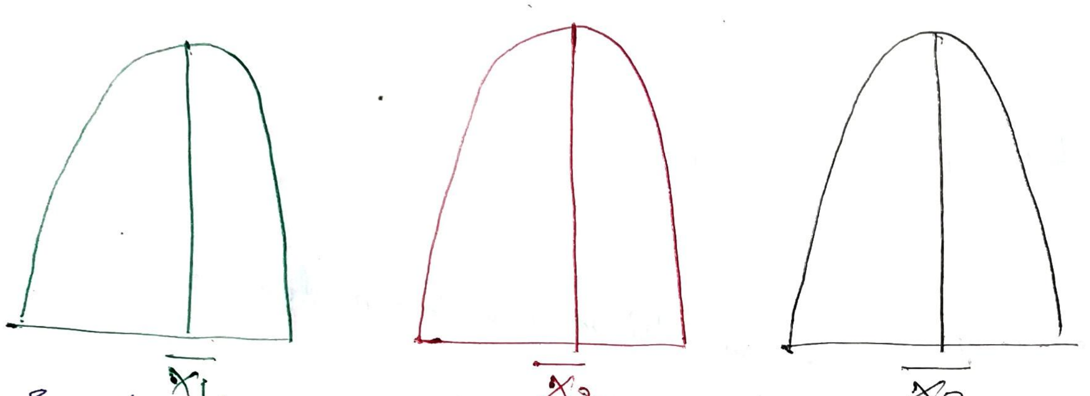
ANOVA TEST (F-test)

Variance

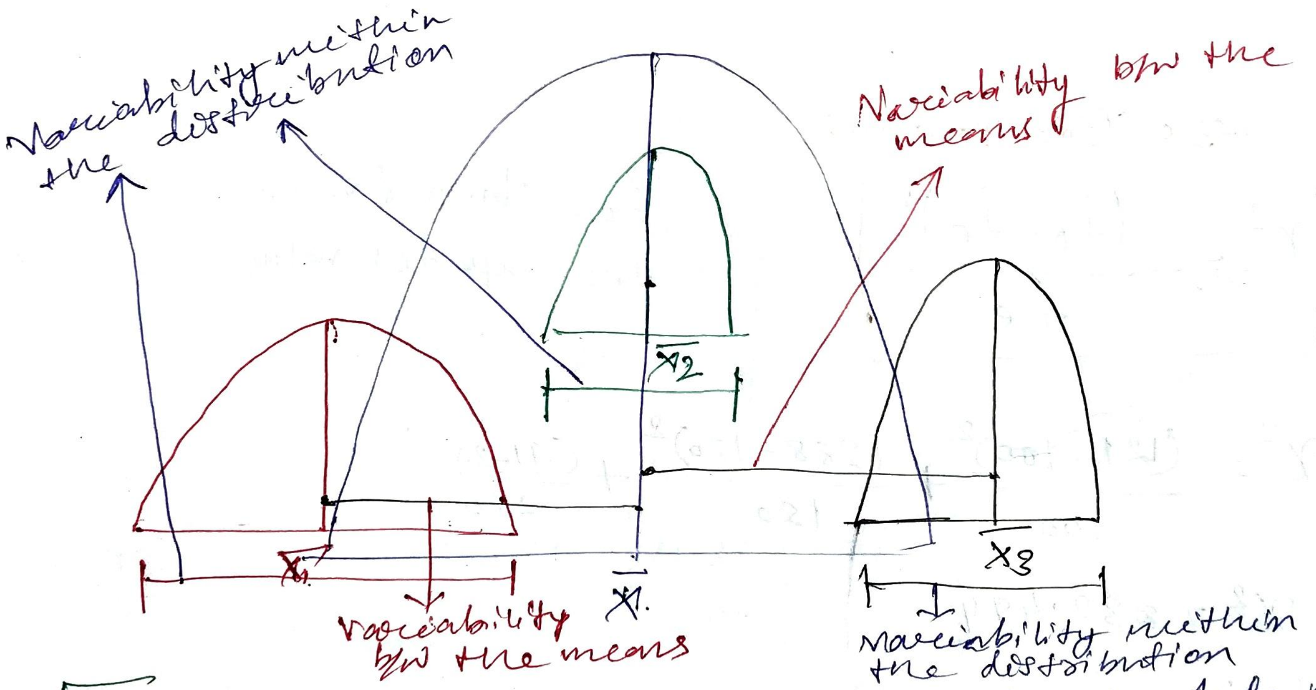
It is defined as the expection of the squared deviation of a random variable from its mean i.e., σ^2 or S^2.

