Analysis of Variance (ANOVA)

Dr. Md. Israt Rayhan
Professor
Institute of Statistical Research and Training (ISRT)
University of Dhaka
Email: israt@isrt.ac.bd

One-Way Analysis of Variance

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

Examples:	Average recovery time for three types of drugs.
	Expected mileage for five brands of tires.

- Assumptions
  - Populations are normally distributed
  - Populations have equal variances
  - Samples are randomly and independently drawn

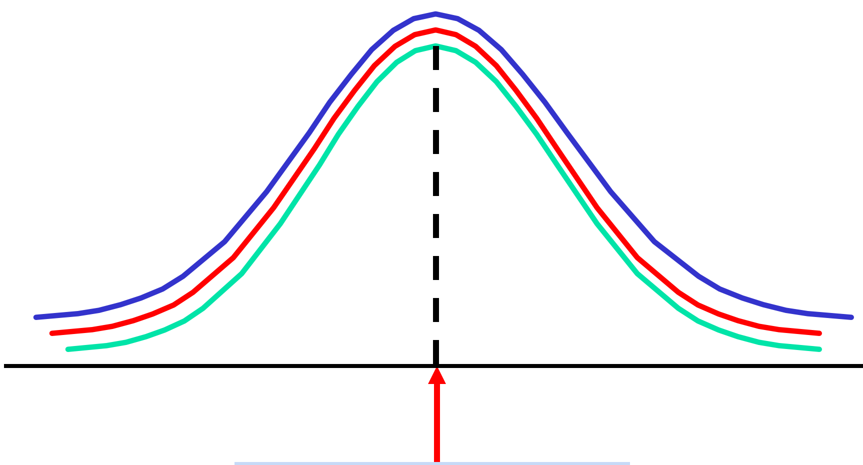
Hypotheses of One-Way ANOVA

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_K$
  - All population means are equal
  - i.e., no variation in means between groups
- $H_1 : \mu_i \neq \mu_j \quad \text{for at least one } i, j \text{ pair}$
  - At least one population mean is different
  - i.e., there is variation between groups
  - Does not mean that all population means are different (some pairs may be the same)

One-Way ANOVA

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

$$H_1 : \text{Not all } \mu_i \text{ are the same}$$



All Means are the same:
The Null Hypothesis is True
(No variation between groups)

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

One-Way ANOVA

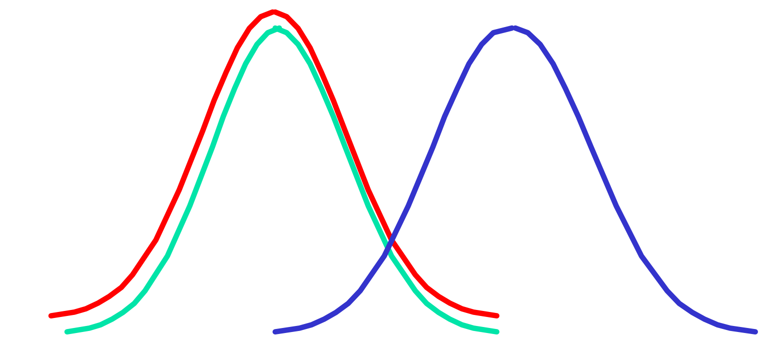
(continued)

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_K$$
$$H_1 : \text{Not all } \mu_i \text{ are the same}$$

At least one mean is different:

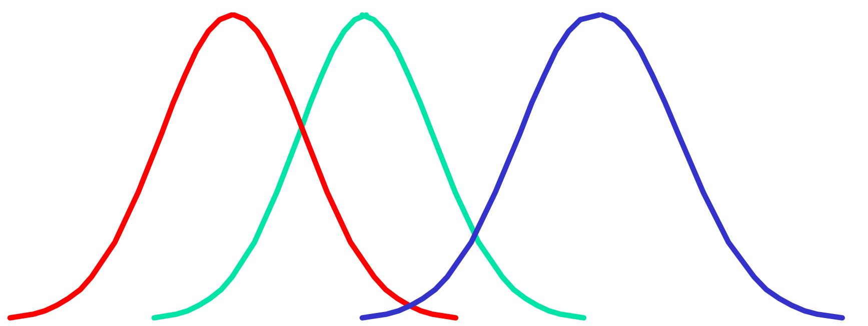
The Null Hypothesis is NOT true

(Variation is present between groups)



