Unit-5 (Lec-2)

(Maths-4)

Topic: Student t-test of Significance (Part-1)

Test of Significance ✓

→ t-test ✓ → F-Test ✓

→ Chi-square Test (χ^2-Test) ✓ → One way ANOVA ✓

Test of Significance of Small samples :→

when sample size less than 30 ✓ ⇒ small sample

t-distribution is used when

* sample size ≤ 30 ✓✓
* Population S.D. is unknown

$$\boxed{t = \frac{\bar{x} - \mu}{S/\sqrt{n}}}, \quad S = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$$

$$S^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$

$\bar{x}$ → sample mean ✓

μ → Population mean ✓

n → sample size ✓✓

S → S.D. of Population ✓✓

If S.D. of sample is given 's' then

$$\boxed{t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}}$$

s and S relation is $\boxed{(n-1)S^2 = s^2 n}$

t-test for single sample Mean ✓✓

$$t = \frac{\bar{x} - \mu}{\ldots} \qquad S = \sqrt{\frac{\sum (x-\bar{x})^2}{\ldots}}$$

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}, \quad S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

$\nu = d.f = n-1$

↳ degree of freedom

$$\boxed{d.f. = \nu = n - k, \quad k \text{ is no. of independent constraints}}$$

t- test for two sample means

$$t_{cal} = \frac{(\bar{x} - \bar{y})}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$\bar{x}$ = mean of sample 1

$\bar{y}$ = " " " 2

n_1, n_2 size of sample 1 and sample 2

$$\boxed{S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

$\nu_1 = n_1 - 1, \quad \nu_2 = n_2 - 1$

$$\boxed{\nu = n_1 + n_2 - 2}$$

Note :- $|t_{cal}| < t_{tab} \Rightarrow$ Null hypothesis H_0 accepted

$|t_{cal}| > t_{tab} \Rightarrow$ Null hypothesis H_0 rejected

$\left.\begin{array}{l} t \\ \nu = \\ 5\% \to .05 \end{array}\right] \quad t_{tab}(.05)$

Level of Sig. → (5%, 1%)

↑ (By default)

Table A.2

t Distribution: Critical Values of t

Degrees of freedom	Significance level — Two-tailed test: 10% / One-tailed test: 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1	6.314	12.706	31.821	63.657	318.309	636.619
2	2.920	4.303	6.965	9.925	22.327	31.599
3	2.353	3.182	4.541	5.841	10.215	12.924
4	2.132	2.776	3.747	4.604	7.173	8.610
5	2.015	2.571	3.365	4.032	5.893	6.869
6	1.943	2.447	3.143	3.707	5.208	5.959
7	1.894	2.365	2.998	3.499	4.785	5.408
8	1.860	2.306	2.896	3.355	4.501	5.041
9	1.833	2.262	2.821	3.250	4.297	4.781
10	1.812	2.228	2.764	3.169	4.144	4.587
11	1.796	2.201	2.718	3.106	4.025	4.437
12	1.782	2.179	2.681	3.055	3.930	4.318
13	1.771	2.160	2.650	3.012	3.852	4.221
14	1.761	2.145	2.624	2.977	3.787	4.140
15	1.753	2.131	2.602	2.947	3.733	4.073
16	1.746	2.120	2.583	2.921	3.686	4.015
17	1.740	2.110	2.567	2.898	3.646	3.965

18	1.734	2.101	2.552	2.878	3.610	3.922
19	1.729	2.093	2.539	2.861	3.579	3.883
20	1.725	2.086	2.528	2.845	3.552	3.850
21	1.721	2.080	2.518	2.831	3.527	3.819
22	1.717	2.074	2.508	2.819	3.505	3.792
23	1.714	2.069	2.500	2.807	3.485	3.768
24	1.711	2.064	2.492	2.797	3.467	3.745
25	1.708	2.060	2.485	2.787	3.450	3.725
26	1.706	2.056	2.479	2.779	3.435	3.707
27	1.703	2.052	2.473	2.771	3.421	3.690
28	1.701	2.048	2.467	2.763	3.408	3.674
29	1.699	2.045	2.462	2.756	3.396	3.659
30	1.697	2.042	2.457	2.750	3.385	3.646
32	1.694	2.037	2.449	2.738	3.365	3.622
34	1.691	2.032	2.441	2.728	3.348	3.601
36	1.688	2.028	2.434	2.719	3.333	3.582
38	1.686	2.024	2.429	2.712	3.319	3.566
40	1.684	2.021	2.423	2.704	3.307	3.551
42	1.682	2.018	2.418	2.698	3.296	3.538
44	1.680	2.015	2.414	2.692	3.286	3.526
46	1.679	2.013	2.410	2.687	3.277	3.515
48	1.677	2.011	2.407	2.682	3.269	3.505
50	1.676	2.009	2.403	2.678	3.261	3.496
60	1.671	2.000	2.390	2.660	3.232	3.460
70	1.667	1.994	2.381	2.648	3.211	3.435
80	1.664	1.990	2.374	2.639	3.195	3.416
90	1.662	1.987	2.368	2.632	3.183	3.402
100	1.660	1.984	2.364	2.626	3.174	3.390
120	1.658	1.980	2.358	2.617	3.160	3.373

Ex. The lifetime of electric blubs for a random sample of 10 from a large consignment gave the following data :

$n=10$

Item	1	2	3	4	5	6	7	8	9	10
Life in '000 hrs'	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average lifetime of bulb is 4000 hrs ?

Sol $H_0 : \mu = 4000$ hrs There is no significant diff b/w sample mean and Population Mean

$\boxed{n=10}$

$H_1 : \mu \neq 4000$ hrs (Two tailed Test)

$$\boxed{t = \frac{\bar{x}-\mu}{S/\sqrt{n}}}$$

x	$x-\bar{x} = x-4.4$	$(x-4.4)^2$
4.2	$-.2$	.04
4.6	.2	.04
3.9	$-.5$	.25
4.1	$-.3$	.09
5.2	.8	.64
3.8	$-.6$	.36
3.9	$-.5$	.25

$$S = \sqrt{\frac{3.12}{9}} \Leftarrow S = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}}$$

$$\boxed{S = .589}$$

$$\bar{x} = \frac{\Sigma x}{n}$$

$$\bar{x} = \frac{44}{10} = 4.4$$

10

4·3	− ·1	·01
4·4	+ 0	0
5·6	1·2	1·44
$\Sigma x = 44$		$\Sigma(x_i - \bar{x})^2 = 3·12$

$$t = \frac{4·4 - 4}{·589/\sqrt{10}} = \frac{·4 \times \sqrt{10}}{·589}$$

$$\boxed{t_{cal} = 2·213}$$

$\nu = n-1 = 10-1 = 9$

$\boxed{\nu = 9}$

level of significance = 5%

$\nu = 9,\ t_{0·05} = 2·262$

$|t_{cal}| < t_{tab} \Rightarrow H_0$ accepted

i.e Avg life time could be 4000 hrs.

Ex. A sample of 18 items has mean 24 units and S.D. 3 units. Test the hypothesis that it is random sample from Normal population with mean 27 Units.

Sol

$$\left\{ \begin{array}{l} n = 18 \\ \bar{x} = 24 \\ s = 3 \end{array} \right\}$$

$H_0 : \mu = 27$, There is no significant difference b/w population mean and sample mean

$H_1 : \mu \neq 27$ (Two tailed Test)

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{24 - 27}{3/\sqrt{17}} = \frac{-\cancel{3}\sqrt{17}}{\cancel{3}}$$

$$t_{cal} = -\sqrt{17} = -4·123$$

$|t_{cal}| = 4.123$

Level of Sig $\rightarrow 5\%$ $\quad .05$

$\nu = d.f = n-1 = 18-1 = 17$

$d.f = 17, \quad t_{tab} = t_{0.05} = 2.110$

Conclusion $\quad \underline{|t_{cal}| > t_{tab}} \Rightarrow \underline{H_0 \text{ rejected}}$

Unit-5 (Lec-3)

(Maths-4)

Topic: Student t-test of Significance (Part-2)

t-test for significance between two sample Mean

	Size		Sample Mean	Pop. Mean	Pop Var.
Sample (1)	n_1	$x_1, x_2, x_3 \dots x_{n_1}$	$\bar{x}$	μ_1	σ
Sample (2)	n_2	$y_1, y_2, y_3 \dots y_{n_2}$	$\bar{y}$	μ_2	σ

$$\boxed{t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

If s_1, s_2 S.D. are given for samples then

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

OR

$$S^2 = \frac{\Sigma (x_i - \bar{x})^2 + \Sigma (y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

Degree of freedom

$\rightarrow \nu_1 = n_1 - 1 \Rightarrow$ for sample (1)

$\rightarrow \nu_2 = n_2 - 1 \Rightarrow$ for " (2)

$$\boxed{d.f. = \nu = n_1 + n_2 - 2}$$

level of significance = (5%, 1%)

Note:- $|t_{cal}| < t_{tab} \Rightarrow$ Ho Accepted | Ho

$\Rightarrow$ Ho rejected | H_1

$|t_{cal}| > t_{tab}$ ⇒

Table A.2

t Distribution: Critical Values of t

Degrees of freedom	Two-tailed test: 10% / One-tailed test: 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
	Significance level					
1	6.314	12.706	31.821	63.657	318.309	636.619
2	2.920	4.303	6.965	9.925	22.327	31.599
3	2.353	3.182	4.541	5.841	10.215	12.924
4	2.132	2.776	3.747	4.604	7.173	8.610
5	2.015	2.571	3.365	4.032	5.893	6.869
6	1.943	2.447	3.143	3.707	5.208	5.959
7	1.894	2.365	2.998	3.499	4.785	5.408
8	1.860	2.306	2.896	3.355	4.501	5.041
9	1.833	2.262	2.821	3.250	4.297	4.781
10	1.812	2.228	2.764	3.169	4.144	4.587
11	1.796	2.201	2.718	3.106	4.025	4.437
12	1.782	2.179	2.681	3.055	3.930	4.318
13	1.771	2.160	2.650	3.012	3.852	4.221
14	1.761	2.145	2.624	2.977	3.787	4.140
15	1.753	2.131	2.602	2.947	3.733	4.073
16	1.746	2.120	2.583	2.921	3.686	4.015
17	1.740	2.110	2.567	2.898	3.646	3.965
18	1.734	2.101	2.552	2.878	3.610	3.922
19	1.729	2.093	2.539	2.861	3.579	3.883
20	1.725	2.086	2.528	2.845	3.552	3.850
21	1.721	2.080	2.518	2.831	3.527	3.819
22	1.717	2.074	2.508	2.819	3.505	3.792
23	1.714	2.069	2.500	2.807	3.485	3.768
24	1.711	2.064	2.492	2.797	3.467	3.745
25	1.708	2.060	2.485	2.787	3.450	3.725
26	1.706	2.056	2.479	2.779	3.435	3.707
27	1.703	2.052	2.473	2.771	3.421	3.690
28	1.701	2.048	2.467	2.763	3.408	3.674
29	1.699	2.045	2.462	2.756	3.396	3.659
30	1.697	2.042	2.457	2.750	3.385	3.646
32	1.694	2.037	2.449	2.738	3.365	3.622
34	1.691	2.032	2.441	2.728	3.348	3.601
36	1.688	2.028	2.434	2.719	3.333	3.582
38	1.686	2.024	2.429	2.712	3.319	3.566
40	1.684	2.021	2.423	2.704	3.307	3.551
42	1.682	2.018	2.418	2.698	3.296	3.538
44	1.680	2.015	2.414	2.692	3.286	3.526
46	1.679	2.013	2.410	2.687	3.277	3.515
48	1.677	2.011	2.407	2.682	3.269	3.505
50	1.676	2.009	2.403	2.678	3.261	3.496
60	1.671	2.000	2.390	2.660	3.232	3.460
70	1.667	1.994	2.381	2.648	3.211	3.435
80	1.664	1.990	2.374	2.639	3.195	3.416
90	1.662	1.987	2.368	2.632	3.183	3.402
100	1.660	1.984	2.364	2.626	3.174	3.390
120	1.658	1.980	2.358	2.617	3.160	3.373

Ex. Samples of sizes 10 and 14 were taken from two Normal populations with S.D 3.5 and 5.2. The sample means were found to be 20.3 and 18.6. Test whether the means of two populations are same at 5% level.

