ii. Fit a second-degree parabola to the following data: **5**

x	1	1.5	2	2.5	3	3.5	4
y	1.1	1.3	1.6	2	2.7	3.4	4.1

iii. Calculate the Karl Pearson's coefficient of correlation between x and y for the following data: **5**

x	23	27	28	28	29	30	31	33
y	18	20	20	27	21	29	27	29

Q.6 Attempt any two:

i. In a test examination given to two groups of students, the marks obtained were as follows: **5**

First group: 18, 20, 36, 50, 49, 36, 34, 49, 41

Second group: 29, 28, 26, 35, 30, 44, 46

Examine the significance of difference between the arithmetic averages of the marks secured by the students of the above two groups. ($t_{0.05,14} = 2.14$)

ii. The following table gives a classification of a sample of 160 plants of their flower colour and flatness of leaf: **5**

	Flat leaves	Coloured leaves	Total
White flower	99	36	135
Red flower	20	5	25
Total	119	41	160

Test whether the flower colours is independent of the flatness of leaves. (Given ${\chi^2}_{0.05,1} = 3.841$)

iii. A manufacturing company has purchased 3 new machines (A, B, C) of different makes and wishes to determine whether one of them is faster than the other in producing a certain item. From hourly production figures are observed at random from each machine and results are given below: **5**

A	B	C
20	18	25
21	20	28
23	17	22
16	25	28
20	15	32

Use one –way ANOVA to test whether machines differ significantly. (Table value of F at 5% level for $v_1 = 2$ and $v_2 = 12$ is 3.89)

******

Total No. of Questions: 6 *Total No. of Printed Pages:4*

Enrollment No......................................

MEDI-CAPS UNIVERSITY
Knowledge is Power

Faculty of Engineering
End Sem Examination May-2023
CS3EL11 / IT3CO29
Statistical Analysis / Computational Statistics

Programme: B.Tech. Branch/Specialisation: CSE All / IT

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. If the first and third quartile of the distributions are 22.19 and 44.63 respectively, then the value of quartile deviation is equal to- **1**

(a) 22.44 (b) 11.22 (c) 66.82 (d) None of these

ii. The relation between Arithmetic Mean (A), Harmonic Mean (H) and Geometric Mean (G) is given by the formula- **1**

(a) $AH = G^2$ (b) $AH < G^2$ (c) $AH > G^2$ (d) None of these

iii. The probability function of a discrete random variable is as follows: **1**

$X = x$	0	1	2	3	4
$P(X = x)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

Then the expected value of random variable X i.e $E(X) =$ ______.

(a) 0 (b) 1 (c) 2 (d) None of these

iv. The value of k for the given probability density function **1**

$f(x) = \begin{cases} kx^2, & 0 \le x \le 3 \\ 0 & ,otherwise \end{cases}$ is-

(a) 1 (b) $\frac{1}{4}$ (c) $\frac{1}{9}$ (d) None of these

v. For a normal distribution, if $x = 0.748$, mean= 0.7515 and standard deviation = 0.002, then the standard normal variate is equal to **1**

(a) – 1.75 (b) 1.75 (c) 0.5 (d) None of these

vi. In a Binomial Distribution, if mean is 4 and variance is 3, then number of trials $(n) =$ ______. **1**

(a) 18 (b) 16 (c) 12 (d) None of these

vii. If regression lines are perpendicular to each other, then the regression coefficients are- **1**

(a) Zero (b) Identical (c) Both (a) and (b) (d) None of these

viii. Coefficient of Correlation lies between- **1**

(a) $-3 \le r \le 3$ (b) $-2 \le r \le 2$

(c) $-1 \le r \le 1$ (d) None of these

ix. Which of the following test is used, if size of the sample is less than 30 ($n \le 30$)? **1**

(a) Chi-square test (b) F-test

(c) Fisher's z-test (d) None of these

x. When we reject the null hypothesis H_0, though it is true, then it is known as _______. **1**

(a) Type I error (b) Type II error

(c) Both (a) and (b) (d) None of these

Q.2 Attempt any two:

i. The following table gives the frequency distribution of married women by age at marriage: **5**

