25/11

Unit - 4

Small Sample

Test of Significance for Small Samples:

The relation b/w sample S.D (s) & population S.D ($\underline{S}$) is $ns^2 = (n-1)\underline{S}^2$

Test of significance for small sample single mean = $t_{cal} = \dfrac{\bar{x} - \mu}{\dfrac{\underline{S}}{\sqrt{n}}}$ (or) $\dfrac{\bar{x}-\mu}{\dfrac{s}{\sqrt{n-1}}}$

1) A mechanist is making engine parts with axle diameters of 0.7 inch. A random sample of 10 parts shows a mean diameter of 0.742 with a S.D 0.040. Compute the statistic you would use to test whether work is meeting the specification at 0.05 level of significance

$n = 10$ (<30 small sample)

$\mu = 0.7$

$\bar{x}$ = sample mean = 0.742

$s = 0.040$

$\alpha = 0.05$

Sample (Small) Single mean problem

Null Hypothesis : $\mu = 0.7$

Alternate hypothesis : $\mu \neq 0.7$ (two tail)

level of significance $\alpha = 0.05$

Test Statistic $t_{cal} = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$ $\left[\because ns^2 = (n-1)S^2\right]$

$$= \frac{0.742 - 0.7}{\dfrac{0.040}{\sqrt{10-1}}} = 3.15$$

$t_{tab} = t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}$ with degrees of freedom $n-1$ i.e $10-1 = 9$ is 2.262

$|t_{cal}| > t_{tab}$

reject null hypothesis
accept alternate hypothesis $\mu \neq 0.7$

2) A sample of 26 bulbs gives a mean life of 990 hrs with a S.D of 20 hrs. The manufacturer claims that the mean life of bulbs is 1000 hrs. Is the sample not upto the standards.

$n = 26$ (Small Smaple)

$\bar{x} = 990$

$s = 20$ (Small Sample Single mean)

$\mu = 1000$

Null hypothesis $H_0 : \mu = 1000$

Alternate hypothesis $H_1 : \mu < 1000$ (left tail)

level of significance $\alpha = 0.05$

test statistic $t_{cal} = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$

$$= \frac{990 - 1000}{\frac{20}{\sqrt{26-1}}}$$

$$= -2.5$$

$|t_{cal}| = |-2.5| = 2.5$

$t_{tab} = t_\alpha = t_{0.05}$ with degrees of freedom $n-1 = 26-1 = 25$ is 1.708

reject null hypothesis $t_{cal} > t_{tab}$

accept alternate hypothesis

mean life of bulbs is less than 1000 hrs

3) The avg breaking strength of steel rods is specified to be 18.5000 pounds. To test this sample of 14 rods were tested. The mean & S.D obtained were 17.85 and 1.955 respectively. Is the result of experiment significant

$\mu = 18.5000$

$n = 14$ (Sample small)

$\bar{x} = 17.85$

$s = 1.955$

$\alpha = 0.05$

Null hypothesis $H_0 : \mu = 18.5000$

Alternate hypothesis $H_1 : \mu \neq 18.5000$

level of significance $\alpha = 0.05$

test statistics $t_{cal} = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$

$= \dfrac{17.85 - 18.5000}{\dfrac{1.955}{\sqrt{14-1}}}$

$= -1.198$

$|t_{cal}| = |-1.1981| = 1.198$

$t_{tab} = t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}$ with degress of freedom $n-1$ i.e $14-1 = 13$ is 2.16

$t_{cal} < t_{tab}$

accept null hypothesis i.e $\mu = 18.5000$

4) A Sample of 15 members has mean 67 & S.D 5.2. Is this sample has been taken from large population of mean 70

$n = 15 \quad \bar{x} = 67$

$\mu = 70 \quad s = 5.2$

Null hypothesis $H_0 : \mu = 70$

alternate hypothesis $H_1 : \mu \neq 70$

level of significance $\alpha = 0.05$

test statistics $t_{cal} = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$

$= \dfrac{67 - 70}{\dfrac{5.2}{\sqrt{15-1}}}$

$= -2.15$

$|t_{cal}| = |-2.15| = 2.15$

$t_{tab} = t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}$ with degrees of freedom $n-1$ i.e $15-1 = 14$ is 2.145

$t_{cal} > t_{tab}$

reject null hypothesis

accept alternate hypothesis

A sample of 100 iron bars is said to be drawn from a)

5) A machine is designed to produce insulating washers for electrical devices of avg thickness of 0.025 cm. A random sample of 10 washers was found to have a thickness of 0.024 cm and S.D of 0.02 cm. Test the significance of the deviation.

2.262

$\mu = 0.025$

$n = 10$

$\bar{x} = 0.024$

$s = 0.02$

Null hypothesis $H_0 : \mu = 0.025$

Alternate hypothesis $H_1 : \mu \neq 0.025$

level of significance $\alpha = 0.05$

test statistics $t_{cal} = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$

$= \dfrac{0.024 - 0.025}{\frac{0.02}{\sqrt{10-1}}}$

$= -0.15$

$|t_{cal}| = |-0.15| = 0.15$

$t_{tab} = t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}$ with degrees of freedom $n-1$ i.e $10-1 = 9$ is 2.262

$t_{cal} < t_{tab}$

accept null hypothesis

Test for difference of means

Let $\bar{x}_1$ & $\bar{y}$ be the means of two independents samples of sizes n_1 & n_2 ($n_1 < 30$, $n_2 < 30$) drawn from two normal populations having means μ_1 & μ_2. Then test statistic t is $\dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$t_{cal} = \frac{\bar{x} \cdot \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{where } S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

where S_1^2 & S_2^2 are sample variance (or)

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

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

1 Samples of two types of electrical bulbs were tested for length of life and following data were obtained

	Type I	Type II
Sample no/ Size	8	7
Sample mean	1234	1036
Sample Standard deviation	36	40

is the difference in the means sufficient to waret that type I is superior to type II regarding length life

$n_1 = 8 < 30$ (Small Sample)

$n_2 = 7 < 30$ (Small Sample)

$\bar{x} = 1234$

$\bar{y} = 1036$

$S_1 = 36$

$S_2 = 40$

The given problem is Small Sample difference of means

Null hypothesis $(H_0) : \mu_1 = \mu_2$

Alternate hypothesis $(H_1) : \mu_1 > \mu_2$

Level of Significance $\alpha = 0.05$

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

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

$$S^2 = \frac{8 \times (36)^2 + 7(40)^2}{8 + 7 - 2} = 1659.07$$

$S = 40.73$

$$t_{cal} = \frac{1234 - 1036}{40.73\sqrt{\frac{1}{8} + \frac{1}{7}}} = 9.39$$

$|t_{cal}| = 9.39$ — (1)

$\therefore t_{tab} = t_\alpha = t_{(0.05)}$ with degrees of freedom

$n_1 + n_2 - 2 = 13$

$t_{tab} = t_{(0.05)} = 1.771$ — (2)

$|t_{cal}| > t_{tab}$

$\therefore$ reject null hypothesis

$\therefore$ Accept alternate hypothesis

2) Measuring Specimens of Nytern yavon from 2 machines it was found that 8 Specimens from 1st machine had a mean deniyed of 9.67 with a Standard deviation of 1.81 while 10 Specimens from 2nd machine had a mean deniyer of 7.43 with a Standard deviation of 1.48. Assuming the proportions are normal. Test the hypothesis $\mu_1 - \mu_2 = 1.5$ against H_1 $\mu_1 - \mu_2 > 1.5$ at 0.05 level of Significance.

$n_1 = 8 < 30$ (Small Sample)

$\bar{x} = 9.67$

$S_1 = 1.81$

$n_2 = 10 < 30$ (Small Sample)

$\bar{y} = 7.43$

$S_2 = 1.48$

$\alpha = 0.05$

The given

of mean

Null hy

Alternat

level of

Test S

where

$S^2 = n_1 s$

n_1

$S =$

t_{cal}

$|t_{c}$

t_{tab}

$n_1 + n_2$

$|t$

$\therefore$ Acc

3. The IQ

City S

deviati

from

Is the

IQ's

Signifi

$n_1 =$

$\bar{x} =$

$S_1 =$

$n_2 =$

The g

of

$\bar{y} = 7.43$
$S_2 = 1.48$
$\alpha = 0.05$

The given problem is small sample difference of means problem

Null hypothesis (H_0) : $\mu_1 - \mu_2 = 1.5$

Alternate hypothesis (H_1) : $\mu_1 - \mu_2 > 1.5$

level of significance $= 0.05$

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

where

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} = \frac{8\,(1.81)^2 + 10\,(1.48)^2}{8 + 10 - 2} = 3.00705$$

$S = 1.734$

$$t_{cal} = \frac{9.67 - 7.43 - 1.5}{1.734\sqrt{\frac{1}{8} + \frac{1}{10}}} = 0.89$$

$|t_{cal}| = 0.89$ — (1)

$t_{tab} = t_\alpha = t_{0.05}$ with degrees of freedom

$n_1 + n_2 - 2 = 16$ is $t_{tab} = t_{(0.05)} = 1.746$ — (2)

$|t_{cal}| < t_{tab}$

$\therefore$ Accept null hypothesis

$\mu_1 - \mu_2 = 1.5$

3. The IQ's of 16 students from one area of a city showed a mean of 107 with a standard deviation of 10. While the IQ's of 14 students from another area of city showed a mean of 8. Is there any significant difference between the IQ's of two different group at 0.05 level of significance

$n_1 = 16$ $\quad$ $\bar{y} = 112$
$\bar{x} = 107$ $\quad$ $S_2 = 8$
$S_1 = 10$ $\quad$ $\alpha = 0.05$
$n_2 = 14$

