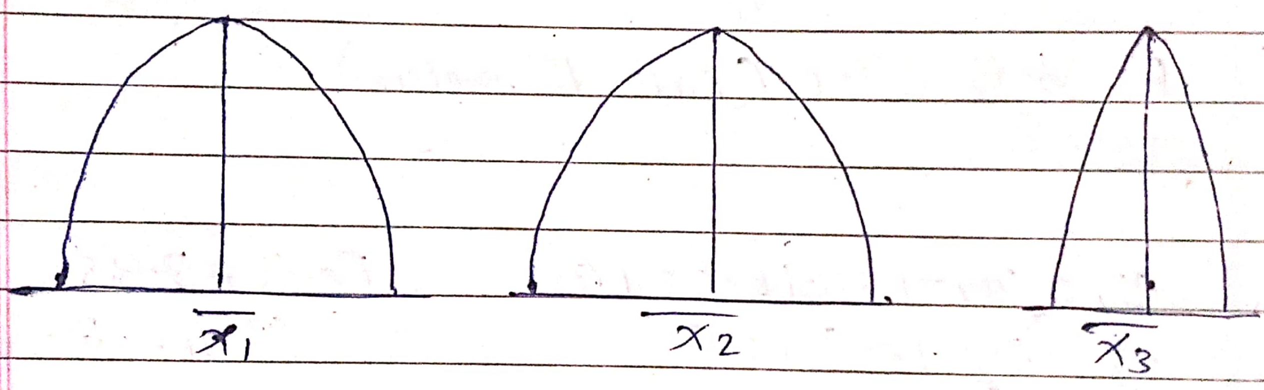
Analysis of Variance
(ANOVA)

→ Developed by R.A. Fischer

<u>Variance</u>: It is defined as the exception of the sq. deviation of a random variable from its mean i.e. σ^2 or s^2

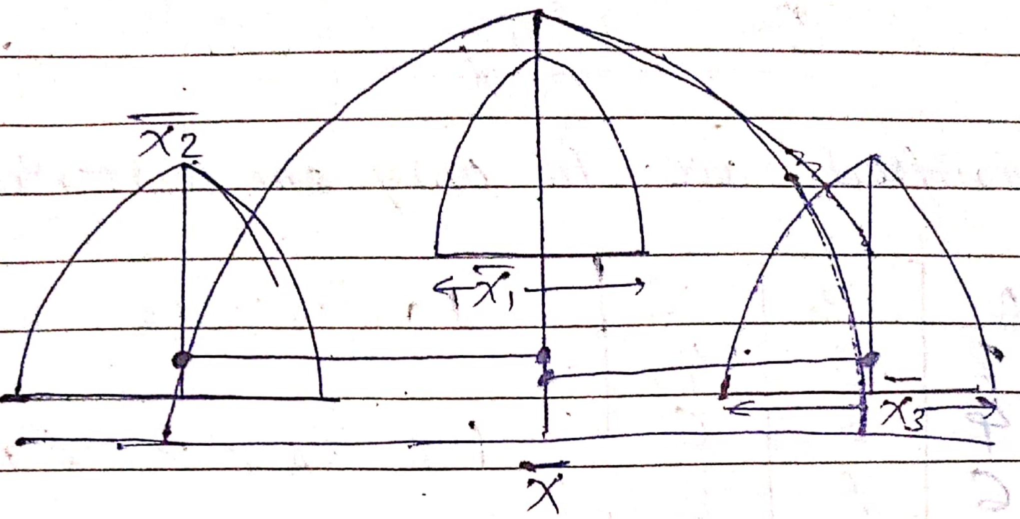
→ For comparison of ~~2 pop.~~ more than 2 pop. or pop. having more than 2 subgroup we will use ANOVA technique.



Are all these 3 means coming from the same pop.?

$$\text{ANOVA} = \frac{\text{Variability b/w the means}}{\text{Variability within the distr.}}$$

Total variance = Variability b/w means + Variability within the distr.



Assumptions:

- Each pop. is having normal distribution
- The pop. from which the sample are drawn have equal variance

  i.e. $S_1^2 = S_2^2 = S_3^2 \cdots\cdots S_k^2$ for k samples

- Each sample is drawn randomly & they are independent

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

$$H_a: \mu_1 \neq \mu_2 \neq \mu_3 \cdots\cdots = \mu_n$$

Classification: → One way ANOVA (one factor)
→ Two way ANOVA (2 factor)

One Way ANOVA

→ Classified acc. to only one criterion.

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

$H_0 : \overline{x}_A = \overline{x}_B = \overline{x}_C$

$H_a : \overline{x}_A \neq \overline{x}_B \neq \overline{x}_C$

① To state the null hyp. & alternate hyp.

② Cal. the variance b/w the samples

ⓐ Cal. mean of each sample

$\overline{x}_A = \frac{12}{3} = 4$, $\overline{x}_B = \frac{15}{3} = 5$, $\overline{x}_C = \frac{18}{3} = 6$

ⓑ Cal. of Grand avg. of means

$\overline{\overline{x}} = \frac{\overline{x}_A + \overline{x}_B + \overline{x}_C}{3} = \frac{4+5+6}{3} = \frac{15}{3} = 5$

ⓒ Take the diff. b/w means of var. samples & $\overline{x}$ & square it

Date.

