::Analysis of Variance(ANOVA)::

Understanding of

Between and Within Group Variation With Diagram

Preview

- Analysis of variance (ANOVA) is a method for testing the hypothesis that three or more population means are equal.

- For example:

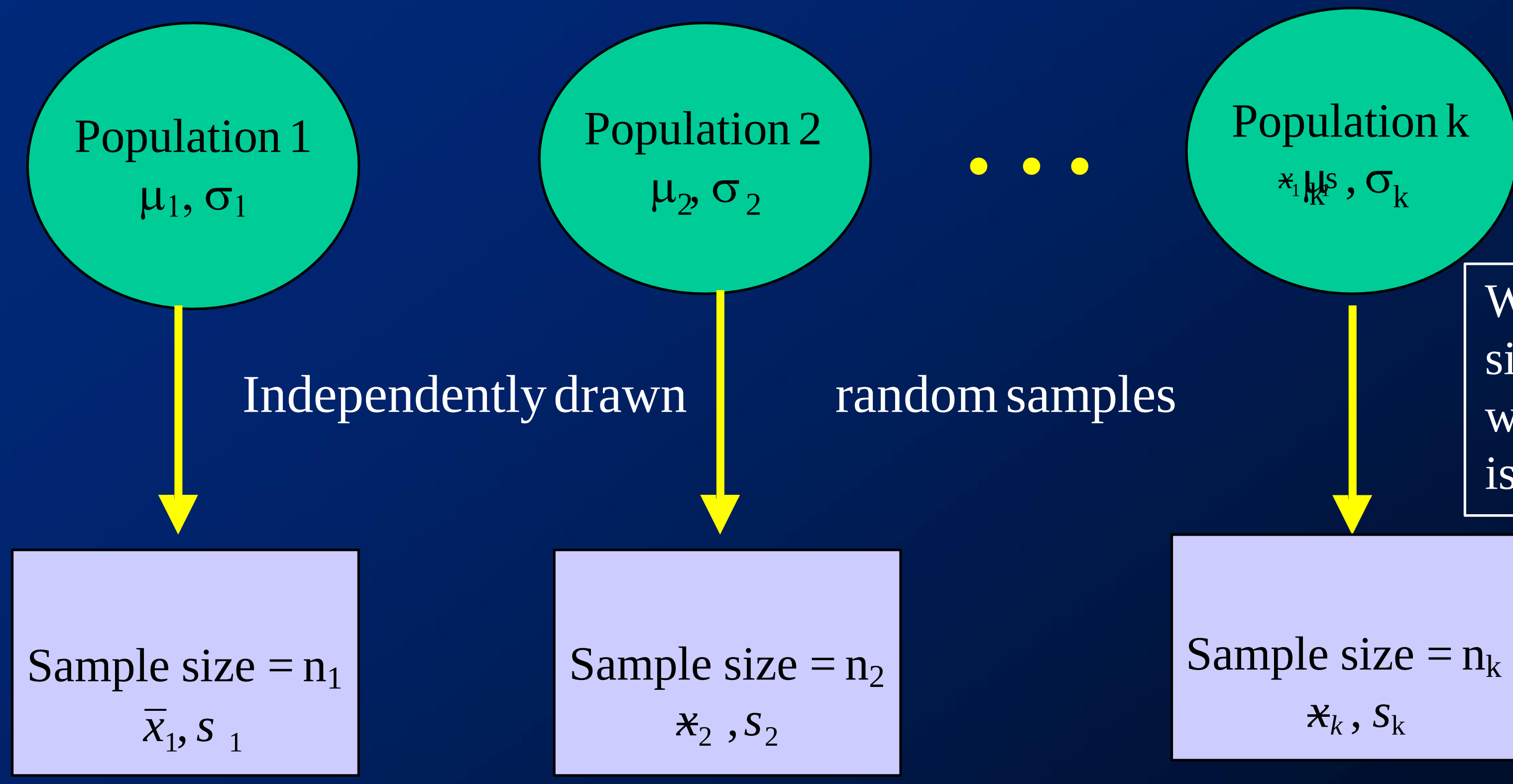
  H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots \mu_k$

  H_1: At least one mean is different

- F test is used

Oneway ANOVA

If an experiment is conducted which randomly allocate units to the *k* “treatment” groups, and independent samples are taken.