The given problem is small sample difference of means

Null hypothesis (H_0) : $\mu_1 = \mu_2$

Alternate hypothesis $(H_1) = \mu_1 \neq \mu_2$

level of Significance $(\alpha) = 0.05$

Test statistic $t_{cal} = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$.

where $S^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \dfrac{16(10)^2 + 14(8)^2}{16 + 14 - 2}$

$S^2 = 89.14$

$S = 9.44$

$t_{cal} = \dfrac{107 - 112}{9.44\sqrt{\frac{1}{16} + \frac{1}{14}}} = -1.447$

$t_{tab} = t_{\frac{0.05}{2}} = t_{0.025}$ with degrees of freedom $16 + 14 - 2 = 28$ is 2.048

$|t_{cal}| < t_{tab}$

$\therefore$ Accept null hypothesis

$\mu_1 = \mu_2$

4. The means of two random samples of sizes 9 & 7 are 196.42 & 198.82 respectively. The Sum of squares of deviations of means 26.94 & 18.73. Can the sample be considered to have been drawn from the same normal population

$n_1 = 9$

$n_2 = 7$ $\qquad \bar{y} = 198.82$

$\bar{x} = 196.42$ $\qquad \Sigma(x_i - \bar{x})^2 = 26.94$

$\Sigma(y_i - \bar{y})^2 = 18.73$

The given problem is small sample difference of means problem

$\alpha = 0.05$

Null hypothesis (H_0) : $\mu_1 = \mu_2$

Alternate hypothesis (H_1) : $\mu_1 \neq \mu_2$

level of Significance $(\alpha) = 0.05$

Test Stastistic $t_{cal} = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

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

$$= \frac{(26.94)+(18.73)}{9+7-2} = 3.26$$

$$s = 1.805$$

$$t_{cal} = \frac{196.42-198.82}{1.805\sqrt{\frac{1}{9}+\frac{1}{7}}} = -2.63$$

$$|t_{cal}| = 2.63$$

$t_{tab} = t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}$ with degrees of freedom $n_1+n_2-2 = 9+7-2 = 14$ is 2.145

$$|t_{cal}| > t_{tab}$$

$\therefore$ Reject null hypothesis

$\therefore$ Accept alternate hypothesis

$$\mu_1 \neq \mu_2$$

5) A group of five patients treated with medicine A weigh 42, 39, 48, 60, 41 kgs. Second group of 7 patients from the same hospital treated with medicine B weigh 38, 42, 56, 64, 68, 69, 62. Do you agree with that claim that medicine B increases the weight significantly.

$$\bar{x} = \frac{\sum x_i}{n_1} = \frac{42+39+48+60+41}{5} = 46$$

$$\bar{y} = \frac{\sum y_i}{n_2} = \frac{38+42+56+64+68+69+62}{7} = 57$$

x_i	$\bar{x}$	(x_i-x)	$(x_i-\bar{x})^2$	y_i	$\bar{y}$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
42	46	-4	16	38	57	-19	361
39	46	-7	49	42	57	-15	225
48	46	2	4	56	57	-1	1
60	46	14	196	64	57	7	49
41	46	-5	25	68	57	11	121
				69	57	12	144
				62	57	5	25
			$\sum(x_i-x)^2 = 290$				$\sum(y_i-\bar{y})^2 = 926$

$n_1 = 5$

$n_2 = 7$

Null hypothesis (H_0) : $\mu_1 = \mu_2$
Alternate hypothesis (H_1) : $\mu_1 < \mu_2$
Level of significance $\alpha = 0.05$

$$S^2 = \frac{\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2}{n_1+n_2-2} = \frac{290+296}{5+7-2} = 121.6$$

$S = 11.03$

$$\text{test statistic } (t_{cal}) = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{46-57}{(11.03)\sqrt{\frac{1}{5}+\frac{1}{7}}} = -1.7$$

$|t_{cal}| = 1.704$

$|t_{cal}| = 1.7$

$t_{tab} = t_\alpha = t_{0.05}$ degrees of freedom

$5+7-2 = 10$ is 1.81

$t_{tab} = 1.81$

$|t_{cal}| < t_{tab}$

$\therefore$ Accept null hypothesis

$\mu_1 = \mu_2$

6) Two horses A & B were according to run on a particular track with following results

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whether two horses have same running capacity

$n_1 = 7$

$n_2 = 6$

$$\bar{x} = \frac{\sum x_i}{n_1} = \frac{28+30+32+33+33+29+34}{7} = 31.28$$

$$\bar{y} = \frac{\sum y_i}{n_2} = \frac{29+30+30+24+27+29}{6} = 28.166$$

x_i	$\bar{x}$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	y_i	$\bar{y}$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
28	31.286	-3.286	10.747	29	28.166	0.834	0.695
30	31.286	-1.286	1.653	30	28.166	1.834	3.363
32	31.286	0.714	0.509	30	28.166	1.834	3.363
33	31.286	1.714	2.937	24	28.166	-4.166	17.355
33	31.286	1.714	2.937	27	28.166	-1.166	1.359

29	31.286	-2.286	5.225	29	28.166	0.834	0.695
34	31.286	2.714	7.365				
			=31.42				=26.83

Null hypothesis (H_0) : $\mu_1 = \mu_2$

Alternate hypothesis (H_1) : $\mu_1 \neq \mu_2$

level of significance $\alpha = 0.05$

$$S^2 = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2} = \frac{31.42 + 26.83}{7+6-2} = 5.295$$

$S = 2.301$

$$\text{test statistics } (t_{cal}) = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{31.286 - 28.166}{2.301\sqrt{\frac{1}{7} + \frac{1}{6}}}$$

$$|t_{cal}| = 2.437 \quad -① $$

$t_{tab} = t_{\frac{0.05}{2}} = t_{0.025}$ with degrees of freedom

$n_1 + n_2 - 2 = 7+6-2 = 11$ is $t_{tab} = 2.2$

$|t_{cal}| > t_{tab}$

$\therefore$ Reject null hypothesis

$\therefore$ Accept alternate hypothesis

$\mu_1 \neq \mu_2$

7) Below are given the gain in weights of pigs fed on two diets A and B

Diet A	25	32	30	34	24	14	32	24	30	31	35	25	-	-	-
Diet B	44	34	22	10	47	31	40	30	32	35	18	21	35	29	22

Test if two diets differ significantly as regards their effects on increase in weights?

$$\bar{x} = \frac{25+32+ - - - +25}{12} = 28$$

$$\bar{y} = \frac{44+34+ - - - - + 22}{15} = 30$$

$n_1 = 12$

$n_2 = 15$

x_i	$\bar{x}$	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	y_i	$\bar{y}$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
25	128	3	9	44	30	14	196
32	128	4	16	34	30	4	16
30	128	2	4	22	30	-8	64
34	128	6	36	10	30	-20	400
24	128	-4	16	47	30	17	289
14	128	-14	196	31	30	1	1
32	126	4	16	40	30	10	100
24	128	-4	16	30	30	0	0
30	128	2	4	32	30	2	4
31	128	3	9	35	30	5	25
35	128	7	49	18	30	-12	144
25	128	-3	9	21	30	-9	81
				35	30	5	25
				29	30	-1	1
				22	30	-8	64
			318				1410

Null hypothesis (H_0) : $\mu_1 = \mu_2$

Alternate hypothesis (H_1) : $\mu_1 \neq \mu_2$

level of Significance $(\alpha) = 0.05$

$$S^2 = \frac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1+n_2-2} = \frac{3880+1410}{12+15-2} = 71.6$$

$S = 8.46$

$$\text{test Statistic } (t_{cal}) = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{28-30}{8.46\sqrt{\frac{1}{12}+\frac{1}{15}}} = 0.61$$

$|t_{cal}| = 0.61$

$t_{tab} = 2.06$

$|t_{cal}| < t_{tab}$

$\therefore$ Accept null hypothesis

$\mu_1 = \mu_2$

8) To examine the hypothesis that the husbands are more intelligent than the wifes. An investigator took a sample of 10 Couples & administered them a test which measures the IQ. The results are as follows

Husband	117	10
wifes	106	98

Test the hypoth

the level of Si

$\bar{x} = \frac{117+105+}{10}$

$\bar{y} = \frac{106+982}{1}$

$n_1 = 10$ $n_2 = 1($

x_i	$\bar{x}$	$x_i-\bar{x}$
117	103	14
105	103	2
97	103	-6
105	103	2
123	103	20
109	103	6
86	103	-17
78	103	-25
103	103	0
107	103	4

Null hypothe

Alternate hy

level of Sig

$S^2 = \Sigma(x$

$S = 13.51$

test Statis

$|t_{cal}$

$t_{tab} = t_{0.05}$

is $t_{tab} = 1$

$\therefore t_{cal} <$

$\therefore$ Accept

Husband	117	105	97	105	123	109	86	78	103	107
Wifes	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with a reasonable test at the level of significance of 0.05?

$$\bar{x} = \frac{117+105+ - - +107}{10} = 103$$

$$\bar{y} = \frac{106+98+ - - +85}{10} = 95.8$$

$n_1 = 10 \quad n_2 = 10$

x_i	$\bar{x}$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	y_i	$\bar{y}$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
117	103	14	196	106	95.8	10.2	104.04
105	103	2	4	98	95.8	2.2	4.84
97	103	-6	36	87	95.8	-8.8	77.44
105	103	2	4	104	95.8	8.2	67.24
123	103	20	400	116	95.8	20.2	408.04
109	103	6	36	95	95.8	-0.8	0.64
86	103	-17	289	90	95.8	-5.8	33.64
78	103	-25	625	69	95.8	-26.8	18.24
103	103	0	0	108	95.8	12.2	148.84
107	103	4	16	85	95.8	-10.8	116.64
			1606				1679.6