	Size	S.D.	Sample Mean
Sample ①	$10 = n_1$	$3.5 = s_1$	$20.3 = \bar{x}$
			$18.6 = \bar{y}$

Sample-2 | 14 n_2 | $5.2 = s_2$ | 18.6

H_0: There is no significant difference, $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$ (Two tailed)

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{10(3.5)^2 + 14(5.2)^2}{10 + 14 - 2}$$

$$S^2 = \frac{501.06}{22} = 22.775$$

$$\boxed{S = 4.772}$$

$$t = \frac{20.3 - 18.6}{4.772\sqrt{\frac{1}{10} + \frac{1}{14}}}$$

$$= .8604$$

$$\boxed{t_{cal} = .8604}$$

$d.f = n_1 + n_2 - 2 = 10 + 14 - 2 = 22$

t at 5% level of significance for 22 d.f

$$t_{tab} = t_{0.05} = 2.074$$

Conclusion:

$$|t_{cal}| = .8604 < t_{0.05}$$

$\Rightarrow$ Null hypothesis H_0 is accepted

$\Rightarrow$ No significant difference b/w their means

Ex. The height of 6 randomly chosen sailors in inches are 63, 65, 68, 69, 71 and 72. Those of 9 randomly chosen soldiers are 61, 62, 65, 66, 69, 70, 71, 72 and 73. Test whether the sailors are on average taller than soldiers.

$X_1 \to$ Sailors $\quad n_1 = 6$

$X_2 \to$ Soldiers $\quad n_2 = 9$

H_0: There is no significant difference i.e $\mu_1 = \mu_2$

H_1: $\mu_1 > \mu_2$ (One tail test)

$$S^2 = \frac{\sum(x_1-\bar{x}_1)^2 + \sum(x_2-\bar{x}_2)^2}{n_1+n_2-2}$$

$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}$$

Sample 1 Sailors

x_1 :	63	65	68	69	71	72
$x_1-\bar{x}_1$:	-5	-3	0	1	3	4
$(x_1-\bar{x}_1)^2$:	25	9	0	1	9	16

$\sum x_1 = 408$

$\sum(x_1-\bar{x}_1)^2 = 60$

mean $\bar{x}_1 = \frac{\sum x_1}{n_1} = \frac{408}{6} = 68$, $\sum(x_1-\bar{x}_1)^2 = 60$

Sample 2 Soldiers

x_2 :	61	62	65	66	69	70	71	72	73
$x_2-\bar{x}_2$:	-6.66	-5.66	-2.66	-1.66	1.34	2.34	3.34	4.34	5.34
$(x_2-\bar{x}_2)^2$:	44.36	32.035	7.0756	2.7556	1.7956	5.4756	11.1556	18.8356	28.5156

$\sum x_2 = 609$

$$\bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{609}{9} = \underline{\underline{67.66}} \qquad \Sigma(x_2-\bar{x}_2)^2 = 152.0002$$

$$S^2 = \frac{\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2} = \frac{60+152.0002}{6+9-2}$$

$$= 16.3077$$

$$S = 4.038$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{68-67.66}{4.038\sqrt{\frac{1}{6}+\frac{1}{9}}} = 0.1569$$

$$\boxed{t_{cal} = .1569}$$

$$d.f. = n_1+n_2-2 = 6+9-2 = \underline{\underline{13}} \leftarrow$$

level of signif $= 5\% \leftarrow$

$$t_{0.05} = \underline{\underline{1.771}}$$

Conclusion :

$$|t_{cal}| = .1569 < t_{0.05}$$

$\Rightarrow$ H_0 hypothesis accepted.

$\Rightarrow$ No significant difference b/w their average

Sailors are not average taller than soldiers

Practice Question

Q→ The mean life of 10 electric motors was found to be 1450 hr with S.D of 423 hrs. A second sample of 17 motors chosen from different batch showed mean life 1280 hrs with S.D of 398 hrs. Is there a significant difference b/w means of two samples.

Unit-5 (Lec-4)

(Maths-4)

Topic: Chi-Square Test (χ^2 – Test) [Most Imp. Topic]

χ^2 – test

$$\boxed{\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i}\right]}$$

O_i = Observed frequency ← $(O_i - E_i)^2 / E_i$

E_i = Expected frequency ←

for simple data $\boxed{E_i = \frac{\sum O_i}{n}}$ $n \to$ No of observated

Degree of freedom $d.f = \nu = n-1$ ⇐

n = No of observations.

⇒ $\nu = (p-1)(q-1)$
$(R-1)(C-1)$

for tabular data
$p \to$ No of Rows
$q \to$ No. of columns

Level of Significance (α) ⟶ 5% or 1%

Note:

$\boxed{\chi^2_{cal} < \chi^2_{tab}}$ ⇒ Null hypothesis H_0 accepted

$\boxed{\chi^2_{cal} > \chi^2_{tab}}$ ⇒ Null hypothesis H_0 Rejected

$\boxed{d.f = \nu = n-k}$ k = No. of independent Constraints.

Ex. :– The following table gives the number of accidents that took place in an industry during various days of week. Test if accidents are uniformly distributed over the week.

Day :-	MON	TUE	WED	THUR	FRI	SAT
No of Accidents :-	14	18	12	11	15	14 ← O_i

Sol :- H_0 : Accidents are Uniformly distributed

$n = 6$ $\boxed{\nu = n-1 = 6-1 = 5}$

level of significance (α) = 5%

Day	Observed (O_i)	Expected (E_i) = $\Sigma O_i / n$	$(O_i - E_i)^2 / E_i$
MON	14	$84/6 = 14$	0
TUE	18	$84/6 = 14$	$16/14 = 1.1428$
WED	12	14	$4/14 = .2857$
THUR	11	14	$9/14 = .6428$
FRI	15	14	$1/14 = .0714$
SAT	14	14	0
	$\Sigma O_i = 84$		$\Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right] = 2.1427$

$$E_i = \frac{\Sigma O_i}{n} \quad , \quad \chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$\boxed{\chi^2_{cal} = 2.1427}$$

level of sig (α) = 5% , $\nu = 5$

$$\boxed{\chi^2_{(5\%, 5)} = 11.07} \quad \text{(Table)}$$

Conclusion

$\chi^2_{cal} = 2.1427 < \chi^2_{tab}$

H_0 (Null hypothesis Accepted) ⇒ Accidents are

Uniformly distributed.

Ex. A die is thrown 276 times and results of these throws are given below

No. appreared on die	1	2	3	4	5	6	
freq	40	32	29	59	57	59	← O_i

Test whether the die is biased OR Not

AKTU-2019

Sol^n

H_0: Die is Unbiased.

$n = 6$

$d.f = \nu = n-1 = 6-1 = 5$

Level of sig $(\alpha) = 5\%$

No appeared	Observed O_i	Expected $(E_i) = \Sigma O_i/n$	$(O_i-E_i)^2/E_i$
1	40	$276/6 = 46$	$36/46$
2	32	46	$(14)^2/46 = 196/46$
3	29	46	$(17)^2/46 = 289/46$
4	59	46	$(13)^2/46 = 169/46$
5	57	46	$(11)^2/46 = 121/46$
6	59	46	$(13)^2/46 = 169/46$

$\Sigma O_i = 276$ $\qquad \Sigma\left[\frac{(O_i-E_i)^2}{E_i}\right] = \frac{980}{46}$

$$\chi^2 = \Sigma\left[\frac{(O_i-E_i)^2}{E_i}\right] = \frac{980}{46} = 21.30$$

$$\boxed{\chi^2_{cal} = 21.30}$$

$\nu = n-1 = 5$

Level of sig $(\alpha) = 5\%$

$$\chi^2_{(5\%,5)} = 11.07 = \chi^2_{tab}$$

<u>Conclusion</u>

$$\chi^2_{cal} = 21.30 > \chi^2_{tab}$$

$\Rightarrow$ Null hypo. H_0 rejected

$\Rightarrow$ Die is biased

TABLE A.4

χ^2 (Chi-Squared) Distribution: Critical Values of χ^2

Degrees of freedom	Significance level 5%	1%	0.1%
1	3.841	6.635	10.828
2	5.991	9.210	13.816
3	7.815	11.345	16.266
4	9.488	13.277	18.467
5	$\boxed{11.070}$	15.086	20.515
6	12.592	16.812	22.458
7	14.067	18.475	24.322
8	15.507	20.090	26.124
9	16.919	21.666	27.877
10	18.307	23.209	29.588

v

Practice Question

Q → Demand for a particular part in a factory was found to vary from day to day. In a sample study,

Day	MON	TUE	WED	THUR	FRI	SAT
No. of Parts demand	1124	1125	1110	1120	1126	1115

O_i

Test the hypothesis that the No. of parts demanded does not depend on the the day of week.