When sample sizes are equal we say design is **balanced**.

Questions:

1) ***Do the k population means differ somehow, i.e. are at least two treatment means differ?***

2) ***If so, which pairs of means differ?***

Motivating Example:
Treating Anorexia Nervosa

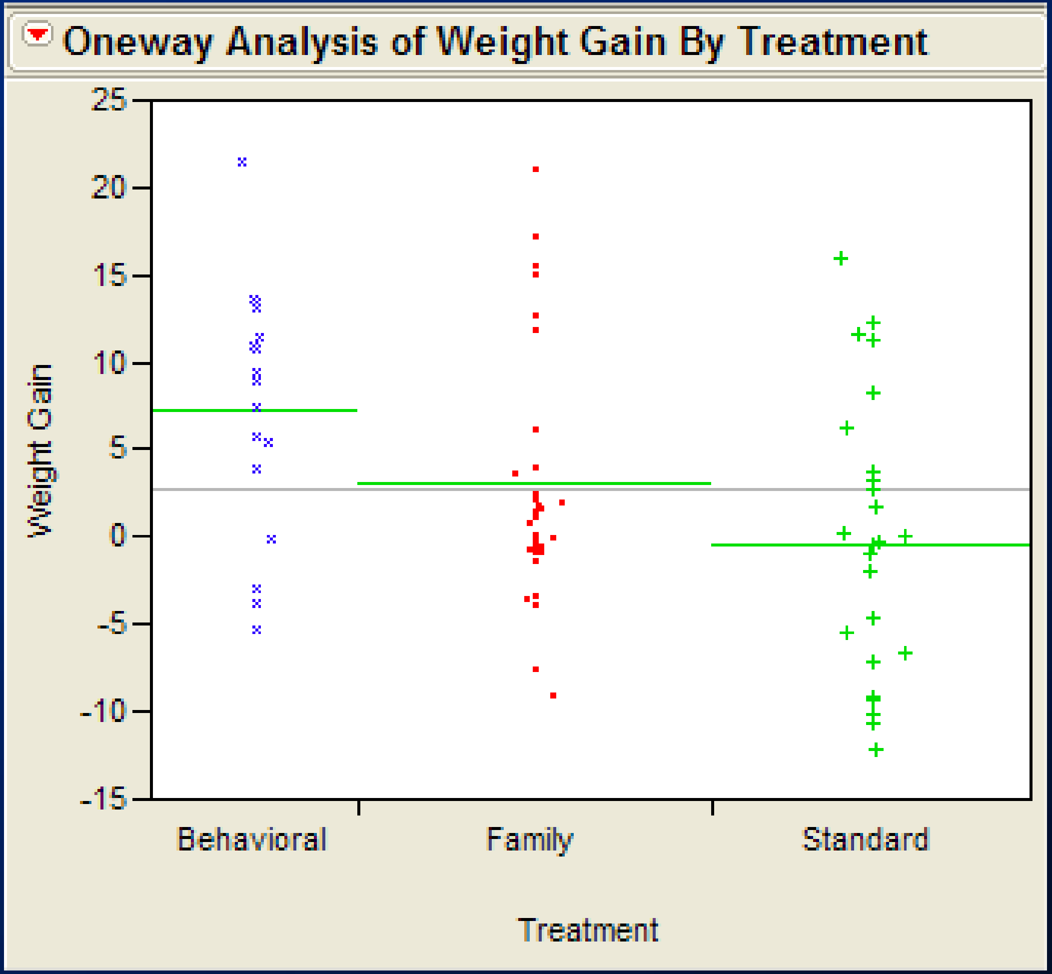
Anorexia patients were randomly assigned to receive one of three different therapies:

- Standard – current accepted therapy
- Family – family therapy
- Behavior – behavioral cognitive therapy

For each patient weight change during the course of the treatment was recorded. The duration of treatment was the same for all patients, regardless of therapy received.