- For comparison of more than two population or population having more than two subgroups we will use ANOVA technique.



-> Sample distributions.

Do all these 3 means are coming from the same population?



Note: $\bar{x}_1$, $\bar{x}_2$ and $\bar{x}_3$ may be close to each other, but it depends on the arrangement of their position in the population.

distance b/w $\bar{x}$ and $\bar{x}_3$ > distance b/w $\bar{x}$ and $\bar{x}_1$

distance b/w $\bar{x}$ and $\bar{x}_2$ < distance b/w $\bar{x}$ and $\bar{x}_1$

distance b/w $\bar{x}$ and $\bar{x}_2$ < distance b/w $\bar{x}$ and $\bar{x}_2$

Depends on the position

$$\boxed{\text{ANOVA} = \frac{\text{Variability b/w the means}}{\text{Variability within the distribution}}}$$

Variability b/w the means = distance b/w $\bar{x}_1$ and $\bar{\bar{x}}$ or
distance b/w $\bar{x}_2$ and $\bar{\bar{x}}$ or
distance b/w $\bar{x}_3$ and $\bar{\bar{x}}$;

$$\boxed{\text{Total Variance} = \text{Variability b/w the means} + \text{Variability within the distribution}}$$

Assumptions of ANOVA

1. Each population is having normal distribution.
2. The population from which the sample are drawn have the equal variance i.e., $S_1^2 = S_2^2 = S_3^2$ for K samples.
3. Each sample is drawn randomly and they are independent.

$H_0 : \mu_1 = \mu_2 = \mu_3 \dots \mu_n$
$H_a : \mu_1 \neq \mu_2 \neq \mu_3 \dots \neq \mu_n$

classification

1. One factor - One way ANOVA
2. Two factor - Two way ANOVA

One-way ANOVA

Que To assess the significance of possible variation in performance in a certain test between the convent schools of a city, a common test was given to a number of students taken at random from the fifth class of the 3 schools concerned the results given below.

→ one factor

A	B	C
9	12	14
11	12	13
13	10	17
9	15	7
8	5	9

one factor → One way ANOVA.

Make the analysis of variance for the given data.

Ans

A	B	C
9	13	14
11	12	13
13	10	17
9	15	7
8	5	9
50	55	60

$\bar{X}_A = \frac{50}{5} = 10$, $\bar{X}_B = \frac{55}{5} = 11$, $\bar{X}_C = \frac{60}{5} = 12$

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

Calculation of SSC (Between the samples)

$(\bar{X}_A - \bar{\bar{X}})$	$(\bar{X}_A - \bar{\bar{X}})^2$	$(\bar{X}_B - \bar{\bar{X}})$	$(\bar{X}_B - \bar{\bar{X}})^2$	$(\bar{X}_C - \bar{\bar{X}})$	$(\bar{X}_C - \bar{\bar{X}})^2$
$(10-11)=-1$	$(-1)^2$	$(11-11)=0$	0	$(12-11)=1$	1
$(10-11)=-1$	$(-1)^2$	$(11-11)=0$	0	$(12-11)=1$	1
$(10-11)=-1$	$(-1)^2$	$(11-11)=0$	0	$(12-11)=1$	1
$(10-11)=-1$	$(-1)^2$	$(11-11)=0$	0	$(12-11)=1$	1
$(10-11)=-1$	$(-1)^2$	$(11-11)=0$	0	$(12-11)=1$	1
$\Sigma(\bar{X}-\bar{\bar{X}})$	5		0		5

$$SSC = \Sigma(\bar{X}_A - \bar{\bar{X}})^2 + \Sigma(\bar{X}_B - \bar{\bar{X}})^2 + \Sigma(\bar{X}_C - \bar{\bar{X}})^2$$
$$= 5+0+5 = 10$$