(Given $\chi^2_{(5\%, 5)} = 11{\cdot}07$, $\chi^2_{(5\%, 6)} = 12{\cdot}59$,

$$\chi^2_{(5\%, 7)} = 14.07$$

Unit-5 (Lec-5)

(Maths-4)

Topic: Chi-Square Test (Part - 2) Most Imp Topic

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$\boxed{E_i = \frac{\sum O_i}{n}}$$

O_i = Observed frequency → (Given)

E_i = Expected frequency →

ν = d.f = $n-1$ simple data

n → No of observation

In case of Binomial distribution d.f = $n-1$ ←

" " Poission " " = $n-2$ ←

" " Normal " " = $n-3$ ←

Level of significance (α) − (5% OR 1%)

If $(\chi^2_{cal}) > \chi^2_{tab}$ ⇒ H_0 Rejected

$\chi^2_{cal} < \chi^2_{tab}$ ⇒ H_0 accepted

TABLE A.4

χ^2 (Chi-Squared) Distribution: Critical Values of χ^2

	Significance level		
Degrees of freedom	5%	1%	0.1%
1	3.841	6.635	10.828
2	5.991	9.210	13.816
3	7.815	11.345	16.266

4	9.488	13.277	18.467
5	$\boxed{11.070}$	15.086	20.515
6	12.592	16.812	22.458
7	14.067	18.475	24.322
8	15.507	20.090	26.124
9	16.919	21.666	27.877
10	18.307	23.209	29.588

Ex. :→ A survey of 320 families with 5 children shows the following distribution :

No of children	5B 0G	4B 1G	3B 2G	2B 3G	1B 4G	0B 5G	Total
No of families	18 (O_1)	56 (O_2)	110 (O_3)	88 (O_4)	40 (O_5)	8 (O_6)	320

Test hypothesis that data are Binomialy distributed and Male and female birth have equal probability. (Given $\chi^2_{(5\%, 5)} = 11{\cdot}07$)

Sol p = prob for B. $\boxed{p = q = \frac{1}{2}}$ $P(r) = {}^nC_r\, p^r q^{n-r}$

q = " " G. $p \to$ prob of success, $q \to$ " " failure

$\boxed{N = 320}$

$r = 0, 1, 2, 3, 4, 5$

E_6 $P(r=0) = P(0B, 5G) = N.\, {}^nC_r\, p^r q^{n-r} = 320\, {}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5$

$= 320 \cdot 1 \cdot \frac{1}{2^5} = 10$ family

E_5 $P(r=1) = P(1B, 4G) = 320 \cdot {}^5C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = 320 \cdot 5\, \frac{1}{2^5}$

$= 50$ families

E_4 $P(r=2) = P(2B, 3G.) = 320 .\, {}^5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = 320 . \frac{5 \times \cancel{4}}{\cancel{2} \times 1} \frac{1}{2^5}$

$= 100$ families

E_3 $P(r=3) = P(\underline{3}B\,2G) = 320\ {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = 320.\frac{5\times \cancel{4}^{2}\times\cancel{3}}{\cancel{3}\times\cancel{2}} \left(\frac{1}{2^5}\right)$

$= 100$ families ✓

E_2 $P(r=4) = P(\underline{4}B, 1G) = 320\ {}^5C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 = 320 \cdot 5\ \frac{1}{2^5}$

$= 50$ ✓

$E_1 = P(r=5) = P(\underline{5B, 0G}) = 320 \cdot {}^5C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 = 320 \cdot 1 \cdot \frac{1}{2^5}$

$= 10$ ✓

$\boxed{H_0 : \text{Data are Binomially distributed}}$

No of children	Observed freq O_i	Expected freq E_i ✓	$(O_i - E_i)^2/E_i$
5B, 0G	18 ✓	10 ✓	$64/10 = 6.4$
4B, 1G ✓	56 ✓	50 ✓	$36/50 = .72$
3B, 2G ✓	110 ✓	100 ✓	$100/100 = 1$
2B, 3G ✓	88 ✓	100 ✓	$144/100 = 1.44$
1B, 4G	40 ✓	50 ✓	$100/50 = 2$
0B, 5G	8 ✓	10 ✓	$4/10 = .4$
	$\Sigma O_i = 320$	$\Sigma E_i = 320$	11.96

$\chi^2_{cal} = 11.96$

$\nu = 6-1 = 5$

At 5% level of sig, $\chi^2_{tab} = 11.07$

$\chi^2_{cal} = 11.96 > \chi^2_{tab}$

χ^2_{cal} ...

$\Rightarrow H_0$: Rejected

Unit-5 (Lec-6)
(Maths-4)

Topic: Chi-Square Test (Part - 3)

χ^2 - Test As a test of Independence

* Sample data is given in tabular form, called Contingency table
* By Using χ^2 - test, we can find whether or not two attributes are related.
* $d.f = \nu = (p-1)(q-1)$ OR $(R-1)(C-1)$

p = No of Rows
q = No of col.

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i}\right]$$

Contingency Table

a	b
c	d

Observed frequency

a O_1	b O_2	a+b
c O_3	d O_4	c+d
a+c	b+d	N = a+b+c+d

← Sum

Sum ↑

Total

Expected frequencies

$E_1 = \frac{(a+b)(a+c)}{a+b+c+d}$	$E_2 = \frac{(a+b)(b+d)}{a+b+c+d}$
$E_3 = \frac{(c+d)(a+c)}{a+b+c+d}$	$E_4 = \frac{(c+d)(b+d)}{a+b+c+d}$

$O_1 \to E_1$ $O_2 \to E_2$ $O_3 \to E_3$ $O_4 \to E_4$

↑ contingency table

TABLE A.4

χ^2 (Chi-Squared) Distribution: Critical Values of χ^2

Degrees of freedom	5%	1%	0.1%
	Significance level		
1	3.841	6.635	10.828
2	5.991	9.210	13.816
3	7.815	11.345	16.266
4	9.488	13.277	18.467
5	11.070	15.086	20.515
6	12.592	16.812	22.458
7	14.067	18.475	24.322
8	15.507	20.090	26.124
9	16.919	21.666	27.877
10	18.307	23.209	29.588

Ex. From the following table regarding the color of eyes of father and son, test if the colour of son's eye is associated with that of the father

[2020-21]

Eye colour of father	Eye Colour of Son	
	Light	Not Light
Light	471	51
Not Light	140	230

Sol

H_0 : The colour of son's eye is not associated with father's eye colour i.e they are independent.

	Son Light	Not Light	
Father Light	O_1 471	O_2 51	$\boxed{522}$
Not Light	O_3 148	O_4 230	$\boxed{378}$
	619	281	900

2x2

619 281 | 900

Expected frequencies

Eye color of FATHER \ Eye color of SON	Light	Not Light	Total
Light	$E_1 = \frac{522 \times 619}{900} = 359.02$	$E_2 = \frac{522 \times 281}{900} = 162.98$	$\Rightarrow$ 522
Not Light	$E_3 = \frac{378 \times 619}{900} = 259.98$	$E_4 = \frac{378 \times 281}{900} = 118.02$	$\Rightarrow$ 378
	619	281	900

Calculation of χ^2

O_i	471 ✓	51	148	230	
E_i	359.02 ✓	162.98	259.98	118.02	
$\frac{(O_i - E_i)^2}{E_i}$	$\frac{(471-359.02)^2}{359.02} = 34.927$	76.939	48.232	106.249	$\Sigma = 266.347$

$$\chi^2_{cal} = \boxed{266.347}$$

$$\nu = (R-1)(C-1) = (2-1)(2-1) = 1$$

$\nu = 1$

Level of sig (α) = 5%

$$\chi^2_{(.05,1)} = 3.841$$

$\chi^2_{cal} = 266.347 > \chi^2_{tab}$

$\Rightarrow$ Ho Rejected

$\Rightarrow$ Color of son's eye associated with father's eye colour.

Ex. To test effectiveness of inoculation against cholera the following table was obtained

	Attacked	Not Attacked	Total
Inoculated	30	160	190
Not Inoculated	140	460	600
	170	620	790

Use χ^2-test to defend or refute the statement that the inoculation prevents attack from cholera.

[2009, 2018]

Sol

Ho : The inoculation does not prevent attack.

Under Ho

observed 2X2

O_1 30	O_2 160	Total 190
O_3 140	O_4	

Expected Frequencies

$E_1 = \frac{190 \times 170}{790} = 40.886$	$E_2 = \frac{190 \times 620}{790} = 149.11$
$E_3 = 600 \times 170$	$E_4 = \frac{600 \times 620}{790}$

	460	600	$\frac{\dots}{790} = 129.11$	$\frac{\dots}{790} = 470.89$
170	620	N = 790		

Calculation of χ^2

	O_1	O_2	O_3	O_4	
O_i	30	160	140	460	
E_i	40.886	149.11	129.11	470.89	
$\frac{(O_i-E_i)^2}{E_i}$	$\frac{(30-40.886)^2}{40.886} = 2.898$	.795	.918	.252	$\sum = 4.863$

$$\boxed{\chi^2_{cal} = 4.863}$$

$$\nu = (R-1)(C-1) = (2-1)(2-1) = 1$$

$$\alpha = 5\%$$

$$\chi^2_{(5\%,1)} = 3.841$$

$$\chi^2_{cal} > \chi^2_{tab}$$

$\Rightarrow$ H_0 Rejected

$\Rightarrow$ we defend that inoculation prevents attack.

Unit-5 (Lec-7)
(Maths-4)

Topic: F-Test for significance of variance

Let
Sample 1 $x_1, x_2, x_3 \dots x_{n_1}$
Sample 2 $y_1, y_2, y_3 \dots y_{n_2}$
} be two random samples drawn from sample population or population with equal variance.

F- test uses variance ratio to test the significance of difference between variances of two samples

$$\boxed{F = \frac{S_1^2}{S_2^2} \quad (S_1 > S_2)}$$

	Size	Variance
Sample 1	n_1	S_1^2
Sample 2	n_2	S_2^2

$$S_1 = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1}}$$

$$S_2 = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1}}$$

$$\bar{x} = \Sigma x_i / n_1$$

$$\bar{y} = \Sigma y_i / n_2$$

Degree of freedom d.f.

n_1 for sample 1 $\quad \nu_1 = n_1 - 1$

n_2 " sample 2 $\quad \nu_2 = n_2 - 1$

Level of significance $\rightarrow$ (5% or 1%)

If $F_{cal} < F_{tab} \Rightarrow$ Null hypothesis is Accepted

$F_{cal} > F_{tab} \Rightarrow$ Null hypothesis is Rejected

F-Table

The F-distribution for various degrees of freedom v_1 and v_2 is given in the following table :

Table 12.3 : Values of F (F-distribution) for 5% and 1% level, where v_1 is the number of degree of freedom for greater estimate of variance and v_2 for the smaller.

v_2 \ v_1		1	2	3	4	5	6	8	12	24	∞
1	5%	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.0
	1%	4052	4999	5403	5625	5764	5849	5981	6016	6234	6366
2	5%	18.51	19.00	19.16	19.25	19.30	19.32	19.37	19.41	19.45	19.50
	1%	98.49	99.00	99.17	99.25	99.30	99.33	99.36	99.42	99.46	99.50
3	5%	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
	1%	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	5%	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
	1%	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	5%	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
	1%	16.26	13.27	12.06	11.39	10.97	10.67	10.27	9.89	9.47	9.02
6	5%	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
	1%	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	5%	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
	1%	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	5%	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
	1%	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	5%	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
	1%	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	5%	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
	1%	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
12	5%	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
	1%	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
14	5%	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
	1%	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
16	5%	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
	1%	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
18	5%	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
	1%	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.01	2.57
20	5%	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
	1%	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42

Ex: $\Rightarrow$ Two Independent sample of sizes 7 and 6 had the following values :

Sample A	28	30	32	33	31	29	34
Sample B	29	30	30	24	27	28	

examine whether the samples have been drawn from Normal population having the same variance

Sol

$H_0 : \sigma_1^2 = \sigma_2^2$, samples have been drawn from Normal population with same variance.