Question: Do the patients in the three treatments/ therapies have different mean weight gains?

Motivating Example: Treating Anorexia Nervosa



Do the patients in the three treatments/therapies have different mean weight gains?

Analysis of Variance

- Analysis of Variance is a widely used statistical technique that partitions the total variability in our data into components of variability that are used to test hypotheses.

- In One-way ANOVA, we wish to test the hypothesis:

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

against:

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

Analysis of Variance

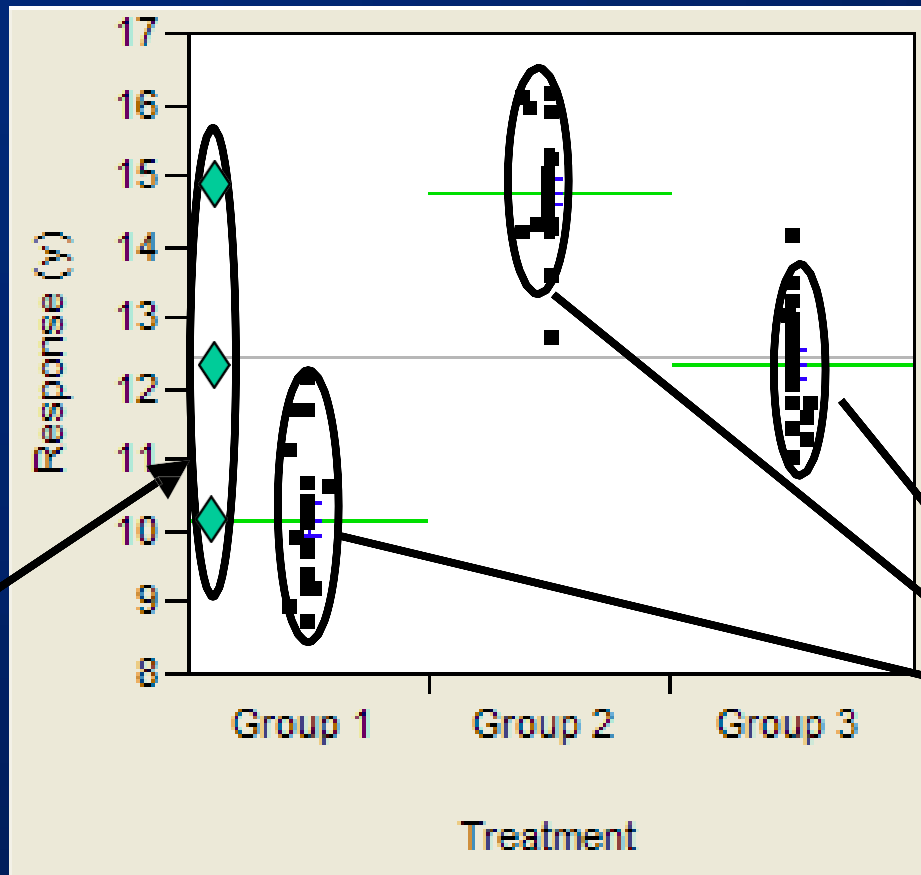
- In ***ANOVA***, we compare the ***between-group variation*** with the ***within-group variation*** to assess whether there is a difference in the population means.

- Thus by comparing these two measures of variance (spread) with one another, we are able to detect if there are true differences among the underlying group population means.

Analysis of Variance

- If the variation between the sample means is large, relative to the variation within the samples, then we would be likely to detect significant differences among the sample means.