Null hypothesis $(H_0) : \mu_1 = \mu_2$

Alternate hypothesis $(H_1) : \mu_1 > \mu_2$

level of significance $(\alpha) = 0.05$

$$S^2 = \frac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1+n_2-2} = 182.533$$

$S = 13.51$

$$\text{test Statistic } (t_{cal}) = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{103-95.8}{13.51\sqrt{\frac{1}{10}+\frac{1}{10}}} = 1.191$$

$|t_{cal}| = 1.91$

$t_{tab} = t_{0.05}$ with degrees of freedom $n_1+n_2-2 = 18$

is $t_{tab} = 1.734$

$t_{cal} < t_{tab}$

$\therefore$ Accept null hypothesis $\mu_1 = \mu_2$

9) Two independent Samples of 8 & 7 items respectively had the following Values

Sample 1	11	11	13	11	18	9	12	14
Sample 2	9	11	10	13	9	8	10	

Is the difference between the mean of the sample is Significant

$n_1 = 8 \quad n_2 = 7$

$$\bar{x} = \frac{11+11+ --- +14}{8} = 12$$

$$\bar{y} = \frac{9+11+ -- +10}{7} = 10$$

x_i	$\bar{x}$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	y_i	$\bar{y}$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
11	12	-1	1	9	10	-1	1
11	12	-1	1	11	10	1	1
13	12	1	1	10	10	0	0
11	12	-1	1	13	10	3	9
15	12	3	9	9	10	-1	1
9	12	-3	9	8	10	-2	4
12	12	0	0	10	10	0	0
14	12	2	4				
			=26				=16

Null hypothesis $(H_0) : \mu_1 = \mu_2$

Alternate hypothesis $(H_1) : \mu_1 \neq \mu_2$

level of Significance $(\alpha) = 0.05$

$$S^2 = \frac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1+n_2-2} = 2.1$$

$S = 1.449$

test Statistic $(t_{cal}) = \dfrac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 1.84$

ems respe-

12	14
10	

the Sam-

$(y_i-\bar{y})^2$
1
1
0
9
1
4
0
$=16$

Distribution

To test the equality of population variances we use F-distribution. Let s_1^2 & s_2^2 are the variances of two samples of sizes n_1 & n_2 then population variances are given by $n_1 s_1^2 = (n_1-1)S_1^2$ and $n_2 s_2^2 = (n_2-1)S_2^2$. The quantities $v_1 = n_1-1$, $v_2 = n_2-1$ are called degrees of freedom of these estimates.

$$S_1^2 = \frac{n_1 s_1^2}{n_1-1} = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1}$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2-1} = \frac{\Sigma(y_i-\bar{y})^2}{n_2-1}$$

$$F = \frac{\text{Greater Variance}}{\text{Smaller Variance}} = \frac{S_1^2}{S_2^2} \quad \text{if } S_1 > S_2$$

Note: $F_\alpha(v_1, v_2)$ is the value of F with v_1 & v_2 degrees of freedom such that area under F distribution to right of F_α is α

1) In one sample of eight observations from a normal population of the sum of squares of deviation from sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level whether the populations have same variance

$n_1 = 8$

$\Sigma(x_i-\bar{x})^2 = 84.4$

$n_2 = 10$

$\Sigma(y_i-\bar{y})^2 = 102.6$

$\alpha = 0.05$

Null hypothesis : $\sigma_1^2 = \sigma_2^2$

Alternate hypothesis $(H_1) = \sigma_1^2 \neq \sigma_2^2$

level of significance $(\alpha) = 0.05$

$$\text{Test statistic} = \frac{\text{Greater Variance}}{\text{Smaller Variance}}$$

$$S_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1} = \frac{84.4}{8-1} = 12.057$$

$$S_2^2 = \frac{\Sigma(y_i-\bar{y})^2}{n_2-1} = \frac{102.6}{10-1} = 11.4$$

1) $F_{cal} = \frac{\text{Greater Variance}}{\text{Smaller Variance}} = \frac{S_1^2}{S_2^2} = \frac{12.057}{11.4} = 1.057$

$F_{tab} = F_\alpha = F_{(0.05)}(\upsilon_1, \upsilon_2)$

$= F_{0.05}(n_1-1, n_2-1)$

$= F_{0.05}(8-1, 10-1)$

$= F_{0.05}(7, 9)$

$F_{tab} = 3.29$

$|F_{cal}| < F_{tab}$

$\therefore$ Accept null hypothesis

$\therefore \sigma_1^2 = \sigma_2^2$

Hence, two variances are equal

2) In one sample of 10 observations from a normal population the sum of squares of deviations of sample values from sample mean is 102.4 in another of 12 observations from another normal population the sum of squares of deviation of sample values from sample mean is 120.5 . Examine whether both populations have sample variances

$n_1 = 10 \quad n_2 = 12$

$\Sigma (x_i - \bar{x})^2 = 102.4$

$\Sigma (y_i - \bar{y})^2 = 120.5$

$\alpha = 0.05$

Null hypothesis : $\sigma_1^2 = \sigma_2^2$

Alternate hypothesis: $\sigma_1^2 \neq \sigma_2^2$

level of significance $(\alpha) = 0.05$

Test statistic $(F) = \frac{\text{Greater Variance}}{\text{Smaller variance}}$

$S_1^2 = \frac{\Sigma (x_i - \bar{x})^2}{n_1 - 1} = \frac{102.4}{10-1} = 11.37$

$S_2^2 = \frac{\Sigma (y_i - \bar{y})^2}{n_2 - 1} = \frac{120.5}{12-1} = 10.95$

$F_{cal} = \frac{S_1^2}{S_2^2} = \frac{11.37}{10.95} = 1.038$

$F_{tab} = F_{\alpha} = F_{0.05}(\nu_1, \nu_2)$

$= F_{0.05}(n_1-1, n_2-1)$

$= F_{0.05}(9, 11)$

$F_{tab} = 2.90$

$|F_{cal}| < F_{tab}$

$\therefore$ Accept null hypothesis

$\therefore \sigma_1^2 = \sigma_2^2$, Hence two variances are equal

3) In a sample of 10 observation the sum of squares of deviations of sample values from sample mean was 120 and in another sample of 12 observations it was 314. Test whether the difference is significance at 5% level of significance

$n_1 = 10 \quad \Sigma(x_i - \bar{x})^2 = 120$

$n_2 = 12 \quad \Sigma(y_i - \bar{y})^2 = 314$

$\alpha = 0.05$

Null hypothesis $\sigma_1^2 = \sigma_2^2$

Alternate hypothesis $\sigma_1^2 \neq \sigma_2^2$

level of significance $(\alpha) = 0.05$

Test Statistic $F_{cal} = \dfrac{\text{Greater Variance}}{\text{Smaller Variance}}$

$$S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} = \frac{120}{10-1} = 13.33$$

$$S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1} = \frac{314}{12-1} = 28.54$$

$$F_{cal} = \frac{28.54}{13.33} = 2.14$$

$$F_{tab} = F_{0.05}(\nu_2, \nu_1) = F_{0.05}(n_2 - 1, n_1 - 1)$$

$$= F_{0.05}(12-1, 10-1)$$

$$= F_{0.05}(11, 9)$$

4) It is known that mean diameter of rivets produced by two firms A and B are practically the same but the standard deviation may differ. For 22 rivets produced by firm A the standard deviation is 2.9 mm while for 16 rivets manufactured by the firm the S.D is 3.8mm. Compute the statistic you would test whether the products of firm A has same variability as those of firm B and test its significance

$n_1 = 22 \qquad n_2 = 16$

$s_1 = 2.9 \qquad s_2 = 3.8$

Null hypothesis (H_0) : $\sigma_1^2 = \sigma_2^2$

Alternate hypothesis (H_1) : $\sigma_1^2 \neq \sigma_2^2$

Test Statistic $(F) = \dfrac{\text{Greater variance}}{\text{Smaller variance}}$

We know that $n_1 s_1^2 = (n_1 - 1) S^2$

$$S_1^2 = \frac{n_1 s_1^2}{(n_1-1)} = \frac{22\,(2.9)^2}{(22-1)}$$

$$S_1^2 = 8.810$$

We know that $n_2 s_2^2 = (n_2 - 1) S_2^2$

$$S_2^2 = \frac{n_2 s_2^2}{(n_2-1)} = \frac{16\,(3.8)^2}{16-1}$$

$$S_2^2 = 15.402$$

$$F_{cal} = \frac{\text{Greater variance}}{\text{Smaller Variance}}$$

$$F_{cal} = \frac{S_2^2}{S_1^2} = \frac{15.402}{8.810} = 1.74$$

$$F_{tab} = F_\alpha(\nu_2, \nu_1) = F_{0.05}(n_2-1, n_1-1)$$

$$= F_{0.05}(16-1, 22-1)$$

$$= F_{0.05}(15, 21)$$

$$= 2.18$$

$|F_{cal}| < F_{tab}$

$\therefore$ Accept null hypothesis

$\sigma_1^2 = \sigma_2^2$

5) Pumpkins were ... conditions. Two ... show the sampl... 9.5 respectively ... distribution are ... true Variances a...

$n_1 = 11 \qquad s_1 = 0$

$n_2 = 9 \qquad s_2 = 0$

Null hypothesis

Alternate hypo-

level of signif

Test Statistic

we know that

$S_1^2 =$

$n_2 s_2$

S_2^2

$F_{cal} = 2.5$

$F_{tab} = F_\alpha ($

$|F_{cal}| < F$

$\therefore$ Accept

6) An instruc... ular subject ... Class B ha... tion althou... in the me... Can we con... that the ... that of A

$n_1 = 16 \qquad n_2$

Null hypo-

Altern...

