

ANOVA: Analysis of Variance

$$F = \frac{MSB}{MSW} = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{\text{Variance bet}^n \text{ groups}}{\text{Variance within groups}}$$

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} \quad MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}}$$

$$df_{\text{between}} = k - 1$$

k = no of groups

$$df_{\text{within}} = n_T - k$$

n_T = no of elements across all groups

$$SS_{\text{between}} = \sum \frac{(\sum x)^2}{n} - \frac{(\sum \sum x)^2}{n_T}$$

$$SS_{\text{within}} = \sum \sum x^2 - \sum \frac{(\sum x)^2}{n}$$

For each group

Across groups

n $\xrightarrow{\text{Add for all groups}}$

n_T

$\sum x$ \longrightarrow

$\sum \sum x \rightarrow \frac{(\sum \sum x)^2}{n_T}$

$\frac{(\sum x)^2}{n}$ \longrightarrow

$\sum \frac{(\sum x)^2}{n}$

$\sum x^2$ \longrightarrow

$\sum \sum x^2$

/ ANOVA is a statistical technique that is used to check if the means of two or more groups are significantly different /

Ex.1: A recent study claims that using music in a class enhances the concentration and consequently helps students absorb more information. To figure this out, it was decided to implement it on a smaller group of randomly selected students from three different classes. Three different groups of ten randomly selected students (all of the same age) from three different classrooms and was provided with a different environment for students to study. Classroom A had constant music being played in the background, classroom B had variable music being played and classroom C was a regular class with no music playing. After one month, a test was conducted for all the three groups and collected their test scores. The test scores that obtained were as follows:

	Test scores of students (out of 10)									
Class A (constant sound)	7	9	5	8	6	8	6	10	7	4
Class B (variable sound)	4	3	6	2	7	5	5	4	1	3
Class C (no sound)	6	1	3	5	3	4	6	5	7	3

Test whether at least one of the three samples have significantly different means? $[F(2,27) = 3.354]$

Solⁿ $H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: \mu_i \neq \mu_j \quad i, j = 1, 2, 3$

Within Groups

	n	$\sum x$	$\frac{(\sum x)^2}{n}$	$\sum x^2$
class A	10	70	490	520
class B	10	40	160	190
class C	10	43	184.9	215

Across Groups

$n_T = n_1 + n_2 + n_3 = 30 \quad \sum \frac{(\sum x)^2}{n} = 834.9$

$\sum \sum x = 153 \quad \sum \sum x^2 = 925$

$\frac{(\sum \sum x)^2}{n_T} = 780.3$

$SS_b = \sum \frac{(\sum x)^2}{n} - \frac{(\sum \sum x)^2}{n_T} = 834.9 - 780.3 = 54.6$

$SS_w = \sum \sum x^2 - \sum \frac{(\sum x)^2}{n} = 925 - 834.9 = 90.1$

$df_b = 3 - 1 = 2, \quad df_w = 30 - 3 = 27$

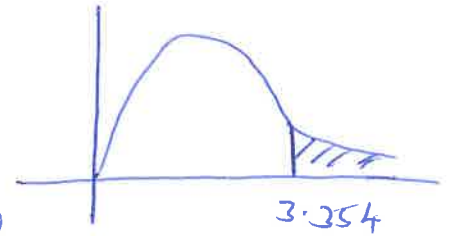
$$F = \frac{MS_B}{MS_W}$$

$$\text{Next, } MS_B = \frac{SS_B}{df_B} = \frac{54.6}{2} = 27.3$$

$$MS_W = \frac{SS_W}{df_W} = \frac{90.1}{27} = 3.34$$

$$F = \frac{MS_B}{MS_W} = \frac{27.3}{3.34} = 8.17 \in$$

critical
region



$\therefore H_0$ is rejected

Hence, one of the three samples have significantly different mean.

Ex. 2. The three drying techniques for curing a glue were studied & the following zones were observed

Formula A : 13 10 8 11 8 -

" B : 13 11 14 14 - -

" C : 4 1 3 4 2 4

Test the hypothesis that the average zones for the three formulae are same @ 1% level of significance

Solⁿ $H_0: M_1 = M_2 = M_3$

$H_1: M_1 \neq M_2 \neq M_3$ ~~$M_1 \neq M_2 \neq M_3$~~

Within Groups

	<u>n</u>	<u>$\sum x$</u>	<u>$(\sum x)^2/n$</u>	<u>$\sum x^2$</u>
A :	5	50	500	518
B :	4	52	676	682
C :	6	18	154	62

Across Groups

$n_T = n_A + n_B + n_C = 15$ $\sum (\sum x)^2/n = 1230$

$\sum \sum x = 120$

$\sum \sum x^2 = 1262$

$\frac{(\sum \sum x)^2}{n_T} = 960$

$df_B = 3 - 1 = 2$

$df_W = 15 - 3 = 12$

$$MS_B \leftarrow \frac{SS_B}{df_B}$$

$$SS_B = \sum \frac{(\sum x)^2}{n} - \frac{(\sum \sum x)^2}{n_T} = 1230 - 960 = 270$$

$$SS_W = \sum \sum x^2 - \sum \frac{(\sum x)^2}{n} = 1262 - 1230 = 32$$

$$\therefore MS_B = \frac{SS_B}{df_B} = \frac{270}{2} = 135$$

$$MS_W = \frac{SS_W}{df_W} = \frac{32}{12} = 2.67$$

$$F = \frac{MS_B}{MS_W} = \frac{135}{2.67} = 50.58 \in$$

Critical
region

H_0 is rejected

Hence Avg zones for 3-formulas are not same