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

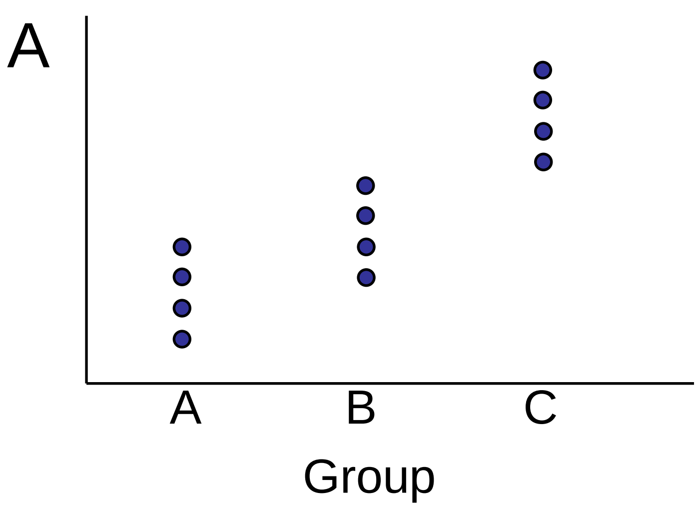
or



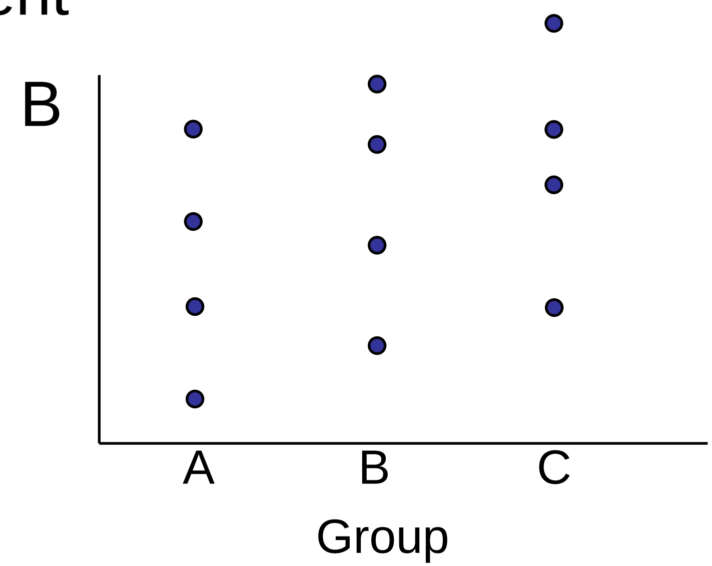
$$\mu_1 \neq \mu_2 \neq \mu_3$$

Variability

- The variability of the data is key factor to test the equality of means
- In each case below, the means may look different, but a large variation within groups in B makes the evidence that the means are different



Small variation within groups



Large variation within groups

Partitioning the Variation

- Total variation can be split into two parts:

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

SST = Total Sum of Squares

Total Variation = the aggregate dispersion of the individual data values across the various groups

SSW = Sum of Squares Within Groups

Within-Group Variation = dispersion that exists among the data values within a particular group

SSG = Sum of Squares Between Groups

Between-Group Variation = dispersion between the group sample means

Partition of Total Variation

Total Sum of Squares (SST)

= **Variation due to random sampling (SSW)** + **Variation due to differences between groups (SSG)**

Total Sum of Squares

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

Where:

$$\boxed{SST = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2}$$

SST = Total sum of squares

K = number of groups (levels or treatments)