5) Pumpkins were grown under 2 experimental conditions. Two random samples at 11 & 9 pumpkin show the sample S.D of their weights as 0.8 & 0.5 respectively. Assuming that the weight distribution are normal. Test hypothesis that the true variances are equal

$n_1 = 11 \quad s_1 = 0.8$

$n_2 = 9 \quad s_2 = 0.5$

Null hypothesis (H_0) : $\sigma_1^2 = \sigma_2^2$

Alternate hypothesis (H_1) : $\sigma_1^2 \neq \sigma_2^2$

level of significance $\alpha = 0.05$

Test statistic (F) : $\dfrac{\text{Greater Variance}}{\text{Smaller Variance}}$

we know that $n_1 s_1^2 = (n_1 - 1) S_1^2$

$$S_1^2 = \frac{n_1 s_1^2}{(n_1-1)} = \frac{11 \times (0.8)^2}{10} = 0.704$$

$$n_2 s_2^2 = (n_2 - 1) S_2^2$$

$$S_2^2 = \frac{n_2 s_2^2}{(n_2-1)} = \frac{9 \times (0.5)^2}{8} = 0.28125$$

$F_{cal} = 2.5$

$F_{tab} = F_\alpha(v_1, v_2) = F_{0.05}(n_1 - 1, n_2 - 1)$

$= F_{0.05}(10, 8)$

$F_{tab} = 3.25$

$|F_{cal}| < F_{tab}$

$\therefore$ Accept null hypothesis

6) An instructor has 2 classes A & B in a particular subject. Class A has 16 students while class B has 25 students. On the same examination although there was no significant difference in the mean grades class A has a S.D of 12. Can we conclude at 0.01 level of significance that the variability of class B is greater than that of A

$n_1 = 16 \quad n_2 = 25 \quad s_1 = 9 \quad s_2 = 12 \quad \alpha = 0.01$

Null hypothesis (H_0) : $\sigma_1^2 = \sigma_2^2$

Alternate hypothesis (H_1) : $\sigma_1^2 < \sigma_2^2$

level of significance $(\alpha) = 0.01$

$$\text{Test Statistic } (F) = \frac{\text{Greater Variance}}{\text{Smaller Variance}}$$

We know that $n_1 s_1^2 = (n_1 - 1) S_1^2$

$$S_1^2 = \frac{n_1 s_1^2}{(n_1-1)} = \frac{16(9)^2}{15}$$

$$S_1^2 = 86.4$$

$$n_2 s_2^2 = (n_2 - 1) S_2^2$$

$$S_2^2 = \frac{n_2 s_2^2}{(n_2-1)} = \frac{25(12)^2}{24}$$

$$S_2^2 = 150$$

$$F_{cal} = \frac{G.V}{S.V} = \frac{150}{86.4} = 1.736$$

$$F_{tab} = F_\alpha(\upsilon_2, \upsilon_1) = F_{0.05}(n_2-1, n_1-1)$$

$$= F_{0.05}(24, 15)$$

$$= 2.29$$

$|F_{cal}| < F_{tab}$

$\therefore$ Accept null hypothesis

$\therefore \sigma_1^2 = \sigma_2^2$

7) Two independent samples of 8 & 7 items respectively had the following values of variables

Sample I	9	11	13	11	16	10	12	14
Sample II	11	13	11	14	10	8	10	

Do the estimates of population variances differ significantly

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{9+11+ - - +14}{8} = 12$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{11+13+ - - +10}{7} = 11$$

x_i	$\bar{x}$	$x_i - \bar{x}$	$(x_i-\bar{x})^2$	y_i	$\bar{y}$	$y_i - \bar{y}$	$(y_i-\bar{y})^2$
9	12	-3	9	11	11	0	0
11	12	-1	1	13	11	2	4
13	12	1	1	11	11	0	0
11	12	-1	1	14	11	3	9
16	12	4	16	10	11	-1	1
10	12	-2	4	8	11	-3	9

		0	0	10	11	-1	1
12	12	2	4				
14	12		36				28

$$S_1^2 = \frac{\Sigma (x_i - \bar{x})^2}{n_1 - 1} = \frac{36}{8-1} = \frac{36}{7} = 5.14$$

$$S_2^2 = \frac{\Sigma (y_i - \bar{y})^2}{n_2 - 1} = \frac{24}{7-1} = \frac{24}{6} = 4$$

F_{cal} = Greater variance

Null hypothesis $(H_0) : \sigma_1^2 = \sigma_2^2$

Alternate hypothesis $(H_1) : \sigma_1^2 \neq \sigma_2^2$

level of significance $(\alpha) = 0.05$

Test Statistic $(F_{cal}) = \frac{\text{Greater Variance}}{\text{Smaller Variance}} = \frac{5.14}{4}$

$= 1.285$

$F_{tab} = F_{0.05}(v_1, v_2) = F_{0.05}(7,6) = 4.21$

$|F_{cal}| < F_{tab}$

$\therefore$ Accept null hypothesis

$\therefore \sigma_1^2 = \sigma_2^2$

8) The measurement of 10 outputs of two units have given the following results. Assuming that both samples have been obtained from the normal population at 1% significant level. Test whether two populations have same variance

Unit A	14.1	10.1	14.7	13.7	14
Unit B	14	14.5	13.7	12.7	14.1

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{(14.1 + 10.1 + - - + 14)}{5} = 13.32$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{14 + 14.5 + - - + 14.1}{5} = 13.8$$

x_i	$\bar{x}$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	y_i	$\bar{y}$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
14.1	13.32	0.78	0.6084	14	13.8	0.2	0.04
10.1	13.32	-3.22	10.3634	14.5	13.8	0.7	0.49
14.7	13.32	1.38	1.9044	13.7	13.8	-0.1	0.01
13.7	13.32	0.38	0.1444	12.7	13.8	-1.1	1.21
14	13.32	0.68	0.4624	14.1	13.8	0.3	0.09
			13.342				1.84

$$S_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1} = \frac{13.342}{4} = 3.372$$

$$S_2^2 = \frac{\Sigma(y_i-\bar{y})^2}{n_2-1} = \frac{1.84}{4} = 0.46$$

Null hypothesis : $\sigma_1^2 = \sigma_2^2$

Alternate hypothesis ; $\sigma_1^2 \neq \sigma_2^2$

$$F_{cal} = \frac{\text{Greater Variance}}{\text{Smaller Variance}} = \frac{3.372}{0.146} = 7.33$$

$$F_{tab} = F_{0.01}(\nu_1, \nu_2) = F_{0.01}(n_1-1, n_2-1)$$
$$= F_{0.01}(5-1, 5-1)$$
$$= F_{0.01}(4,4)$$
$$= 6.39$$

9) The time taken by workers in performing a big method 1 & method 2 are given below

method 1	20	16	26	27	23	22	
method 2	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differs significantly

$n_1 = 6 \quad n_2 = 7$

$\bar{x} = 22.33$

$\bar{y} = 34.42$

$(y_i-\bar{y})^2$

0.04
0.49
0.01
1.21
0.09

1.84

33

rming a
below

22	
34	38

s of time
ch these
ignificantly

x_i	$\bar{x}$	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	y_i	$\bar{y}$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
20	22.33	-2.33	5.4289	27	34.42	-7.42	55.05
16	22.33	-6.33	40.0689	33	34.42	-1.42	2.0164
26	22.33	3.67	13.4689	42	34.42	7.58	57.4564
27	22.33	4.67	21.808	35	34.42	0.58	0.3364
23	22.33	0.67	0.4489	32	34.42	-2.42	5.8564
22	22.33	-0.33	0.1089	34	34.42	-0.42	0.1764
				38	34.42	3.58	12.8164
			81.3344				133.7148

$$S_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1} = \frac{81.3344}{5} = 16.26668$$

$$S_2^2 = \frac{\Sigma(y_i-\bar{y})^2}{n_2-1} = \frac{133.7148}{6} = 22.2855$$

$$F_{cal} = \frac{\text{Greater Variance}}{\text{Smaller Variance}} = \frac{22.2858}{16.26668} = 1.3700$$

$$F_{tab} = F_\alpha(\upsilon_2,\upsilon_1) = (F_{0.05})(7-1, 6-1)$$
$$= F_{0.05}(6,5)$$
$$= 4.95$$

$|F_{cal}| < F_{tab}$

$\therefore$ Accept null hypothesis

$\therefore \sigma_1^2 = \sigma_2^2$

10) The nicotin content in milligrams in two Samples of tobacco were found to be as follows

Sample A	24	27	26	21	25	
Sample B	27	30	28	31	22	36

Can it be Said that two Samples have come from same normal population. To test whether the two Sample have came from same normal population have to test (i) equality of variances by using F test
(ii) equality of means by using t Test

$n_1 = 5 \quad n_2 = 6 \quad \alpha = 0.05$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{24 + - - + 25}{5} = 24.6$$

$\bar{y} = \frac{\sum y_i}{n} = \frac{27 + - - + 36}{6} = 29$

9)

x_i	$\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	y_i	$\bar{y}$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
24	24.6	-0.6	0.36	27	29	-2	4
27	24.6	2.4	5.76	30	29	1	1
26	24.6	1.4	1.96	28	29	-1	1
21	24.6	-3.6	12.96	31	29	2	4
25	24.6	0.4	0.16	22	29	-7	49
				36	29	7	49
			21.2				108

F test

$S_1^2 = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1} = \frac{21.2}{4} = 5.3$

