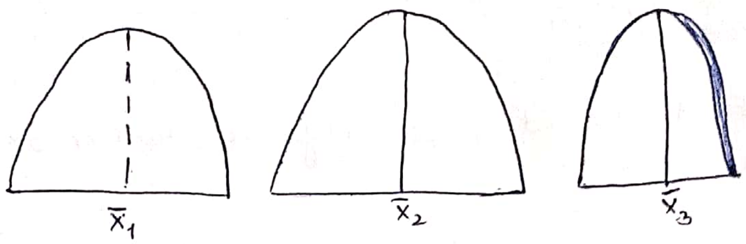
Analysis of Variance (ANOVA)

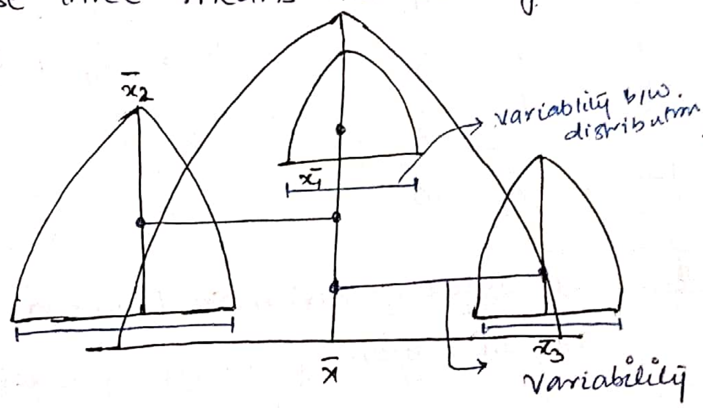
Introduction & Basics

- For comparison of more than two population or population having more than two subgroups, ANOVA technique should be used.



$\therefore$ 3-different means
$\bar{x}_1 \neq \bar{x}_2 \neq \bar{x}_3$

Q. Do all these Three means are coming from same population?



$\therefore$ $\bar{x}_1 \simeq \bar{x}$
$\bar{x}_2$ too far from $\bar{x}$
$\bar{x}_3$ is less far $\bar{x}$

ANOVA : $\dfrac{\text{Variability between the means}}{\text{Variability within the distribution}}$

Total Variance = Variability between the means + Variability within the distribution.

Assumptions:

1: Each population is having normal distribution.

2: The population from which the sample are drawn have the equal variance, i.e., $s_1^2 = s_2^2 = s_3^2 \dots s_k^2$ for k samples.

3: Each sample is drawn randomly & they are independent.

$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_n.$

$H_a: \mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \neq \mu_n.$

Classification: One factor - one way ANOVA; Two-factors - Two way ANOVA.

Example:
the farmer wants to know if there is a significant difference in the yield of three different types of corn seed. He plants equal amounts of each seed type in separate fields, and at harvest time, he measures the yield of each field. He can use a one-way ANOVA to analyze the data and determine if there is a significant difference in yield between the three types of seed. The independent variable in this case is the type of seed (A, B, and C) and the dependent variable is the yield.

ONE-WAY ANOVA

- It is classified according to only one factor or one criteria.

A one-way ANOVA is a statistical method used to determine whether there is a significant difference between the means of two or more groups. It is used when you have one independent variable and one dependent variable.

In summary, one-way ANOVA is used to compare means of one independent variable and one dependent variable.

Example:

A	B	C
2	3	4
4	5	6
6	7	8
12	15	18

① STEP-I

$H_0: \bar{X}_A = \bar{X}_B = \bar{X}_C$

$H_a: \bar{X}_A \neq \bar{X}_B \neq \bar{X}_C$

② STEP-II: Calculate the variance between the samples.

(a) calculation of mean of each sample.