$(\overline{x}_A - \overline{\overline{x}})$	$(\overline{x}_A - \overline{\overline{x}})^2$	$(\overline{x}_B - \overline{\overline{x}})$	$(\overline{x}_B - \overline{\overline{x}})^2$	$(\overline{x}_C - \overline{\overline{x}})$	$(\overline{x}_C - \overline{\overline{x}})^2$
4-5=-1	1	5-5	0	6-5=1	1
4-5=-1	1	5-5	0	6-5=1	1
4-5=-1	1	5-5	0	6-5=1	1
	3		0		3

Sum of sq. b/w samples: $(\Sigma(\overline{x} - \overline{\overline{x}}) = 3+0+3 = 6)$

③ Cal. the variance within sample

ⓐ Cal. of mean for each sample
ⓑ Take the dev. of the various items in a sample from mean values of the resp. samples & sqd. it

$(A - \overline{x}_A)$	$(A - \overline{x}_A)^2$	$(B - \overline{x}_B)$	$(B - \overline{x}_B)^2$	$(C - \overline{x}_C)$	$(C - \overline{x}_C)^2$
2-4=-2	4	3-5=-2	4	4-6=-2	4
4-4=0	0	5-5=0	0	6-6=0	0
6-4=2	4	7-5=2	4	8-6=2	4
	8		8		8

Sum of sq. within the sample,

$$(\Sigma(x-\overline{x})^2 = 8+8+8 = 24$$

ⓒ Cal. the ratio of F

Source of var.	Sum of squares	D.O.F. (ν)	Mean sum of sq.	F
b/w the sample	SSC = 6	$\nu_1 = c-1$ $3-1=2$	$MSC = SSC/c-1$ $= 6/2 = 3$	$F = \frac{MSC}{MSE}$ $F = 3/4$ $= 0.75$
Within the sample	SSE = 24	$\nu_2 = n-c$ $= 9-3 = 6$	$MSE = SSE/n-c$ $= 24/6 = 4$	

SSC = Sum of sq. b/w samples (col.)

SSE = Sum of sq. within samples (rows)

MSC = Mean sum of sq. b/w samples

MSE = " " within "

(d) Comp. the cal. F-value with tabulated F value

$F_{tab} = F_{0.05} = 5.14$

(e) Ho is correct.

Q.1) To assess the sig. of possible variance in perf. in a certain test b/w the convent schools of a city, a common test was conducted of a no. of students taken at random from the 5th class of the 3 schools concerned.

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

Make the analysis of variance for the given data

Soln:

~~A~~ ~~B~~ C

$\bar{x}_A = \frac{50}{5} = 10$ $\quad$ $\bar{x}_C = \frac{60}{5} = 12$

$\bar{x}_B = \frac{55}{5} = 11$

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

Cal. of SSC

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

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

$$= 5 + 0 + 5 = 10$$

Source of var.	Sum of sq.	D.O.F.	Mean sq.	F
b/w the sample	$SSC = 10$	$\nu_1 = c-1 = 3-1 = 2$	$MSC = \frac{SSC}{\nu_1} = \frac{10}{5} = 2$	$\frac{MSC}{MSE} = 5/11.5$
Within the sample	$SSE = 138$	$\nu_2 = n-c = 15-3 = 12$	$MSE = \frac{SSE}{\nu_2} = \frac{138}{12} = 11.5$	$= 0.435$

Cal. of SSE

Cal. of SSE

$(A-\bar{x}_A)$	$(A-\bar{x}_A)^2$	$(B-\bar{x}_B)$	$(B-\bar{x}_B)^2$	$(C-\bar{x}_C)$	$(C-\bar{x}_C)^2$
9-10=-1	1	13-11=2	4	14-12=2	4
11-10=1	1	12-11=1	1	13-12=1	1
13-10=3	9	10-11=-1	1	17-12=5	25
9-10=-1	1	15-11=4	16	7-12=-5	25
8-10=-2	4	5-11=-6	36	9-12=-3	9
	16		58		64

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

Cal. F value = 0.435
Tab. — " — = 3.89

~~Null passed~~ Null Hyp. is passed &
no sig. diff.

Two Way ANOVA

Classified acc. to 2 factors / criteria

Days	A	B	C	D
Mon (M)	2	3	4	5
Tue (T)	4	5	6	7
Wed (W)	6	7	8	9

Two way ANOVA can be applied & var. can be det.

(1) b/w the col. (b/w A, B, C, D)

(2) " —— rows (b/w M, T, W)

(S-1) Cal. of Grand Total & Corr. factor

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

Taking 5 as mean value & rule:

$$\text{Corr. factor} = \frac{T^2}{N} = \frac{6^2}{12} = \frac{36}{12} = 3$$

② Cal. of SSC (Sum of sq. b/w col.)