Age (in years)	Frequency	Age (in years)	Frequency
15-19	53	40-44	9
20-24	140	45-49	5
25-29	98	50-54	3
30-34	32	55-59	3
35-39	12	60-64	2

Calculate the median.

ii. The expenditure of 100 families are given below: **5**

Expenses in Rs	No. of families
0-10	14
10-20	?
20-30	27
30-40	?
40-50	15

The mode for the distribution is 24. Find the missing frequencies.

iii. Calculate the standard deviation and coefficient of variation for the following table. **5**

Class	Frequency	Class	Frequency
0-10	5	40-50	30
10-20	10	50-60	20
20-30	20	60-70	10
30-40	40	70-80	5

Q.3 Attempt any two:

i. Define Random Variable, Probability Mass Function, Probability Density Function and Cumulative Distribution Function. **5**

ii. In a college 25% students in Mathematics, 15% students in Physics and 10% students in Mathematics and Physics both are failed, A student is selected at random: **5**

(a) If he is failed in Physics, then find the chance of his failure in Mathematics.

(b) If he is failed in Mathematics, then find the chance of his failure in Physics.

(c) Find the chance of his failure in Mathematics and Physics.

iii. If $f(x) = \begin{cases} 0 & , x < 2 \\ \frac{1}{18}(3+2x), & 2 \le x \le 4 \\ 0, & x > 4 \end{cases}$ **5**

Prove that it is a probability density function. Find the probability that a variate having this density will fall in the interval $2 \le x \le 3$.

Q.4 Attempt any two:

i. Fit a Poisson's Distribution to the following and calculate the expected theoretical frequencies. **5**

x	0	1	2	3	4
f	122	60	15	2	1

ii. In a sample of 1000 cases, the mean of a certain test is 14 and the standard deviation is 2.5. Assuming the distribution to be normal, find: **5**

(a) How many students score between 12 and 15?

(b) How many students score above 18?

(c) How many students score below 8?

Given $P(-0.8 < z < 0) = 0.2881,\ P(0 < z < 0.4) = 0.1554$,

$P(0 < z < 1.6) = 0.4452, P(0 < z < -2.4) = 0.4918$

iii. Derive the mean and variance for a Binomial Distribution. **5**

Q.5 Attempt any two:

i. If regression equation of y on x be $4x - 5y + 33 = 0$ and regression equation of x on y be $20x - 9y = 107$ are two lines of regression, then find the following: **5**

(a) The mean values of x and y(b) The regression coefficients

(c) The correlation coefficient (d) The value of y for $x = 3$

(e) The standard deviation of x if the variance of x is 9

Faculty of Engineering
End Sem Examination, May-2023
CS3EL11/IT3CO29

Scan

(1)

Statistical Analysis / Computational Statistics

Programme - B. Tech   Branch - CSEA11/IT

Q.1 MCQ's

Ans (i) (b) 11.22 +1

Ans (ii) (a) $AH = G^2$ +1

Ans (iii) (c) 2 +1

Ans (iv) (c) 1/9 +1

Ans (v) (a) −1.75 +1

Ans (vi) (b) 16 +1

Ans (vii) (a) Zero +1

Ans (viii) (c) $-1 \leq r \leq 1$ +1

Ans (ix) (d) None of these +1

Ans (x) (a) Type I error +1

Q.2 Attempt any two:-

Sol^n (i)

Age (in Years)	(f) Frequency	C.F
14.5 - 19.5	53	53
19.5 - 24.5	140	193
24.5 - 29.5	98	291
29.5 - 34.5	32	323
34.5 - 39.5	12	335
39.5 - 44.5	9	344
44.5 - 49.5	5	349
49.5 - 54.5	3	352
54.5 - 59.5	3	355
59.5 - above (64.5)	2	357

+2

$N = \Sigma f = 357$

Here $N = 357$,
Median is the measure of $\frac{1}{2}(N+1)$ +1
$= 179$th term which lies in the
class 19.5 - 24.5 → median class

$$\text{Median } (M_d) = l + \frac{\frac{N}{2} - F}{f} \times i$$ +1

$$= 19.5 + \frac{178.5 - 53}{140} \times 5$$

$$= 19.5 + 4.48$$

Solⁿ 2(ii) Let F_1 & F_2 be the unknown frequencies of classes 10-20 & 30-40 resp^t.