$\bar{x}_A = 12/3 = 4;\quad \bar{x}_B = \frac{15}{3} = 5;\quad \bar{x}_C = 18/3 = 6$

(b) Calculation of grand average of means.

$$\bar{\bar{X}} = \frac{\bar{X}_A + \bar{X}_B + \bar{X}_C}{3} = \frac{4+5+6}{3} = \frac{15}{3} = 5$$

(c) Take the difference between the means of various samples & $\bar{\bar{x}}$ & square it.

$(\bar{x}_A - \bar{\bar{x}})$	$(\bar{x}_A - \bar{\bar{x}})^2$	$(\bar{x}_B - \bar{\bar{x}})$	$(\bar{x}_B - \bar{\bar{x}})^2$	$(\bar{x}_C - \bar{\bar{x}})$	$(\bar{x}_C - \bar{\bar{x}})^2$
-1	1	0	0	1	1
-1	1	0	0	1	1
-1	1	0	0	1	1
	3		0		3

$3+0+3 = 6$

(3) STEP-III : Calculate the variance within the Sample.

(a) Calculation of mean for each sample.

$\bar{X}_A = 4$, $\bar{X}_B = 5$, & $\bar{X}_C = 6$

(b) Take the deviations of the various items in a sample from the mean values of the respective sample & squared it.

$(A-\bar{X}_A)$	$(A-\bar{X}_A)^2$	$(B-\bar{X}_B)$	$(B-\bar{X}_B)^2$	$(C-\bar{X}_C)$	$(C-\bar{X}_C)^2$
-2	4	-2	4	-2	4
0	0	0	0	0	0
2	4	2	4	2	4
$\Sigma(A-\bar{X}_A)^2 =$	8		8		8

Sum of The Square within the sample $(\Sigma(x-\bar{x})^2 = 24$

(c) Calculation of The ratio of F.

Source of Variance	Sum of Squares	Degree of Freedom	Mean Sum of square	F
Between The Sample	SSC = 6	$v_1 = c-1$ $\Rightarrow 3-1=2$	MSC = SSC/c-1 6/2 = 3	$\frac{MSC}{MSE}$
Within The Sample	SSE = 24	$v_2 = n-c$ $\Rightarrow 9-3=6$	MSE = SSE/n-c 24/6 = 4	3/4 = 0.75

∴ SSC = Sum of sq. b/w The Samples (columns)

SSE = Sum of sq. within the samples (rows)

MSC = Mean sum of sq. b/w the samples

MSE = Mean sum of sq. within The samples.

(d) Compare the F (calculated value) with The F (tabulated value)

$F_{tab} = 5.14$

H_0 is passed & accepted.

QUESTION 1:

Solution.

A	B	C
9	13	14
11	12	13
13	10	17
9	15	7
8	5	9
$\Sigma_A = 50$	$\Sigma_B = 55$	$\Sigma_C = 60$

$\bar{x}_A = 50/5 = 10$; $\bar{x}_B = 55/5 = 11$; $\bar{x}_C = 60/5 = 12$

$$\bar{\bar{x}} = \frac{\bar{x}_A + \bar{x}_B + \bar{x}_C}{3} = \frac{10+11+12}{3} = 33/3 = 11$$

- Calculation of SSC

$(\bar{x}_A - \bar{\bar{x}})$	$(\bar{x}_A - \bar{\bar{x}})^2$	$(\bar{x}_B - \bar{\bar{x}})$	$(\bar{x}_B - \bar{\bar{x}})^2$	$(\bar{x}_C - \bar{\bar{x}})$	$(\bar{x}_C - \bar{\bar{x}})^2$
-1	1	0	0	1	1
-1	1	0	0	1	1
-1	1	0	0	1	1
-1	1	0	0	1	1
-1	1	0	0	1	1
$\Sigma(x-\bar{\bar{x}})$	5		0		5

$$SSC = \Sigma(\bar{x}_A - \bar{\bar{x}})^2 + \Sigma(\bar{x}_B - \bar{\bar{x}})^2 + \Sigma(\bar{x}_C - \bar{\bar{x}})^2$$

$$\Rightarrow 5 + 0 + 5 = 10$$

- Calculation of SSE:

$(A - \bar{x}_A)$	$(A - \bar{x}_A)^2$	$(B - \bar{x}_B)$	$(B - \bar{x}_B)^2$	$(C - \bar{x}_C)$	$(C - \bar{x}_C)^2$
-1	1	2	4	+2	4
1	1	1	1	1	1
3	9	-1	1	5	25
-1	1	4	16	-5	25
-2	4	-6	36	-3	9
$\Sigma(x-\bar{x})^2$ 16			58		64

$$SSE = \Sigma(A - \bar{x}_A)^2 + \Sigma(B - \bar{x}_B)^2 + \Sigma(C - \bar{x}_C)^2 = 16 + 58 + 64 = 138$$

Source of Variance	Sum of Square	Degree of Freedom	Mean Square	F
Between the Sample	SSC = 10	$v_1 = C-1 = 3-1 = 2$	$MSC = SSC/v_1 = 10/2 = 5$	$\frac{MSC}{MSE} = \frac{5}{11.5}$
Within the Sample	SSE = 138	$v_2 = n-C = 15-3 = 12$	$MSE = SSE/v_2 = 138/12 = 11.5$	$= \boxed{0.435}$

Calculated F-Value = 0.435
Tabulated F-Value = 3.89

Ho is passed & accepted. All the schools are having same output. No significantee variation in the schools.

QUESTION NO: 02

Given Data:

A	B	C
8	8	17
10	6	10
7	11	12
14	8	12
11	8	15
16	13	12

$\Sigma_A = 66$ $\Sigma_B = 54$ $\Sigma_C = 78$

$\bar{x}_A = \frac{66}{6} = 11$; $\bar{x}_B = \frac{54}{6} = 9$; $\bar{x}_C = \frac{78}{6} = 13$

$\bar{\bar{x}} = \frac{\bar{x}_A + \bar{x}_B + \bar{x}_C}{3} = \frac{11+9+13}{3} = 11$

- Calculation of SSC:

$(\bar{x}_A - \bar{\bar{x}})$	$(\bar{x}_A - \bar{\bar{x}})^2$	$(\bar{x}_B - \bar{\bar{x}})$	$(\bar{x}_B - \bar{\bar{x}})^2$	$(\bar{x}_C - \bar{\bar{x}})$	$(\bar{x}_C - \bar{\bar{x}})^2$
0	0	-2	4	2	4
0	0	-2	4	2	4
0	0	-2	4	2	4
0	0	-2	4	2	4
0	0	-2	4	2	4
0	0	-2	4	2	4
$\Sigma(x-\bar{x})^2$	0		24		24

$$SSC = \Sigma(\bar{X}_A - \bar{\bar{X}})^2 + \Sigma(\bar{X}_B - \bar{\bar{X}})^2 + \Sigma(\bar{X}_C - \bar{\bar{X}})^2$$
$$= 0 + 24 + 24 = 48$$

Calculation of SSE

$(A - \bar{X}_A)$	$(A - \bar{X}_A)^2$	$(B - \bar{X}_B)$	$(B - \bar{X}_B)^2$	$(C - \bar{X}_C)$	$(C - \bar{X}_C)^2$
-3	9	-1	1	4	16
-1	1	-3	9	-3	9
-4	16	+2	4	-1	1
3	9	-1	1	-1	1
0	0	-1	1	2	4
5	25	4	16	-1	1
$\Sigma(A - \bar{X}_A)^2$	60		32		32

$$SSE = \Sigma(A - \bar{X}_A)^2 + \Sigma(B - \bar{X}_B)^2 + \Sigma(C - \bar{X}_C)^2$$
$$= 60 + 32 + 32 = 124$$

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F
Between the Samples	SSC = 48	$\nu_1 = c - 1 = 2$	$MSC = SSC/\nu_1$ $48/2 = 24$	$\frac{MSC}{MSE}$
Within the Samples	SSE = 124	$\nu_2 = n - c = 15$	$MSE = SSE/\nu_2$ $= \frac{124}{15} = 8.27$	$= \frac{24}{8.27}$ $= \boxed{2.90}$

Calculated F value = 2.90

Tabulated F " = 3.68

H_0 is correct and accepted.