n_i = number of observations in group i

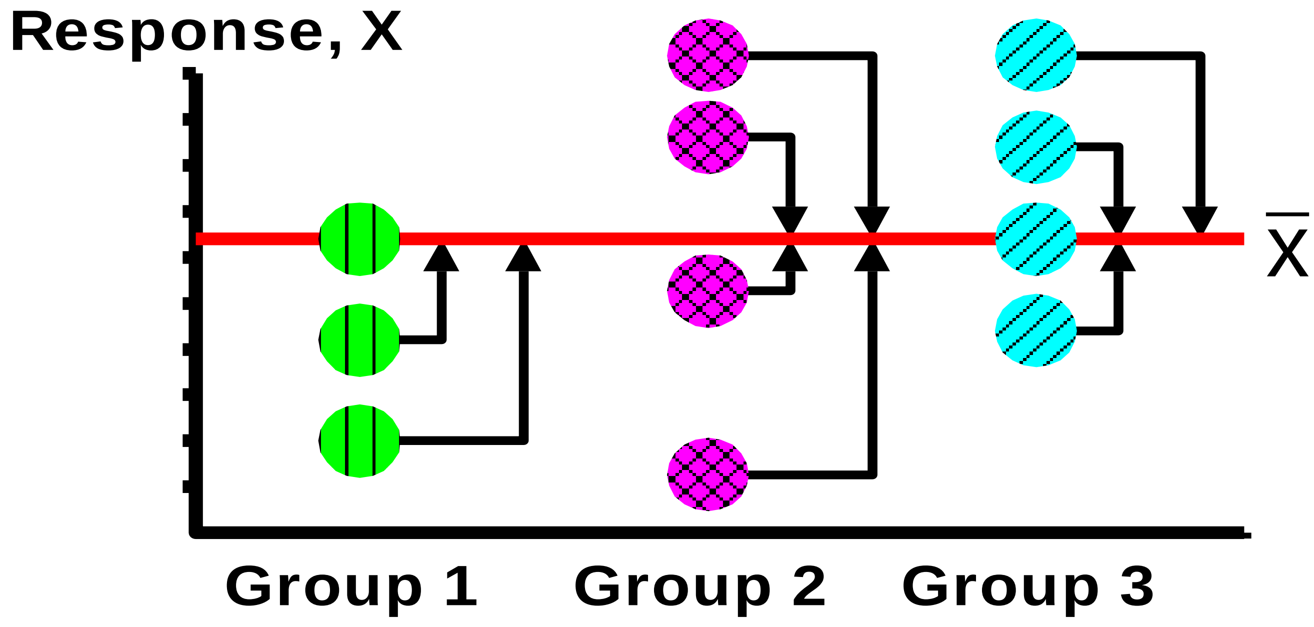
x_{ij} = j^{th} observation from group i

$\bar{x}$ = overall sample mean

Total Variation

(continued)

$$SST = (x_{11} - \bar{x})^2 + (X_{12} - \bar{x})^2 + \ldots + (x_{Kn_K} - \bar{x})^2$$



Within-Group Variation

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

$$\boxed{SSW = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}$$

Where:

SSW = Sum of squares within groups

K = number of groups

n_i = sample size from group i

$\bar{X}_i$ = sample mean from group i

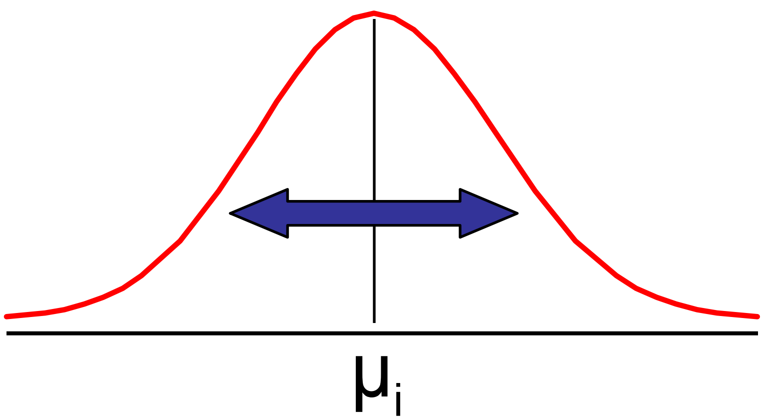
X_{ij} = j^{th} observation in group i

Within-Group Variation

(continued)

