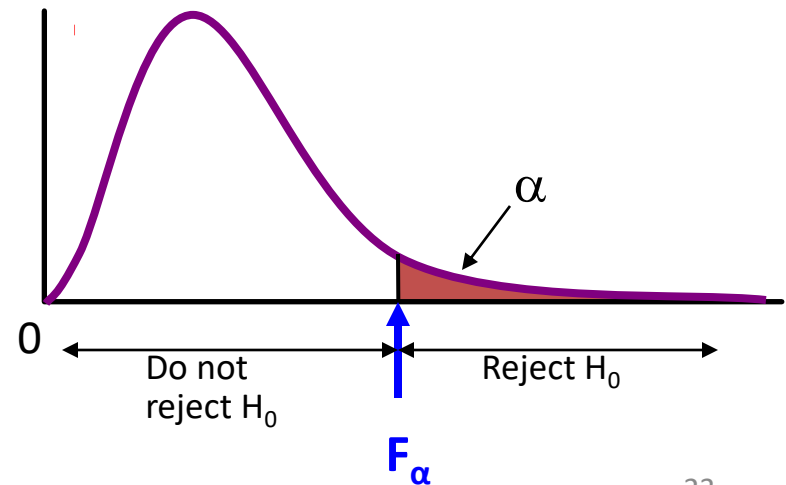
- $df_2 = n - c$  (n = sum of sample sizes from all populations)

# Interpreting One-Way ANOVA F Statistic

- 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

Anda ingin mengetahui apakah jarak bola yang dipukul oleh tiga stick golf yang berbeda akan berbeda. Anda memilih lima pemukulan bola secara acak untuk masing-masing stick golf tersebut. Pada tingkat signifikansi 0,05, apakah terdapat bukti yang menunjukkan perbedaan pada rata-rata jarak pukulan bola?

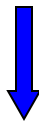
<u>Club 1</u>	<u>Club 2</u>	<u>Club 3</u>
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204