The given means are equal as per one-way ANOVA.

TWO WAY ANOVA

- It is classified according to two factors or two criteria.

\ DAYS	A	B	C	D
MON	2	3	4	5
TUE	4	5	6	7
WED	6	7	8	9

(Four students) study hours.

Two way ANOVA can be applied & the variance can be determined

1: Between the columns (Among the students).
2: Between the rows (among the days).

Source of Variation	Sum of Squares	Degree of Freedom	Mean sum of squares	Ratio of F
Between the columns	SSC =15	$\nu_1 = (c-1)$ $4-1=3$	$MSC = SSC/\nu_1$ $15/3=5$	MSC/MSE $5/0=\infty$
Between the rows	SSR =32	$\nu_2 = (r-1)$ $=3-1=2$	$MSR = SSR/\nu_2$ $=32/2=16$	MSR/MSE $16/0=\infty$
Residual on Error	SSE =0	$\nu_3 = \nu_1 * \nu_2$ $= 3\times 2=6$	$MSE = SSE/\nu_3$ $=0/6=0$	
	SST =47	$\nu = n-1$ $=12-1=11$		

∴ SST : Total sum of squares.

STEP-I : Calculation of Grand Total & correction factor.

∴ Let's take the mean of total = 5

Day	A	B	C	D	Total
M	-3	-2	-1	0	-6
T	-1	0	1	2	2
W	+1	+2	3	4	10
Total	-3	0	3	6	6

6 → Grand total (T)

Example:
A researcher wants to determine the effect of two different teaching methods (Method A and Method B) and two different class sizes (Small and Large) on students' test scores. He randomly assigns students to four groups: Method A-Small, Method A-Large, Method B-Small, and Method B-Large. He then administers a test to all students and uses a two-way ANOVA to analyze the data. He can use this analysis to determine the main effects of teaching method and class size on test scores, and also to see if there is an interaction between the two independent variables. The independent variables in this case are the teaching method and class size, and the dependent variable is the students' test scores.

A two-way ANOVA is an extension of the one-way ANOVA and is used when you have two independent variables and one dependent variable. It allows you to determine the effect of two independent variables on the dependent variable, and also to see if there is an interaction between the two independent variables.

In summary, Two-way ANOVA is used to compare means of two independent variables and one dependent variable.

Correction factor $= \frac{T^2}{N} = \frac{(6)^2}{12} = \frac{36}{12} = 3$

STEP-II : Calculation of SSC (sum of sq. b/w The columns)

$$SSC = \frac{A^2}{n_A} + \frac{B^2}{n_B} + \frac{C^2}{n_C} + \frac{D^2}{n_D} - T^2/N$$

$$= \frac{\cancel{9}^{3}}{\cancel{3}} + \frac{0}{3} + \frac{\cancel{9}^{3}}{\cancel{3}} + \frac{36}{3} - \frac{36}{12}$$

$$= 3+0+3+12-3$$

$$= 15$$

STEP-III : Calculation of SSR (sum of sq. b/w the rows)

$$SSR = \frac{M^2}{n_M} + \frac{T^2}{N_T} + \frac{W^2}{n_W} - \frac{T^2}{N}$$

$$= \frac{36}{4} + \frac{4}{4} + \frac{100}{4} - 3$$

$$= 9 + 1 + 25 - 3$$

$$= 32$$

STEP-IV : Calculation of SST (Total Sum of Sq.)