Exp. in (Rs)	No. of families (f)	C.F
0-10	14	14
10-20	F_1	$14+F_1$
20-30	27	$41+F_1$
30-40	F_2	$41+F_1+F_2$
40-50	15	$56+F_1+F_2$
	$N=\Sigma F = 56+F_1+F_2$	

+2

$N = 100 \Rightarrow 56 + F_1 + F_2 = 100$

$\Rightarrow F_1 + F_2 = 44$ — (1)

$$M_0 = l + \frac{f - f_{-1}}{2f - f_{-1} - f_1} \times i$$ +1

∵ Mode is 24 which lies in the class 20-30

$l = 20,\ f = 27,\ f_{-1} = F_1,\ f_1 = F_2,\ i = 10.$

+1

$$24 = 20 + \frac{27 - F_1}{54 - F_1 - F_2} \times 10$$

$\Rightarrow 3F_1 - 2F_2 = 27$ — (2)

on solving (1) & (2), we get

+2

Solⁿ 2 iii

Class	Mid value x	f	$u = \frac{x-35}{10}$	fu	u^2	fu^2
0-10	5	5	-3	-15	9	45
10-20	15	10	-2	-20	4	40
20-30	25	20	-1	-20	1	20
30-40	35	40	0	0	0	0
40-50	45	30	1	30	1	30
50-60	55	20	2	40	4	80
60-70	65	10	3	30	9	90
70-80	75	5	4	20	16	80
		$\Sigma f = 140$		$\Sigma fu = 65$		$\Sigma fu^2 = 385$

Assume mean $= 35 = A$, $h = 10$

A.M $\quad M = A + h \frac{\Sigma fu}{N}$

$$= 35 + 10\left(\frac{65}{140}\right) \qquad +1$$

$$M = 39.64$$

$$S.D\ (\sigma) = h\sqrt{\frac{\Sigma fu^2}{N} - \left(\frac{\Sigma fu}{N}\right)^2} \qquad +1$$

$$= 10\sqrt{\frac{385}{140} - (0.464)^2}$$

$$S.D = 15.9 \quad \text{Ans}$$

Q3. Attempt any two:-

Solⁿ 3 (i) Define

Random Variable:- Let S be the sample space associated with a given random experiment. Then a real valued fuctⁿ X which assigns to each outcome $x \in S$ to a unique real number $X(x)$ is called a random variable +1

Probability mass function:-
If $x_1, x_2, \dots x_n$ are n different values of a discrete random variable X and $p(x_1), p(x_2), \dots p(x_n)$ be their respective probabilities such that

(i) $p(x_i) \geq 0$, $i = 1, 2, \dots n$.

(ii) $\sum p(x_i) = 1$, $i = 1, 2, \dots n$. +1

then,
$p(x)$ is known as p.m.f of the variable X.

Probability density function:-
Let $f(x)$ be a probability fuctⁿ in the interval $[a, b]$, then the probability that the variate value X to lie in the interval $[a, b]$ is given by +1

The fut^n $f(x)$ is p.d.f. if it holds two properties

(i) $f(x) \geq 0$

(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$, i.e. total probability is unity.

Cumulative Distribution function :-

For Discrete random variable :-

Let X be a discrete r.v, then distributive fut^n of X is given by

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i$$ +1

s.t. $p_i \geq 0$ & $\sum_{i=1}^{\infty} p_i = 1$

For continuous random variable -

Let X be a continuous r.v., then distributive fut^n of X is given by +1

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx.$$

Sol^n 3 (ii)

Let E_1 & E_2 be the events of failure in Mathematics and physics respt.