Between Group Variation is Large Compared to Within Group Variation



Between-group variation (group means are diamonds)

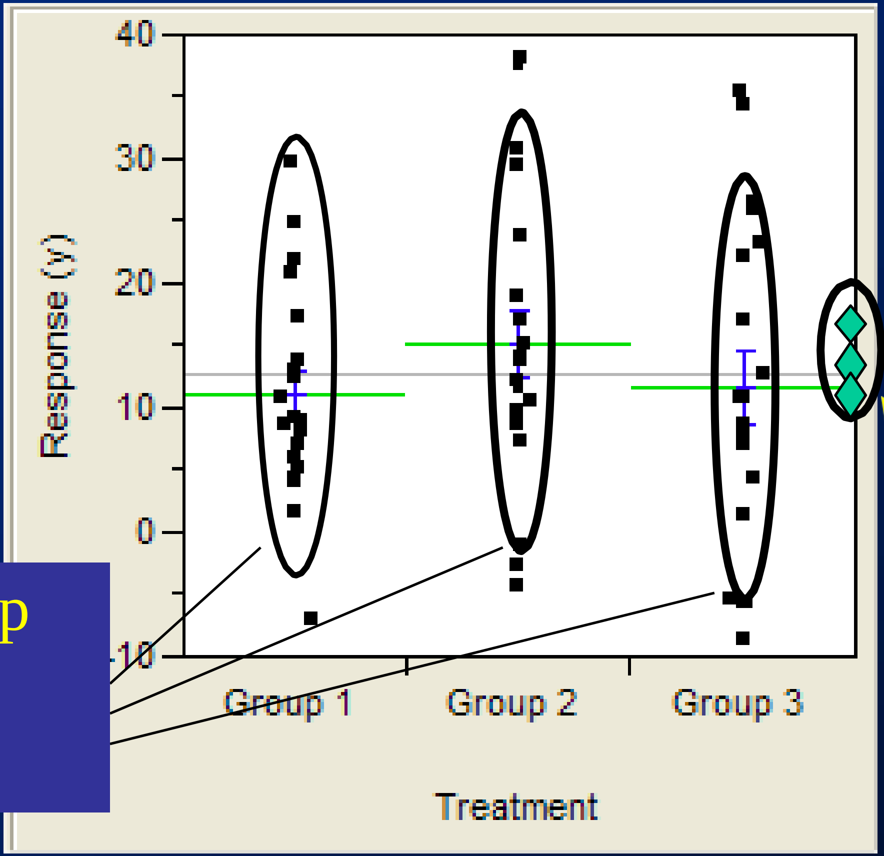
Within-group variation

Here we would almost certainly reject the null hypothesis.

Analysis of Variance

- If the variation between the sample means is small, relative to the variation within the samples, then there would be considerable overlap of observations in the different samples, and we would be unlikely to detect any differences among the population means.

Between Group Variation is Small Compared to Within Group Variation



Within-group variation is large.

Between-group variation is small

Here we would fail to reject the null hypothesis.

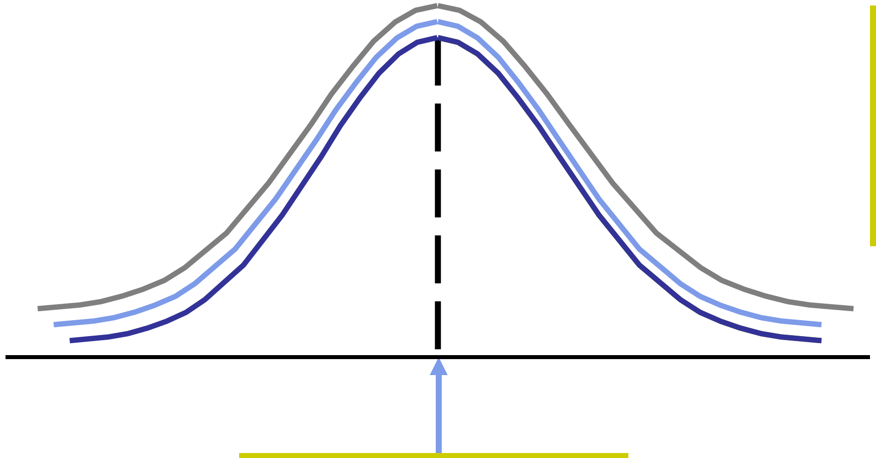
Hypotheses: One-Way ANOVA

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$
  - All population means are equal
  - i.e., no treatment effect (no variation in means among groups)
- H_1 : Not al of the population means are the same
  - At least one population mean is different
  - i.e., there is a treatment (groups) effect
  - Does not mean that all population means are different (at least one of the means is different from the others)