$$SST = (-3)^2 + (-1)^2 + (1)^2 + (-2)^2 + 0^2 + (2)^2 + (-1)^2 + (1)^2 + (3)^2 + (0)^2 + (2)^2 + (4)^2 - 3$$

$$= 9+1+1+4+4+1+1+9+4+16-3$$

$$= 47$$

STEP V : Calculation of SSE (Total Sum of sq. due to error)

$$SSE = SST - (SSC + SSR)$$

$$= 47 - (15 + 32) = 47 - 47 = 0$$

F-Value$_{Tab}$ for $v_1 = 6, v_2 = 3, f_{0.05} = 4.76$
F-Value for $v_1 = 6, v_2 = 2, f_{0.05} = 5.14$
} Tabulated value.

H_a is passed and we can say that there is significant amount of variation b/w the students as well as b/w the days in terms of hours.

QUESTION NO: 2

Sol:-

Water Temp/Det	A	B	C
Cold	47	45	50
Warm	39	42	52
Hot	44	36	48

(1) Grand Total & correction factor:

- Data is coded by subtracting any guessed mid value (i.e., 40) for easy calculation.

Temp \ DET	A	B	C	Total
Cold	+7	+5	10	22
Warm	-1	2	12	13
Hot	4	-4	8	8
Total	10	3	30	43

43 → Grand Total (T)

Correction Factor

$$\frac{T^2}{N} = \frac{(43)^2}{9} = \frac{1849}{9} = \boxed{205.44}$$

P.T.O.

Step 2 :

Calculation of SSC

$$SSC = \frac{A^2}{n_A} + \frac{B^2}{n_B} + \frac{C^2}{n_C} - T^2/N$$

$$= \frac{100}{3} + 9/3 + \frac{(30)^2}{3} - 205.44$$

$$= 33.33 + 3 + 300 - 205.44$$

$$= \boxed{130.89}$$

Step 3: SSR = ? (calculation of SSR)

$$SSR = \frac{C^2}{n_C} + \frac{W^2}{n_W} + \frac{H^2}{n_H} - T^2/N$$

$$= \frac{(22)^2}{3} + \frac{(13)^2}{3} + \frac{(8)^2}{3} - 205.44$$

$$= \frac{484}{3} + \frac{169}{3} + \frac{64}{3} - 205.44$$

$$= 161.33 + 56.33 + 21.33 - 205.44 = 33.55$$

Step 4: Calculation of SST

$$SST = (7)^2 + (-1)^2 + (4)^2 + (5)^2 + (2)^2 + (-4)^2 + (10)^2 + (12)^2 + (8)^2 - 205.44$$

$$= 49 + 1 + 16 + 25 + 4 + 16 + 100 + 144 + 64 - 205.44$$

$$= 419 - 205.44 = 213.56$$

Step 5: Calculation of SSE :

$$SSE = SST - (SSC + SSR)$$

$$= 213.56 - (130.89 + 33.55)$$

$$= 49.12$$

Scanned with CamScanner

Source of Variation	Sum of Squares	Degree of freedom	Mean sum of Squares	Ratio of F
Between the Columns	SSC $= 130.89$	$v_1 = (c-1)$ $= 3-1 = 2$	$MSC = SSC/v_1$ $= 130.89/2$ $= 65.45$	MSC/MSE $= \frac{65.45}{12.28} = 5.33$
Between the rows	$SSR = 33.55$	$v_2 = (r-1)$ $= 3-1 = 2$	$MSR = SSR/v_2$ $= 33.55/2 = 16.78$	MSR/MSE $= \frac{16.78}{12.28} = 1.37$
Residual or Errors	SSE $= 49.12$	$v_3 = v_1 \times v_2$ $= 4$	$MSE = SSE/v_3$ $= 49.12/4 = 12.28$	
	$SST = 213.5$	$v = n-1$ $= 8$		

Tabulated F-Value, $v_1 = 4$, $v_2 = 2$, $f_{0.05} = 6.94$

Tabulated F-Value, $v_1 = 4$, $v_2 = 2$, $f_{0.05} = 6.94$