Let $n(S) = 100$

∵ 25% Students are failed in Maths,

so $n(E_1) = 25$

$$P(E_1) = \frac{n(E_1)}{n(S)} = \frac{25}{100} = \frac{1}{4}$$ $+\frac{1}{2}$

∵ 15% Students are failed in physics

So, $n(E_2) = 15$ $+\frac{1}{2}$

$$P(E_2) = \frac{n(E_2)}{n(S)} = \frac{15}{100} = \frac{3}{20}$$

Again, 10% Students are failed in physics & Maths, so

$n(E_1 \cap E_2) = 10$

$$P(E_1 \cap E_2) = \frac{n(E_1 \cap E_2)}{n(S)} = \frac{10}{100} = \frac{1}{10}$$ $+1$

(i) $P(E_1/E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{1/10}{3/20} = \frac{2}{3}$ Ans $+1$

(ii) $P(E_2/E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{1/10}{1/4} = \frac{2}{5}$ Ans $+1$

Q of 3 (iii) (i) The given fu^n $f(x) \geq 0$ and +1

(ii) $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{2} f(x)\,dx + \int_{2}^{4} f(x)\,dx + \int_{4}^{\infty} f(x)\,dx$

$= 0 + \int_{2}^{4} \left(\frac{3+2x}{18}\right) dx + 0$ +1

$= \frac{1}{18}\left(3x + x^2\right)_2^4$

$\int_{-\infty}^{\infty} f(x)\,dx = 1$ +1

So, the given fu^n is p.d.f.