$S_2^2 = \frac{\sum (y_i - \bar{y})^2}{n_2 - 1} = \frac{108}{5} = 21.6$

$F_{cal} = \frac{\text{Greater variance}}{\text{smaller variance}} = \frac{21.6}{5.3} = 4.07$

$F_{tab} = F_\alpha (v_2, v_1) = F_{0.05} (n_2 - 1, n_1 - 1)$

$= F_{0.05} (5, 4) = 6.26$

$|F_{cal}| < F_{tab}$

$\therefore$ Accept null hypothesis

$\therefore \sigma_1^2 = \sigma_2^2$

$\therefore$ Two populations have same Variance

T test

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

$S^2 = \frac{21.2 + 108}{5 + 6 - 2} = 14.35$

$S = 3.78$

Null hypothesis $(H_0) : \mu_1 = \mu_2$

Alternate hypothesis: $\mu_1 \neq \mu_2$

level of Significance $\alpha = 0.05$

Test Statistic $(t_{cal}) = \frac{\bar{x} - \bar{y}}{S \sqrt{\frac{1}{n} + \frac{1}{n}}}$

$t_{cal} = \frac{24.}{3.7}$

$t_{tab} = t_{\frac{\alpha}{2}}$

freedom n_1

$|t_{cal}|$

$\therefore$ Accept

$\therefore \mu_1 = \mu_2$

$\therefore$ Two sa

population

$\therefore$ Two sam

population

11) Two rando

results

Sample
1
2

Test whe

normal po

$n_1 = 10$

$\bar{x} = 15$

$\sum (x_i - \bar{x})$

F test

Null hyp

Alternate

$\alpha = 0.0$

$S_1^2 =$

$S_2^2 = \sum$

$F_{cal} =$

$F_{tab} =$

$$t_{cal} = \frac{24.6 - 29}{3.78\sqrt{\frac{1}{5}+\frac{1}{6}}} = -1.92$$

$t_{tab} = t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.25}$ with degrees of freedom $n_1+n_2-2 = 9$ is 2.26

$|t_{cal}| < t_{tab}$

$\therefore$ Accept null hypothesis

$\therefore \mu_1 = \mu_2$

$\therefore$ Two samples have same population & same population variances

$\therefore$ Two samples are drawn from same normal population

11) Two random samples reveal the following results

Sample	Size	Sample mean	Sum of squares of deviation from mean
1	10	15	90
2	12	14	108

Test whether the samples come from same normal population.

$n_1 = 10 \quad n_2 = 12$

$\bar{x} = 15 \quad \bar{y} = 14$

$\Sigma(x_i-\bar{x})^2 = 90 \quad \Sigma(y_i-\bar{y})^2 = 108$

F test

Null hypothesis : $H_0 \; \sigma_1^2 = \sigma_2^2$

Alternate hypothesis : $H_1 = \sigma_1^2 \neq \sigma_2^2$

$\alpha = 0.05$

$$S_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1} = \frac{90}{9} = 10$$

$$S_2^2 = \frac{\Sigma(y_i-\bar{y})^2}{n_2-1} = \frac{108}{11} = 9.81$$

$$F_{cal} = \frac{\text{Greater Variance}}{\text{Smaller Variance}} = \frac{10}{9.81} = 1.01$$

$$F_{tab} = F_{0.05}(v_2, v_2) = F_{0.05}(n_1-1, n_2-1)$$

$$= F_{0.05}(9, 11)$$

$$= 2.90$$

$|F_{cal}| < F_{tab}$

$\therefore$ Accept null hypothesis

$\sigma_1^2 = \sigma_2^2$

T test

$$S^2 = \frac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1+n_2-2} = \frac{90+108}{10+12-2} = 9.9$$

$S = 3.14$

Null hypothesis (H_0) : $\mu_1 = \mu_2$

Alternate hypothesis (H_1) : $\mu_1 \neq \mu_2$

level of significance $(\alpha) = 0.05$

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

$$= \frac{15-14}{3.14\sqrt{\frac{1}{10}+\frac{1}{12}}} = 0.743$$

$t_{tab} = t_{\frac{\alpha}{2}} = t_{0.025}$ with degrees of freedom $n_1+n_2-2 = 10+12-2 = 20$ is 2.086

$|t_{cal}| < t_{tab}$

$\therefore$ Accept null hypothesis

$\therefore \mu_1 = \mu_2$

$\therefore$ Two samples have same population mean & same population variances

$\therefore$ Two samples are drawn from same normal population.

2/12

Chisquare Test

If a set of events $A_1, A_2 --- A_n$ are observed to occur with frequencies $O_1, O_2 --- O_n$ respectively and according to the probability rules $A_1, A_2 --- A_n$ are expected to occur with frequencies $E_1, E_2 --- E_n$ then $\chi^2 = \Sigma\left(\frac{(O_i-E_i)^2}{E_i}\right)$ with degrees of freedom $(r-1)(c-1)$ where r is the number of rows & c is the number of columns.

Conditions for validity of χ^2 test

The following are the conditions which should be satisfied before χ^2 test can be applied

* The sample observation should be independent
* N the total frequency is large i.e $N > 50$
* The constraints on the cell frequencies if any are linear
* No theoretical (or) expected frequency should be less than linear
* It is only dependent on degrees of freedom

χ^2 test for independence of attributes

* An attribute means a quality (or) characteristic

Ex: Drinking, Smoking, beauty, honesty etc.

The 2×2 contigency table is given by

a	b
c	d

* The observed frequencies are a, b, c, d

a	b	a+b
c	d	c+d
a+c	b+d	a+b+c+d

→ Grand total

* Expected frequency of each cell is given by

$$\frac{\text{row total} \times \text{Column total}}{\text{Grand total}}$$

$$\therefore E(a) = \frac{(a+b)(a+c)}{(a+b+c+d)}$$

$$E(b) = \frac{(a+b)(b+d)}{(a+b+c+d)}$$

$$E(c) = \frac{(c+d)(a+c)}{(a+b+c+d)}$$

$$E(d) = \frac{(c+d)(b+d)}{(a+b+c+d)}$$

1. On the basis of information given below about the treatment 200 patients suffering from a disease. State whether the new treatment is comparatively superior to the conventional treatment

	Favourable	Not Favourable	Total
New	60	30	90
Conventional	40	70	110
Total	100	100	200

Expected frequency

$E_{11}(60) = \frac{90 \times 100}{200} = 45$ $\quad E_{21}(40) = \frac{110 \times 100}{200} = 55$

$E_{12}(30) = \frac{40 \times 100}{200} = 45$ $\quad E_{22}(70) = \frac{110 \times 100}{200} = 55$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
60	45	15	225	5
30	45	-15	225	5
40	55	-15	225	4·09
70	55	15	225	4·09

$$\chi^2_{cal} = \Sigma\left(\frac{(O_i - E_i)^2}{E_i}\right) = 18{\cdot}18$$

Null hypothesis : The new & conventional frequencies are same

Alternate hypothesis : new is superior to conventional treatment

level of significance $\alpha = 0{\cdot}05$

Test Statistic $(\chi^2_{cal}) = \Sigma\left(\frac{(O_i - E_i)^2}{E_i}\right) = 18{\cdot}18$

$\chi^2_{tab} = \chi^2_{0{\cdot}05}$ with degrees of freedom $(r-1)(c-1) = (2-1)(2-1) = 1$ is $3{\cdot}841$

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

$\therefore$ Reject null hypothesis

$\therefore$ Accept alternate hypothesis i.e new treatment is superior to the conventional treatment

2) The following table gives the classification of 100 workers according to the sex and nature of work. Test whether the nature of work is independent of the sex of the worker

	Stable	Unstable	Total
males	40	20	60
Females	10	30	40
Total	50	50	100

Exp

$E_{11}(40) = \frac{60 \times 50}{100} = 30$ $\quad E_{21}(10) = \frac{40 \times 50}{100} = 20$

$E_{12}(20) = \frac{60 \times 50}{100} = 30$ $\quad E_{22}(30) = \frac{40 \times 50}{100} = 20$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
40	30	10	100	3.33
20	30	-10	100	3.33
10	20	-10	100	5
30	20	10	100	5

$$\chi^2_{cal} = \sum \left(\frac{(O_i - E_i)^2}{E_i} \right) = 16.66$$

Null hypothesis: Nature of work is independent of the sex of the worker

Alternate hypothesis: Nature of work is dependent of sex of the workers

level of Significance $(\alpha) = 0.05$

Test statistic $(\chi^2_{cal}) = \sum \left(\frac{(O_i - E_i)^2}{E_i} \right) = 16.66$

$\chi^2_{tab} = \chi^2_{0.05}$ with degrees of freedom $(r-1)(c-1)$ $= (2-1)(2-1) = 1$ is 3.841

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

$\therefore$ Reject null hypothesis

$\therefore$ Accept alternate hypothesis

$\therefore$ Nature of work is dependent on sex of the workers

3) Given the following contigency table for hair color & eye colour. Find the value of χ^2 is there any good association b/w the two

Eye color	Hair Color: Fair	Brown	Black	Total
Blue	15	5	20	40
Grey	20	10	20	50
Brown	25	15	20	60
Total	60	30	60	150

$E_{11}(15) = \frac{40 \times 60}{150} = \frac{240}{15} = 16$ $\quad$ $E_{23}(20) = \frac{50 \times 60}{150} = \frac{300}{15} = 20$

$E_{12}(5) = \frac{40 \times 30}{150} = \frac{120}{15} = 8$ $\quad$ $E_{31}(25) = \frac{60 \times 60}{150} = \frac{360}{15} = 24$

$E_{13}(20) = \frac{40 \times 60}{150} = \frac{240}{15} = 16$ $\quad$ $E_{32}(15) = \frac{60 \times 30}{150} = 12$