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

$$SSC = \frac{(-3)^2}{3} + \frac{0^2}{3} + \frac{3^2}{3} + \frac{6^2}{3} - 3 = 3+0+3+12-3$$

$$= 15$$

$$\boxed{SSC = 15}$$

③ Cal. of SSR (Sum of sq. b/w rows)

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

$$SSR = \frac{(-6)^2}{4} + \frac{0^2}{4} + \frac{10^2}{4} - 3 = 9 + 1 + 25 - 3 = 32$$

$$\boxed{SSR = 32}$$

④ Cal. of SST (Total sum of ~~sq~~ sq.)

$$SST = A[-3^2 + (-1)^2 + 1^2] + B[-2^2 + 0^2 + 2^2] +$$
$$C[-1^2 + 1^2 + 3^2] + D[0^2 + 2^2 + 4^2] - T^2/N$$

$\boxed{SST = 47}$

⑤ Cal. of SSE (Total sum of sq. due to error)

$$SSE = SST - (SSC + SSR) = 47 - (15+32) = 0$$

~~F value for $v_1 = 6, v_2 = 3$~~

Src of variation	Sum of sq.	D.O.F.	Mean sum of sq.	Ratio of F
b/w col.	$SSC = 15$	$v = c-1 = 4-1 = 3$	$MSC = \frac{SSC}{c-1} = 15/3 = 5$	$MSC/MSE = 5/0 = \infty$
b/w rows	$SSR = 32$	$v = r-1 = 3-1 = 2$	$MSR = \frac{SSR}{r-1} = \frac{32}{2} = 16$	$MSR/MSE = \frac{16}{0} = \infty$
Residual or errors	$SSE = 0$	$v = (c-1)\times(r-1) = (4-1)(3-1) = 6$	$MSE = \frac{SSE}{D.O.F.} = 0/6 = 0$	
	$SST = 47$	$v = n-1 = 11$		

F value for $v_1 = 6$ & $v_2 = 3$, $f_{0.05} = 4.76$

F-value for $v_1 = 6$ & $v_2 = 2$, $f_{0.05} = 5.14$

(Tab. value)

As $T_{cal} = \infty$ for both cases, there is sig. diff. b/w students & days

1) The foll. data represents the no. of units of tablet prod. (in thousands) per day by 5 diff. technicians by using 4 diff. types of machines.

Worker	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

a) Test whether the mean prod. of the diff. machines are same?

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

Soln:

Taking mean as 50.

1)

	A	B	C	D	Total
P	4	-2	+7	-4	+5
Q	6	0	+12	+3	+21
R	-6	-4	+4	-8	-14
S	+3	-2	+6	-6	+1
T	-2	+2	+9	-2	+7
Total	5	-6	+38	-17	20

$\rightarrow T$

$N = n_C \times n_R$
$= 4 \times 5$
$= 20$

Corr. factor $= \frac{T^2}{N} = \frac{20^2}{20} = 20$

2) Cal. of SSC

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

$$\cancel{SSC = \frac{25}{5} + \frac{36}{5} +}$$

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

$$\boxed{SSC = 338.8}$$

(3) Cal. of SSR

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

$$SSR = \frac{5^2}{4} + \cancel{44}\ \frac{21^2}{4} + \frac{(-14)^2}{4} + \frac{1^2}{4} + \frac{(7)^2}{4} - 20$$

$$\boxed{SSR = 158}$$

(4) Cal. of SST

$$SST = 4^2 + 6^2 + (-6)^2 + 3^2 + (-2)^2 + (-2)^2 + 0^2 + (-4)^2 + (-2)^2 + 2^2 + 7^2 + 12^2 + 4^2 + 16^2 + 9^2 + (-4)^2 + 3^2 + (-8)^2 + (-6)^2 + (-2)^2 - 20$$

$$\boxed{SST = 564}$$

$v_1 = 12$, $v_2 = 3$ $F_{0.05} = 3.49$ } Tab.

— 11 —, $v_2 = 4$ $F_{0.05} = 3.26$

Src of variation	Sum of sq.	D.O.F.	Mean sum of sq.	Ratio of F
1) b/w the col.	$SSC = \cancel{3368}$ 336.8	$v = c-1$ $= 4-1 = 3$	$MSC = \frac{SSC}{c-1}$ $= \frac{336.8}{3} =$ 112.57	$\frac{MSC}{MSE} = \frac{112.57}{5.6}$ $= 20.1$
2) b/w the rows	$SSR = 158$	$v = r-1$ $5-1 = 4$	$MSR = \frac{SSR}{r-1}$ $= \frac{158}{4} = 39.5$	MSR/MSE $= \frac{39.5}{5.6} = 7.05$
Residual or Error	$SSE =$ 67.2 $SST-(SSC+SSR)$	$v = (c-1)(r-1)$ $(4-1)(5-1)$ $= 12$	$MSE = \frac{SSE}{(c-1)(r-1)}$ $= \frac{67.2}{12}$ $= 5.6$	
	$SST =$ 564	$v = n-1$		

There is signif. diff. betn the mean prod. of workers