Now, $P(2 \leq x \leq 3) = \int_{2}^{3} f(x)\,dx$

$= \int_{2}^{3} \left(\frac{3+2x}{18}\right) dx$

$= \frac{1}{18}\left[3x + x^2\right]_2^3$ +1

$P(2 \leq x \leq 3) = \frac{4}{9}$ Ans +1

Q.4 Attempt any two:-

Sol^n 4(i) Mean $= \frac{\Sigma fx}{\Sigma f} = \frac{100}{200} = 0.5$ +1

The theoretical frequencies of r-success

$f(r) = N P(r)$

$= N \times \cdot \frac{e^{-m} m^r}{\lfloor r}$ +1

$= 200 \times \frac{e^{-(0.5)} (0.5)^r}{\lfloor r}, \; r = 0,1,2,3,4.$

For $r=0$, $f(0) = 121.3 \approx 121$

For $r=1$, $f(1) = 60.65 \approx 61$

For $r=2$, $f(2) = 15.25 \approx 15$ +2

For $r=3$, $f(3) = 2.53 \approx 3$

For $r=4$, $f(4) = 0.316 \approx 0$

+1

Thus theoretical frequencies are

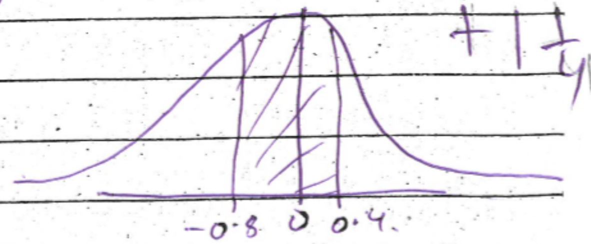
r	0	1	2	3	4
fe	121	61	15	3	0

Sol^n 4 (ii) $n = 1000$, $\bar{x} = 14$, $\sigma = 2.5$

(i) Score bet^n 12 & 15

$z_1 = \frac{x - \bar{x}}{\sigma} = \frac{12 - 14}{2.5} = -0.8$

$z_2 = \frac{15 - 14}{2.5} = 0.4$



+1 1/4

$P(12 \le x \le 15) = P(-0.8 \le z \le 0.4)$

$= P(-0.8 \le z \le 0) + P(0 \le z \le 0.4)$

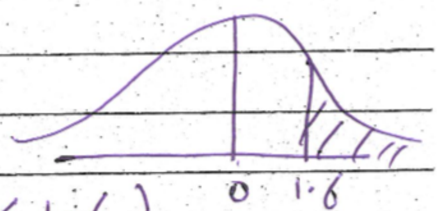
$= 0.2881 + 0.1554$

$= 0.4435$

$\therefore$ required no. of students $= 1000 \times 0.4435$ +1 1/4

$= 443.5 \approx 444$ (approx) Ans

(ii) $z = \frac{18 - 14}{2.5} = 1.6$



$P(x > 18) = P(z > 1.6)$

$= 0.5 - P(0 < z < 1.6)$

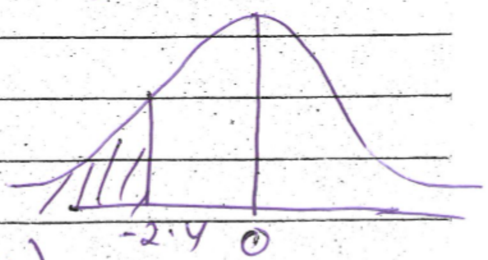
$= 0.5 - 0.4452$

$= 0.0548$

$\therefore$ required no. of students $= 1000 \times 0.0548$ +1 1/4

$= 54.8 \approx 55$ Ans

(iii) $z = \frac{8 - 14}{2.5} = -2.4$



$P(x < 8) = P(z < -2.4)$

$= 0.5 - P(0 < z < 2.4)$

Solⁿ 4 (iii) Mean and Variance for a Binomial distribution

The first moment about origin is given by

$$\mu_1' = \sum_{r=0}^{n} r \cdot P(r)$$ +1

Where $P(r) = {}^nC_r \, q^{n-r} p^r$ is Binomial distⁿ.

$$\mu_1' = \sum_{r=0}^{n} r \cdot {}^nC_r \, q^{n-r} p^r$$

$$= \sum_{r=1}^{n} r \cdot \frac{\lfloor n}{\lfloor r \lfloor n-r} \, q^{n-r} p^r$$

$$= \sum_{r=1}^{n} \frac{\lfloor n}{\lfloor r-1 \lfloor n-r} \, q^{n-r} p^r$$

$$= p \sum_{r=1}^{n} \frac{n \lfloor n-1}{\lfloor r-1 \lfloor (n-1)-(r-1)} \, q^{(n-1)-(r-1)} p^r$$

$$= np \sum_{r=1}^{n} {}^{n-1}C_{r-1} \, q^{(n-1)-(r-1)} \cdot p^{(r-1)}$$

$$= np\,(q+p)^{n-1}$$ +1

$$\boxed{\mu_1' = np} \rightarrow \text{mean}$$

Second moment about origin is given by

$$= \sum_{r=0}^{n} \{r(r-1)+r\}\, {}^nC_r q^{n-r} p^r$$

$$= \sum_{r=0}^{n} r(r-1)\, {}^nC_r q^{n-r} p^r + \sum_{r=0}^{n} r.\, {}^nC_r q^{n-r} p^r$$

$$= \sum_{r=2}^{n} r(r-1) \frac{\lfloor n}{\lfloor r \lfloor n-r} q^{n-r} p^r + np$$

$$= \sum_{r=2}^{n} \frac{r(r-1)\, n(n-1) \lfloor n-2}{r(r-1) \lfloor (r-2) \lfloor n-r} q^{n-r} p^r + np$$

$$= n(n-1) p^2 \sum_{r=2}^{n} \frac{\lfloor n-2}{\lfloor r-2 \lfloor (n-2)-(r-2)} q^{(n-2)-(r-2)} p^{r-2} + np$$

$$= n(n-1) p^2 \sum_{r=2}^{n} {}^{n-2}C_{r-2}\, q^{(n-2)-(r-2)} p^{r-2} + np$$

$$= n(n-1) p^2 (q+p)^{n-2} + np$$

$$= n^2p^2 - np^2 + np$$

$$\mu_2' = n^2p^2 + npq \qquad \text{---(ii)}$$

$$\text{Variance} = \mu_2' - (\mu_1')^2$$

$$= n^2p^2 + npq - (np)^2$$

$$\underline{\underline{\text{Variance} = npq}} \qquad \text{---(iii)}$$

Q.5 Attempt any two:-

Soln 5 (i) (a) The mean values of x and y i.e. $(\bar{x}, \bar{y})$ lie on both lines.