Hypotheses: One-Way ANOVA

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

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



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

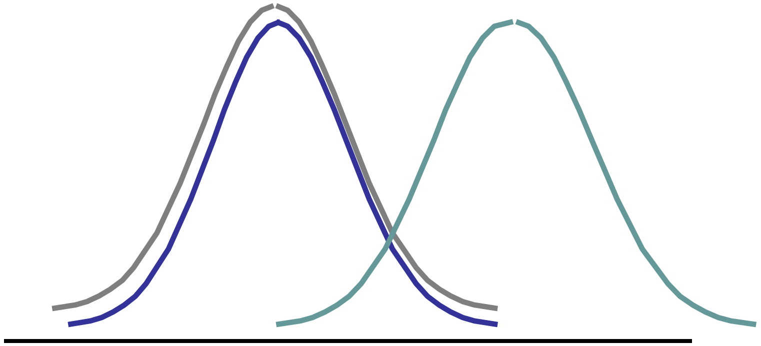
$$\mu_1 = \mu_2 = \mu_3$$

Hypotheses: One-Way ANOVA

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

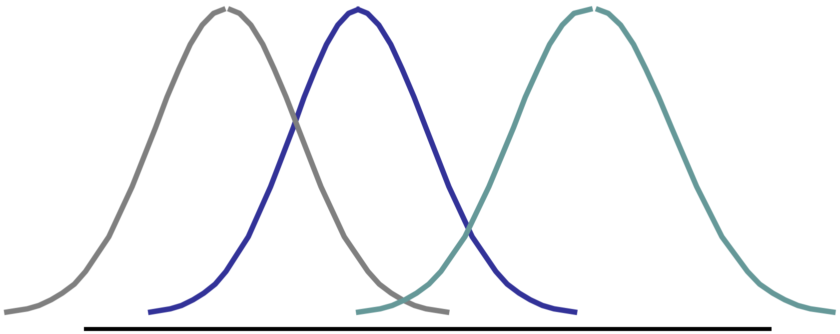
H_1 : Not all μ_j are the same

At least one mean is different:
The Null Hypothesis is NOT true
(Treatment Effect is present)