$H_1 : \sigma_1^2 \neq \sigma_2^2$

$$\boxed{F = \frac{S_1^2}{S_2^2} \quad (S_1 > S_2)}$$

$$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1}, \quad S_2 = \frac{\sum(X_2 - \bar{X}_2)^2}{n_2 - 1}$$

Sample A			Sample B		
X_1	$X_1-\bar{X}$ X_1-31	$(X_1-\bar{X})^2$	X_2	$(X_2-\bar{X}_2)$ X_2-28	$(X_2-\bar{X}_2)^2$
28	-3	9	29	1	1
30	-1	1	30	2	4
32	1	1	30	2	4
33	2	4	24	-4	16
31	0	0	27	-1	1
29	-2	4	28	0	0
34	3	9			
217		28	168		26

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{217}{7} = 31$$

$$\bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{168}{6} = 28$$

$$S_1^2 = \frac{\Sigma(X_1-\bar{X}_1)^2}{n_1-1} = \frac{28}{6} = \boxed{4.666}$$

$$S_2^2 = \frac{\Sigma(X_2-\bar{X}_2)^2}{n_2-1} = \frac{26}{5} = \boxed{5.2}$$

$$F = \frac{S_2^2}{S_1^2} = \frac{5.2}{4.666}$$

II^nd

I^st

$$\boxed{F_{cal} = 1.1158}$$

$$\nu_2 = n_1 - 1 = 7-1 = 6$$

$$\nu_1 = n_2 - 1 = 6-1 = 5$$

level of sig = 5%

$$\boxed{F_{tab} = 4.39}$$

$F_{cal} = 1.1158 < F_{tab} \Rightarrow H_0$ accepted

there is no significant diff b/w variance

Ex. Two Random samples drawn from 2 Normal Populations are as follows

A	17	27	18	25	27	29	13	17
B	16	16	20	27	26	25	21	

Test whether the samples are drawn from same Normal Population.

[Hint → To test samples are drawn from same population we test

(i) equality of Means by t-test

(ii) equality of Population Variance by F-test

First F-Test

H_0 : There is no significant difference b/w Population variance

i.e $\sigma_1^2 = \sigma_2^2$

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1-1}$$

H_1 : $\sigma_1^2 \neq \sigma_2^2$

$\bar{X} = 21.625$, $S_1^2 = 36.267$

$\bar{Y} = 21.57$ $S_2^2 = 20.94$

$$F = \frac{S_1^2}{S_2^2} = 1.7319 \quad (S_1 > S_2)$$

Ans :- H_0 accepted

T- test

$H_0 : \mu_1 = \mu_2$ Population means are equal

$H_1 : \mu_1 \neq \mu_2$

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S^2 = \frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1 + n_2 - 2}$$

$S^2 = 29.196$

$S = 5.403$

$t = .2449$

$d.f = n_1 + n_2 - 2$

$t_{tab\,(5\%,\, d.f)} =$ ___

Ans :- H_0 Accepted

Final Ans :→ Two samples have been drawn from same Normal Population.

By - Monika Mittal
(MM)

Unit-5 (Lec-8)

(Maths-4)

Topic : ANOVA (Analysis of Variance)

ANOVA is statistical tool that can be used for comparision among more than two groups.
It is used for testing the hypothesis of equality of more than two Normal population means.

Analysis of Variance' is carried on the basis of Ratio between the variances.

The Variance Ratio is obtained by dividing the variance between samples by variance within samples.

$$F = \frac{\text{Variance between the sample}}{\text{Variance within the samples}}$$

This ratio forms F- statistic

Remember that in F- Statistic, variance b/w the samples greater than variance within samples

Techniques of ANOVA

* One way ANOVA → used to Analysis of effects of one independent factor on dependent variable.

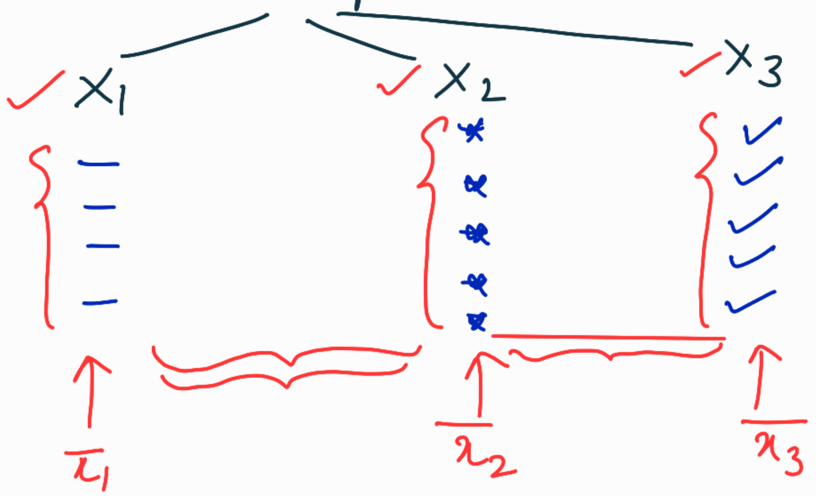
* Two way ANOVA → used to Analysis of effects of two independent factors on dependent variable

Assumptions :→

* Each sample is randomly drawn.
* Populations, from which the samples are drawn, are Normally distributed.
* Each sample is independent of other sample
* Each of population has same variations and identical means.

ONE WAY ANOVA One independent factor effects different samples

Independent Var



Total Variation

Variance b/w Samples

Variance within Samples (error variance)

Null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$, i.e means of Populations from which samples are drawn are equal

Alternate hypothesis :→ H_1 : at least one mean is different

Short cut Method :→

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
$n_1 = 4$		$n_2 = 4$		$n_3 = 4$	
ΣX_1	ΣX_1^2	ΣX_2	ΣX_2^2	ΣX_3	ΣX_3^2

* Grand Total $\quad G.T. = \Sigma X_1 + \Sigma X_2 + \Sigma X_3$

* Correction factor $\quad C.F. = \dfrac{(G.T.)^2}{n}$ $\quad n \to$ total No. of items.

* Sum of squares between samples (SSC)

$$\frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - C.F.$$

$d.f = c-1$ $\qquad d.f = 3-1 = 2$

* Total Sum of squares (SST)

$$\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - C.F.$$

$d.f = n-1$

* Sum of squares within the samples (SSE) error

$$\Rightarrow \quad SST = SSC + SSE$$

$$SSE = SST - SSC$$

$d.f = n-c$

* Mean Sum of squares between samples (Variance)

$$MSC = \frac{SSC}{d.f} = \frac{SSC}{c-1}$$

* Mean Sum of squares within samples (Variance)

$$MSE = \frac{SSE}{n-c} = \frac{SSE}{d.f}$$

* F- Statistic $\quad F = \dfrac{MSC}{MSE}$ $\quad F_{cal}$. d.f. $\quad F_{(n, \alpha)}$

* level of significance $(\alpha) = 5\%$

If $F_{cal} < F_{tab}$ at 5% level of significance $\Rightarrow$ H_0 Accepted

$F_{cal} > F_{tab}$ $\Rightarrow$ H_0 Rejected

ANOVA TABLE

Variance

Source of Variation	Sum of Squares	Degree of freedom	Mean sum of squares	F
Between Samples	SSC	$c-1$	$MSC = SSC/(c-1)$	$F = \frac{MSC}{MSE}$ $MSC > MSE$
Within Samples	SSE	$n-c$	$MSE = SSE/(n-c)$	
Total	SST	$n-1$		

$$F = \frac{\text{Variance b/w samples}}{\text{Variance within Samples}} = \frac{MSC}{MSE}$$

$F_{(N, D)}$

N: D.f. of N^r

D: d.f. of D^r

Ex. If is desired to compare three hospitals with regards to number of deaths per month. A sample of death records were selected from the records of each hospital and the number of deaths was as given below. From these data, suggest a difference in the no. of deaths per months among 3 hospitals

[2020-21]
AKTU

Hospitals

A	B	C
3	6	7
4	3	3
3	3	4
5	4	6
0	4	5

Given $F_{2,12} = 3.89$ at 5%

Sol. Null hypothesis H_0: There is no significant difference

Sol Null hypothesis ... in No. of deaths among three hospitals ✓

H_1 : There is significant difference in No. of deaths ✓

A: x_1	x_1^2	B: x_2	x_2^2	C: x_3	x_3^2
3	9	6	36	7	49
4	16	3	9	3	9
3	9	3	9	4	16
5	25	4	16	6	36
0	0	4	16	5	25
$\Sigma x_1 = 15$	$\Sigma x_1^2 = 59$	$\Sigma x_2 = 20$	$\Sigma x_2^2 = 86$	$\Sigma x_3 = 25$	$\Sigma x_3^2 = 135$

$$G.T = \Sigma x_1 + \Sigma x_2 + \Sigma x_3 = 15 + 20 + 25 = 60$$ ✓

$$C.F. = \frac{(G.T)^2}{n} = \frac{(60)^2}{15} = \frac{3600}{15} = 240$$

correction factor

Sum of squares b/w samples (SSC) ✓

$$\Rightarrow \frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} - C.F.$$

$$= \frac{(15)^2}{5} + \frac{(20)^2}{5} + \frac{(25)^2}{5} - 240 = 10 = SSC$$

Total sum of squares (SST) ✓

$$= \Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 - C.F.$$

$$= 59 + 86 + 135 - 240 = 40 = SST$$

ANOVA Table

Source of Variation	Sum of Squares	Degree of freedom	Mean sum of squares (Variance)	F
Between Samples	$SSC = 10$	$c-1 = 3-1 = 2$	$MSC = SSC/(c-1) = 10/2 = 5$	$F = \frac{MSC}{MSE} = 5/2.5 = 2$
Within Samples	$SSE = 30$	$n-c = 12$	$MSE = SSE/(n-c) = 30/12 = 2.5$	
Total	$SST = 40$	$n-1 = 15-1 = 14$		

$F_{cal} = 2$

$F_{(2,12)\ cal} = 2$ $\qquad F_{cal} < F_{tab}$ Accepted H_0

$F_{(2,12)\ tab} = 3.89$

There is no significant diff.