$$SSW = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

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



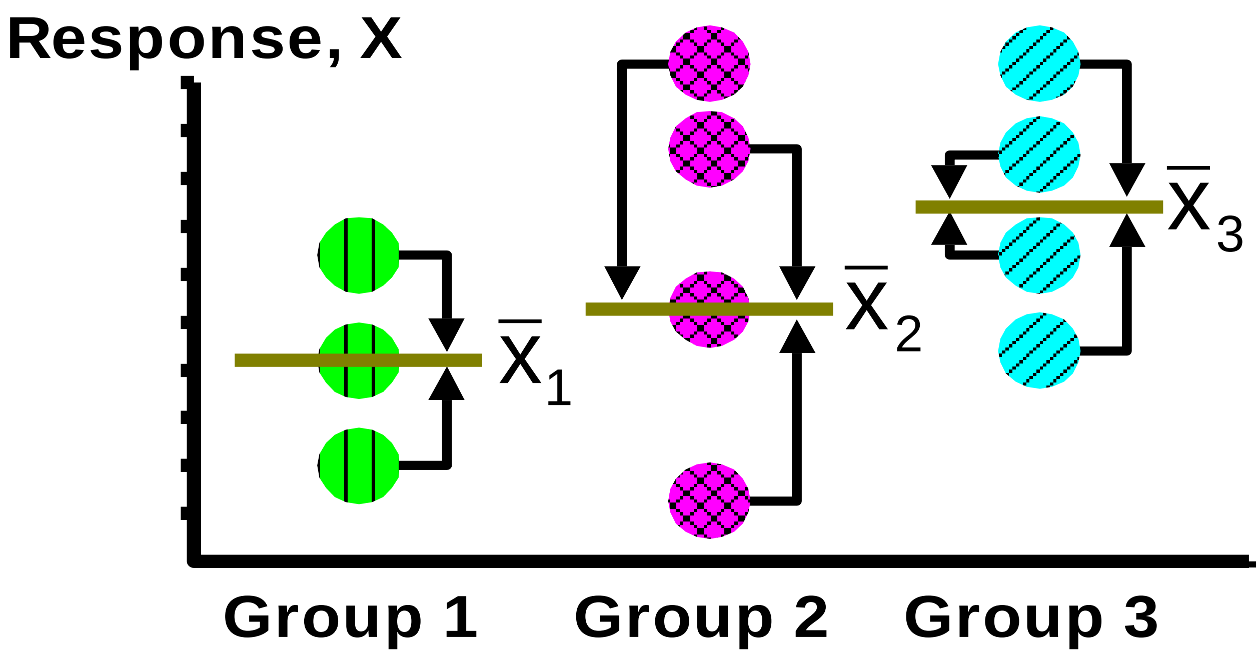
$$MSW = \frac{SSW}{n-K}$$

Mean Square Within = SSW/degrees of freedom

Within-Group Variation

(continued)

$$SSW = (x_{11} - \bar{x}_1)^2 + (x_{12} - \bar{x}_1)^2 + \ldots + (x_{Kn_K} - \bar{x}_K)^2$$



Between-Group Variation

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

$$\boxed{SSG = \sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x})^2}$$

Where:

SSG = Sum of squares between groups

K = number of groups

n_i = sample size from group i

$\bar{x}_i$ = sample mean from group i

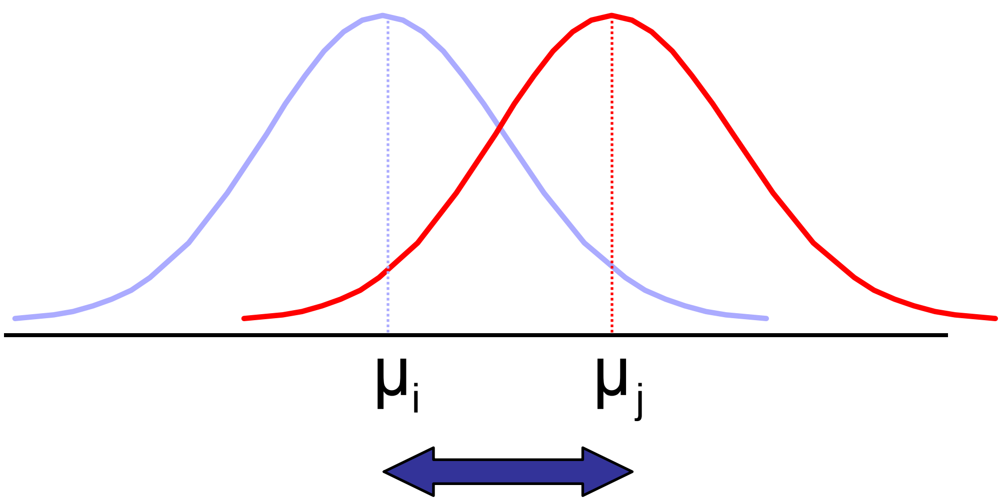
$\bar{x}$ = grand mean (mean of all data values)