$\mu_1 = \mu_2 \neq \mu_3$

or



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

Partitioning the Variation

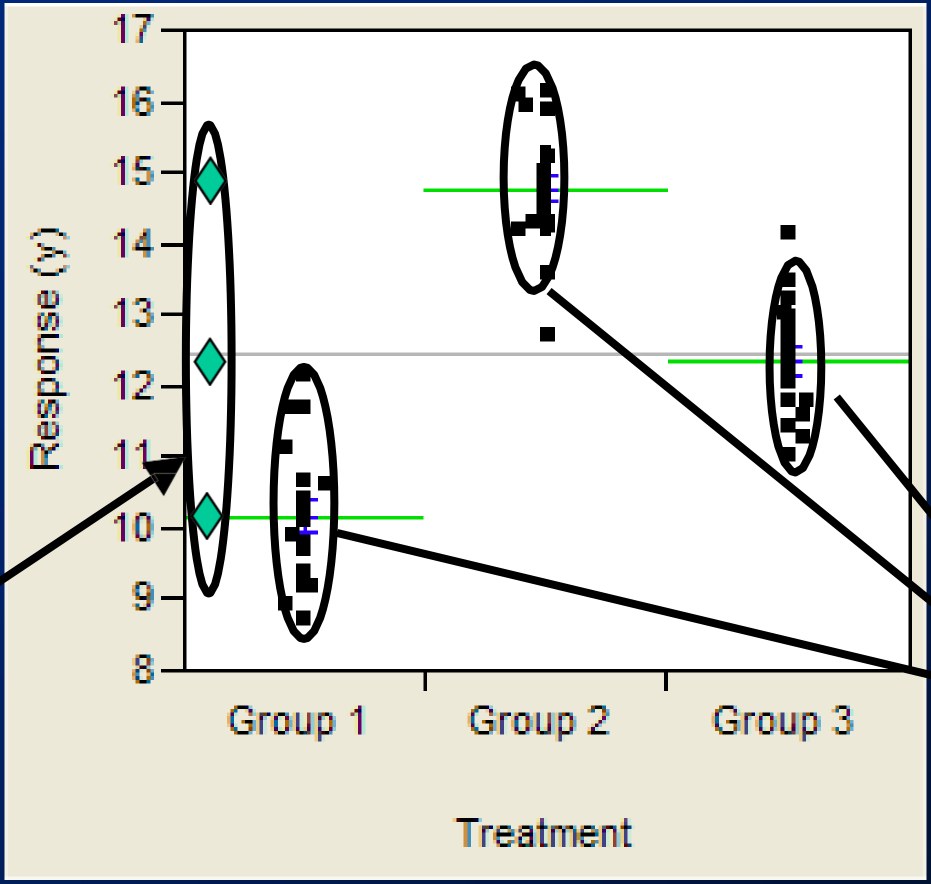
SST = SSA + SSW

Total Sum of Squares = Total Variation (SST)

Sum of Squares Among Groups = Among-Group Variation **=** dispersion between the factor sample means (SSA)

Sum of Squares Within Groups = Within-Group Variation = dispersion that exists among the data values within the particular factor levels (SSW)

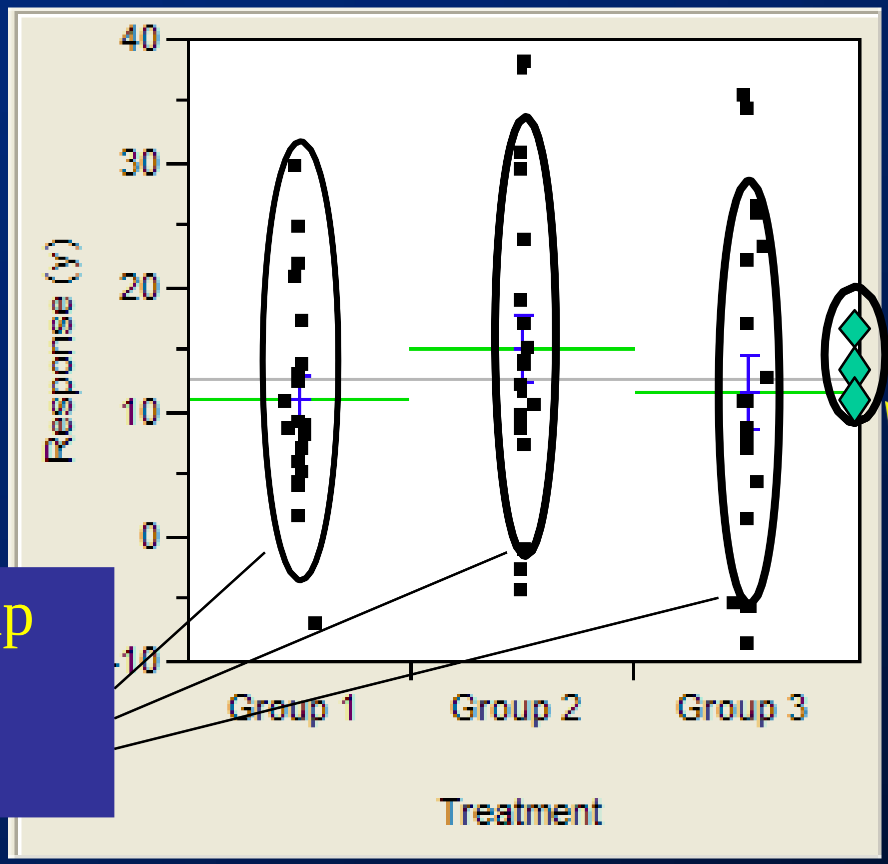
SSA - Among Group Variation



Between-group variation (group means are diamonds)

Within-group variation

SSW - Within Group Variation



Within-group variation is large.