# One-Way ANOVA Example: Scatter Plot

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

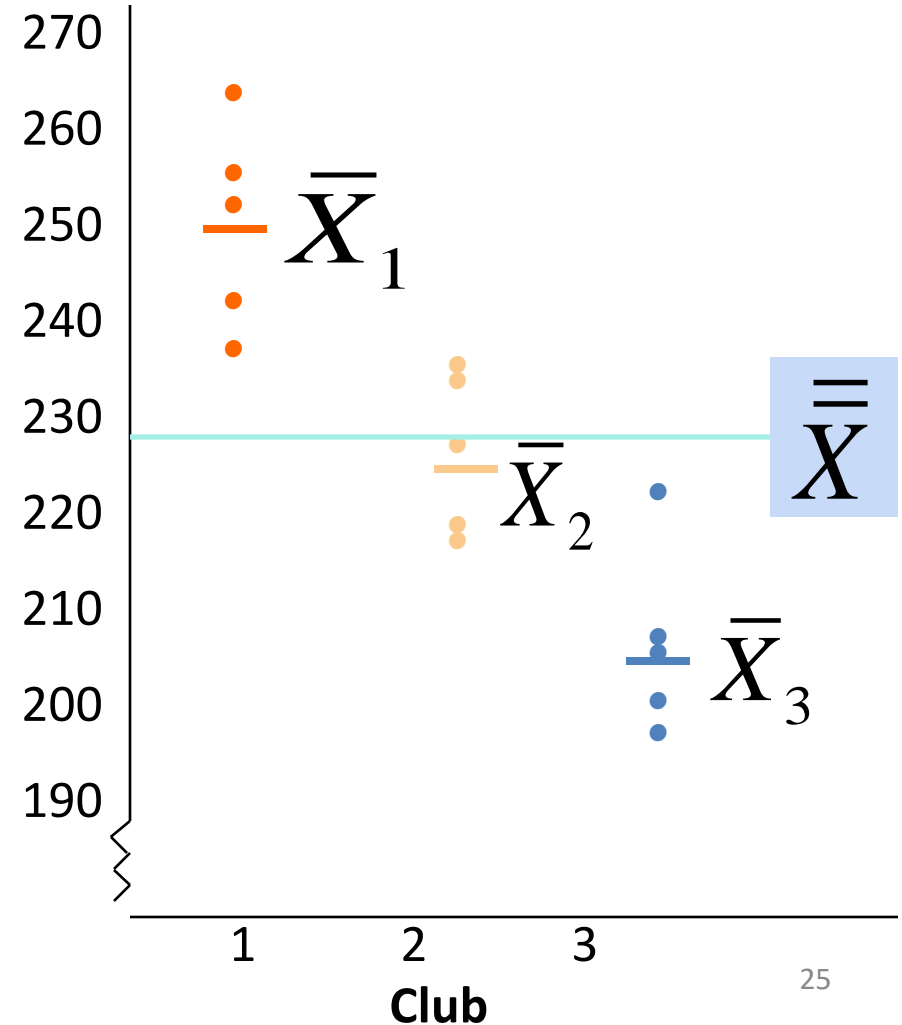


$\bar{X}_1 = 249.2$	$\bar{X}_2 = 226.0$	$\bar{X}_3 = 205.8$
---------------------	---------------------	---------------------