Between-Group Variation

(continued)

$$SSG = \sum_{i=1}^{K} n_i (\bar{x}_i - \bar{x})^2$$

Variation Due to
Differences Between Groups



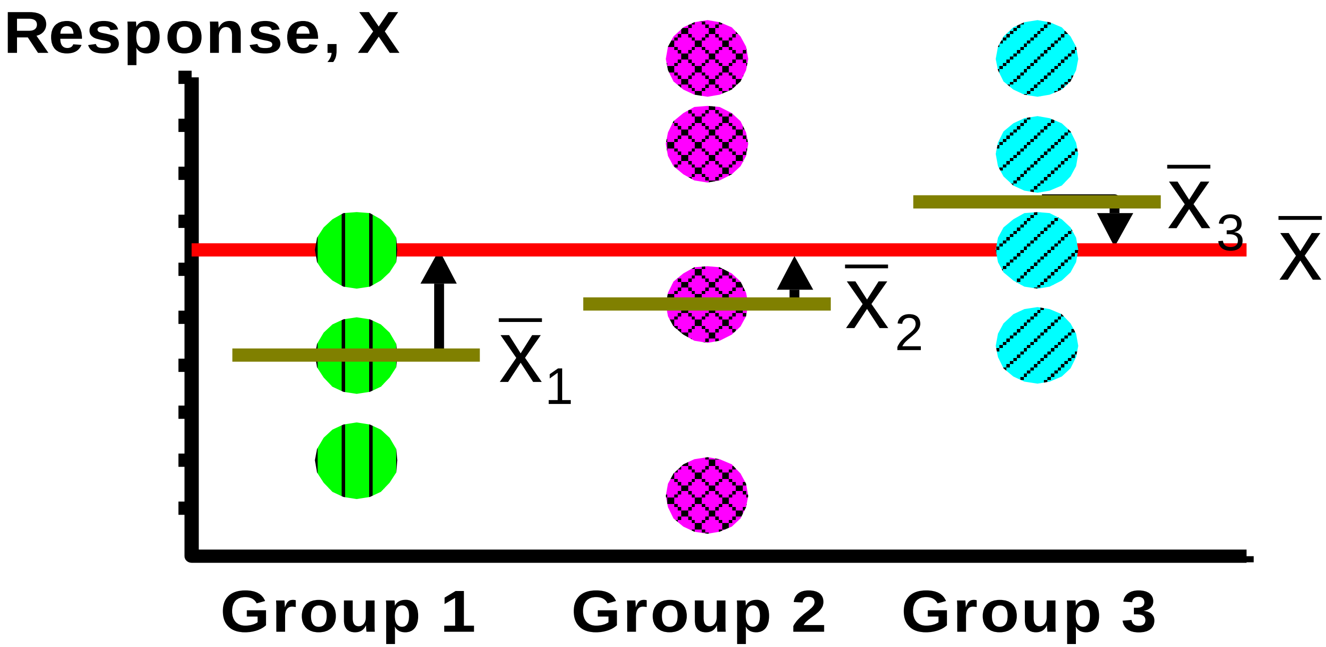
$$MSG = \frac{SSG}{K-1}$$

Mean Square Between Groups
= SSG/degrees of freedom

Between-Group Variation

(continued)

$$SSG = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_K(\bar{x}_K - \bar{x})^2$$



Obtaining the Mean Squares

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

$$\boxed{MSW = \frac{SSW}{n-K}}$$

$$\boxed{MSG = \frac{SSG}{K-1}}$$

One-Way ANOVA Table

Source of Variation	SS	df	MS (Variance)	F ratio
Between Groups	SSG	K - 1	$MSG = \frac{SSG}{K - 1}$	$F = \frac{MSG}{MSW}$
Within Groups	SSW	n - K	$MSW = \frac{SSW}{n - K}$	
Total	SST = SSG+SSW	n - 1		