Between-group variation is small

Obtaining the Mean Squares

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

Mean Squares Among Variance

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

Mean Squares Within

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

Mean Squares Total

c = number of groups

n = sum of the sample sizes from all groups

One-Way ANOVA Table

Source of Variation	df	SS	MS (Variance)	F-Ratio
Among Groups (Regression)	c-1	SSA	MSA	$F = \frac{MSA}{MSW}$
Within Groups (Residual)	n-c	SSW	MSW	
Total	n-1	SST = SSA+SSW		

c = number of groups

n = sum of the sample sizes from all groups

df = degrees of freedom

One-Way ANOVA Test Statistic

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

H_1: At least two population means are different

- Test statistic

$$\boxed{F = \frac{MSA}{MSW}}$$

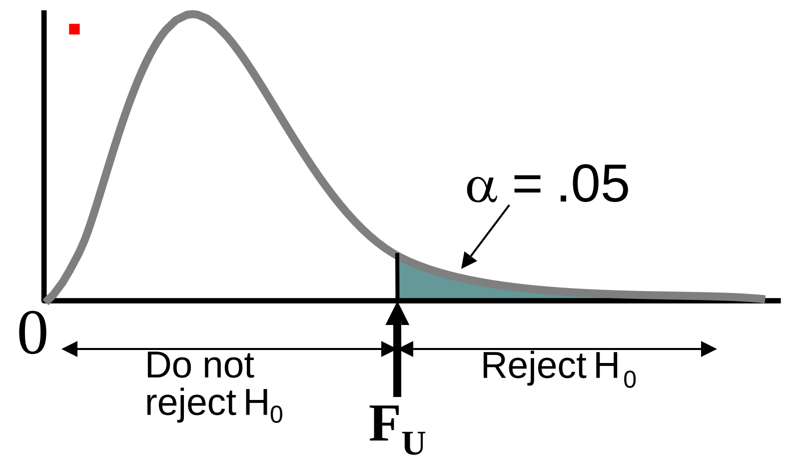
  - *MSA* is mean squares among variances
  - *MSW* is mean squares within variances
- Degrees of freedom
  - $df_1 = c - 1$ (c = number of groups)
  - $df_2 = n - c$ (n = sum of all sample sizes)

One-Way ANOVA Test Statistic

- The F statistic is the ratio of the among variance to the within 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 $F > F_U$, otherwise do not reject H_0



One-Way ANOVA 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 .05 significance level, is there a difference in mean distance?

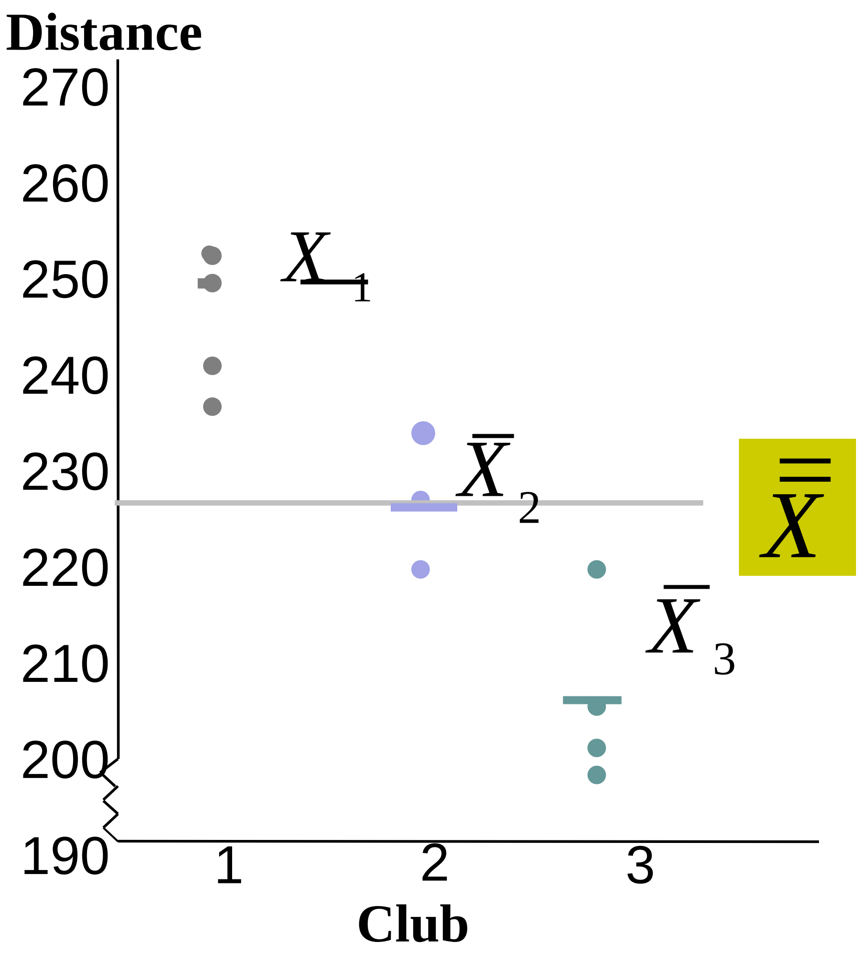
Club 1	**Club 2**	**Club 3**
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