Calculation of Degree of freedom

$\nu_1 = c-1$
$=$ no. of category -1
$= 3-1 = 2$

Between the sample

$\nu_2 = n-c$
$=$ no. of data $-$ number of category
$= \cancel{4}\cancel{18}\ 15-3$
$= \cancel{6}\ 12$

within the sample

Calculation of SSE (within the samples)

$(A-\bar{X}_A)$	$(A-\bar{X}_A)^2$	$(B-\bar{X}_B)$	$(B-\bar{X}_B)^2$	$(C-\bar{X}_C)$	$(C-\bar{X}_C)^2$
$9-10=-1$	$(-1)^2$	$13-11=2$	$(2)^2$	$14-12=2$	$(2)^2$
$11-10=1$	$(1)^2$	$12-11=1$	$(1)^2$	$13-12=1$	$(1)^2$
$13-10=3$	$(3)^2$	$10-11=-1$	$(-1)^2$	$17-12=5$	$(5)^2$
$9-10=-1$	$(-1)^2$	$15-11=4$	$(4)^2$	$7-12=-5$	$(-5)^2$
$8-10=-2$	$(-2)^2$	$5-11=-6$	$(-6)^2$	$9-12=-3$	$(-3)^2$
$\Sigma(X-\bar{X})^2$	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 Variation	Sum of Square	Degree of freedom	Mean Square	F
Between the Sample	SSC = 10	$v_1 = 2$	$MSC = \frac{SSC}{v_1} = \frac{10}{2}$ $= 5$	$F = \frac{MSC}{MSE}$ $= \frac{5}{11.5}$ $= 0.435$
Within the Sample	SSB = 138	$v_2 = 12$	$MSB = \frac{SSB}{v_2}$ $= \frac{138}{12} = 11.5$	

from F-tables.

Tabulated F-value for $v_1 = 2$ and $v_2 = 12$ = 3.89

calculated F-value = 0.435

Conclusion

Hence, calculated F-value < tabulated F-value,

we will accept the Null Hypothesis. It means there is no significant variation in performance b/w the schools.

Two-way Anova

Que The following data represents the number of units of tablet production (in thousands) per day five different technicians by using four different type of machines.

1st factor → Workers; 2nd factor → A B C D

Workers	A	B	C	D
P	54	48	57	46
Q	56	50	62	53
R	44	46	54	42
S	53	48	56	44
T	48	52	59	48

Two factor
Two-way ANOVA

(a) Test whether the mean productivity of the different machines are same?

(b) Test whether the 5 technicians differ with respect to the mean productivity?

Soln:

Machines / Technicians	A	B	C	D
P	54	48	57	46
Q	56	50	62	53
R	44	46	54	42
S	53	48	56	44
T	48	32	59	48

⇒ To calculate Grand total, assume a mean value. Here it is 50.

(1) Calculation of Grand Total and Correction Factor

	A	B	C	D	Total
P	54-50=4	48-50	57-50	46-50	+5
Q	56-50=6	50-50	62-50	53-50	+21
R	44-50=-6	46-50	54-50	42-50	-14
S	53-50=3	48-50	56-50	44-50	+1
T	48-50=-2	32-50	59-50	48-5	+7
Total	+5	-6	+38	-17	$\boxed{20}$ → Grand Total, T

$$\boxed{\text{Correction Factor} = \frac{T^2}{N}}$$

T = Grand Total
N = Number of data

$$\text{Correction factor} = \frac{(20)^2}{20} = 20$$

② Calculation of SSC (Between the columns)