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

One-Factor ANOVA F Test Statistic

H_0: $\mu_1 = \mu_2 = \ldots = \mu_K$

H_1: At least two population means are different

- Test statistic

$$F = \frac{MSG}{MSW}$$

MSG is mean squares between variances

MSW is mean squares within variances

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

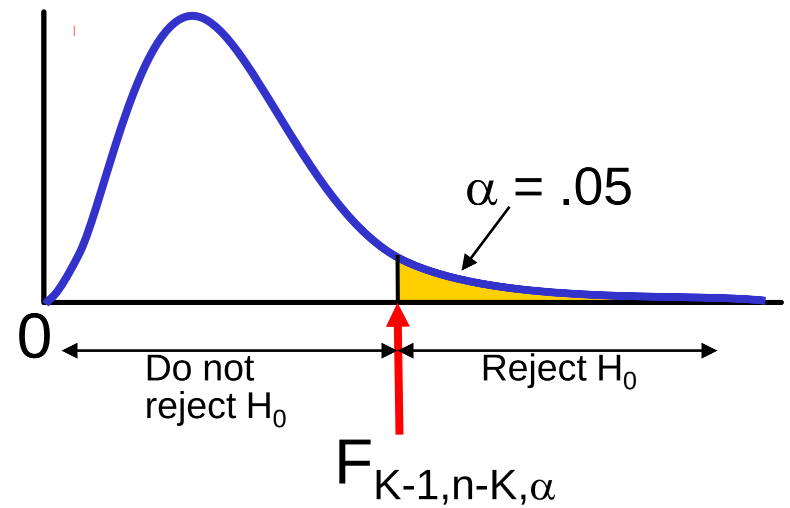
Interpreting the F Statistic

- The F statistic is the ratio of the between estimate of variance and the within estimate of variance
  - The ratio must always be positive
  - $df_1 = K - 1$ will typically be small
  - $df_2 = n - K$ will typically be large

Decision Rule:

- Reject H_0 if

  $F > F_{K-1,n-K,\alpha}$



One-Factor ANOVA F Test Example

You want to see if different three drugs yield different recovery time in hours. You randomly select five patients those gone under treatment for each drug. At the .05 significance level, is there a difference in mean recovery time in hours?

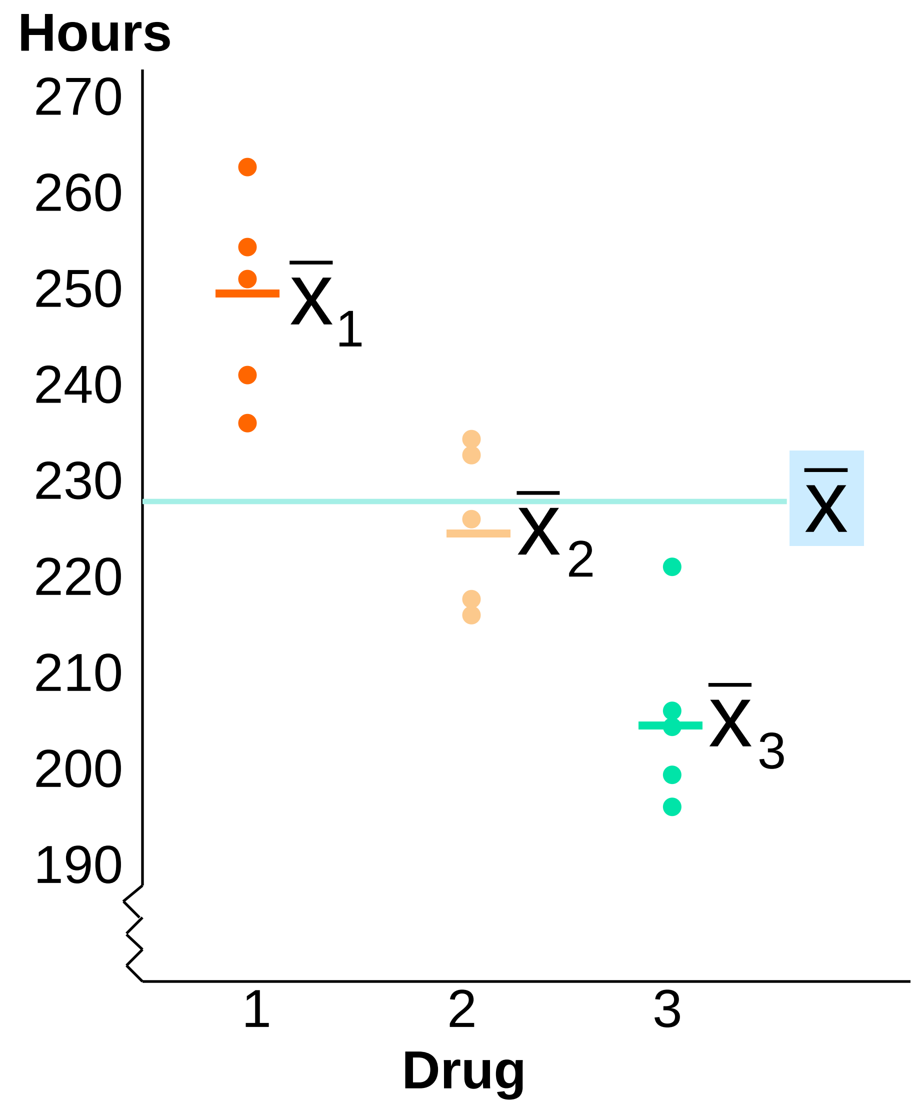
Drug 1	**Drug 2**	**Drug 3**
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