One-Way ANOVA Example

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

$\bar{X}_1 = 249.2$	$n_1 = 5$
$\bar{X}_2 = 226.0$	$n_2 = 5$
$\bar{X}_3 = 205.8$	$n_3 = 5$

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

$n = 15$

$c = 3$

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

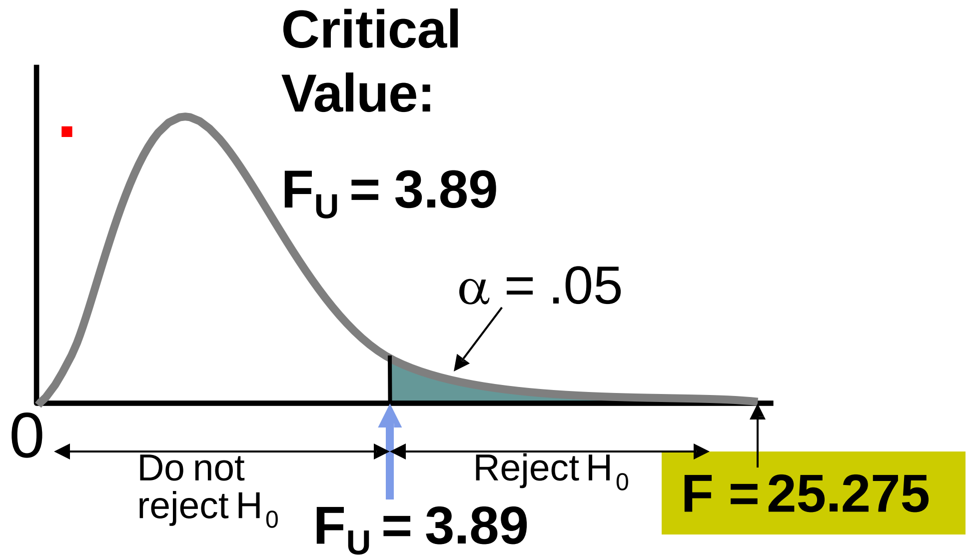
$$SSW = (254 - 249.2)^2 + (263 - 249.2)^2 + \ldots + (204 - 205.8)^2 = 1119.6$$

One-Way ANOVA Example

MSA = 4716.4 / (3-1) = 2358.2

MSW = 1119.6 / (15-3) = 93.3

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



One-Way ANOVA Example

- H_0: $\mu_1 = \mu_2 = \mu_3$
- H_1: μ_j not all equal
- $\alpha = .05$
- $df_1 = 2$ $df_2 = 12$

Decision:

Reject H_0 at $\alpha = 0.05$

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

Two-Way Analysis of Variance

Two-Way ANOVA involves two factors.

- **Test performance of clubs along with weight of the ball**
- **2 factors**

There is an interaction between two factors if the effect of one of the factors changes for different categories of the other factor.

Two-Way ANOVA

- **Test performance of clubs along with weight of the ball**
- **2 factors**
- **Calculate interaction between these factors to find out whether one factor effect the other one**