$E_{21}(20) = \frac{50 \times 60}{150} = \frac{300}{15} = 20$ $\quad$ $E_{33}(20) = \frac{60 \times 60}{150} = 24$

$E_{22}(10) = \frac{50 \times 30}{150} = 10 = 10$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
15	16	-1	1	0.0625
5	8	-3	9	1.125
20	16	4	16	1
20	20	0	0	0
10	10	0	0	0
20	20	0	0	0
25	24	1	1	0.0416
15	12	3	9	0.75
20	24	-4	16	0.666

$$\chi^2_{cal} = \sum \left(\frac{(O_i - E_i)^2}{E_i} \right) = 3.6446$$

Null hypothesis : Hair and Eye colour are independent

Alternate hypothesis : Hair & eye colour are not independent

level of significance $(\alpha) = 0.05$

Test statistic $(\chi^2_{cal}) = \sum \left(\frac{(O_i - E_i)^2}{E_i} \right) = 3.64$

$\chi^2_{tab} = \chi^2_{0.05}$ with degrees of freedom $(r-1)(c-1) = (3-1)(3-1) = 4$ is 9.488

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

$\therefore$ Accept null hypothesis

i.e Hair & eye colour are independent

4) Four methods are under development for making disks of a Super conducting material. 50 disks are made by each method and they are checked for Super conductivity when cooled with liquid

	1st method	2nd method	3rd method	4th method	T
Super Conductivity	31	42	22	25	120
Failures	19	8	28	25	80
Total	50	50	50	50	200

$E_{11}(31) = \frac{120 \times 50}{200} = 30$ $\quad E_{21}(19) = \frac{80 \times 50}{200} = 20$

$E_{12}(42) = \frac{120 \times 50}{200} = 30$ $\quad E_{22}(8) = \frac{80 \times 50}{200} = 20$

$E_{13}(22) = \frac{120 \times 50}{200} = 30$ $\quad E_{23}(28) = \frac{80 \times 50}{200} = 20$

$E_{14}(25) = \frac{120 \times 50}{200} = 30$ $\quad E_{24}(25) = \frac{80 \times 50}{200} = 20$

O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
31	30	1	1	0.033
42	30	12	144	4.8
22	30	-8	64	2.133
25	30	-5	25	0.833
19	20	-1	1	0.05
8	20	-12	144	7.2
28	20	8	64	3.2
25	20	5	25	1.25

$$\chi^2_{cal} = \sum \frac{(O_i - E_i)^2}{E_i} = 18.249$$

5/12

Corelation & Regression

Rank Corelation Coefficient or Spearman's rank Corelation Coefficient

Spearman's rank Corelation Coefficient

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$$ where D the difference of two ranks, n is no. of paird observations

→ when the ranks are repeated then

$$\rho = 1 - \frac{6\left(\sum\left[D^2 + \frac{m(m^2-1)}{12} + \cdots\right)\right.}{N(N^2-1)}$$

where m = the no. of times that number is repeated

1) A random sample of 5 college students is selected and their grades in mathematics & Statistics are found to be

	1	2	3	4	5
mathematics	85	60	73	40	90
Statistics	93	75	65	50	80

Calculate Spearman's rank Corelation Coefficient

mathematics x	rank of x r_1	Statistics y	rank of y (r_2)	D $r_1 - r_2$	D^2
85	2	93	1	1	1
60	4	75	3	1	1
73	3	65	4	-1	1
40	5	50	5	0	0
90	1	80	2	1	1
					$\sum D^2 = 4$

Spearman's rank Correlation is

$$\rho = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

$$= 1 - \left[\frac{6 \times 4}{5(5^2-1)}\right]$$

$$= 1 - \frac{24}{24 \times 5} = \frac{4}{5} = 0.8$$

2) Following ... ents in ... To what ext... in 2 Subj...

Statistics	
mathematic...	
rank of x r_1	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

3) The ra... Statis... (1,1) (8,6) (15,16... Coeffi... in m...

2) Following are the ranks obtained by 10 students in 2 subjects mathematics & statistics. To what extent the knowledge of students in 2 subjects in related

statistics	1	2	3	4	5	6	7	8	9	10
mathematics	2	4	1	5	3	9	7	10	6	8

rank of x r_1	rank of y r_2	D $r_1 - r_2$	D^2
1	2	-1	1
2	4	-2	4
3	1	2	4
4	5	-1	1
5	3	2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	3	9
10	8	2	4
			$\Sigma D^2 = 40$

$$\rho = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 40}{10(10^2-1)} = 1 - \frac{6 \times 40}{10(100-1)}$$

$$= 0.76$$

3) The ranks of 16 students in mathematics & statistics are given below
(1,1) (2,10) (3,3) (4,4) (5,5) (6,7) (7,2) (8,6) (9,8) (10,11) (11,15) (12,9) (13,14) (14,12) (15,16) (16,13) Calculate rank corelation coefficient for proficiencies of this group in mathematics & statistics

rank of mathematics r_1	rank of statistics r_2	$D = r_1 - r_2$	D^2
1	1	0	0
2	10	-8	64
3	3	0	0
4	4	0	0
5	5	0	0
6	7	-1	1
7	2	5	25
8	6	2	4
9	8	1	1
10	11	-1	1
11	15	-4	16
12	9	3	9
13	14	1	1
14	12	2	4
15	16	-1	1
16	13	3	9
			$\Sigma D^2 = 136$

$$\rho = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6(136)}{16(16^2-1)}$$

$$= 0.8$$

4) For the following data calculate the corelation Coefficient

X	48	33	40	9	16	16	65	24	16	57
Y	13	13	24	6	15	4	20	9	6	19

X	rank of X r_1	Y	rank of Y r_2	$D = r_1 - r_2$	D^2
48	3	13	5.5	-2.5	6.25
33	5	13	5.5	-0.5	0.25
40	4	24	1	3	9
9	10	6	8.5	1.5	2.25

	8	15	4	4	16
16	8	4	10	-2	4
16	1	20	2	-1	1
65	6	9	7	-1	1
24	8	6	8.5	0.5	0.25
16	2	19	3	-1	1
57					$\Sigma D^2 = 41$

Correction factor in x Series:
In x Series 16 is repeated 3 times
$m = 3$
$\therefore$ C.F in x Series is $= \frac{1}{12} m(m^2-1) = \frac{1}{12} \times 3(3^2-1) = 2$

Correction factor in Y Series:
In Y-Series 13 is repeated 2 times, 6 is repeated 2 times
C.F in Y Series is $\frac{1}{12} 2(2^2-1) + \frac{1}{12} 2(2^2-1)$
$= \frac{1}{2} + \frac{1}{2} = 1$

rank Correlation coefficient is

$$\rho = 1 - \frac{6\,[\Sigma D^2 + \text{C.F in x Series} + \text{C.F in Y Series}]}{N(N^2-1)}$$

$$= 1 - \frac{6\,[41+2+1]}{10(10^2-1)}$$

$$= 0.733$$

5) Obtain the rank Corelation Coefficient for the following data

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

X	rank of $x r_1$	Y	rank of Y r_2	$D = r_1 - r_2$	D^2
68	4	62	5	-1	1
64	6	58	7	-1	1
75	2.5	68	3.5	1	1
50	9	45	10	-1	1
64	6	81	1	5	25
80	1	60	6	-5	25

	68 2.5	68	3.5	-1	1
75	48 10	48	9	1	1
40	8	50	8	0	0
55	6	70	2	4	16
64					

$\Sigma D^2 = 72$

Correction factor in x Series

In x Series 75 is repeated 2 times and 64 is repeated 3 times

C.F in x Series is $= \frac{1}{12} 2(2^2-1) + \frac{1}{2} 3(3^2-1)$

$= 2.5$

Correction factor in Y Series

In Y Series 68 is repeated 2 times

C.F in Y Series $= \frac{1}{12} 2(2^2-1)$

$= 0.5$

rank Corelation Coefficient

$$\rho = 1 - \frac{6\,[\Sigma D^2 + C.F \text{ in x Series} + C.F \text{ in Y Series}]}{N(N^2-1)}$$

$$= 1 - \frac{6\,[72 + 2.5 + 0.5]}{10(10^2-1)}$$

$= 0.545$

The following table gives the Score obtained by 11 Students in English & Telugu translation Find the rank Corelation Coefficient

Scores in English	40	56	54	60	70	80	82	85	85	90	95
Scroes in Telugu	45	45	50	43	40	75	55	72	65	42	70

x	rank of x r_1	y	rank of y r_2	$D = r_1 - r_2$	D^2
40	11	45	7.5	3.5	12.25
56	9	45	7.5	1.5	2.25
54	10	50	6	4	16
60	8	43	9	-1	1
70	7	40	11	4	16

	6	75	1	5	25
80	5	55	5	0	0
82	3.5	72	2	1.5	2.25
85	3.5	65	4	-0.5	0.25
85	2	42	10	-8	64
90	1	70	3	-2	4
95					