$$4\bar{x} - 5\bar{y} = -33$$
$$20\bar{x} - 9\bar{y} = 107$$

on solving

$$\bar{x} = 13 \ \& \ \bar{y} = 17.$$

so mean value of x & y are 13 & 17. +1

(b) The regression line of y on x is

$$4x - 5y + 33 = 0$$

$$y = \frac{4}{5}x + \frac{33}{5}$$

Regression coeff. y on x $(b_{yx}) = \frac{4}{5}$ + ½

Regression line of x on y is

$$20x - 9y = 107$$
$$x = \frac{9}{20}y + \frac{107}{20}$$

Regression coeff ~~x~~ on y (b_{xy})

$$b_{xy} = \frac{9}{20}.$$ + ½

(c) Correlation coefficient

$$r = \pm\sqrt{b_{yx} \times b_{xy}}$$

$$r = \pm\sqrt{\frac{9}{...} \times \frac{4}{...}}$$

(d) put $x = 3$ in $4x - 5y + 33 = 0$.

$y = 9$ +1

(e) Variance of $x = 9$

$\sigma_x^2 = 9 \Rightarrow \sigma_x = 3$.

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

$\frac{4}{5} = \frac{3}{5} \times \frac{\sigma_y}{3}$ +1

$\Rightarrow \sigma_y$ (S.D. of y) $= 4$ Ans

Solⁿ 5 (ii) $m = 7$, and changing the origin

$u = \frac{x - 2.5}{0.5} = 2x - 5$, $v = y$

Let the second degree parabola of the curve be

$V = a + bu + cu^2$

then the normal equations are

$$\left.\begin{aligned} \Sigma v &= ma + b\Sigma u + c\Sigma u^2 \\ \Sigma uv &= a\Sigma u + b\Sigma u^2 + c\Sigma u^3 \\ \Sigma u^2 v &= a\Sigma u^2 + b\Sigma u^3 + c\Sigma u^4 \end{aligned}\right\} +2$$

15

x	y	$u=2x-5$	$v=y$	u^2	uv	u^2v	u^3	u^4
1	1.1	-3	1.1	9	-3.3	9.9	-27	81
1.5	1.3	-2	1.3	4	-2.6	5.2	-8	16
2	1.6	-1	1.6	4	-1.6	1.6	-1	1
2.5	2	0	2	0	0	0	0	0
3	2.7	1	2.7	1	2.7	2.7	1	1
3.5	3.4	2	3.4	4	6.8	13.6	8	16
4	4.1	3	4.1	9	12.3	36.9	27	81
		$\Sigma u=0$		$=28$	14.3	69.9	0	196

+2

Putting the values in normal eqn.

$7a + 28c = 16.2$

$28b = 14.3 \Rightarrow b = 0.51$

$28a + 196c = 69.9.$

On solving, we get

$a = 2.07, \quad c = 0.061, \quad b = 0.51$

Hence,

$v = 2.07 + 0.511u + 0.061u^2$

$y = 2.07 + 0.511(2x-5) + 0.061(2x-5)^2$

$$\boxed{y = 1.04 - 0.193x + 0.243x^2}$$

+1

Ans

Sol^n 5 (iii) Let $u_i = x_i - 28$, $v_i = y_i - 27$, $n = 8$

x_i	y_i	$u_i = x_i - 28$	$v_i = y_i - 27$	$u_i v_i$	u_i^2	v_i^2
23	18	-5	-9	45	25	81
27	20	-1	-7	7	1	49
28	20	0	-7	0	0	49
28	27	0	0	0	0	0
29	21	1	-6	-6	1	36
30	29	2	2	4	4	4
31	27	3	0	0	9	0
33	29	5	2	10	25	4
		$\Sigma u_i = 5$	$\Sigma v_i = -25$	$\Sigma u_i v_i = 60$	$\Sigma u_i^2 = 65$	$\Sigma v_i^2 = 223$

+3

$$r(x,y) = \frac{n \Sigma u_i v_i - (\Sigma u_i)(\Sigma v_i)}{\sqrt{n \Sigma u_i^2 - (\Sigma u_i)^2}\sqrt{n \Sigma v_i^2 - (\Sigma v_i)^2}}$$ +1

$r(x,y) = 0.80$ Ans +1

* Note:- Another formula can also be used.