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

$$SSC = \frac{(+5)^2}{5} + \frac{(-6)^2}{5} + \frac{(38)^2}{5} + \frac{(-17)^2}{5} - 20$$

$$SSC = \frac{25}{5} + \frac{36}{5} + \frac{38 \times 38}{5} + \frac{289}{25} - 20$$

$$SSC = 338.8$$

③ Degree of freedom

$$\left.\begin{aligned} \nu &= (c-1) \\ &= 4-1 \\ &= 3 \end{aligned}\right\} \text{Between the columns}$$

$$\left.\begin{aligned} \nu &= (r-1) \\ &= 5-1 \\ &= 4 \end{aligned}\right\} \text{Between the rows}$$

c = no. of columns
r = no. of rows

8

④ Calculation of SSR (Between the rows)

$$\boxed{SSR = \frac{P^2}{n_P} + \frac{Q^2}{n_Q} + \frac{R^2}{n_R} + \frac{S^2}{n_S} + \frac{T^2}{n_T} - \frac{T^2}{N}}$$

$$SSR = \frac{(+5)^2}{4} + \frac{(+21)^2}{4} + \frac{(-14)^2}{4} + \frac{(+1)^2}{4} + \frac{(+7)^2}{4} - 20$$

$$= \frac{25}{4} + \frac{441}{4} + \frac{196}{4} + \frac{1}{4} + \frac{49}{4} - 20$$

$$= 158$$

⑤ Calculation of MSC (Between the columns)

$$\boxed{MSC = \frac{SSC}{(C-1)}}$$

C = no. of columns

$$MSC = \frac{338.8}{3} = 112.93$$

⑥ Calculation of MSR (Between the rows)

$$\boxed{MSR = \frac{SSR}{(r-1)}}$$

r = no. of rows

$$MSR = \frac{158}{5-1} = 39.5$$

⑦ Calculation of SST

$$SST = (4)^2 + (6)^2 + (-6)^2 + (8)^2 + (-2)^2 + (-2)^2 + (0)^2 + (-4)^2$$
$$+ (-2)^2 + (2)^2 + (+7)^2 + (12)^2 + (4)^2 + (6)^2 + (9)^2$$
$$+ (-4)^2 + (3)^2 + (-8)^2 + (-6)^2 + (-2)^2 - \boxed{20} \to \frac{T^2}{N}$$

$$SST = 564$$

Degree of freedom (SST)

$$\boxed{\mu = n-1}$$

n = no. of data

$$\mu = 20-1 = 19$$

(8) Calculation of Residual or Errors, SSE

$$\boxed{SSE = SST - (SSC + SSR)}$$

$$SSB = 564 - (338.8 + 158) = 67.2$$

(9) Degree of freedom (Residual or Errors)

$$\boxed{\nu = (c-1)(r-1)}$$

$$= (4-1)(5-1)$$

$$= 3 \times 4$$

$$= 12$$

c = no. of columns
r = no. of rows

(10) Calculation of MSE (Residual or Errors)

$$\boxed{MSB = \frac{SSB}{(c-1)(r-1)}}$$

$$MSB = \frac{67.2}{12} = 5.6$$

Source of variation	Sum of Squares	Degree of freedom	Mean Sum of Squares	Ratio of F
Between the columns	SSC = 338.8	$\nu = c-1$ = 3	$MSC = \frac{SSC}{c-1}$ = 112.93	$\frac{MSC}{MSB} = \frac{112.93}{5.6}$ = 20.16
Between the rows	= SSR = 158	$\nu = r-1$ = 4	$MSR = \frac{SSR}{r-1}$ = 39.5	$\frac{MSR}{MSB} = \frac{39.5}{5.6}$ = 7.05
Residual or Error	SSB = 67.2	$\nu = (c-1)(r-1)$ = 12	$MSB = \frac{SSE}{(c-1)(r-1)}$ = 5.6	
	SST = 564	$\nu = n-1$ = 19		

Tabulated value

Between the columns

for, $\nu_1 = 12$, & $\nu_2 = 3$,

$F_{0.05} = 3.49$ (from F-table)

Between the rows

for, $\nu_1 = 12$ & $\nu_2 = 4$

$F_{0.05} = 3.26$ (from F-table)

Conclusion

F calculated > F tabulated } Between the columns

20.16 > 3.49

F calculated > F tabulated } Between the rows

7.05 > 3.26

Hence, we will accept the null hypothesis. ~~It~~

Between the column

It means there is significant variation in the productivity of the machines.

Between the rows

It means there are significant variation among technicians.