$\Sigma D^2 = 143$

Correction factor in x Series

In x Series 85 is repeated 2 times

C.F in x Series $= \frac{1}{12} 2(2^2-1)$

$= 0.5$

Correction factor in Y Series

In Y Series 45 is repeated 2 times

C.F in Y Series $= \frac{1}{12} 2(2^2-1)$

$= 0.5$

rank corelation Coefficient

$$\rho = 1 - \frac{6\,[\Sigma D^2 + C.F \text{ in } x \text{ Series} + C.F \text{ in } Y \text{ Series}]}{N(N^2-1)}$$

$$= 1 - \frac{6\,[143+0.5+0.5]}{11(11^2-1)}$$

$= 0.345$

6/12 Karl Pearson's Corelation Coefficient

Karl Pearson's Corelation Coefficient

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

where $X = x - \bar{x}$, $Y = y - \bar{y}$

1) Calculate Coefficient of Corelation for the following data

X	12	9	8	10	11	13	7
Y	14	8	6	9	11	12	3

$\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{12+9+\cdots+7}{7} = 10$

$\bar{y} = \frac{\Sigma y_i}{n_2} = 14+8+\cdots+3/7 = 9$

x	X $x-\bar{x}=x-10$	y	Y $y-\bar{y}=y-9$	XY	X^2	Y^2
12	2	14	5	10	4	25
9	-1	8	-1	1	1	1
8	-2	6	-3	6	4	9
10	0	9	0	0	0	0
11	1	11	2	2	1	4
13	3	12	3	9	9	9
7	-3	13	-6	18	9	36
				$\Sigma XY=46$	$\Sigma X^2=28$	$\Sigma Y^2=84$

Corelation Coeffient

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} = \frac{46}{\sqrt{28\times 84}} = 0.94$$

2) Find if there is any significant Corelation b/w heights & weights given

Heights in inches	57	59	62	63	64	65	55	58	57
Weight in lbs	113	117	126	126	130	129	111	116	112

$$\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{57+59+ - - - +57}{9} = 60$$

$$\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{113+117+ - - - +112}{9} = 120$$

x	X $x-\bar{x}=x-60$	y	Y $y-\bar{y}=y-120$	XY	X^2	Y^2
57	-3	113	-7	21	9	49
59	-1	117	-3	3	1	9
62	2	126	6	12	4	36
63	3	126	6	18	9	36
64	4	130	10	40	16	100
65	5	129	9	45	25	81
55	-5	111	-9	45	25	81
58	-2	116	-4	8	4	16
57	-3	112	-8	24	9	64
				216	102	472

Corelation Coeff

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} =$$

3) A psycological ... ing ability were ... is a record of ... gence ratio IR & ... Coefficient rel...

Student	A
IR	105
ER	101

$\bar{x} = \frac{\Sigma x_i}{n_1} = 10$...

$\bar{y} = \frac{\Sigma y_i}{n_2} = 10$...

x	X $x-99$
105	6
104	5
102	3
101	2
100	1
99	0
98	-1
96	-3
93	-6
92	-7

Corelation

$r = \Sigma$...

y^2
25
1
9
0
4
9
36
$\Sigma y^2 = 84$

Corelation

55 55 58 57
29 111 116 112

y^2
49
9
36
36
100
81
81
16
64
472

Corelation coefficient

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{216}{\sqrt{102 \times 472}} = 0.98$$

3) A psycological test of intelligence and engineering ability were applied to 10 students. Here is a record of ungrouped data showing intelligence ratio IR & engineering ratio (ER) Calculate coefficient relation

Student	A	B	C	D	E	F	G	H	I	J
IR	105	104	102	101	100	99	98	96	93	92
ER	101	103	100	98	95	96	104	92	97	94

$$\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{105 + --- + 92}{10} = 99$$

$$\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{101 + 103 + --- + 94}{10} = 98$$

x	X $x - 99$	y	Y $x - 98$	XY	X^2	Y^2
105	6	101	3	18	36	9
104	5	103	5	25	25	25
102	3	100	2	6	9	4
101	2	98	0	0	4	0
100	1	95	-3	-3	1	9
99	0	96	-2	0	0	4
98	-1	104	6	-6	1	36
96	-3	92	-6	18	9	36
93	-6	97	-1	6	36	1
92	-7	94	-4	28	49	16
				92	170	140

Corelation coefficient

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} = \frac{92}{\sqrt{170 \times 140}} = 0.59$$

4) Calculate Corelation Coefficient

Wages	100	101	102	102	100	99	97	98	96	95
Cost of living	98	99	99	97	95	92	95	94	90	91

$\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{100+101+ - - +95}{10} = 99$

$\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{98+99+ - - - +91}{10} = 95$

x	X $x-99$	y	Y $y-95$	XY	X^2	Y^2
100	1	98	3	3	1	9
101	2	99	4	8	4	16
102	3	99	4	12	9	16
102	3	97	2	6	9	4
100	1	95	0	0	1	0
99	0	92	-3	0	0	9
97	-2	95	0	0	4	0
98	-1	94	-1	1	1	1
96	-3	90	-5	15	9	25
95	-4	91	-4	16	16	16
				61	54	96

Corelation Coefficient

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} = \frac{61}{\sqrt{54 \times 96}} = 0.84$$

5) Calculate the Karl Pear's Corelation Coefficient

X	28	41	40	38	35	33	40	32	36	33
Y	23	34	33	34	30	26	28	31	36	38

$\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{28+41+ - - - - +33}{10} = 35.6$

$\bar{y} = \frac{\Sigma y_i}{n_2} = 31.3$

Let the assumed means be $\bar{x} = 35$, let $\bar{y} = 31$

x	X $x-35$
28	-7
41	6
40	5
38	3
35	0
33	-2
40	5
32	-3
36	1
33	-2

Corelation

$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 ...}}$

6) Find the

Height of

Height of

$\bar{x} = \Sigma x ...$

$\bar{y} = \Sigma y ...$

Let the

x	X $x-$
65	-3
66	-2
67	-1
67	-1
68	0
69	1
71	3
73	5

97 98 96 95
5 94 90 91

y^2
9
16
16
4
0
9
0
1
25
16
96

efficient

33
38

$\bar{y} = 31$

x	X $x-35$	y	Y $y-31$	XY	X^2	Y^2
28	-7	23	-8	56	49	64
41	6	34	3	18	36	9
40	5	33	2	10	25	4
38	3	34	3	9	9	9
35	0	30	-1	0	0	1
33	-2	26	-5	10	4	25
40	5	28	-3	-15	25	9
32	-3	31	0	0	9	0
36	1	36	5	5	1	25
33	-2	38	7	-14	4	49
				709	162	195

Corelation coefficient

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} = \frac{79}{\sqrt{162 \times 195}} = 0.448$$

6) Find the Suitable Corelation coefficient

Height of father 65 66 67 67 68 69 71 73
Height of son 67 68 64 68 72 70 69 70

$$\bar{x} = \frac{\Sigma x_i}{n} = 68.25$$

$$\bar{y} = \frac{\Sigma y_i}{n} = 68.5$$

Let the assumed mean $\bar{x} = 68$ & $\bar{y} = 68$

x	X $x-68$	y	Y $y-68$	XY	X^2	Y^2
65	-3	67	-1	3	9	1
66	-2	68	0	0	4	0
67	-1	64	-4	4	1	16
67	-1	68	0	0	1	0
68	0	72	4	0	0	16
69	1	70	2	2	1	4
71	3	69	1	3	9	1
73	5	70	2	10	25	4
				22	50	42

Corelation coefficient

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} = \frac{22}{\sqrt{50 \times 42}} = 0.48$$

7) Calculate the Corelation Coefficient for the following data

Fertilizers used	15	18	20	24	30	35	40	50
productivity	85	93	95	105	120	130	150	160

$$\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{15+18+ - - - - - +50}{8} = 29$$

$$\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{85+93+ - - - - + 160}{8} = 117.25$$

Assumed mean $\bar{x} = 29$ $\bar{y} = 117$

x	X $x-29$	y	Y $y-117$	XY	X^2	Y^2
15	-14	85	-32	448	196	1024
18	-11	93	-24	264	121	576
20	-9	95	-22	198	81	484
24	-5	105	-12	60	25	144
30	1	120	3	3	1	9
35	6	130	13	78	36	169
40	11	150	33	363	121	1089
50	21	160	44	924	441	1936
				2338	1022	5431

Corelation Coefficient

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

$$= \frac{2338}{\sqrt{1022 \times 5431}}$$

$$= 0.99$$

Regression:

The study of Corelation measures the direction & Strength of relationship b/w 2 variables

→ In Regretion we can estimate the value of one variable with value of another variable which is known

→ The statistical method which helps to estimate the unknown value of 1 variable from known value of related variable is called regretion.

→ The line described in the avg relationship b/w two variables is known as line of regression

→ Lines of Regression of x on y is

$x-\bar{x} = b_{xy}(y-\bar{y})$ where $b_{xy} = \frac{\Sigma xy}{\Sigma y^2}$

Similarly regression eq of y on x is

$y-\bar{y} = b_{yx}(x-\bar{x})$ where $b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$

→ Always the lines of regression pass through mean

1) Price indices of Cotton & wool are given below for 12 months of an year. Obtain the equations of lines of regressions b/w them.

cotton	78	77	85	88	87	82	81	77	76	83	97	93
wool	84	82	82	85	89	90	88	92	83	89	98	99

$\bar{x} = \frac{\Sigma x_i}{n} = 83.66$ Assumed mean $\bar{x} = 83$, $\bar{y} = 88$

$\bar{y} = \frac{\Sigma y_i}{n} = 88.4$

x	X $\bar{x}-83$	y	Y $\bar{y}-88$	xY	X^2	Y^2
78	-5	84	-4	20	25	16
77	-6	82	-6	36	36	36
85	2	82	-6	12	4	36
88	5	85	-3	15	25	9
87	4	89	1	4	16	1