Ex. The following figures relate to the production in kg. of three varities I, II, III of wheat sown in 12 plots.

Variety I	14	16	18		
Variety II	14	13	15	22	
Variety III	18	16	19	19	20

Is there any significant difference in production of three Varieties? Given $F_{2,9} = 4.26$ at 5% level significance.

... difference in population means

Sol H_0: No significant difference ...

$\mu_1 = \mu_2 = \mu_3$

H_1: There is significant difference

Use Code data by subtrating 12 from each figures

I X_1	X_1^2	II X_2	X_2^2	III X_3	X_3^2
2	4	2	4	6	36
4	16	1	1	4	16
6	36	3	9	7	49
		10	100	7	49
				8	64
$\Sigma X_1 = 12$	$\Sigma X_1^2 = 56$	$\Sigma X_2 = 16$	$\Sigma X_2^2 = 114$	$\Sigma X_3 = 32$	$\Sigma X_3^2 = 214$

$$G.T = \Sigma X_1 + \Sigma X_2 + \Sigma X_3 = 12 + 16 + 32 = 60$$

$$C.F. = \frac{(G.T)^2}{n} = \frac{(60)^2}{12} = 300$$

Sum of squares between Samples (SSC)

$$= \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - C.F.$$

$$= \frac{(12)^2}{3} + \frac{(16)^2}{4} + \frac{(32)^2}{5} - 300 = 16.8 = SSC$$

Total sum of squares (SST)

$$= \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - C.F.$$

$$= 56 + 114 + 214 - 300$$

$$\boxed{SST = 84}$$

Sum of squares within Samples

$$SSE = SST - SSC$$

$$= 84 - 16.8 = 67.2$$

$$\boxed{SSE = 67.2}$$

ANOVA Table

Variance

Source of Variation	Sum of Squares	Degree of freedom	Mean sum of squares	F
Between Samples	$SSC = 16.8$	$c-1 = 2$	$MSC = 16.8/2 = 8.4$	$F = \frac{MSC}{MSE} = 8.4/7.46 = 1.125$
Within Samples	$SSE = 84 - 16.8 = 67.2$	$n-c = 9$	$MSE = 67.2/9 = 7.46$	
Total	$SST = 84$	$n-1 = 11$		

$$F_{cal\,(2,9)} = 1.125$$

$$F_{tab\,(2,9)} = 4.86$$

$$F_{cal} < F_{tab} \Rightarrow H_0 \text{ Accept}$$

Topic: Statistical Quality Control (SQC) ✓

Statistical Quality Control is statistical method to ensure the quality of a product such as

* standards of quality acceptable to customers and economical ✓
* to locate ✓ and identify ✓ the process of faults to control defectives.
* to take corrective measures to maintain quality of products. ✓

Techniques of SQC

There are two broad ways of controlling the quality of product

✓ Process Quality Control ✓

⇒ Control the Quality of product during Production process. ✓

✓ Product Quality Control

⇒ Control the Quality of product by critical examination ✓

⇒ It is done by inspection of goods already produced ✓

Control Charts

The most common working statistical tools in Quality Control are <u>Control charts</u> and it is based upon the fact that <u>variability does exist</u> in all <u>repetitive process</u>.

This Variability is due to <u>chance</u> cause as well as assignable cause. A control chart eliminates error entirely due to assignable causes.

In SQC, Control charts is used to study and control the repetitive process

Control limits are defined within which variations are acceptable.

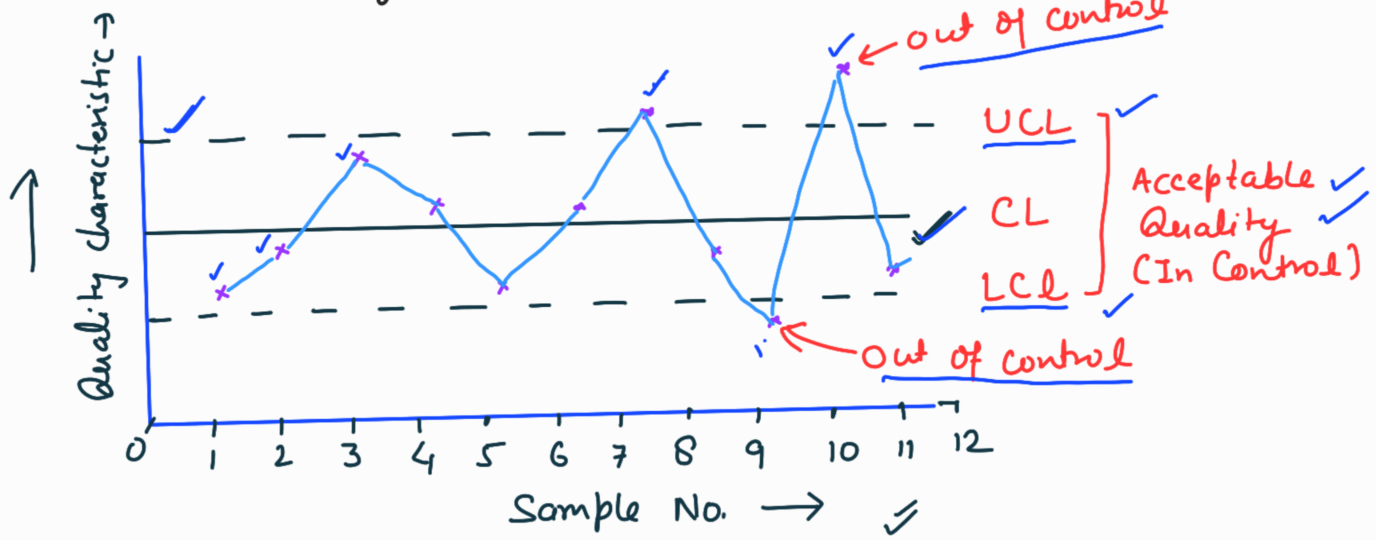
Any variations beyond these limits, can be expected to be caused by assignable cause which need to be controlled and investigated.

A Control chart consists of 3 Horizontal lines

* CL (control line) (solid line) represents desired control level of Process.

* UCL (upper control limit) (Dotted line) represents upper tolerance limit.

* LCL (lower control limit) (Dotted line) represents lower tolerance limit.



Types of Control Charts

* Control charts for Variables → are useful to measure quality characteristics, and control fully automatic process. like diameter of hole, thickness of pipe, length of bolt etc.

They are used for measurable quality characteristics

→ Control chart for Sample Mean ($\bar{X}$ -Chart)

$\Rightarrow$ Control chart for Sample Range (R-chart)

* Control Charts for Attributes $\rightarrow$ Inspection by Attributes,
It is used to maintain and acheive an acceptable quality level. It is classified as "Good" or 'Bad', 'Ok' or 'Not Ok', 'Rejected' or 'Accepted'

$\Rightarrow$ Control charts for fraction defective (p-chart)
$\Rightarrow$ Control chart for number of defective (np-chart)
$\Rightarrow$ Control chart for number of defects (C-chart)

Unit-5 (Lec-10)
(Maths-4)

Topic: Statistical Quality Control (SQC)

Control charts for Variables

A random sample of size n is drawn during manufacturing process

Sample mean $\bar{x} = \dfrac{x_1 + x_2 + x_3 \ldots + x_n}{n}$

Sample Range $R = x_{max} - x_{min}$

Let k consecutive samples are selected

$$\begin{array}{cccc} X_1, & X_2 & \ldots & X_k \\ \downarrow & \downarrow & & \downarrow \\ \bar{x}_1 & \bar{x}_2 & & \bar{x}_k \\ R_1 & R_2 & & R_k \end{array}$$

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \dots + \bar{x}_k}{k} = \frac{\Sigma \bar{x}}{\text{No. of Samples}}$$

$$\bar{R} = \frac{R_1 + R_2 + R_3 + \dots R_k}{k}$$

For X-Chart (Mean chart)

$$CL = \begin{cases} \bar{\bar{x}}, & \text{when tolerance limits are not given} \\ \mu, & \text{when tolerance limits are given} \end{cases}$$

$$\mu = \frac{1}{2}[\underbrace{LCL}_{\text{Lower Control limit}} + \underbrace{UCL}_{\text{Upper Control limit}}]$$

Standard Error of Mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The limits are $UCL = \bar{\bar{x}} + 3\sigma_{\bar{x}}$
$LCL = \bar{\bar{x}} - 3\sigma_{\bar{x}}$

$\boxed{\bar{\bar{x}} \pm 3\sigma_{\bar{x}}}$

Control limits are
$UCL = \bar{\bar{x}} + A_2\bar{R}$
$LCL = \bar{\bar{x}} - A_2\bar{R}$

A_2 are given in table ←

A_2 depends on sample size

For R-chart

$$CL = \bar{R}$$

$$LCL = \bar{R} D_3$$

$$UCL = \bar{R} D_4$$

D_3, D_4 can be found by table or given in Question. It is depend on sample size.

Process Capability is given by $\sigma = \dfrac{\bar{R}}{d_2}$, σ = S.D.

d_2 can be obtained by table and depend on sample size.

Ex. The following are mean length and ranges of lengths of a finished product from 10 samples each of size 5. The specification limits for length are 200 ± 5 cm. Construct $\bar{X}$ and R- chart and examine whether the process is under control and state your recommendation

Sample No	1	2	3	4	5	6	7	8	9	10
Mean ($\bar{X}$)	201	198	202	200	203	204	199	196	199	201
Range (R)	5	0	7	3	3	7	2	8	5	6

Assume for $n = 5$ $A_2 = 0.58$, $D_4 = 2.11$, $D_3 = 0$

Soln Control limits for $\bar{X}$ chart :→

Given limits 200 ± 5 (specification limits / Tolerance limits)

$$\bar{R} = \frac{R_1 + R_2 + R_3 \ldots + R_{10}}{10} = \frac{\Sigma R}{10}$$

$$= \frac{5+0+7+3+3+7+2+8+5+6}{10} = \frac{46}{10} = \boxed{4.6 = \bar{R}}$$

$\boxed{CL = \mu = 200}$

Given
$A_2 = .58$

$= 200 + (.58 \times 4.6)$

$$UCL_x = \bar{\bar{x}} + A_2\bar{R} = \mu + A_2\bar{R} = 200+(\cdot 58\times 4\cdot 6)$$
$$= \boxed{202\cdot 668}$$

$D_3 = 0$
$D_4 = 2\cdot 11$
$n = 5$

$$LCL_x = \bar{\bar{x}} - A_2\bar{R} = \mu - A_2\bar{R} = 200-(\cdot 58\times 4\cdot 6)$$
$$= \boxed{197\cdot 332}$$