One-Factor ANOVA Example: Scatter Diagram

Drug 1	**Drug 2**	**Drug 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-Factor ANOVA Example Computations

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

$\bar{x}_1 = 249.2$ $n_1 = 5$

$\bar{x}_2 = 226.0$ $n_2 = 5$

$\bar{x}_3 = 205.8$ $n_3 = 5$

$n = 15$

$\bar{x} = 227.0$

$K = 3$

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

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

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

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

$$F = \frac{2358.2}{93.3} = 25.275$$

One-Factor ANOVA Example Solution

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

H_1: μ_i not all equal

$\alpha = .05$

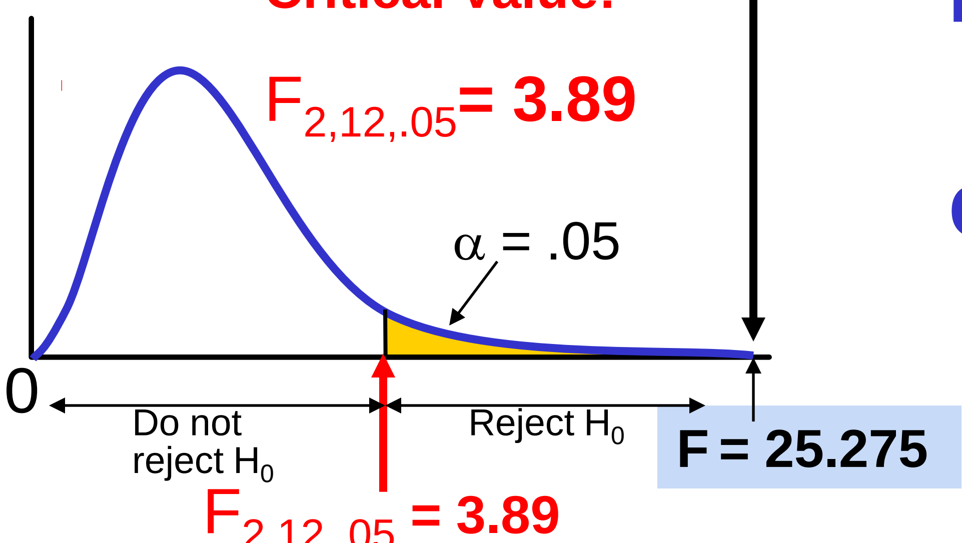
$df_1 = 2 \quad df_2 = 12$

Test Statistic:

$$F = \frac{MSA}{MSW} = \frac{2358.2}{93.3} = 25.275$$

Critical Value:

$F_{2,12,.05} = 3.89$



Decision:

Reject H_0 at $\alpha = 0.05$

Conclusion:

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

Thank you