$\bar{\bar{X}} = 227.0$
-------------------------



Distance



# One-Way ANOVA Example Computations

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



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



$$SSA = 5 (249.2 - 227)^2 + 5 (226 - 227)^2 + 5 (205.8 - 227)^2 = 4,716.4$$

$$SSW = (254 - 249.2)^2 + (263 - 249.2)^2 + \dots + (204 - 205.8)^2 = 1,119.6$$

$$MSA = 4,716.4 / (3-1) = 2,358.2$$

$$MSW = 1,119.6 / (15-3) = 93.3$$

$$F_{STAT} = \frac{2,358.2}{93.3} = 25.275$$

# One-Way ANOVA Example Solution

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

$$H_1: \mu_j \text{ not all equal}$$

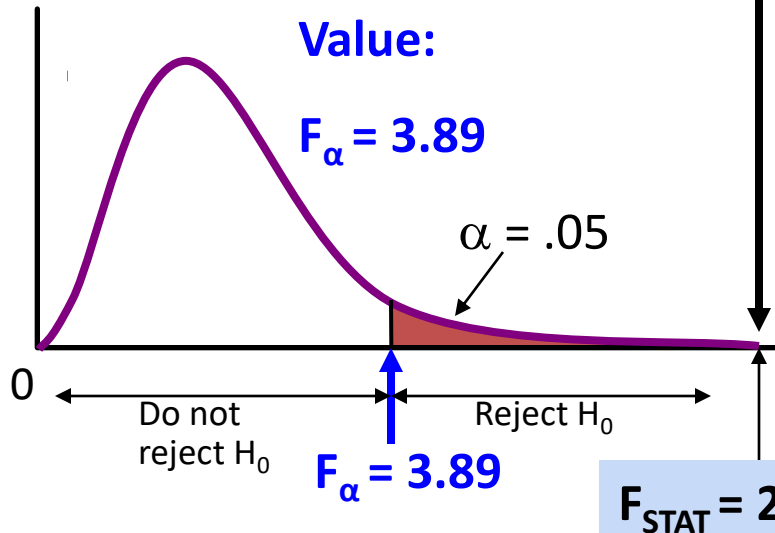
$$\alpha = 0.05$$

$$df_1 = 2 \quad df_2 = 12$$

**Critical Value:**

$$F_{\alpha} = 3.89$$

$$\alpha = .05$$



**Test Statistic:**

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

**Decision:**

Reject  $H_0$  at  $\alpha = 0.05$

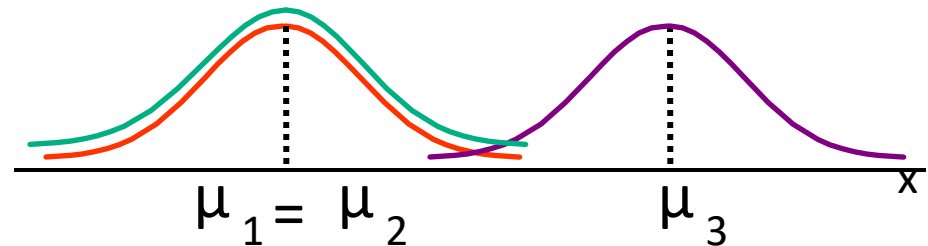
**Conclusion:**

There is evidence that at least one  $\mu_j$  differs from the rest

# **TUKEY-KRAMER OF ANOVA**

# The Tukey-Kramer Procedure

- Tells **which** population means are significantly different
  - e.g.:  $\mu_1 = \mu_2 \neq \mu_3$
  - Done after rejection of equal means in ANOVA
- Allows paired comparisons
  - Compare absolute mean differences with critical range



# Tukey-Kramer Critical Range

$$\textit{Critical Range} = Q_{\alpha} \sqrt{\frac{MSW}{2} \left( \frac{1}{n_j} + \frac{1}{n_{j'}} \right)}$$

where:

$Q_{\alpha}$  = Upper Tail Critical Value from Studentized Range Distribution with  $c$  and  $n - c$  degrees of freedom (see appendix E.7 table)

$MSW$  = Mean Square Within

$n_j$  and  $n_{j'}$  = Sample sizes from groups  $j$  and  $j'$

# The Tukey-Kramer Procedure: Example

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

1. Compute absolute mean differences:

$$|\bar{x}_1 - \bar{x}_2| = |249.2 - 226.0| = 23.2$$

$$|\bar{x}_1 - \bar{x}_3| = |249.2 - 205.8| = 43.4$$

$$|\bar{x}_2 - \bar{x}_3| = |226.0 - 205.8| = 20.2$$

2. Find the  $Q_\alpha$  value from the table in appendix E.7 with  $c = 3$  and  $(n - c) = (15 - 3) = 12$  degrees of freedom:

$$Q_\alpha = 3.77$$



# The Tukey-Kramer Procedure: Example

## 3. Compute Critical Range:

$$\text{Critical Range} = Q_{\alpha} \sqrt{\frac{MSW}{2} \left( \frac{1}{n_j} + \frac{1}{n_{j'}} \right)} = 3.77 \sqrt{\frac{93.3}{2} \left( \frac{1}{5} + \frac{1}{5} \right)} = 16.285$$

## 4. Compare:

5. All of the absolute mean differences are greater than critical range. Therefore there is a significant difference between each pair of means at 5% level of significance. Thus, with 95% confidence we can conclude that the mean distance for club 1 is greater than club 2 and 3, and club 2 is greater than club 3.

$$|\bar{x}_1 - \bar{x}_2| = 23.2$$

$$|\bar{x}_1 - \bar{x}_3| = 43.4$$

$$|\bar{x}_2 - \bar{x}_3| = 20.2$$





# 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

# EXERCISE

## 11.7 (cont'd)

The Computer Anxiety Rating Scale (CARS) mengukur level kecemasan komputer individu, dengan skala dari 20 (tidak ada kecemasan) hingga 100 (level tertinggi kecemasan). Peneliti pada Miami University menyebarkan CARS pada 172 mahasiswa bisnis. Salah satu tujuan penelitian ini adalah untuk menentukan apakah terdapat perbedaan tingkat kecemasan komputer pada mahasiswa dengan jurusan yang berbeda. Mereka mendapatkan data sebagai berikut:

# 11.7 (cont'd)

<b>Sumber variasi</b>	<b>Degrees of Freedom</b>	<b>Sum of Squares</b>	<b>Mean Squares</b>	<b>F</b>
Antar Jurusan	5	3.172		
Dalam Jurusan	166	21.246		
Total	171	24.418		

<b>Jurusan</b>	<b>n</b>	<b>Mean</b>
Marketing	19	44,37
Management	11	43,18
Other	14	42,21
Finance	45	41,8
Accountancy	36	37,56
MIS	47	32,21

# 11.7

- a. Lengkapi tabel ringkasan ANOVA diatas.
- b. Pada tingkat signifikansi 0,05, apakah terdapat bukti adanya perbedaan pada rata-rata tingkat kecemasan komputer yang dialami oleh mahasiswa dengan jurusan yang berbeda?
- c. Jika hasil pada poin (b) menunjukkan ada yang berbeda, gunakan prosedur Tukey-Kramer untuk menentukan jurusan apa yang berbeda tingkat kecemasan komputernya.