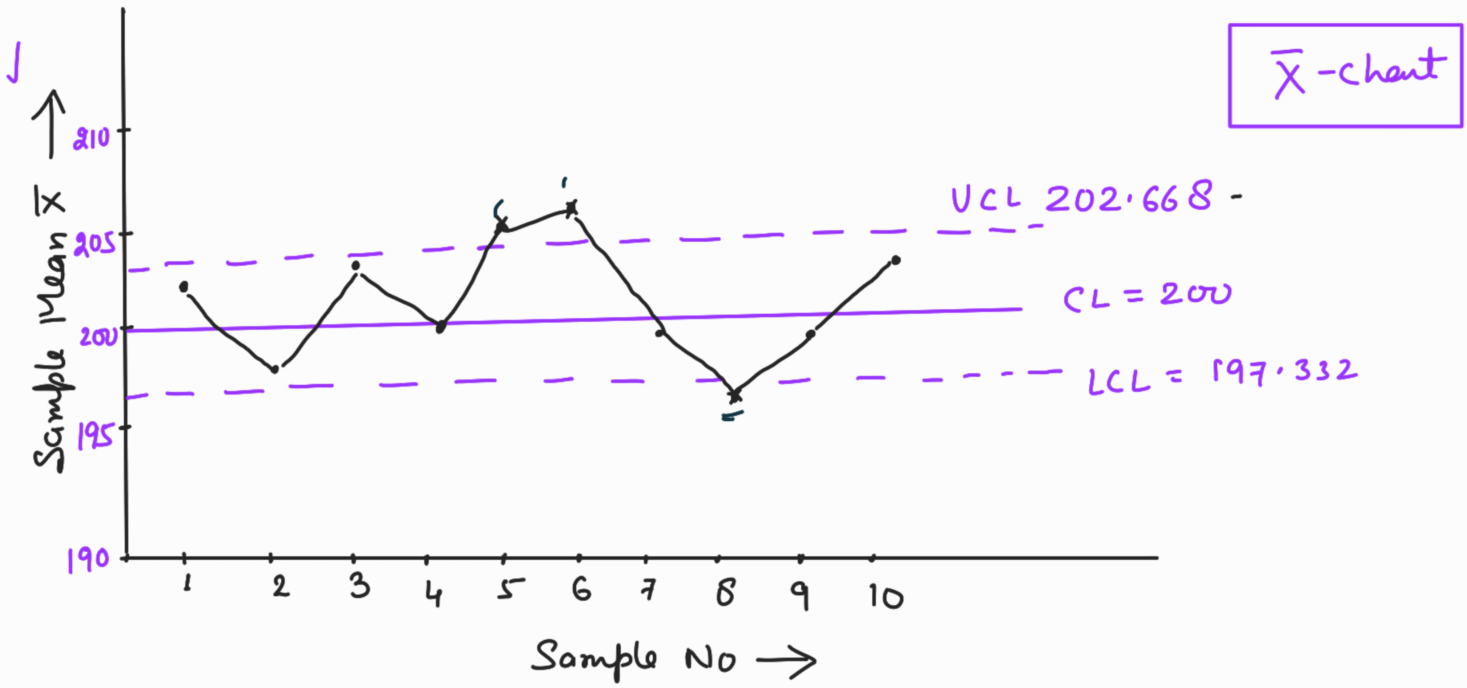
Control limits for R chart

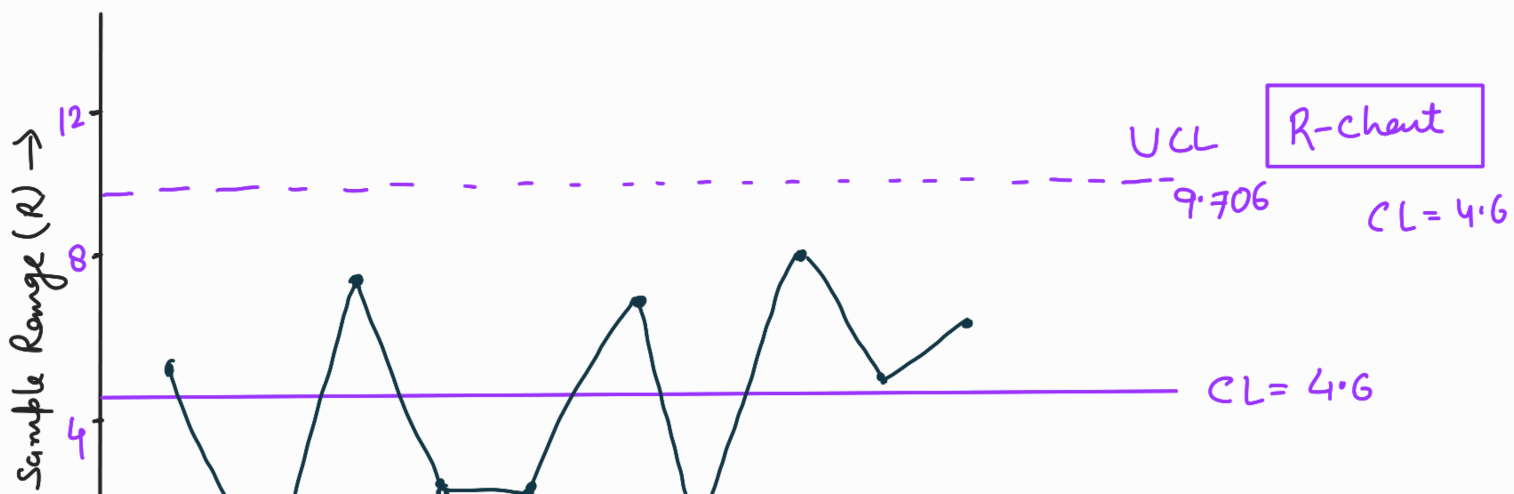
Given
$A_2 = \cdot 58$
$D_3 = 0$
$D_4 = 2\cdot 11$
$n = 5$

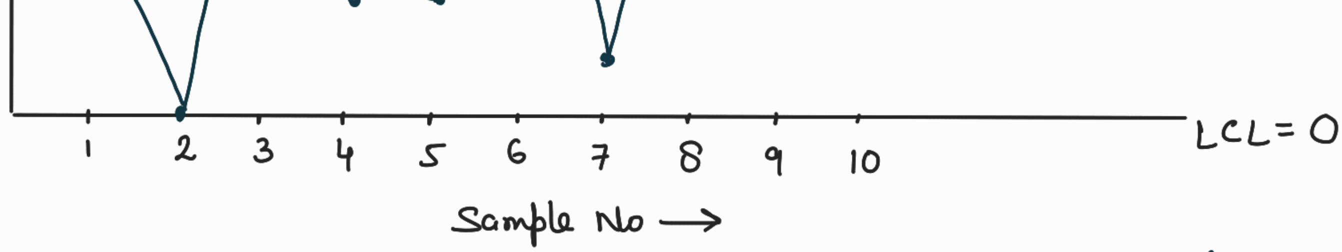
Control limit $CL = \boxed{\bar{R} = 4\cdot 6}$

$$UCL_R = \bar{R}D_4 = 4\cdot 6\times 2\cdot 11 = 9\cdot 706$$

$$LCL_R = \bar{R}D_3 = 4\cdot 6\times 0 = 0$$







Process variability is under control. But $\bar{X}$-chart is not in statistical control. process should be halted.

Ex. A company manufactures screws to a nominal diameter $.500 \pm .030$ cm. Five samples were taken randomly from the manufactured lots and 3 measurements were taken on each sample at different lengths. Following are readings

Sample No ↓	Measurement per sample x (in cm) 1	2	3
1	.488	.489	.505
2	.494	.495	.499
3	.498	.515	.487
4	.492	.509	.514
5	.490	.508	.499

Calculate Control limits for $\bar{X}$, R charts. Draw $\bar{X}$, R Charts and examine whether the process is in Statistical Control? $[A_2 = 1.02,\ D_4 = 2.57,\ D_3 = 0$ for $n = 3]$

Sol

Sample	1 (x_1)	2 (x_2)	3 (x_3)	4 (x_4)	5 (x_5)
	.488	.494	.498	.492	.490
	.489	.495	.515	.509	.508
	.505	.499	.487	.514	.499
Mean	$\frac{1.482}{3}$ = .494	$\frac{1.488}{3}$ = .496	$\frac{1.5}{3}$ = .500	$\frac{1.515}{3}$ = .505	= .499
Range	.505 − .488 = .017	.499 − .494 = .005	.515 − .487 = .028	.514 − .492 = .022	.508 − .490 = .018

Range = Max − min

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4 + \bar{x}_5}{5} = \frac{.494 + .496 + .5 + .505 + .499}{5}$$

$$\boxed{\bar{\bar{x}} = .4988}$$

$$\bar{R} = \frac{R_1 + R_2 + R_3 + R_4 + R_5}{5} = \frac{.017 + .005 + .028 + .022 + .018}{5}$$

$$= \frac{.090}{5} = .018$$

$$\boxed{\bar{R} = .018}$$

Trial Control limits for $\bar{X}$-chart

$$CL = \bar{\bar{x}} = .4988$$

$$UCL = \bar{\bar{x}} + A_2\bar{R} = .4988 + (1.02) \times .018 = .5172$$

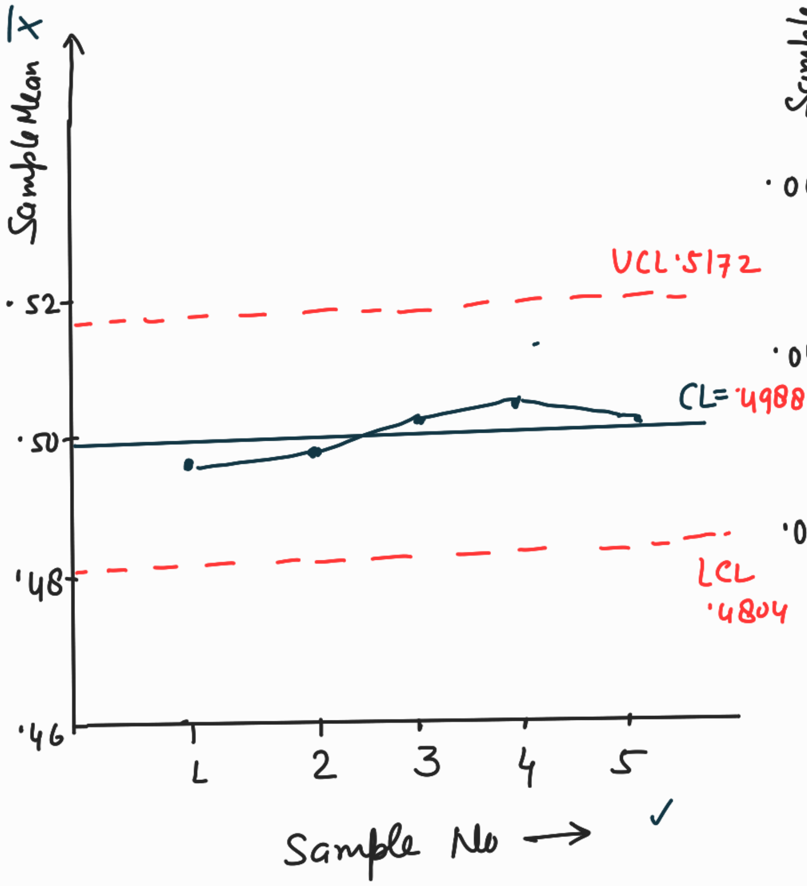
$$LCL = \bar{\bar{x}} - A_2\bar{R} = .4988 - (1.02) \times .018 = .4804$$