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

Q.6. Attempt any two :-

Solⁿ 6(i)

First Group			Second Group		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
18	-19	361	29	-5	25
20	-17	289	28	-6	36
36	-1	1	26	-8	64
50	13	169	35	1	1
49	12	144	30	-4	16
36	-1	1	44	10	100
34	-3	9	46	12	144
49	12	144			
41	4	16			
333	0	1134	238	0	386

+2

Mean of I group $= \bar{x} = \dfrac{\Sigma x}{n_1} = \dfrac{333}{9} = 37$

Mean of II group $= \bar{y} = \dfrac{\Sigma y}{n_2} = \dfrac{238}{7} = 34$

+1

S = standard error of difference

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

$$t = \frac{\bar{x} - \bar{y}}{s} \sqrt{\left(\frac{n_1 \times n_2}{n_1 + n_2}\right)}$$ +1

$$= \frac{37 - 34}{10.242} \sqrt{\left(\frac{9 \times 7}{9 + 7}\right)}$$

$$\boxed{t = 0.577}$$

d.o.f $= v = n_1 + n_2 - 2 = 9 + 7 - 2 = 14$.

∴ tabulated value > calculated value $+\frac{1}{2}$

Hence, the difference between arithmetic averages of the sample is not significant. $+\frac{1}{2}$

Solⁿ 6 (ii) (1) Null Hypothesis – There is no association betⁿ colour flowers and flatness of leaves. +1

(2) Calculation of expected frequencies.

	Flat leaves	Coloured leaves	Total
White flower	$\frac{135 \times 119}{160} = 100.41$	$\frac{135 \times 41}{160} = 34.59$	135
Red flower	$\frac{25 \times 119}{160} = 18.59$	$\frac{25 \times 41}{160} = 6.41$	25

(3) Calculation of χ^2-statistic

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2 = \frac{(99-100.41)^2}{100.41} + \frac{(20-18.59)^2}{18.59} +$$

$$\frac{(36-34.59)^2}{34.59} + \frac{(5-6.41)^2}{6.41}$$

+1

$$\chi^2 = 0.492$$

(4) d.o.f = $(m-1)(n-1)$
$= (2-1)(2-1) = 1$

(5) $\because$ calculated value < tabulated value

(6) Decision → So null hypothesis is accepted

+1

i.e. there is no association betn colour flowers and flatness of leaves.

sol^n 6 (iii) correction factor, $C = \frac{T^2}{N}$,

where T is the grand total in all the samples
N is the total no of values

$$C = \frac{330 \times 330}{15} = 7260$$ +1

Sum of squares bet^n col^m (SSC) $= \frac{(\Sigma x_1)^2 + (\Sigma x_2)^2 + (\Sigma x_3)^2}{5} - C$

$$SSC = \frac{(100)^2 + (95)^2 + (135)^2}{5} - 7260$$

$$\boxed{SSC = 190}$$ +1

Total sum of squares of deviation
(SST) $= (20^2 + 21^2 + 23^2 + 16^2 + 20^2 + 18^2 + 20^2 + 17^2 + 25^2 + 15^2 + 25^2 + 28^2 + 22^2 + 28^2 + 32^2) - 7260$

$$\boxed{SST = 330}$$ +1

Sum of squares within machines

SSE = SST − SSC = 140

ANOVA Table

	Source of Variation	Sum of Squares	Degree of freedom	Mean square	Calculation of F
SSC	Betn samples	190	2	$MSC = \frac{190}{2} = 95$	$F = \frac{MSC}{MSE} = \frac{95}{11.67} = 8.14$
SSE	Within samples	140	12	$MSE = \frac{140}{12} = 11.67$	

$\because$ Calculated value > tabulated value

$\therefore$ Null hypothesis is rejected.

i.e. Machines differ significantly.