## 11.11 (cont'd)

Rata-rata jumlah pengunjung pada masing-masing toko dari sebuah peritel ternama (yang memiliki lebih dari 10.000 toko) akhir-akhir ini selalu tetap pada angka 900 orang pengunjung. Untuk meningkatkan jumlah pelanggan, peritel tersebut berencana untuk menurunkan harga kopinya. Peritel tersebut ingin mengetahui seberapa banyak pengurangan harga yang mampu meningkatkan jumlah pengunjung harian tanpa terlalu besar mengurangi keuntungan kasar dari penjualan kopi tersebut.

## 11.10 (cont'd)

Sebuah produsen pulpen menyewa jasa agen periklanan untuk membuat iklan produk mereka. Untuk mempersiapkan proyek ini, direktur riset melakukan penelitian mengenai pengaruh iklan pada persepsi produk. Sebuah eksperimen didesain untuk membandingkan lima iklan yang berbeda. Iklan *A* sangat tidak menonjolkan karakteristik pulpen. Iklan *B* kurang menonjolkan karakteristik pulpen. Iklan *C* agak melebih-lebihkan karakteristik pulpen. Iklan *D* sangat melebih-lebihkan karakteristik pulpen. Iklan *E* mencoba menggambarkan karakteristik pulpen dengan tepat.

## 11.10 (cont'd)

Sebuah sampel yang terdiri dari 30 orang responden diminta untuk mengevaluasi salah satu iklan tersebut (sehingga terdapat 6 responden untuk masing-masing iklan). Setelah membaca iklan dan dapat membayangkan ekspektasi produk dari iklan tersebut, semua responden diberi pulpen yang sama untuk dievaluasi. Responden diizinkan untuk mencoba pulpen tersebut sesuai dengan yang dijanjikan iklan yang mereka terima. Kemudian responden diminta untuk memberi penilaian dari 1 hingga 7 (terendah hingga tertinggi) untuk karakteristik produk tersebut: penampilan, ketahanan, dan kinerja. Total dari ketiga karakteristik dari masing-masing responden dapat dilihat dari tabel berikut:



# 11.10 (cont'd)

A	B	C	D	E
15	16	8	5	12
18	17	7	6	19
17	21	10	13	18
19	16	15	11	12
19	19	14	9	17
20	17	14	10	14

# 11.10

- a. Pada tingkat signifikansi 0,05, apakah terdapat bukti adanya perbedaan pada rata-rata penilaian pulpen dari iklan yang berbeda tersebut?
- b. Jika dibutuhkan, tentukan iklan mana yang berbeda rata-rata penilaiannya.
- c. Iklan manakah yang harusnya anda gunakan dan iklan manakah yang harusnya anda hindari? Jelaskan.

## 11.12 (cont'd)

Sirkuit terintegrasi (IC) dimanufaktur pada papan sirkuit silikon melalui serangkaian proses. Sebuah eksperimen dilakukan untuk mengetahui efek yang dihasilkan dari tiga metode pada proses pembersihan. Hasilnya adalah sebagai berikut:

# 11.12 (cont'd)

Metode Baru 1	Metode Baru 2	Metode Standar
38	29	31
34	35	23
38	34	38
34	20	29
19	35	32
28	37	30

# 11.12

- a. Pada tingkat signifikansi 0,05, apakah terdapat bukti adanya perbedaan rata-rata hasil antara tiga metode yang digunakan pada proses pembersihan?
- b. Jika dibutuhkan, tentukan metode apa yang memberikan rata-rata hasil yang berbeda.

**THANK YOU**