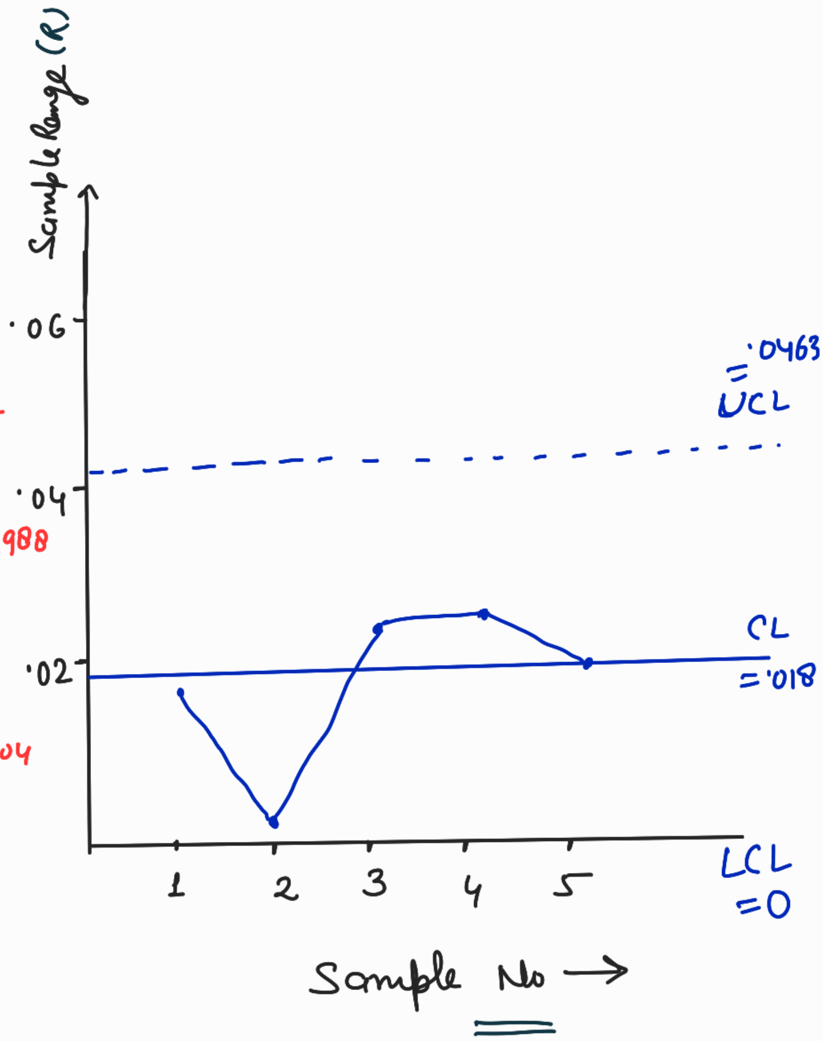
Control limits for R-charts

$$CL = \overline{R} = \cdot 018$$

$$UCL = D_4 \overline{R} = 2\cdot 57 \times \cdot 018 = \cdot 0463$$

$$LCL = D_3 \overline{R} = 0 \times \cdot 018 = 0$$





Process is in statististical Control

Practice Question

A. drilling machine bores holes with a mean diameter of ·5230 cm and a standard deviation of ·0032 cm. Calculate 2σ and 3σ upper and lower control limit for means of sample of 4

[Hint:→ 2σ limit

$$\left.\begin{aligned} CL &= \bar{\bar{x}} &&= .5230 \\ UCL &= \bar{\bar{x}} + 2\frac{\sigma}{\sqrt{n}} &&= .5262 \\ LCL &= \bar{\bar{x}} - 2\frac{\sigma}{\sqrt{n}} &&= .5198 \end{aligned}\right\}$$

3σ limit

$$\left.\begin{aligned} CL &= \bar{\bar{x}} &&= .5230 \\ UCL &= \bar{\bar{x}} + 3\frac{\sigma}{\sqrt{n}} &&= .5278 \\ LCL &= \bar{\bar{x}} - 3\frac{\sigma}{\sqrt{n}} &&= .5182 \end{aligned}\right\}$$

Unit-5 (Lec-11)
(Maths-4)

Topic: Statistical Quality Control (SQC)

Control charts for Attributes (p-chart, np-chart, C-chart)

C-chart → C-chart is used when no. of defects per unit are counted. They are not classified as 'defective' or 'Not defective'.

C = No. of defects per Unit

C - follows Poission distribution

$\bar{C}$ = Mean of C = $\frac{\text{No of defects in all samples}}{\text{Total No of samples}}$

S.D $\sigma_C = \sqrt{\bar{C}}$

Thus 3σ limits for C-charts

$UCL_c = \bar{C} + 3\sqrt{\bar{C}}$ $\qquad \bar{C} + 3\sigma$

$LCL_c = \bar{C} - 3\sqrt{\bar{C}}$ $\qquad \bar{C} - 3\sigma$

$CL = \bar{C}$

Ex. Draw a C-chart for following data pertaining to the number of foreign coloured threads (considered as defects) in 15 piece of cloths of 2m×2m in a certain make of synthetic fibre and state your conclusion

7, 12, 3, 20, 21, 5, 4, 3, 10, 8, 0, 9, 6, 7, 20

Sol No of cloth piece = 15

Total No of defects = 7+12+3+20+21+5+4+3+10+8+0+9+6+7+20 = 135

$\bar{C} = \frac{\text{Total defects}}{\text{total samples}} = \frac{135}{15} = 9$

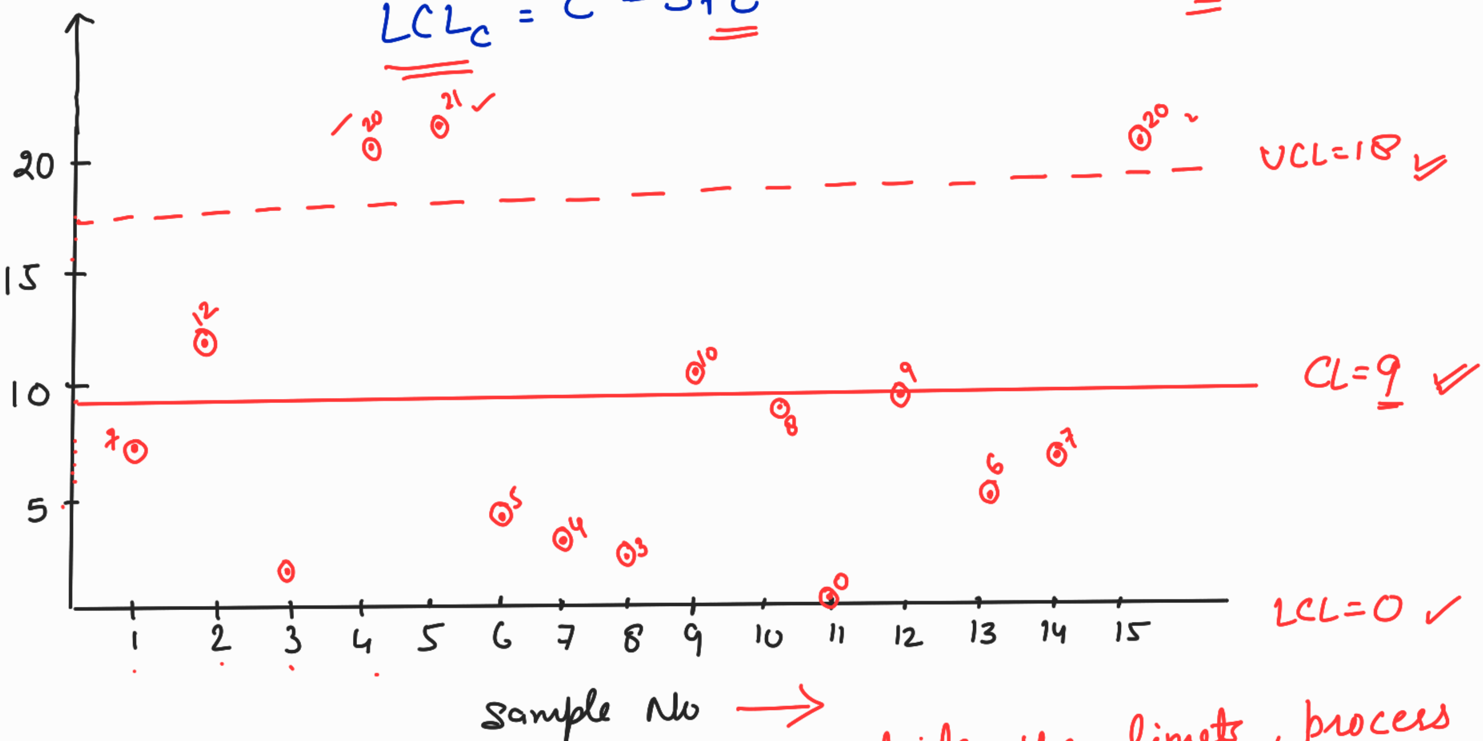
Total samples

3σ control limits $\qquad \sigma = \sqrt{\bar{c}}$

$$CL = \bar{c} = 9$$

$$UCL_c = \bar{c} + 3\sqrt{\bar{c}} = 9 + 3\sqrt{9} = 18$$

$$LCL_c = \bar{c} - 3\sqrt{\bar{c}} = 9 - 3\sqrt{9} = 0$$



Three sample points are outside the limits, process is not under statistical control.

Fraction defective chart (P-chart)

This is designed to control the proportion (p) percentage ($100p$) or defective per sample

$$\text{Fraction defective } (p) = \frac{\text{No. of defective item}}{\text{Sample size}} = d/n$$

Mean fraction defective $= \bar{p}$

$$\bar{p} = \frac{\text{Total No of defectives in all sample}}{\text{Total No of items in all samples}}$$

The statistical theory Binomial distribution is used to construct p- chart.

fraction defective $= \bar{p}$

Mean value of fraction ...