82	-1	90	2	2	1	4
81	-2	88	0	0	4	0
77	-6	92	4	-24	36	16
76	-7	83	-5	35	49	25
83	0	89	1	0	0	1
97	14	98	10	140	196	100
93	10	99	11	110	100	121
				292	492	365

lines of regression of x on y is

$x-\bar{x} = b_{xy}(y-\bar{y})$ where $b_{xy} = \frac{\Sigma XY}{\Sigma Y^2}$

$= \frac{292}{365} = 0.8$

$\therefore x-83 = 0.8(y-88)$

$x-83 = 0.8y - 70.4$

$x - 0.8y = 12.6$

lines of regression of y on x is

$y-\bar{y} = b_{yx}(x-\bar{x})$ where $b_{yx} = \frac{\Sigma XY}{\Sigma X^2}$

$= \frac{292}{492} = 0.59$

$y-88 = 0.59(x-83)$

$y-88 = 0.59x - 48.97$

$0.59x - y = -39.03$

$y - 0.59x = 39.03$

Find the lines of regression of x on y and y on x for the following data

Price	10	12	13	12	16	15
Amount demanded	40	38	43	45	37	43

Estimate the likely demand when price is 20

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{10+12+---+15}{6} = 13$

$\bar{y} = \frac{\Sigma y_i}{n} = \frac{40+38+---+43}{6} = 41$

x	X $x-13$	y	Y $y-41$	XY	X^2	Y^2
10	-3	40	-1	3	9	1
12	-1	38	-3	3	1	9
13	0	43	2	0	0	4
12	-1	45	4	-4	1	16
16	3	37	-4	-12	9	16
15	2	43	2	4	4	4
				-6	24	50

lines of regression of x on y is

$x-\bar{x} = b_{xy}(y-\bar{y})$ where $b_{xy} = \frac{\Sigma XY}{\Sigma Y^2}$

$= \frac{-6}{50} = -0.12$

$x-13 = -0.12(y-41)$

$x-13 = -0.12y + 4.92$

$x + 0.12y = 17.92$

lines of regression of y on x is

$y-\bar{y} = b_{yx}(x-\bar{x})$ where $b_{yx} = \frac{\Sigma XY}{X^2}$

$= \frac{-6}{24} = -0.25$

$y-41 = -0.25(x-13)$

$y-41 = -0.25x + 3.25 \Rightarrow 0.25x + y = 44.25$

The heights of mother & daughters are given. From 2 tables of regression. estimate the avg height of daughter when height of mother is 64.5

mother	62	63	64	64	65	66	68	70
daughter	64	65	61	69	67	68	71	65

$\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{62+63+ - - - - +70}{8} = 65.25$

$\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{64+ - - - - +65}{8} = 66.25$

Assumed mean $\bar{x} = 65$ $\bar{y} = 66$

x	X $x-65$	y	Y $y-66$	XY	X^2	Y^2
62	-3	64	-2	6	9	4
63	-2	65	-1	2	4	1
64	-1	61	-5	5	1	25
64	-1	69	3	-3	1	9
65	0	67	1	0	0	1
66	1	68	2	2	1	4
68	3	71	5	15	9	25
70	5	65	-1	-5	25	1
				22	50	70

lines of regression of x on y is

$x-\bar{x} = b_{xy}(y-\bar{y})$ where $b_{xy} = \frac{\Sigma XY}{\Sigma Y^2}$

$= \frac{22}{70} = 0.31$

$x-65 = 0.31(y-66)$

$x-65 = 0.31y - 20.46$

$x - 0.31y = 44.54$

lines of regression of y on x is

$y-\bar{y} = b_{yx}(x-\bar{x})$ where $b_{yx} = \frac{\Sigma XY}{\Sigma X^2}$

$= \frac{22}{50} = 0.44$

$y-66 = 0.44(x-65)$

$y-66 = 0.44x - 28.6$

$0.44x - y = -37.4$

$0.44(64.5)$

8/12

1) Find the mean values of the variables x & y and corelation coefficient from following regression eq $2y-x-50=0$, $3y-2x-10=0$

Given lines of regression eq are

$2y-x-50=0$ — ①

$3y-2x-10=0$ — ②

mean $\bar{x}$ &
of regressi
$\therefore$ $2\bar{y}-\bar{x}$
$\bar{y}=90$ &
From eq (
$2y-x-5$
$2y = x+$
$y=\frac{1}{2}($
$b_{yx} = \frac{1}{2}$
$3\bar{y}-2$
2
b_{xy}
$r \cdot \frac{\sigma_y}{\sigma_x} \cdot r$
2) The regre
$x = 0.7$
mean of
tion b/w
$x = 0$
$y = 0$
$\therefore$ mean
$\bar{x} =$
$\bar{y} =$
$\bar{x} =$
From e
$x =$

mean $\bar{x}$ & $\bar{y}$ always passes through the lines of regression

$\therefore 2\bar{y} - \bar{x} - 50 = 0$ $\quad$ $3\bar{y} - 2\bar{x} - 10 = 0$

$\bar{y} = 90$ & $\bar{x} = 130$

From eq ①

$2y - x - 50 = 0$

$2y = x + 50$

$y = \frac{1}{2}(x + 50) = \frac{1}{2}x + 25$

$b_{yx} = \frac{1}{2} = r\frac{\sigma_y}{\sigma_x}$

$3\bar{y} - 2x - 10 = 0$

$2x = 3\bar{y} - 10$

$x = \frac{3}{2}\bar{y} - 5$

$b_{xy} = \frac{3}{2} = r\frac{\sigma_x}{\sigma_y}$

$r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y} = \frac{1}{2} \times \frac{3}{2} = \frac{3}{4}$

2) The regression eq of two variables x & y are $x = 0.7y + 5.2$ $\quad$ $y = 0.3x + 2.8$ Find the mean of the variables & Coefficient of Corelation b/w

$x = 0.7y + 5.2$ — ①

$y = 0.3x + 2.8$ — ②

$\therefore$ mean always passes through lines of regression

$\bar{x} = 0.7\bar{y} + 5.2$

$\bar{y} = 0.3\bar{x} + 2.8$

$\bar{x} = 9.06$ $\quad$ $\bar{y} = 5.51$

From eq ①

$x = 0.7y + 52$

$b_{xy} = 0.7$

$$\frac{r\sigma_x}{\sigma_y} = 0.7$$

From eq (2) $y = 0.3x + 2.8$

$b_{yx} = 0.3$

$$\frac{r\sigma_y}{\sigma_x} = 0.3$$

$$r \cdot \frac{\sigma_x}{\sigma_y} \cdot r\frac{\sigma_y}{\sigma_x} = 0.7 \times 0.3$$

$$r^2 = 0.21$$

3) Given that $x = 4y+5$ and $y = kx+4$ are the regression lines of x on y & y on x respectively. Show that $0 \le k \le 0.25$. If $k = 0.1$ find the mean the variables x and y and also the coefficient of corelation

$x = 4y+5 \qquad y = kx+4$

$b_{xy} = 4 \qquad b_{yx} = k$

$$\frac{r\sigma_x}{\sigma_y} = 4 \qquad \frac{r\sigma_y}{\sigma_x} = k$$

$$r \cdot \frac{\sigma_x}{\sigma_y} \cdot r\frac{\sigma_y}{\sigma_x} = 4 \times k$$

$$r^2 = 4k$$

we know that $-1 \le r \le 1$ (S.O.B.S)

$$0 \le r^2 \le 1$$

$$0 \le 4k \le 1$$

$$0 \le k \le \frac{1}{4}$$

$$0 \le k \le 0.25$$

When $K = 0.1$ $\quad x = 4y+5 \qquad y = 0.1x+4$

$x - 4y - 5 = 0 \qquad 0.1x - y + 4 = 0$

$\bar{x} - 4\bar{y} - 5 = 0 \qquad 0.1\bar{x} - \bar{y} + 4 = 0 \qquad \bar{x} = 18.33$, $\bar{y} = 5.83$

$x = 4y+5$

$b_{xy} = 4$

$$\frac{r\sigma_x}{\sigma_y} = 4$$

$y = 0.1x + 4$

$b_{yx} = 0.1$

$$\frac{r\sigma_y}{\sigma_x} = 0.1$$

$$r\frac{\sigma_x}{\sigma_y} \times r\frac{\sigma_y}{\sigma_x} = 4 \times 0.1$$

$$r^2 = 0.4$$

4) The corelation coefficient b/w variables x & y is $r = 0.6$ if $\Sigma x = 1.5$ $\Sigma y = 2$ $\bar{x} = 10$ & $\bar{y} = 20$ find regression lines of y on x & x on y

Regression lines of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$$x - 10 = 0.6 \frac{1.5}{2}(y - 20)$$

$$x - 10 = 0.45(y - 20)$$

$$x - 10 = 0.45y - 9$$

$$x - 0.45y = 1$$

$$x - 0.45y - 1 = 0$$

Regression lines of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 20 = r\frac{\sigma_y}{\sigma_x}(x - 10)$$

$$y - 20 = 0.6 \frac{2}{1.5}(x - 10)$$

$$y - 20 = 0.8(x - 10)$$

$$y - 20 = 0.8x - 8$$

$$0.8x - y + 12 = 0$$

5) Test whether the eq $2x + 3y = 4$ and $x - y = 5$ be the regression (valid) lines

$$2x + 3y = 4 \qquad\qquad x - y = 5$$

$$2x = -3y + 4 \qquad\qquad y = x - 5$$

$$x = \frac{-3}{2}y + 2 \qquad\qquad b_{yx} = 1$$

$$b_{xy} = \frac{-3}{2}$$

$$r\frac{\sigma_x}{\sigma_y} = \frac{-3}{2} \qquad\qquad r\frac{\sigma_y}{\sigma_x} = 1$$

$$r\frac{\sigma_x}{\sigma_y} \cdot r\frac{\sigma_y}{\sigma_x} = \frac{-3}{2} \times 1$$

$$r^2 = \frac{-3}{2}$$

The two lines are not lines of regression

9/12