" " S.D $= \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sigma_p$

$UCL_p = \bar{p} + 3\sigma_p = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ $\qquad \bar{p} + 3\sigma_p$

$LCL_p = \bar{p} - 3\sigma_p = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ $\leftarrow$ $\qquad \bar{p} - 3\sigma_p$

Note → • In sometime, LCL comes out negative, it is taken as zero.

* It is used to control fraction defectives when sample size does not remain uniform or it varies.

np-Chart :→ Use when sample size is constant.

$$\bar{p} = \frac{\sum np}{\sum n}$$

$$\sigma_{np} = n\sigma_p = n\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{n\bar{p}(1-\bar{p})}$$

$CL = n\bar{p}$

$UCL_{np} = n\bar{p} + 3\sigma_{np} = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$

$LCL_{np} = n\bar{p} - 3\sigma_{np} = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$

Note • if LCL comes out negative, it is taken as zero.

Ex. The average percentage of defectives in 27 samples of size 1500 each was found to be 13.7%. Construct a suitable control chart for this situation. Discuss how the control chart can be used to control quality.

$k = 27$

$n = 1500$

$\bar{p} = .137$

$\sigma_p = \sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}$

$CL = \bar{p} = .137$

$UCL = \bar{p} + 3\sigma_p = .137 + 3\sqrt{\dfrac{.137(1-.137)}{1500}}$

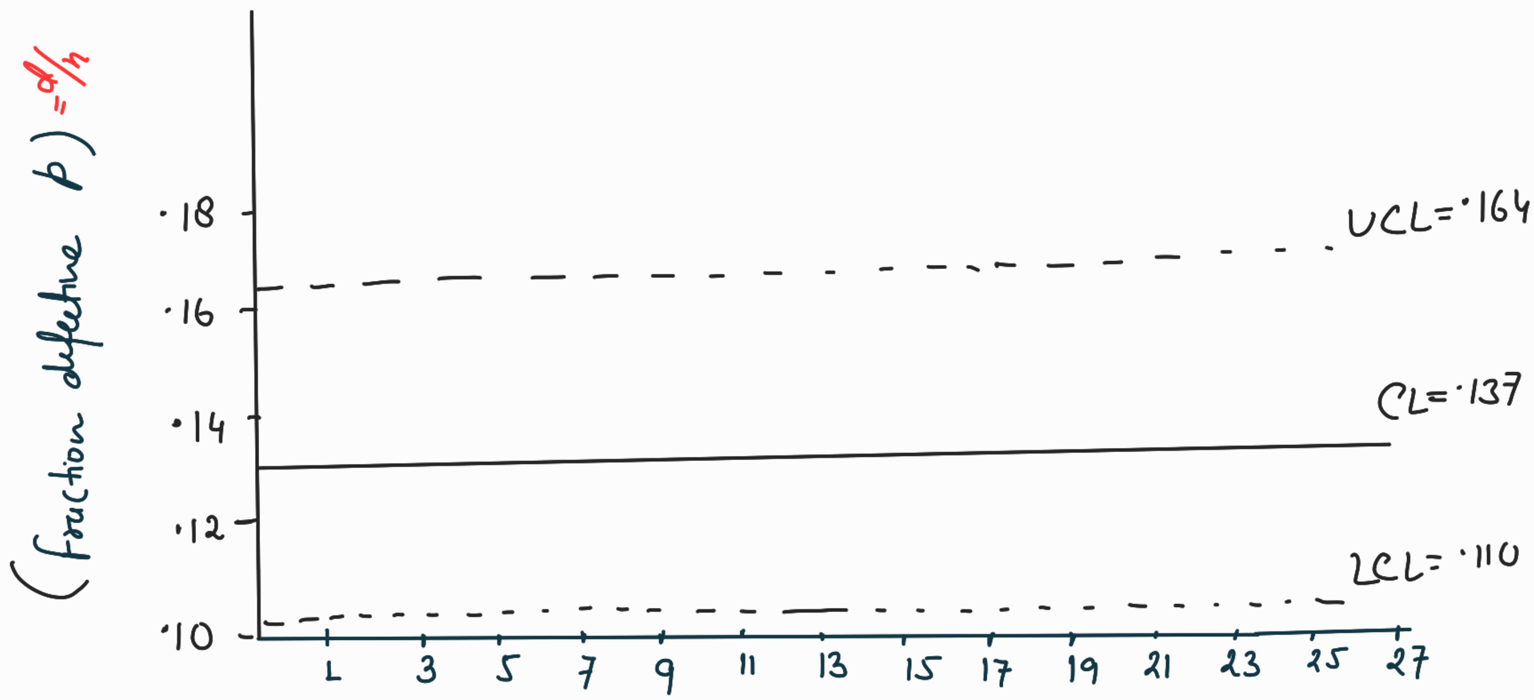
$= .137 + .027$

$\boxed{UCL = .164}$

$LCL = \bar{p} - 3\sigma_p = .137 - 3\sqrt{\dfrac{.137(1-.137)}{1500}}$

$= .137 - .027$

$\boxed{LCL = .110}$



Sample No →

Ex. In a blade manufacturing factory. 1000 blades are examined daily. Draw the np chart for the following table and examine whether the

the following table and examine whether the process is under control? [2010]

Date :→	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No of defective blades :→	9	10	12	8	7	15	10	12	10	8	7	13	14	15	16

Sol^n

$n = 1000$

Total No of items = $1000 \times 15 = 15000$

Total No of defective = Σd

$= 9+10+12+8+7+15+10+12+10+8+7+13+14+15+16$

$= 166$

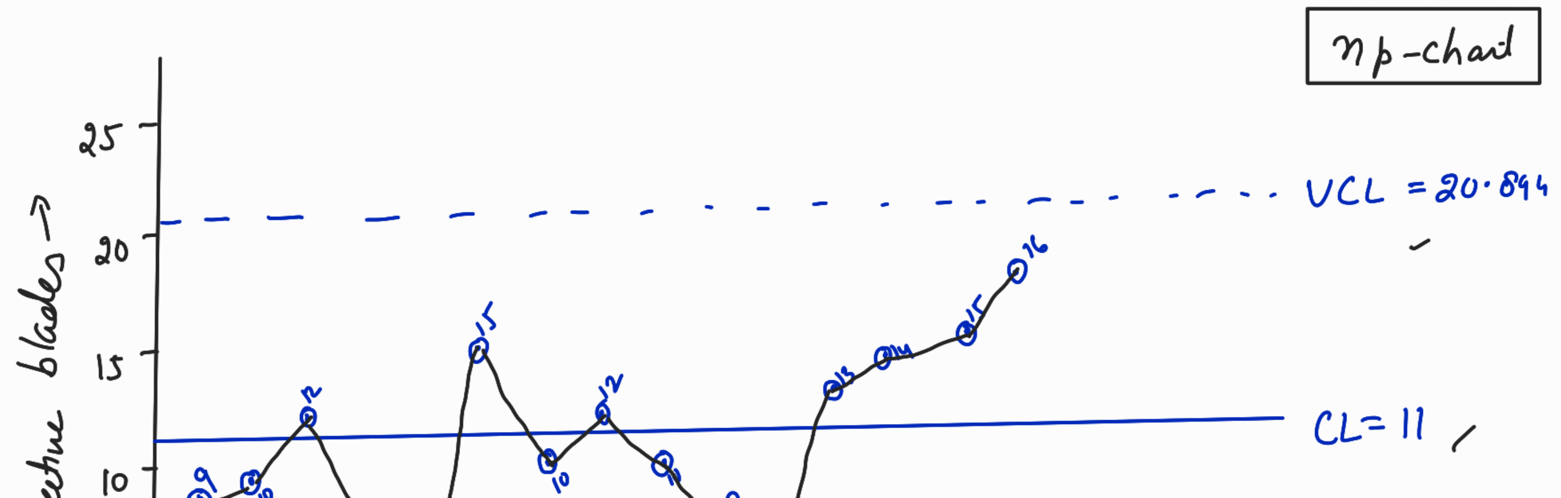
$$\bar{p} = \frac{\Sigma d}{N} = \frac{166}{15000} = .011$$

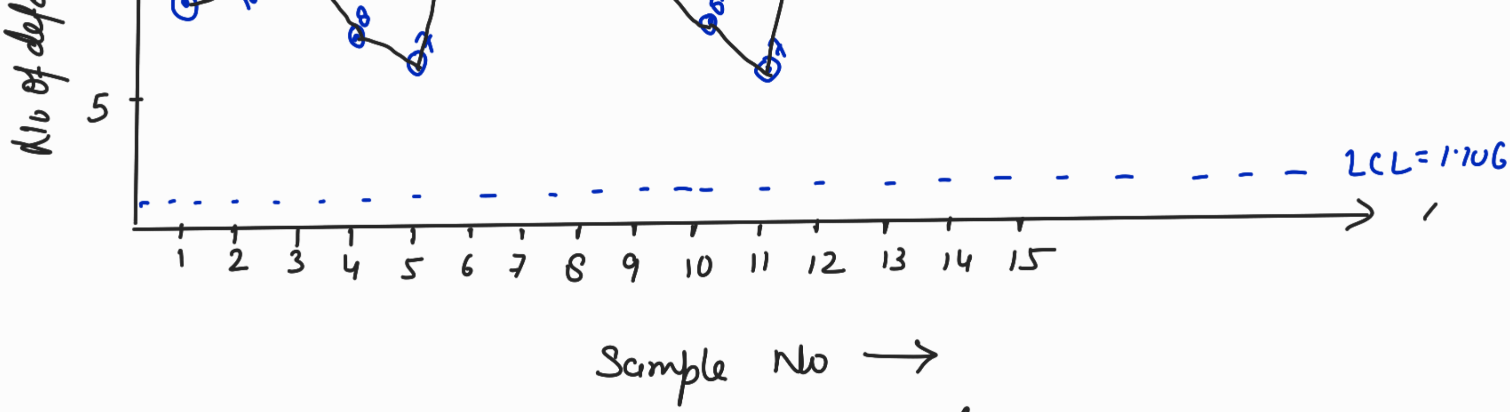
$n\bar{p} = 1000 \times .011 = 11$

$CL = n\bar{p} = 11$

$$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})} = n\bar{p} + 3\sigma_{np} = 11 + 3\sqrt{11(1-.011)}$$
$$= 20.894$$

$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})} = n\bar{p} - 3\sigma_{np} = 11 - 3\sqrt{11(1-.011)}$$
$$= 1.106$$





process is under control

Practice Question

Q→ Distinguish between p-chart and np-chart. Following is the data of defective of 10 samples of size 100 each. Construct np-chart and give your comment

Sample No	1	2	3	4	5	6	7	8	9	10
No of defectives	6	9	12	5	12	8	8	16	13	7

[GBTU 2010, 2020-21]

[Hint $n = 100$, $\bar{p} = .096$, $n\bar{p} = 9.6$

$CL = 9.6$, $UCL = 18.44$, $LCL = .762$

Process is in statistical control]